



HAL
open science

Utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats

Paul Savary

► **To cite this version:**

Paul Savary. Utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats. Sciences agricoles. Université Bourgogne Franche-Comté, 2021. Français. NNT : 2021UBFCK045 . tel-03714898

HAL Id: tel-03714898

<https://theses.hal.science/tel-03714898>

Submitted on 6 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de doctorat de l'établissement Université Bourgogne Franche-Comté

Préparée dans l'UMR 6049 ThéMA - Université Bourgogne Franche-Comté - CNRS
et dans l'UMR 6282 Biogéosciences - Université Bourgogne Franche-Comté - CNRS

Financée par l'entreprise ARP-Astrance

École Doctorale n° 554 Environnements - Santé

Spécialité : Biologie des populations et écologie

Par

Paul SAVARY

Utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats

Thèse présentée et soutenue à Besançon le 7 décembre 2021

Directeur de thèse : **Jean-Christophe Foltête**

Co-directeur de thèse : **Stéphane Garnier**

Co-encadrant de thèse : **Hervé Moal**

Jury

Mme Stéphanie Manel,	Directrice d'études	Rapporteuse
M. Éric Petit,	Directeur de recherches	Rapporteur
Mme Laurence Després,	Professeure des universités	Présidente, Examinatrice
M. Laurent Bergès,	Ingénieur - chercheur	Examinateur

ARP-Astrance, 75008 Paris, France

UMR 6049 ThéMA, Université Bourgogne Franche-Comté - CNRS, 25000 Besançon, France

UMR 6282 Biogéosciences, Université Bourgogne Franche-Comté - CNRS, 21000 Dijon, France



Table des matières

	Page
Table des matières	ix
Liste des figures	xiii
Liste des tableaux	xv
Résumés	xvii
Remerciements	xx
Avant-propos	xxi
Introduction générale	1
1 La connectivité écologique des habitats : un enjeu pour la conservation de la biodiversité	1
1.1 La biodiversité : différents niveaux d'organisation du vivant affectés par les activités humaines	1
1.2 Distribution spatiale de la biodiversité	4
1.2.1 La relation aire-espèces	4
1.2.2 La théorie de la biogéographie insulaire	4
1.2.3 Les modèles de métapopulations	5
1.3 L'écologie du paysage	6
1.3.1 Origines de la discipline	6
1.3.2 Hétérogénéité du paysage	8
1.3.3 Quantité et configuration spatiale des habitats	9
1.4 Connectivité des habitats et biodiversité	14
1.4.1 Origines et évolutions de la définition de la connectivité des habitats	14
1.4.2 Influence de la connectivité des habitats sur la biodiversité	16
1.4.3 Implications pour la conservation de la biodiversité	18
2 Les graphes paysagers, outils de modélisation de la connectivité des habitats	25
2.1 Modélisation spatiale de la connectivité des habitats	25
2.2 Les graphes paysagers	29
2.2.1 La théorie des graphes en écologie	29
2.2.2 Modéliser les réseaux d'habitat sous la forme de graphes paysagers	36
2.2.3 Intérêts et limites des graphes paysagers	44
2.3 Modalités d'utilisation conjointe de données biologiques et de graphes paysagers	46
3 La structure génétique des populations, une réponse biologique influencée par la connectivité des habitats	55

3.1	La structure génétique des populations	55
3.1.1	Diversité génétique locale et différenciation génétique	55
3.1.2	Dérive et flux génétique : principaux déterminants des variations génétiques neutres	58
3.2	Distribution spatiale des variations génétiques	61
3.2.1	Influence de la subdivision des populations et de l'hétérogénéité de leurs tailles	62
3.2.2	L'isolement par la distance	63
3.3	L'apport de la génétique du paysage pour analyser l'influence de la connectivité sur la structure génétique	65
3.3.1	Origines et objectifs de la génétique du paysage	65
3.3.2	Prise en compte de l'hétérogénéité du paysage	65
3.3.3	Plusieurs niveaux de mise en relation des réponses génétiques et des variables paysagères	68
3.3.4	La boîte à outils de la génétique du paysage	68
3.4	Limites actuelles des études de génétique du paysage	71
4	Les graphes génétiques, des outils pour l'analyse de la structure génétique des populations	74
4.1	Objectifs de l'utilisation des graphes génétiques	74
4.2	Construction et analyse des graphes génétiques	75
4.2.1	Méthodes de construction des graphes génétiques	75
4.2.2	Méthodes d'analyse des graphes génétiques	77
4.3	Limites de l'utilisation des graphes génétiques	79
5	Enjeux de l'utilisation conjointe des graphes génétiques et paysagers	81
5.1	Comparer graphes génétiques et paysagers pour traiter des problématiques de génétique du paysage	82
5.1.1	Analyser les relations entre les différentes composantes de la structure du paysage et de la structure génétique	82
5.1.2	Évaluer les effets d'échelle	84
5.1.3	Identifier des barrières dans le paysage	85
5.2	Intégrer graphes génétiques et paysagers pour bénéficier de leur complémentarité	86
	Organisation du manuscrit de thèse	87

Partie I : Comparaison de graphes génétiques et paysagers 89

Chapitre 1 Comment construire et analyser des graphes génétiques pour étudier l'effet du paysage sur les réseaux de dispersion et de flux génétique ? 91

1	Introduction	93
2	Matériel et méthodes	95
2.1	Données paysagères	95
2.2	Simulations de flux génétiques	95
2.3	Graphes génétiques	96
2.3.1	Méthodes d'élagage	96
2.3.2	Distances génétiques	97
2.4	Analyses des graphes	97

	2.4.1	Analyses de similarité des topologies	97
	2.4.2	Analyses basées sur les distances	97
	2.4.3	Analyse factorielle des résultats des simulations	98
3		Résultats	100
	3.1	Résultats des simulations	100
	3.2	Graphes génétiques	101
	3.2.1	Analyses de similarité des topologies	101
	3.2.2	Analyses basées sur les distances	103
4		Discussion	107
	4.1	Quand et pourquoi élaguer un graphe génétique ?	107
	4.2	Comment élaguer un graphe génétique pour identifier la topologie précise du réseau de dispersion ?	108
	4.3	Comment élaguer un graphe génétique pour inférer la résistance du paysage au flux génétique ?	108
	4.4	Quelle distance génétique utiliser et comment ?	110
	4.5	Limites et perspectives	110

Chapitre 2 graph41g : un package R pour construire et analyser des graphes en génétique du paysage 115

1		Introduction	117
2		Organisation des fonctions du package	117
	2.1	Traitements des données brutes	118
	2.1.1	Données génétiques	118
	2.1.2	Données spatiales	118
	2.1.3	Analyses préliminaires	119
	2.2	Construction et analyse de graphes génétiques	120
	2.2.1	Construction de graphes génétiques	120
	2.2.2	Analyses de graphes génétiques	120
	2.3	Construction et analyse de graphes paysagers	121
	2.4	Comparaisons de graphes génétiques et paysagers	121
3		Limites et conclusion	125

Chapitre 3 Analyse de l’influence de la quantité d’habitat atteignable sur la structure génétique à l’aide de graphes génétiques et paysagers 129

1		Introduction	131
2		Matériel et méthodes	133
	2.1	Espèce étudiée et zone d’échantillonnage	133
	2.2	Analyse des données génétiques	133
	2.2.1	Indices de structure génétique	133
	2.3	Estimation de la quantité d’habitat atteignable (ARH)	135
	2.3.1	Construction des graphes paysagers	135
	2.3.2	Métriques d’ARH	136
	2.4	Analyses de la relation entre les métriques d’ARH et les indices de structure génétique	137
	2.4.1	Analyses de corrélation	137

	2.4.2	Régressions Partial Least Squares (PLS)	138
3		Résultats	138
	3.1	Corrélations entre métriques d'ARH et indices génétiques	138
	3.2	Régressions Partial Least Squares (PLS)	141
4		Discussion	143
	4.1	Les métriques d'ARH sont-elles adaptées à l'analyse de la structure génétique ?	143
	4.2	Est-ce que l'ARH influence la diversité et la différenciation génétiques de façon aussi importante et à la même échelle spatiale ?	144
	4.3	Est-ce que le coût de déplacement à travers la matrice paysagère affecte la diversité et la différenciation génétiques de la même manière ?	145
	4.4	Limites et perspectives de recherche	146
Chapitre 4 Validation de graphes paysagers à l'aide de données génétiques			149
1		Introduction	151
2		Matériel et méthodes	152
	2.1	Zone et modèle d'étude	152
	2.2	Modélisation de la connectivité écologique à l'aide de graphes paysagers	153
	2.2.1	Données spatiales et climatiques	153
	2.2.2	Dires d'experts	155
	2.2.3	Indices de spécialisation	155
	2.2.4	Modèle de distribution de l'espèce	155
	2.2.5	Graphes paysagers	156
	2.3	Analyse de données génétiques	158
	2.3.1	Échantillonnage et génotypage	158
	2.3.2	Indices de structure génétique	158
	2.4	Validation de la modélisation basée sur les graphes paysagers	158
	2.4.1	Validation des distances-coût associées aux liens des graphes paysagers	159
	2.4.2	Validation des métriques de connectivité	160
3		Résultats	160
	3.1	Graphes paysagers	160
	3.1.1	Paramètres de construction des graphes paysagers	160
	3.1.2	Caractéristiques des éléments des graphes	162
	3.2	Structure génétique des populations	162
	3.3	Validation des distances-coût associées aux liens	162
	3.4	Validation des métriques de connectivité	167
4		Discussion	169
	4.1	Les graphes paysagers sont validés empiriquement par les données génétiques .	169
	4.2	Quelles relations entre la validité écologique des graphes, leur besoin en données et la complexité des méthodes de construction et d'analyse ?	170
	4.3	Implications pour la conservation de la biodiversité	172

Chapitre 5	Analyse de la sensibilité des chemins de moindre coût et des distances-coût aux scénarios de coût	179
1	Introduction	181
2	Méthodes	183
2.1	Sélection des paysages	183
2.2	Création des scénarios de coût	183
2.3	Modélisation des chemins de moindre coût	183
2.4	Coïncidences spatiales des LCPs et corrélations entre matrices de CDs	185
2.5	Description des contextes paysagers et des scénarios de coût	185
2.6	Analyses statistiques des facteurs de la sensibilité de la modélisation aux scénarios de coût	185
3	Résultats	186
3.1	Structure des paysages sélectionnés	186
3.2	Influence relative des valeurs de coût et de la structure du paysage sur la sensibilité de la modélisation	186
3.3	Influence de la structure du paysage sur la sensibilité de la modélisation	188
3.4	Influence des caractéristiques des scénarios de coût sur la sensibilité de la modélisation	188
3.5	Relation entre la coïncidence spatiale entre LCPs et la corrélation entre matrices de CDs	190
4	Discussion	193
4.1	Sensibilité des LCPs et CDs aux scénarios de coût	193
4.2	Implications pour l'inférence des valeurs de coût et la modélisation basée sur les LCPs	195
Chapitre 6	Inférer la résistance du paysage au flux génétique en intégrant les nœuds et les liens de graphes génétiques et paysagers dans des modèles gravitaires	199
1	Introduction	201
2	Méthodes	203
2.1	Simulations	203
2.1.1	Simulations des paysages et des populations	203
2.1.2	Simulations de flux génétiques	207
2.1.3	Scénarios de coût alternatifs	208
2.2	Modèles gravitaires	208
2.3	Évaluation de la performance des modèles	209
3	Résultats	210
3.1	Résultats des simulations	210
3.2	Modèles gravitaires	210
3.3	Arbres de régression	212
4	Discussion	214
4.1	L'inférence de valeurs de coût à partir de données génétiques est-elle correcte ?	214

4.2	La SHNe, la distribution spatiale des populations et le taux de migration influencent-ils l'inférence ?	215
4.3	Quand devrait-on inclure des variables intra-populationnelles dans des modèles gravitaires pour inférer des valeurs de coût ?	216
4.4	Limites et perspectives	217
5	Conclusion	218

Discussion générale **221**

1	Les graphes génétiques et paysagers permettent-ils d'améliorer notre compréhension des relations entre connectivité des habitats et structure génétique ?	222
1.1	Enseignements tirés de la mise en relation de métriques de connectivité et d'indices génétiques issus de graphes	222
1.1.1	Mise en évidence empirique de l'influence de la quantité d'habitat atteignable	222
1.1.2	La quantité d'habitat atteignable affecte différemment la diversité et la différenciation génétiques	224
1.2	Intérêt des graphes pour l'analyse de la relation entre distances génétiques et paysagères	225
1.2.1	Élaguer les graphes pour ajuster l'échelle de l'analyse	225
1.2.2	Considérer les nœuds et les liens des graphes pour intégrer l'hétérogénéité spatiale de la dérive	227
1.3	Obstacles à la compréhension des relations entre connectivité des habitats et structure génétique	229
1.3.1	Influence de la méthode de modélisation des liens entre taches d'habitat	229
1.3.2	Questions relatives à la modélisation de la structure génétique sous la forme d'un graphe	231
2	Intérêt des graphes génétiques et paysagers pour la conservation de la connectivité des habitats	237
2.1	Les graphes paysagers peuvent-ils orienter correctement la conservation de la connectivité ?	238
2.1.1	Validation écologique des graphes paysagers à l'aide de graphes génétiques	238
2.1.2	Rapport coûts-bénéfices associé à la modélisation par les graphes paysagers	238
2.2	Comment orienter la conservation à l'aide de graphes génétiques et paysagers ?	239
2.2.1	Utilisation des graphes paysagers pour conserver la quantité d'habitat atteignable	240
2.2.2	Intégrer graphes génétiques et paysagers pour conserver la variabilité génétique	243
2.3	Obstacles à l'utilisation des graphes génétiques et paysagers pour conserver la connectivité	244
2.3.1	Le coût associé à l'obtention des données génétiques	244
2.3.2	Le passage difficile de la modélisation à l'action	248
3	Perspectives de recherche	250

3.1	Vers une diversification de l'utilisation des graphes en génétique du paysage? .	250
3.1.1	Peut-on construire des graphes génétiques avec d'autres marqueurs génétiques?	250
3.1.2	Comment dépasser le graphe de populations statique et non-dirigé? .	251
3.2	Applications des méthodes développées à d'autres disciplines	252
3.2.1	Comprendre l'enchevêtrement des réseaux d'hôtes, de parasites et de leurs habitats à l'aide de graphes	252
3.2.2	Contribuer à la compréhension et à la conservation de la diversité spécifique à l'aide de graphes	253
	Conclusion	257
	Glossaire	263
	Bibliographie	264

Table des figures

1	Niveaux d'organisation du vivant et menaces pesant sur la biodiversité	3
2	Relation aire-espèces	4
3	Théorie de la biogéographie insulaire	5
4	Le concept d'hétérogénéité du paysage en écologie du paysage	9
5	Processus temporels et patrons spatiaux liés à la quantité et à la fragmentation de l'habitat	10
6	Richesse spécifique en fonction de la fragmentation du paysage selon l'HAH	13
7	Positionnement disciplinaire de la thèse	19
8	Modélisation de la connectivité structurelle à l'aide de la méthode de dilatation-érosion	26
9	Création d'une surface de résistance à partir de données géographiques	27
10	Calcul des chemins de moindre coût	27
11	Méthodes guidant le choix des coûts associées à chaque type d'occupation du sol	28
12	Compromis entre la qualité et la complexité des modèles de connectivité	30
13	Différents types de graphes	32
14	Principe de la partition d'un graphe à l'aide du calcul de la modularité	33
15	Étapes de construction d'un graphe paysager	37
16	Fonction de transformation d'une distance de dispersion en probabilité de dispersion	39
17	Métriques de connectivité des habitats issues d'un graphe paysager	40
18	Grille d'analyse des articles couplant graphes paysagers et données biologiques	47
19	Caractéristiques des études couplant graphes paysagers et données biologiques	48
20	Résultat de l'ACM des liens entre données biologiques, objectifs, implications des gestionnaires et méthodes de construction et d'analyse des graphes paysagers	50
21	CAH réalisée à partir de l'ACM des liens entre données biologiques, objectifs, implications des gestionnaires et méthodes de construction et d'analyse des graphes paysagers	51
22	Utilisation conjointe de données biologiques et de graphes paysagers	52
23	Diversité génétique locale et différenciation génétique	56
24	Forces évolutives à l'origine des variations génétiques	59
25	Influence de la dérive et du flux génétique sur la diversité génétique locale	60
26	Typologie des patrons d'isolement par la distance	64
27	Influence de la connectivité des habitats sur la dérive et le flux génétique	66
28	Différents niveaux d'analyse en génétique du paysage	69
29	Marqueurs génétiques utilisés en génétique du paysage	70
30	Graphe génétique, flux de dispersion direct et flux génétiques pluri-générationnels	75
31	Élagage d'un graphe à l'aide d'un seuil de percolation	76
32	Élagage d'un graphe à l'aide du principe d'indépendance conditionnelle	77
33	Signification écologique de la moyenne de l'inverse des poids des liens génétiques	78
34	Calcul de la distance génétique conditionnelle	79

35	Enjeux de l'utilisation conjointe des graphes génétiques et paysagers	83
36	Complémentarité des graphes génétiques et paysagers	86
37	Quatre combinaisons contrastées d'un paysage et d'une distribution spatiale de populations (A, B, C, D)	100
38	ACP des huit variables décrivant les résultats des simulations	101
39	Nuages de points représentant la relation entre les distances génétiques et les distances paysagères	102
40	Corrélations de Mantel entre les distances-coût et les distances génétiques séparant les populations directement connectées sur les graphes génétiques	103
41	Corrélations de Mantel entre les distances-coût et les distances génétiques calculées sur le graphe génétique le long des chemins les plus courts entre populations	105
42	Guide méthodologique pour la construction et l'analyse des graphes génétiques	112
43	Organisation des fonctions du package <code>graph4lg</code>	123
44	Figures de diagnostic et d'analyse de graphes génétiques réalisées à l'aide de <code>graph4lg</code>	124
45	Criquet ensanglanté (<i>Stethophyma grossum</i>)	133
46	Localisation des populations de Criquet ensanglanté échantillonnées	134
47	Indices génétiques et métriques d'ARH calculés dans les graphes génétiques et paysagers	135
48	Coefficients de corrélation de Spearman entre les indices MIW et la métrique F	140
49	Résultats des régressions PLS modélisant les indices génétiques en fonction des métriques d'ARH	141
50	Méthode de validation des graphes paysagers à l'aide de données génétiques	153
51	Topographie de l'île de Guadeloupe	153
52	La Paruline caféïette (<i>Setophaga plumbea</i>)	154
53	Occupation du sol de la Guadeloupe et modèle de distribution de la Paruline caféïette	154
54	Cartes des coûts de déplacement de la Paruline caféïette obtenues en Guadeloupe selon chaque scénario de coût	163
55	Ajustement des modèles expliquant les distances génétiques entre des populations de Paruline caféïette de Guadeloupe en fonction de distances-coût calculées le long des liens de graphes paysagers	165
56	Cartes des erreurs de prédiction des distances génétiques entre les populations de Paruline caféïette en Guadeloupe	166
57	Différences entre les rangs des taches d'habitat de la Paruline caféïette (Guadeloupe) en termes de connectivité selon les paramètres utilisés pour calculer les métriques	168
58	Protocole de l'analyse de sensibilité des LCPs et CDs aux scénarios de coût	184
59	Résultats de la sélection de paysages	187
60	Distribution statistique des coïncidences spatiales entre LCPs et des corrélations de Mantel entre CDs selon le scénario ou le paysage	189
61	Résultats des régressions PLS-R2 évaluant l'influence de la structure du paysage sur la sensibilité de la modélisation	190
62	Arbres de régression évaluant l'influence des caractéristiques des scénarios de coût sur la sensibilité de la modélisation	191

63	Relation entre la coïncidence spatiale entre LCPs et la corrélation entre matrices de CDs	192
64	Simulations et évaluation de la qualité des inférences de valeurs de coût basées sur les modèles gravitaires	204
65	Identification des situations dans lesquelles l'inclusion de variables locales dans les modèles améliore l'inférence	205
66	Calcul de la capacité des taches dans les simulations permettant d'évaluer la qualité de l'inférence des coûts	206
67	Arbre de régression expliquant les valeurs de la variable D qui mesure l'intérêt des modèles gravitaires pour l'inférence des valeurs de coût	213
68	Distribution statistique des valeurs de D dans les feuilles de l'arbre de régression	214
69	Optimisation d'un scénario de coût basée sur la comparaison des modules de graphes génétiques et paysagers	233
70	Indice de diversité génétique potentielle	245
71	Exemple de simulation génétique destinée à orienter les mesures de conservation	247

Liste des tableaux

1	Paramètres de construction des graphes génétiques	99
2	Topologies des graphes de dispersion et des graphes génétiques	106
3	Distances génétiques entre populations calculées par la fonction <code>mat_gen_dist</code>	119
4	Scénarios de coût utilisés pour construire les graphes paysagers	136
5	Coefficients de corrélation de Spearman entre métriques d'ARH et indices génétiques .	139
6	Résultats des régressions PLS des indices génétiques en fonction des métriques d'ARH	142
7	Définition des éléments des graphes paysagers créés en Guadeloupe selon différentes méthodes	157
8	Indices de spécialisation de la Paruline caféïette et valeurs de coût associés à chaque type d'occupation du sol selon les différents scénarios	161
9	Indices décrivant la structure génétique des populations de Paruline caféïette en Guadeloupe	164
10	Coefficients de corrélation de Spearman entre les indices décrivant la structure génétique des populations de Paruline caféïette (Guadeloupe) et les métriques de connectivité . .	168
11	Différences entre les modélisations obtenues avec le vrai scénario et les scénarios alternatifs	193
12	Ajustement des modèles gravitaires	211
13	Capacité des modèles gravitaires à distinguer le vrai scénario de coût des scénarios alternatifs	212

Résumé

Titre : Utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats

Mots clés : connectivité des habitats, génétique du paysage, écologie du paysage, génétique des populations, théorie des graphes

La connectivité écologique des habitats est nécessaire aux processus écologiques assurant le maintien de la biodiversité. Des méthodes ont donc été développées pour la modéliser afin de comprendre précisément son influence et d'orienter les mesures de conservation de la biodiversité. Parmi ces méthodes, les graphes paysagers modélisent un réseau d'habitat sous la forme d'un ensemble de taches d'habitat (nœuds) reliées par des chemins de dispersion potentiels (liens). La validité écologique de ces outils nécessitait néanmoins d'être évaluée à l'aide de données reflétant les réponses biologiques des populations à la connectivité de leurs habitats. Les données génétiques permettent cette validation car la structure génétique des populations dépend notamment des flux génétiques entre leurs taches d'habitat. La structure génétique peut également être modélisée par un graphe génétique dont les nœuds correspondent à des populations et dont les liens sont pondérés par le degré de différenciation génétique entre populations. L'objectif de cette thèse était d'utiliser conjointement des graphes génétiques et paysagers pour (i) évaluer la validité écologique des graphes

paysagers et (ii) améliorer notre compréhension de la relation entre connectivité et structure génétique. Après avoir identifié les méthodes de construction et d'analyse des graphes génétiques les plus adaptées à chaque contexte et développé un outil informatique permettant l'utilisation conjointe des graphes génétiques et paysagers, nous les avons comparés dans le cadre de deux études empiriques. Elles ont permis (i) d'évaluer l'influence respective des différentes composantes de la connectivité des habitats sur la diversité et la différenciation génétiques et (ii) de confirmer la validité écologique des graphes paysagers. Nous avons ensuite montré que l'intégration de variables associées à la fois aux nœuds et aux liens de ces deux types de graphes améliorerait l'estimation de l'influence des éléments du paysage sur la connectivité. Les méthodes développées dans cette thèse pourraient trouver d'autres applications dans ce champ d'étude comme dans d'autres. Nous espérons que les résultats de cette thèse et l'outil informatique développé y contribueront.

Abstract

Title: Bridging landscape graphs and genetic graphs for analysing habitat ecological connectivity

Key words: habitat connectivity, landscape ecology, population genetics, graph theory

Several key ecological processes for maintaining biodiversity rely upon the ecological connectivity of habitat. Accordingly, connectivity modelling methods have been developed for understanding precisely the influence of connectivity and deriving sound biodiversity conservation measures. Among them, landscape graphs represent habitat networks as sets of habitat patches (nodes) connected by potential dispersal paths (links). Yet, the ecological relevance of these tools required validation from biological data reflecting closely the influence of habitat connectivity. Genetic data allow for such validation as population genetic structure partly depends on dispersal-driven gene flow between habitat patches. Genetic structure can also be modelled as a genetic graph whose nodes correspond to populations while its links are weighted by pairwise genetic differentiation measures. The objective of this PhD project was to bridge landscape and genetic graphs in or-

der to (i) assess the ecological relevance of landscape graphs and (ii) gain knowledge regarding the relationship between habitat connectivity and population genetic structure. After identifying genetic graph construction and analysis methods fitting several research contexts and developing a software package for the joint use of both landscape and genetic graphs, we compared them in two empirical studies. We thereby (i) assessed the respective influence of several components of the habitat connectivity pattern on both genetic diversity and differentiation and (ii) validated the ecological relevance of landscape graphs. We then evidenced that integrating variables deriving from the nodes and links of both types of graphs could improve the inference of the effect of every landscape feature on connectivity. The methods we have developed could find new applications in this field and others. We hope that the results of this thesis will contribute to this.

Remerciements

Quel doctorant aurais-je été sans les liens tissés avec les personnes qui m'ont entouré pendant ces années de thèse ? Une tache isolée ? vouée à la dérive ? à la circularité consanguine de mes idées ? Très probablement. Ma thèse a concerné des réseaux d'habitat et de populations mais ce sont avant tout des connexions humaines qui m'ont permis de la mener à bien. Malgré cette métaphore théorique, trop tentante pour que je ne la fasse pas, ce sont des sentiments bien personnels qui me poussent à présent à remercier l'ensemble des personnes qui ont contribué à cette thèse, parfois sans même s'en rendre compte.

La légende raconte que c'est dans un restaurant de la boucle bisontine que tout a commencé. Un soir, Jean-Christophe, Stéphane et Hervé se sont dit qu'il serait intéressant de comparer des graphes paysagers et des graphes génétiques. L'idée a rapidement fait son chemin et je les ai rencontrés le 27 juillet 2017 au 9 avenue Percier à Paris. L'aventure a commencé là et je tiens à vous remercier tous les trois chaleureusement pour vos contributions parfaitement complémentaires à tout ce qui a suivi.

Jean-Christophe, en géographe modeste, tu as souvent prétendu ne rien comprendre à la génétique des populations. J'en ai toujours fortement douté et bien que tu aies parfois considéré ta contribution à ma thèse comme mineure, laisse moi te dire que c'est loin d'être le cas. Tes remarques pertinentes m'ont non seulement aidé à avancer, quitte à rectifier le tir parfois, mais elles m'ont aussi rassuré et guidé dans les choix qu'impose la réalisation d'une thèse. Tu accordes une place importante à la qualité des relations humaines dans toute aventure scientifique et c'est une chance d'y participer à tes côtés. C'est donc autant pour la transmission de tes connaissances et de ton expérience de chercheur que pour le cadre de confiance instauré autour de ma thèse que je te remercie.

Stéphane, tes avis, guidés par un œil affuté et peu faillible, m'ont été extrêmement utiles. J'ai rapidement compris que lorsque tes remarques commençaient par "*je vais peut-être dire une connerie...*", cela signifiait que tu avais déniché un point clé que j'avais totalement négligé. Les nombreuses heures passées dans le bureau 322 du bâtiment Gabriel ont été véritablement formatrices. Elles ont fait émerger beaucoup d'idées et ont parfois donné lieu à des débats presque philosophiques sur l'équilibre entre la migration et la dérive. Tu as toujours su accorder ta rigueur à ta bienveillance et je te remercie également pour ta confiance qui m'a permis d'explorer librement le champ des possibles.

Hervé, sans ton dynamisme et ta curiosité sans limites, cette thèse n'aurait jamais vu le jour. Tu m'as toujours soutenu et incité à sortir des sentiers battus. Nos rendez-vous réguliers ont donné un véritable rythme à ce travail et m'ont permis de prendre le recul nécessaire même lorsque j'étais accaparé par des détails.

Gilles, plus fin connaisseur des algorithmes génétiques que de la biologie moléculaire initialement, je crois que tu ne peux plus dire que tu ne sais pas ce qu'est un allèle. Tu as su répondre à beaucoup de mes interrogations, donné un grand nombre d'idées et rendu une infinité de services pour réaliser des calculs parfois assez pénibles dans un langage que tu n'apprécies guère. Pourtant, résumer ta contribution à des réflexions théoriques voire à du dépannage informatique serait extrêmement réducteur. Merci pour tous les moments passés ensemble et pour ton soutien sans faille.

Le laboratoire ThéMA a constitué un cadre de travail extrêmement agréable (hors confinement). Je dois pour cela remercier mes camarades de thèse Anissa, Claire, Hanae, les Julien(s), Léo, Marion, Maud, Morgan ainsi que Pauline et Jérôme et tous les autres camarades, collègues et membres de ThéMA que j'ai pris plaisir à rencontrer et que je ne peux hélas pas lister exhaustivement ici. Un excellent accueil m'a également été réservé à Biogéosciences. Je remercie particulièrement Antoine, pour les discussions riches et rassurantes que nous avons eues lors de mes passages à Dijon, Alexandrine, avec qui j'ai beaucoup apprécié travailler dans la dernière ligne droite de cette thèse, ainsi qu'Aurélie, Bruno et Nicolas pour leurs remarques et conseils.

J'adresse un immense merci à Béatrice, Tolga, Gwennaële et Benoît. Sans leur soutien, en particulier lors de ces derniers mois, le contenu de cette thèse aurait été fortement "élagué". Ils m'ont permis de travailler dans d'excellentes conditions durant quatre années. À toute l'équipe d'ARP-Astrance, merci pour vos attentions à chacun de mes rares voyages à Paris.

Le sentiment d'appartenir à "l'équipe Graphab" a largement contribué à ma motivation tout au long de cette aventure. Les échanges et les moments passés avec vous, en particulier lors de nos séminaires annuels, m'ont beaucoup apporté. Je remercie donc les membres de cette équipe (que je n'ai pas déjà remerciés), à savoir Yohan, Céline, Marc et Xavier.

Catherine Labruère, Ahmed Jebrane, Catherine Larédo, Catherine Matias, merci de m'avoir répondu et aidé lorsque j'ai été confronté à des problèmes mathématiques dépassant largement mes compétences. Je remercie également Maarten van Strien, dont le travail a largement inspiré, à son insu, le premier chapitre de cette thèse, et qui a ensuite directement contribué au troisième chapitre et m'a gentiment reçu dans son laboratoire. Marie-Josée Fortin et Eve Afonso, merci beaucoup pour vos contributions lors des comités de thèse. Stéphanie Manel, Éric Petit, Laurent Bergès et Laurence Després, merci d'avoir accepté de faire partie de mon jury.

À l'ensemble de mes proches, merci. Je vous ai parfois ennuyés avec des histoires de graphes et vous m'avez souvent aidé à "déconnecter". Peio, je t'adresse un merci bien particulier pour ton soutien durant ces quatre années. Sans mes parents (éloignés géographiquement mais génétiquement très proches!), je n'aurais pas pu réaliser les études dont cette thèse est l'aboutissement. Merci à vous. Enfin, je remercie mes grands parents Christian et Odette qui m'ont abonné à Science & Vie Junior il y a quelques années maintenant et appris tant de choses.

Avant-propos

Cette thèse de doctorat s'inscrit dans le champ de la génétique du paysage et concerne l'utilisation conjointe de graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats. Elle a été financée par ARP-Astrance dans le cadre d'une convention CIFRE établie pour la période 2018-2021 par l'Association Nationale Recherche Technologie (ANRT). Ce travail de recherche a impliqué le laboratoire ThéMA (UMR 6049 Université Bourgogne Franche-Comté - CNRS, directeur de thèse : Jean-Christophe Foltête) et le laboratoire Biogéosciences (UMR 6282 Université Bourgogne Franche-Comté - CNRS, co-directeur de thèse : Stéphane Garnier). Hervé Moal et Tolga Coskun (ARP-Astrance) en ont assuré l'encadrement opérationnel.

Le manuscrit qui suit rassemble des résultats ayant fait l'objet d'articles scientifiques rédigés en anglais et intégrés aux annexes. Ainsi, les chapitres 1 et 2 et la section 2.3 de l'introduction se basent sur des articles déjà publiés dans des revues à comité de lecture au moment de la rédaction de ce manuscrit. Les autres articles ont été soumis en vue d'une publication ou sont en cours de finalisation. Cette thèse est intégralement rédigée en français et par souci de clarté, la traduction des articles s'est accompagnée d'une vulgarisation de leur contenu.

Compte tenu de la diversité des champs d'action et disciplines représentés au sein des organismes porteurs de cette thèse, nous avons souhaité que son contenu s'adresse à un large public. Aussi, certaines notions pouvant paraître élémentaires à certains sont parfois expliquées dans le détail. Ce souhait se traduit par l'intégration au corps du texte d'encadrés concernant des notions fondamentales et par la présence d'un glossaire à la fin du manuscrit. Par ailleurs, des annexes présentent de façon détaillée certaines des méthodes d'analyse statistique utilisées lors de ce travail.

Introduction générale

*Le vivant n'est pas une cathédrale, c'est un feu qui s'éteint,
un feu qui se reconstruit tout seul, se déploie, crée mille formes,
dès qu'on lui laisse de la place et du temps.*

Baptiste Morizot

1 La connectivité écologique des habitats : un enjeu pour la conservation de la biodiversité

1.1 La biodiversité : différents niveaux d'organisation du vivant affectés par les activités humaines

Le vivant se rencontre sous une grande diversité de formes que l'on nomme "biodiversité" depuis près d'une quarantaine d'années (Wilson, 1988) (Encadré 1). Toutes les formes du vivant ne se répartissent pas de la même façon à la surface de la planète. Cette distribution spatiale de la biodiversité est à l'origine de processus écologiques permanents, qui l'influencent en retour (Figure 1). Dans ce contexte, les travaux d'écologie visent à identifier les déterminants des interactions entre les patrons spatiaux de la biodiversité et les processus écologiques auxquels ils sont liés.

La biodiversité est actuellement menacée par les activités d'une espèce en particulier : l'espèce humaine (Figure 1) (Díaz *et al.*, 2019 ; Dirzo et Raven, 2003 ; Duraiappah *et al.*, 2005 ; Ceballos *et al.*, 2015 ; Pereira *et al.*, 2012 ; Sala *et al.*, 2000 ; Vitousek *et al.*, 1997). En effet, les activités humaines sont responsables du déclin et de la disparition de nombreuses espèces (Ceballos *et al.*, 2015 ; Díaz *et al.*, 2019 ; Dirzo *et al.*, 2014) et elles bouleversent les patrons de diversités spécifique et génétique

(DiBattista, 2008 ; Garner *et al.*, 2005 ; Miraldo *et al.*, 2016)¹. Alors que la pollution des milieux (Penuelas *et al.*, 2013) et le changement climatique (Bellard *et al.*, 2012) modifient entre autres les aires de distribution potentielle des espèces, l'introduction d'espèces exotiques invasives bouleverse les interactions entre espèces (Vitousek *et al.*, 1996). En outre, lorsque la chasse, la pêche et d'autres types de prélèvements sont peu ou mal régulés, ils engendrent des diminutions irréversibles des effectifs des espèces concernées (Maxwell *et al.*, 2016). Enfin, la destruction des habitats impacte les populations de nombreuses espèces et les conditions nécessaires à leur survie (Díaz *et al.*, 2019). Elle constitue la cause la plus importante de cet impact écologique (Díaz *et al.*, 2019), et continuera de l'être selon des projections faites pour le siècle en cours (Sala *et al.*, 2000).

Nous considérons ici que les zones d'**habitat**² d'une espèce sont celles qui constituent sa niche fondamentale, au sens d'Hutchinson (1957), c'est-à-dire celles qui réunissent l'ensemble des conditions abiotiques (climatiques, pédologiques, topographiques) et biotiques (trophiques, symbiotiques, parasitaires) nécessaires à l'accomplissement du cycle de vie de cette espèce. Bien que toutes ces zones ne soient pas nécessairement occupées par l'espèce considérée, différant alors de sa niche réalisée, leur destruction réduit directement l'espace potentiellement disponible pour cette espèce et de fait le nombre de populations que forme cette dernière (Hughes *et al.*, 1997). De plus, elle perturbe indirectement des processus écologiques nécessaires au maintien et à l'évolution de la biodiversité, tels que les mouvements des espèces (Tucker *et al.*, 2018). Cette influence indirecte fait l'objet de travaux dans le monde entier (Haddad *et al.*, 2015). Ses conséquences s'observent dans le temps, dans un délai plus ou moins long après la modification de la distribution des habitats (Halley *et al.*, 2016 ; Hanski et Ovaskainen, 2002 ; Tilman *et al.*, 1994). Elles sont néanmoins le corollaire de l'influence de la distribution spatiale des habitats sur la biodiversité, qu'il est nécessaire de comprendre pour minimiser les impacts anthropiques.

Encadré 1 : Qu'est-ce que la biodiversité ?

La biodiversité est le terme utilisé pour désigner l'ensemble de la variabilité des formes du vivant (Wilson, 1988). Elle recouvre habituellement la diversité des écosystèmes, des espèces et des gènes (Figure 1) (Dirzo et Raven, 2003). Au niveau le plus large, la diversité des écosystèmes est produite par la variabilité des conditions de coexistence d'individus appartenant aux différentes branches de la classification du vivant. Les extrémités de ces branches correspondent généralement aux espèces. Leur diversité se chiffre en millions (Mora *et al.*, 2011) et constitue le second niveau d'organisation de la biodiversité. L'ensemble de ces formes vivantes ont pour point commun de posséder du matériel génétique dont les variations sont responsables de la diversité du vivant. Elles expliquent pourquoi deux individus de deux espèces différentes ne se ressemblent pas, mais elles sont aussi à l'origine de différences phénotypiques (taille, couleur, comportement, aptitudes

1. Les analyses de Millette *et al.* (2020) et de Leung *et al.* (2020) nuancent ce constat sur le déclin de la biodiversité, insistant sur le fait que les activités humaines bouleversent de façon importante les patrons de diversité mais n'entraînent pas systématiquement la diminution nette du nombre d'espèces ou de la variation génétique.

2. Les termes écrits en gras dans le corps du texte sont définis dans le glossaire la première fois qu'ils sont employés ou lorsqu'ils sont définis précisément dans le texte.

immunitaires, etc.) entre deux individus appartenant à la même espèce. La **diversité génétique** constitue ainsi le niveau d'organisation du vivant le plus fin.

Ainsi, bien que les différences entre deux écosystèmes, deux espèces et deux individus trouvent toutes leur origine dans des variations génétiques, la distinction entre la diversité des écosystèmes, des espèces et des gènes vient du fait qu'on peut se focaliser sur plusieurs niveaux d'organisation de ces variations pour les caractériser. Notons par ailleurs qu'il est fréquent que le niveau génétique du concept de biodiversité fasse implicitement référence aux variations génétiques intra-spécifiques exclusivement (voir par exemple [Harper et Hawksworth \(1994\)](#)). Il s'agit en quelque sorte d'un abus de langage étant donné que la variation génétique s'observe également entre les espèces.

Enfin, outre la variabilité des formes du vivant rencontrée sur la planète (patrons), la biodiversité englobe également la variabilité des processus écologiques auxquels le vivant est lié. En effet, les individus de chaque espèce se déplacent et interagissent avec d'autres individus de la même espèce ou d'espèces différentes et avec le milieu abiotique. Les variations observées dans la façon dont ces processus se déroulent constituent en elles-mêmes une dimension de la biodiversité.

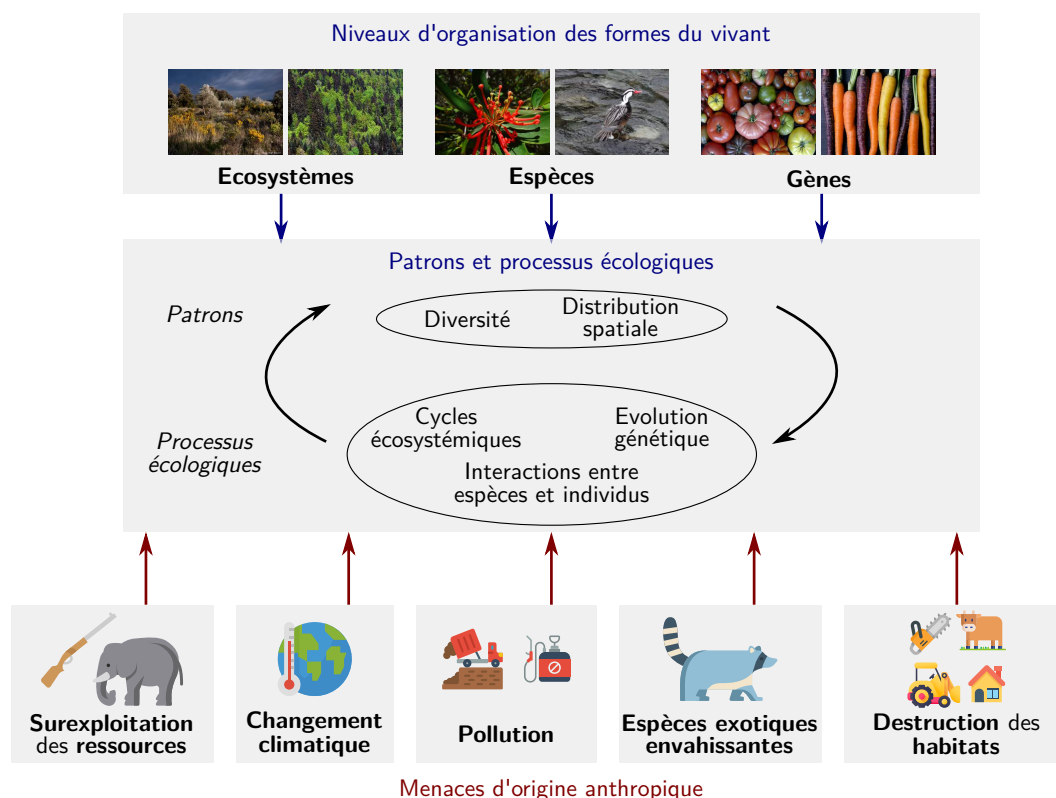


FIGURE 1 – Niveaux d'organisation du vivant et menaces pesant sur la biodiversité

1.2 Distribution spatiale de la biodiversité

1.2.1 La relation aire-espèces

L'identification des différentes formes du vivant et de leur distribution spatiale a commencé avec les travaux des fondateurs de la biogéographie, dont faisaient partie Alexander von Humboldt (1769-1859), Charles Darwin (1809-1882) ou Alfred Russel Wallace (1823-1913). La mise en évidence de l'évolution des espèces (Darwin, 1859) et le concept de niche écologique (Elton, 1927 ; Grinnell, 1917 ; Hutchinson, 1957) ont donné une autre dimension à ces premiers travaux essentiellement descriptifs.

En effet, l'écologie scientifique a dès lors cherché des "lois" générales expliquant la présence et l'abondance des espèces et a testé les hypothèses correspondantes à partir d'observations empiriques. Ces travaux ont mis en évidence un lien entre le nombre d'espèces observées lors d'inventaires, c'est-à-dire la richesse spécifique, et la superficie totale échantillonnée lors de ces inventaires (Arrhenius, 1921). Les écologues ont découvert que cette relation "aire-espèces" obéissait souvent à une loi puissance telle que $S = kA^z$ avec S la richesse spécifique du groupe d'espèces étudié, A l'aire échantillonnée et k et z deux constantes (Preston, 1962) (Figure 2).

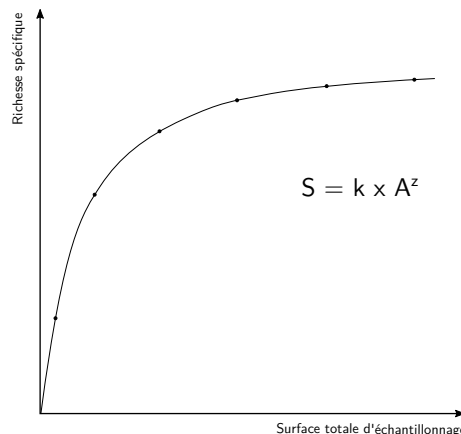


FIGURE 2 – Relation aire-espèces. Évolution de la richesse spécifique en fonction de la surface totale d'échantillonnage selon la relation identifiée par Preston (1962).

1.2.2 La théorie de la biogéographie insulaire

MacArthur et Wilson (1967) ont montré que dans le cas particulier des archipels, la richesse spécifique ne dépend pas uniquement de la surface totale échantillonnée (Figure 3). En effet, selon leur théorie de la biogéographie insulaire (TBI), le nombre d'espèces présentes sur une île dépend à la fois de la surface de l'île et de son éloignement par rapport au continent. L'extinction et la colonisation sont les deux processus écologiques à l'origine du patron spatial de diversité décrit par la TBI. D'une part, plus une île est grande, plus le taux d'extinction de ses espèces est faible. Ce taux d'extinction augmente également à mesure que le nombre d'espèces présentes sur l'île augmente car la quantité de ressources disponibles tend à diminuer dans ce cas, faisant augmenter le risque d'extinction. D'autre part, plus une île est éloignée du continent, plus le taux de colonisation de cette île par une espèce venant du continent est faible. De même, la probabilité de colonisation de cette île par de nouvelles espèces est d'autant plus faible que les espèces déjà présentes sur l'île sont nombreuses car il est difficile pour une nouvelle espèce de trouver les conditions correspondant à sa niche écologique dans

une île déjà occupée par d'autres espèces. Lorsque le taux d'extinction et le taux de colonisation sont égaux, le nombre d'espèces présentes sur une île atteint une valeur d'équilibre. La figure 3A illustre les différences de richesse spécifique à l'équilibre observées sur les îles d'un archipel en fonction de leur surface et de leur distance au continent.

La TBI telle que formulée par MacArthur et Wilson (1967) prévoit également qu'à surface échantillonnée équivalente il y ait plus d'espèces sur une grande île que sur une petite île. Cela se traduit par le fait que la richesse spécifique totale est initialement plus élevée si on commence par comptabiliser les espèces observées au niveau des surfaces situées sur les plus grandes îles (Figure 3B). Notons que cette prédiction dépend en grande partie du fait que les espèces que l'on peut observer sur les petites îles soient un sous-ensemble des espèces présentes sur les grandes îles, et elles-mêmes un sous-ensemble de celles du continent (Simberloff et Abele, 1976).

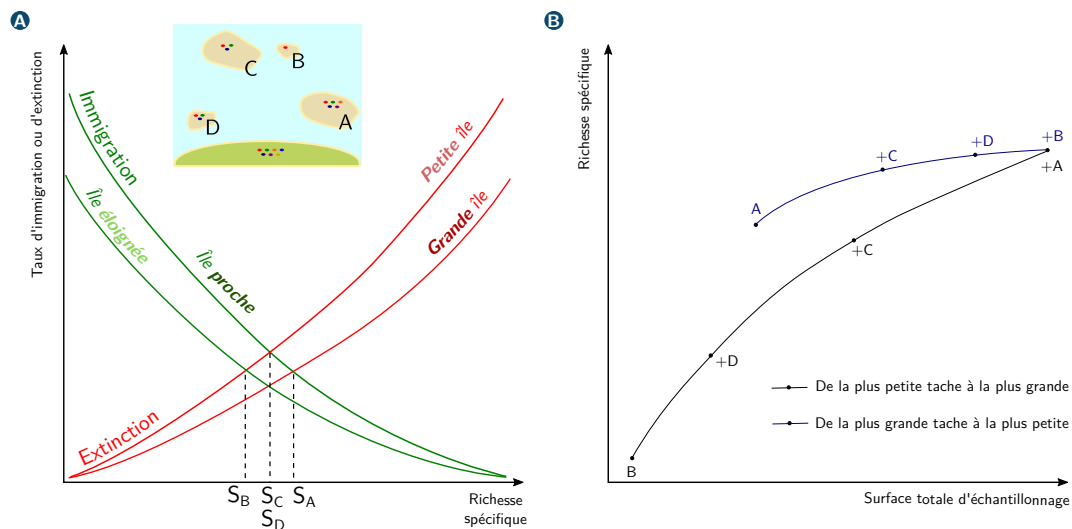


FIGURE 3 – Théorie de la biogéographie insulaire (TBI) de MacArthur et Wilson (1967). (A) Richesses spécifiques à l'équilibre selon la TBI. Lorsque les taux d'extinction et de colonisation sont égaux, la richesse spécifique S est à l'équilibre (S_A , S_B , S_C , S_D). L'évolution de ces deux processus en fonction de la richesse spécifique déjà présente dans des îles plus ou moins grandes et proches du continent est à l'origine des différences de richesse spécifique entre ces îles selon la TBI. (B) Relation aire-espèces dans le cas d'un archipel d'îles séparées du continent qui agit comme une source d'espèces, d'après la TBI. Plus une île est grande, plus le taux d'extinction des espèces y est faible et plus on y trouve d'espèces à surface échantillonnée équivalente. Ainsi, en comptant le nombre total (accumulé) d'espèces en commençant par celles observées sur les plus grandes îles, à surface échantillonnée équivalente, on compte rapidement beaucoup d'espèces puis peu de nouvelles espèces par la suite (courbe bleue). À l'inverse, en commençant par comptabiliser les espèces observées sur les petites îles, à surface équivalente, le nombre d'espèces est initialement faible et croît de façon importante ensuite (courbe noire). Le nombre total d'espèces observées à la fin de l'échantillonnage est logiquement le même dans les deux cas.

1.2.3 Les modèles de métapopulations

Lorsque les zones d'habitat d'une espèce sont subdivisées dans l'espace, elles forment des agrégats que l'on nomme taches d'habitat (*habitat patch* en anglais). Un ensemble de populations peut alors occuper différentes taches d'habitat. Dès le milieu du XX^e siècle, Wright (1940) et Andrewartha *et al.* (1954) ont évoqué la dynamique différente de ces populations subdivisées par rapport à celle d'une population non subdivisée. Au cours du temps, on y observe de façon aléatoire des extinctions locales suivies de recolonisations. Ce fonctionnement basé sur une répartition du risque d'extinction global (*spreading of risk*, Den Boer (1968)) permet la survie d'un ensemble de populations malgré les épisodes d'extinction locale (Reddingius et Den Boer, 1970). En effet, lorsqu'une espèce occupe plusieurs taches d'habitat, le risque que toutes ses populations s'éteignent en même temps est d'autant plus faible que

les taches occupées sont nombreuses. En outre, le rôle important de la colonisation dans ce phénomène explique l'influence prépondérante de la **dispersion** pour la survie des populations dans des environnements changeants ; démontré empiriquement à plusieurs reprises (voir [Cayuela et al. \(2020\)](#) par exemple).

À la fin des années 1970, les ensembles de populations subdivisées occupant des taches d'habitat discrètes, ou "populations de populations", ont été qualifiés de "métapopulations" et leur dynamique modélisée par des équations dépendant des taux d'extinction et de colonisation ([Levins, 1969, 1970](#)). Les modèles de métapopulations et la TBI ont pour point commun qu'ils accordent un rôle majeur à la taille et à l'isolement des taches d'habitat dans les processus d'extinction et de colonisation. En revanche, tandis que la TBI prédit la richesse spécifique d'une île à l'équilibre, le modèle des métapopulations est dynamique et concerne l'extinction et la colonisation d'une seule espèce dans un ensemble de taches d'habitat. Par ailleurs, la TBI distingue les îles du continent tandis que dans le modèle initial des métapopulations de [Levins \(1969\)](#), toutes les taches sont de nature équivalente. Les travaux d'[Hanski et Gilpin \(1991\)](#) et d'[Hanski \(1989\)](#) ont complexifié ce modèle et testé sa validité à partir de données empiriques ([Hanski, 1998](#)) ; issues notamment du suivi à long terme des populations d'une espèce de papillon, la Mélitée du Plantain (*Melitaea cinxia*). Ils ont également introduit des métriques permettant de calculer le risque d'extinction et de colonisation de chaque tache d'habitat à partir de sa surface, de la surface des autres taches et des distances spatiales qui la séparent de ces autres taches.

La principale conclusion de ces premiers travaux portant sur la distribution de la biodiversité et sur la dynamique des populations est que : la structure spatiale des processus d'extinction et de colonisation affecte autant les populations que les taux de natalité et de mortalité, la compétition et la prédation ([Hanski, 1998](#)). Ces travaux s'inscrivent ainsi dans le champ de l'écologie spatiale. Les modèles de métapopulations et la TBI diffèrent également de par la nature de la **matrice**, c'est-à-dire le territoire ou paysage, au sein duquel se situent les taches d'habitat considérées³. En effet, les travaux sur les métapopulations concernent majoritairement des espèces occupant des taches d'habitat situées au sein d'une matrice paysagère terrestre. Dans les modèles, cette matrice était initialement considérée comme étant homogène, limitant ainsi leur réalisme.

1.3 L'écologie du paysage

1.3.1 Origines de la discipline

La discipline qui s'est intéressée à l'**hétérogénéité du paysage** et qui a permis de la prendre en compte dans les études ultérieures est l'**écologie du paysage**. Le terme "écologie du paysage"⁴ a été utilisé pour la première fois en 1939 par Carl Troll, un géographe allemand. À la même époque, Arthur George Tansley, botaniste anglais, a défini un écosystème comme un système formé par un ensemble d'organismes vivants au sein d'un milieu donné ([Tansley, 1935](#)). L'écologie a initialement étudié les écosystèmes de façon verticale, en se focalisant sur la relation entre les facteurs abiotiques (le biotope) et les réponses biotiques (la biocénose) observées localement. [Tansley \(1939\)](#) et [Troll \(1966\)](#) ont ensuite défini indépendamment la notion d'écotope. Pour [Troll \(1966\)](#), il s'agit de la plus petite

3. [Haila \(2002\)](#) a fait remarquer qu'il n'y a qu'un seul paragraphe de l'ouvrage de [MacArthur et Wilson \(1967\)](#) sur la TBI qui compare des habitats subdivisés au sein d'une matrice terrestre à des îles pour extrapoler les conclusions de leur théorie.

4. Landschaftsökologie

unité au sein de laquelle on peut étudier les relations entre la biocénose et le biotope (Schreiber, 1990). On peut l'assimiler au paysage élémentaire de l'écologie du paysage, qui constitue l'expression spatiale élémentaire d'un écosystème. Néanmoins, les définitions accordées au mot paysage dans cette discipline sont multiples et diffèrent également de celles qu'on lui accorde dans d'autres disciplines (Encadré 2). Certains ont parfois distingué la géoécologie et la bioécologie comme des sous-branches de l'écologie du paysage, reflétant le penchant des géographes et des écologues pour l'étude des sphères abiotiques ou biotiques, respectivement (Schreiber, 1990). Cette distinction est oubliée aujourd'hui et c'est l'intégration de ces deux communautés scientifiques qui caractérise le mieux l'approche spatiale, horizontale, de l'étude des écosystèmes qu'a introduite l'écologie du paysage.

À partir de la fin des années 1960, l'écologie du paysage a principalement étudié les paysages du nord et de l'est de l'Europe, largement façonnés par les activités humaines. Une société d'écologie du paysage a été créée dès 1972 aux Pays-Bas, territoire fortement anthropisé (Schreiber, 1990). La discipline était influencée à cette époque par la géographie, l'aménagement du territoire et la sociologie (Risser *et al.*, 1984 ; Wiens, 1997). L'influence qu'exercent les humains sur les paysages était donc au cœur de ces travaux.

Puis, les échanges entre communautés scientifiques qui ont eu lieu lors de congrès suite à la création de l'International Association for Landscape Ecology (IALE) en 1982 en Tchécoslovaquie ont propagé cette discipline en Amérique du Nord et en Australie (Forman, 1990). Cela a renforcé sa dimension proprement écologique (Risser *et al.*, 1984 ; Wiens, 1997). En effet, outre Atlantique, les écologues du paysage étudiaient des processus écologiques se produisant à large échelle dans de grands espaces peu anthropisés (*wilderness*).

Encadré 2 : Le Paysage

Le **paysage** est une notion fortement polysémique. Alors que le dictionnaire Le Robert le définit premièrement comme la "*partie d'un pays que la nature présente à un observateur*", le Larousse le définit comme une "*étendue spatiale, naturelle ou transformée par l'homme, qui présente une certaine identité visuelle ou fonctionnelle*" ou comme une "*vue d'ensemble que l'on a d'un point donné*". Ces définitions rappellent le fait que le paysage caractérise tout d'abord la perception visuelle, esthétique, d'un observateur. En Occident, la notion de paysage apparaît d'ailleurs à la Renaissance en tant qu'objet artistique. À partir du XIX^e siècle, la notion de paysage est utilisée pour caractériser la façon dont les sociétés façonnent leur environnement et s'y adaptent localement (par le géographe Paul Vidal de la Blache notamment ; Donadieu (2012)). On peut dès lors délimiter spatialement des paysages sur une carte, passant de la perception humaine du paysage, depuis le sol, à la vision zénithale de la cartographie. C'est d'ailleurs l'analyse de photographies aériennes qui a conduit Carl Troll à définir l'écologie du paysage. Dans cette discipline, le paysage est l'unité spatiale élémentaire au sein de laquelle on peut étudier les écosystèmes (Schreiber, 1990). L'hétérogénéité spatiale du paysage forme des motifs dont la répétition caractérise le paysage et l'échelle à laquelle on le considère. L'étymologie du mot anglais *landscape*, issu de *land*, une portion délimitée de territoire, et de *scape*, une collection d'objets similaires, nous le rappelle (Burel et Baudry, 2003). Bien que les écologues

aient plus souvent considéré le paysage comme une entité spatiale que comme une perception visuelle, notons qu'en écologie comportementale, le paysage se définit en fonction de la façon dont les espèces étudiées perçoivent leur environnement (Nathan *et al.*, 2008). L'échelle à laquelle on l'étudie diffère donc d'une espèce à une autre (Baguette *et al.*, 2013). Les définitions du paysage sont multiples, et cela au sein même de l'écologie du paysage (Bastian, 2001). Des connexions sont aujourd'hui faites entre ces conceptions, pour étudier par exemple le lien entre la perception esthétique d'un paysage et les processus écologiques qui s'y déroulent (Sahraoui, 2016).

Notons enfin que les analyses d'écologie du paysage ont conduit à une simplification du paysage, étape nécessaire à toute modélisation. En effet, les paysages sont souvent caractérisés dans cette discipline par la structure spatiale de leur occupation du sol, alors que le relief ou la nature des sols sont plus fréquemment négligés. L'intérêt accordé à l'influence humaine sur les processus écologiques explique probablement cela dans la mesure où les activités humaines influencent essentiellement l'occupation du sol. Par ailleurs, les cartes d'occupation du sol se prêtent bien à l'identification des taches de ressources nettement différenciées qui constituent les briques élémentaires du paysage dans cette discipline.

1.3.2 Hétérogénéité du paysage

Paradoxalement, alors que l'écologie du paysage définit et délimite spatialement les paysages comme des zones au sein desquelles on observe des conditions géomorphologiques et climatiques homogènes (Baguette *et al.*, 2013), l'hétérogénéité spatiale des éléments qu'on y rencontre constitue son objet d'étude principal. Cela se retrouve dans l'ouvrage fondateur de Forman et Godron (1986) qui ont défini le paysage comme : "*a heterogeneous land area composed of a cluster of interacting ecosystems that is repeated in similar form throughout*". Le paysage y est conçu comme un espace hétérogène constitué de taches de ressources ayant une certaine configuration spatiale, formant des motifs ou patrons paysagers caractéristiques. Ces **taches** (ou *patches*) se situent au sein d'une matrice paysagère et sont parfois reliées par des éléments appelés corridors. Dans le modèle tache-matrice-corridor initial (Forman et Godron, 1986), la matrice paysagère est considérée comme un "fond de carte" homogène. Néanmoins, sa propre hétérogénéité spatiale a également été prise en compte par la suite. Notons que le caractère homogène ou hétérogène de la configuration des taches de ressources dans le paysage influence de nombreux processus écologiques mais dépend de l'échelle à laquelle on l'analyse. Pour cette raison, les notions d'hétérogénéité spatiale et d'échelle sont clés en écologie du paysage.

De nombreux concepts et théories d'écologie du paysage ont été énoncés (Burel et Baudry, 2003 ; Forman et Godron, 1986 ; Risser *et al.*, 1984 ; Turner, 1989). Ils ont entre autres donné naissance à des méthodes de description du paysage et de modélisation spatiale et contribué à la compréhension théorique des processus écologiques et à l'aménagement du territoire. En somme, on peut aujourd'hui définir l'**écologie du paysage** comme la discipline qui étudie les interactions entre (i) l'hétérogénéité spatiale de paysages plus ou moins anthropisés et (ii) les patrons et processus écologiques qu'on y observe. Dans cette discipline, la structure du paysage et son hétérogénéité sont décrites à plusieurs échelles spatiales en distinguant la composition et la configuration du paysage (Fahrig *et al.*, 2011) (Figure 4). De façon simplifiée, un paysage est essentiellement caractérisé par la structure de

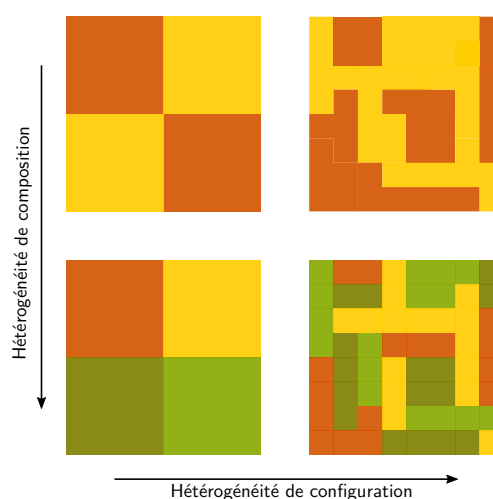
son occupation du sol ([Encadré 2](#)). Ainsi, la composition du paysage dépend de la diversité des types d'occupation du sol et de leurs proportions relatives tandis que la configuration du paysage dépend de l'arrangement spatial de ces types d'occupation du sol.

Lorsqu'on se focalise sur le patron spatial des habitats d'une espèce, la distinction entre composition et configuration se réduit à une distinction entre quantité d'habitat et configuration spatiale de l'habitat. [Hanski et Gilpin \(1991\)](#) écrivaient :

"Given their common concepts and terminology, the fusion of metapopulation studies and landscape ecology should make for an exciting scientific synthesis."

Trente ans et quelques milliers de publications plus tard, la fusion de ces approches est une réalité. Toutefois, les débats autour de l'influence de la structure spatiale des habitats sur la biodiversité sont loin d'être clos. La terminologie et les méthodes employées pour décrire la structure des habitats sont au cœur de ces débats car ils déterminent en grande partie l'influence relative sur la biodiversité attribuée à chacune des propriétés des habitats ([Didham et al., 2012](#) ; [Fahrig, 2003](#) ; [Fardila et al., 2017](#) ; [Hadley et Betts, 2016](#)). Nous nous focalisons dans cette thèse sur la **connectivité écologique des habitats**. Dans ce contexte, il est nécessaire d'indiquer ce qui distingue ou rapproche selon nous la connectivité des habitats de leur quantité ou de propriétés de leur configuration spatiale telles que la fragmentation.

FIGURE 4 – Le concept d'hétérogénéité du paysage en écologie du paysage, d'après [Fahrig et al. \(2011\)](#). Chaque couleur correspond à un type d'occupation du sol.



1.3.3 Quantité et configuration spatiale des habitats

Pour étudier l'influence relative de la quantité et de la configuration spatiale des habitats sur la biodiversité, il faut calculer des variables décrivant ces composantes pour essayer d'expliquer des réponses biologiques à partir d'elles. L'objectif est alors de décrire le patron spatial des habitats pour expliquer un patron spatial de biodiversité (richesse spécifique, abondance d'une espèce, diversité génétique, etc.) ou un processus écologique (interactions entre espèces, mouvements des espèces, etc.). Les mesures de la quantité et de la configuration des habitats sont au cœur des analyses de connectivité des habitats, mais également au cœur de débats scientifiques qu'il est nécessaire d'aborder ici pour justifier nos choix terminologiques et méthodologiques ultérieurs.

La fragmentation des habitats : un débat terminologique

L'étude de l'influence du patron spatial des habitats sur la biodiversité a été motivée par les pertes importantes de surfaces d'habitat et par leur fragmentation à l'échelle mondiale ([Haddad et al., 2015](#)). Dans ce contexte, la perte et la fragmentation des habitats sont des processus temporels et non pas des patrons spatiaux (Figure 5) ([Fahrig, 2003](#)). De plus, la fragmentation d'un habitat est un processus impliquant nécessairement une perte de surface et, pour certains auteurs, cette interdépendance

empêche de distinguer leurs effets respectifs (Didham *et al.*, 2012). Le terme "fragmentation" est donc ambigu puisque dans la littérature, il peut être utilisé pour décrire un processus ou un patron, et peut se confondre avec la perte d'habitat.

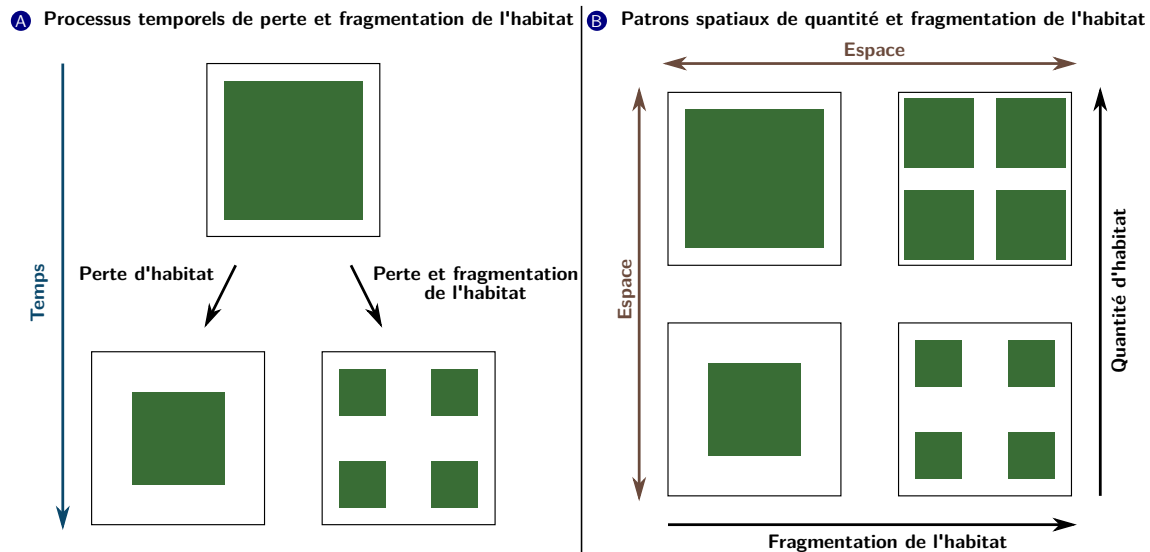


FIGURE 5 – Processus temporels (A) et patrons spatiaux (B) liés à la quantité et à la fragmentation de l'habitat.

Chaque paysage a la même superficie, délimitée par un cadre noir, et contient la même quantité d'habitat. (A) Considérée en tant que processus temporel, la fragmentation de l'habitat implique nécessairement une perte d'habitat (sinon cela impliquerait de "déplacer" les zones d'habitat dans le paysage pour les séparer en conservant leur surface).

(B) Si l'on considère la fragmentation comme un patron spatial, on peut observer des paysages ayant une quantité d'habitat équivalente et des degrés de fragmentation différents.

Ainsi, en 2003, Fahrig (2003) a défini la fragmentation au sens strict comme le degré de séparation des taches d'habitat à surface totale équivalente afin de pouvoir séparer les effets de cette variable de ceux de la quantité d'habitat. Selon son analyse, la plupart des études antérieures à 2003 montrent un effet prépondérant de la quantité d'habitat sur la biodiversité, contrastant avec l'effet très limité de la fragmentation au sens strict. Bien qu'il ait été cité des milliers de fois, cet article n'a pas suffi à faire changer le vocabulaire employé et la façon dont le patron spatial des habitats a été décrit. En 2017, Lenore Fahrig a reproduit la même analyse et a mis en évidence un effet positif de la fragmentation des habitats sur la biodiversité dans 76 % des cas (Fahrig, 2017). Elle y insistait à nouveau sur la distinction entre quantité d'habitat et fragmentation au sens strict, ce qui a provoqué de vives réactions de la part de la communauté de l'écologie du paysage (Fletcher Jr *et al.*, 2018). Néanmoins, décider du caractère positif ou négatif de l'effet de la fragmentation sur la biodiversité reste arbitraire. Par exemple, lorsqu'on étudie l'interaction entre un parasite et son hôte, si la fragmentation fait augmenter la prévalence du parasite, elle a un effet positif sur celui-ci et un effet sur son hôte qui est souvent négatif mais qui peut aussi être positif dans certains cas. Cela illustre le fait que le signe d'une relation statistique entre une variable paysagère et une réponse biologique permet difficilement de conclure de façon dichotomique quant à l'impact écologique de la structure du paysage sur les organismes étudiés. Cela constitue une difficulté supplémentaire dans la recherche d'une issue à ce débat (Miller-Rushing *et al.*, 2019). Notons que ces travaux sur la fragmentation des habitats ont l'intérêt de mettre en évidence l'importance de la définition stricte des variables caractérisant le patron des habitats, et en particulier sa configuration spatiale.

À quelle échelle spatiale décrire le patron des habitats ?

Outre la définition des variables utilisées pour caractériser le patron des habitats, l'échelle spatiale à laquelle elles sont calculées a une importance majeure sur les résultats et leur interprétation. Lorsque l'espèce étudiée forme des populations discrètes au sein de taches d'habitat subdivisées, l'échelle de la tache d'habitat est souvent pertinente⁵. Un tel focus sur la tache d'habitat reflète l'héritage de la théorie de la biogéographie insulaire et des modèles de métapopulations. À cette échelle, la surface et l'isolement d'une tache d'habitat vis-à-vis des autres sont censés déterminer respectivement le taux d'extinction et le taux de colonisation de cette tache, et expliquent donc certaines réponses biologiques (Moilanen et Nieminen, 2002). Néanmoins, les résultats obtenus en étudiant ces variables ont parfois été interprétés au-delà de leur signification. En effet, l'isolement d'une tache d'habitat mesuré en fonction de sa distance aux autres taches est parfois assimilé à une mesure de fragmentation. Or, la fragmentation au sens strict est une propriété du paysage dépendante du nombre de taches d'habitat et indépendante de la surface totale d'habitat. Elle ne peut donc pas se mesurer à l'échelle d'une tache. En outre, l'isolement des taches est étroitement corrélé à la quantité d'habitat dans le paysage (Fahrig, 2003). Ainsi, les études réalisées à l'échelle de taches dans un paysage unique ne peuvent pas évaluer les effets respectifs de la quantité d'habitat et de la fragmentation sur la biodiversité.

L'Habitat Amount Hypothesis : et si la configuration spatiale des habitats n'avait pas d'influence ?

En 2013, un article a remis en question la façon dont l'influence du patron spatial des habitats sur la biodiversité était étudiée (Fahrig, 2013). Lenore Fahrig y a formulé l'*Habitat Amount Hypothesis* (HAH), hypothèse simple selon laquelle la richesse spécifique observée au niveau de différents sites d'échantillonnage devrait augmenter avec la quantité totale d'habitat mesurée dans le paysage local à une distance donnée autour du site. Selon cette hypothèse, la richesse spécifique serait indépendante de la surface de la tache d'habitat à laquelle appartient le site d'échantillonnage. Bien qu'elle n'en ait pas démontré la validité dans cet article, Lenore Fahrig y a détaillé la façon dont cette hypothèse pourrait être testée. Si cette hypothèse était vraie, pour étudier la relation entre la richesse spécifique et le patron spatial des habitats, il serait selon elle inutile (i) de distinguer les effets de l'habitat à l'échelle de la tache et à celle du paysage, (ii) de distinguer l'effet de la taille d'une tache de celui de son isolement, (iii) de considérer la configuration spatiale des habitats indépendamment de la quantité d'habitat et (iv) de distinguer des taches d'habitat.

Cette hypothèse a par la suite été remise en cause (Hanski, 2015). Mais n'étant qu'une hypothèse (Fahrig, 2015), elle a également été testée à partir de données empiriques (Martin, 2018b). Certaines études l'ont confirmée (Archaux *et al.*, 2018 ; De Camargo *et al.*, 2018 ; Melo *et al.*, 2017) tandis que d'autres l'ont rejetée (Haddad *et al.*, 2017a ; Lindgren et Cousins, 2017 ; Torrenta et Villard, 2017). Des méta-analyses ont confirmé l'effet prédominant de la quantité totale d'habitat sur la biodiversité (Fahrig, 2017 ; Martin, 2018b), mais dans le même temps, de nombreuses études ont montré que la configuration spatiale des habitats avait une influence significative (Balkenhol *et al.*, 2013 ; Haddad *et al.*, 2017b,a ; Pérez-Rodríguez *et al.*, 2018). Ces résultats contrastés s'expliquent peut-être par le contexte (espèces, régions, patrons ou processus écologiques étudiés) propre à chacune de ces études. En l'état actuel du débat relatif à l'HAH et de celui relatif aux effets de la fragmentation des habitats au sens strict, l'existence d'une influence de la configuration spatiale des habitats sur la biodiversité

5. Bien qu'une conception du paysage comme un continuum sans distinction nette entre l'habitat et la matrice puisse être plus pertinente pour étudier certaines espèces (Fischer et Lindenmayer, 2006).

indépendamment de la quantité totale d’habitat ne fait pas l’unanimité.

Quantité et configuration spatiale des habitats sont-elles réellement indépendantes ?

Une question se pose néanmoins. Supposons que l’HAH soit vérifiée et que seule la quantité d’habitat autour d’un site détermine la richesse spécifique observée au niveau de ce site. Est-ce que cela signifierait que la configuration spatiale des habitats n’a pas d’influence sur la richesse spécifique observée dans un paysage ? On peut notamment se demander si la façon dont [Fahrig \(2013\)](#) a préconisé de mesurer la quantité d’habitat autour d’un site d’échantillonnage est réellement indépendante de la configuration spatiale des habitats. En effet, considérons des sites d’échantillonnage situés dans deux paysages contenant la même quantité totale d’habitat. Si l’HAH est vérifiée, la richesse spécifique est strictement proportionnelle à la quantité d’habitat mesurée dans un certain rayon autour de chaque site. Selon [Fahrig \(2013\)](#), indépendamment de la configuration des taches d’habitat de ces paysages on devrait alors y comptabiliser le même nombre d’espèces en cumulant l’ensemble des espèces observées dans leurs sites respectifs. Mais est-ce vraiment le cas ? Santiago Saura a récemment apporté une réponse à cette question ([Saura, 2021](#)).

La quantité d’habitat et leur configuration spatiale étant des propriétés paysagères, tester leur influence requiert de mesurer les réponses biologiques qu’elles sont censées influencer à l’échelle du paysage. Ainsi, [Saura \(2021\)](#) a étudié la distribution de valeurs de richesse spécifique (richesse moyenne et maximale, proportion de pixels ayant des richesses maximales) calculées dans des paysages en partant du postulat que l’HAH était vraie. Pour cela, il a simulé des paysages avec la même quantité d’habitat et des configurations spatiales plus ou moins fragmentées. Il a ensuite calculé la richesse spécifique R_i au niveau de chaque pixel i des taches d’habitat comme le produit d’une constante k et de la quantité d’habitat A_i mesurée à une distance d autour de ce pixel, tel que $R_i = k \times A_i$. Ses résultats montrent que la configuration spatiale des habitats a une influence importante sur la richesse spécifique, quand bien même l’HAH est vérifiée (Figure 6). Les valeurs de richesse spécifique sont globalement plus faibles dans les paysages dont les taches d’habitat sont les plus fragmentées, c’est-à-dire les plus nombreuses et les plus éloignées les unes des autres (Figure 6). Ce résultat est d’autant plus marqué que la distance d à laquelle on mesure la quantité d’habitat est faible⁶. En somme, alors que seule la quantité d’habitat influence la richesse spécifique de chaque site dans cette illustration, on arrive à la conclusion que la configuration spatiale influence la richesse spécifique à l’échelle du paysage. Cela montre qu’il est très difficile de rendre totalement indépendantes la quantité d’habitat et sa configuration spatiale dès lors qu’on étudie une réponse biologique à l’échelle du paysage, comme l’avaient déjà montré [Didham et al. \(2012\)](#).

Les mouvements entre taches d’habitat ont-ils été pris en compte dans ces débats ?

Dans les résultats de l’étude de [Saura \(2021\)](#), on note que la distance séparant les taches et la distance à laquelle la quantité d’habitat autour d’un site est mesurée influencent la richesse spécifique.

6. Ces résultats ont néanmoins pour limite qu’ils considèrent la richesse spécifique à l’échelle du paysage à partir de la distribution de valeurs de richesse spécifique locale (diversité α). Or, les différences de composition des assemblages d’espèces, c’est-à-dire la diversité β , déterminent la richesse spécifique totale à l’échelle du paysage (diversité γ). Pour cette raison, si les espèces vivant au niveau des lisières des habitats sont différentes des espèces vivant en leur cœur, le nombre total d’espèces pourrait être supérieur lorsqu’il y a à la fois des lisières et des zones cœur. Ainsi, pour des niveaux de fragmentation élevés, la richesse spécifique à l’échelle du paysage serait supérieure. Cet élément a été mentionné par [Fahrig \(2021\)](#), bien qu’elle ait également reconnu ses erreurs d’interprétation de l’HAH. Cela rappelle l’importance de considérer la diversité au niveau d’un site mais également entre les sites, point sur lequel nous reviendrons dans la section 3.1.1 lorsque nous étudierons la structure génétique des populations.

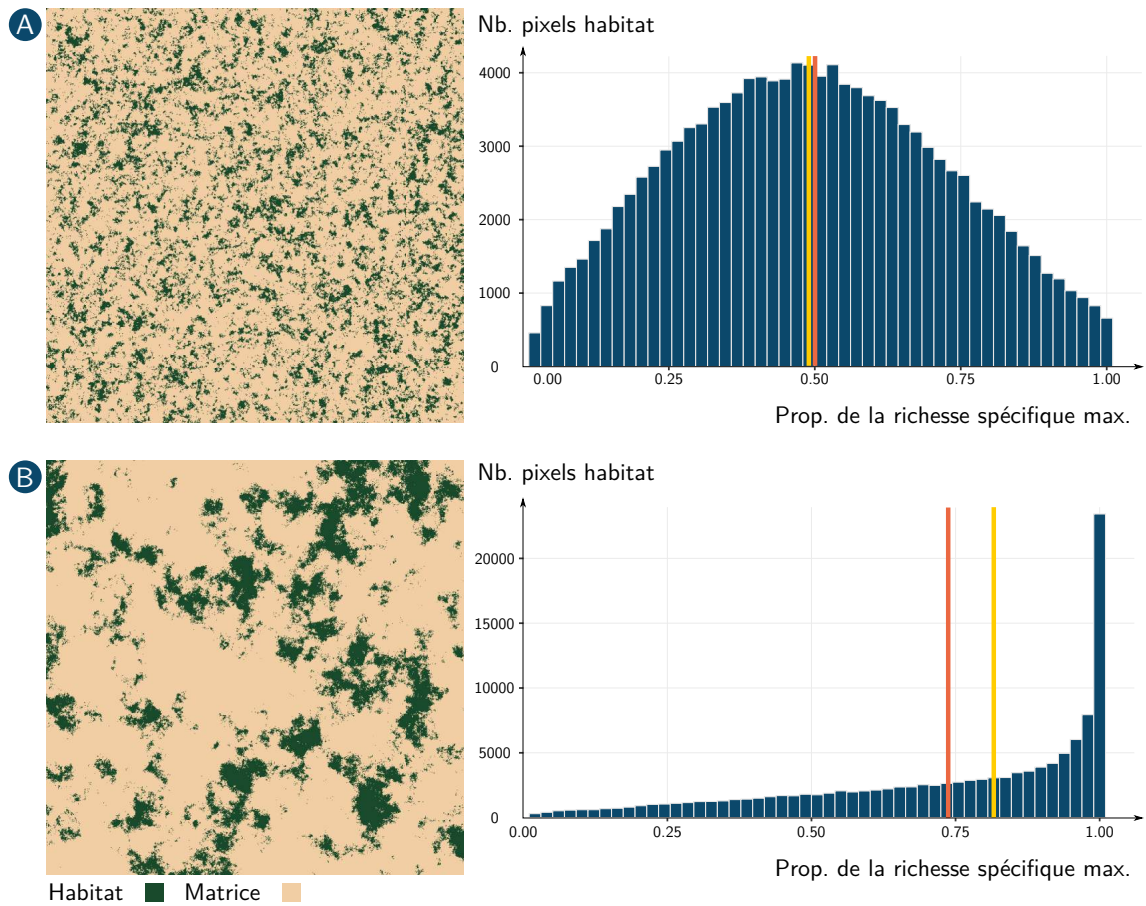


FIGURE 6 – Distribution de la richesse spécifique en fonction de la fragmentation du paysage selon l'*Habitat Amount Hypothesis*. Les paysages simulés A et B contiennent la même quantité d'habitat (20 % de 800×800 pixels) mais le paysage A est plus fragmenté que le paysage B. La richesse spécifique de chaque pixel d'habitat est calculée comme étant strictement proportionnelle au nombre de pixels d'habitat localisés autour du pixel considéré (cercle de 9 pixels de diamètre). Elle est exprimée en proportion de la richesse spécifique maximale observée dans les paysages A et B. Les traits verticaux orange et jaune représentent respectivement la moyenne et la médiane de la richesse spécifique. Cette figure rappelle les résultats obtenus par [Saura \(2021\)](#) en suivant un protocole similaire.

Les principaux déterminants de la richesse spécifique dans la TBI et les modèles de métapopulations sont d'une part l'extinction, et d'autre part la colonisation, qui implique le mouvement des individus. Pour tester l'HAH, le besoin de choisir un rayon d pour calculer la quantité d'habitat est une façon implicite de tenir compte des déplacements individuels et de leur échelle spatiale. Selon [Fahrig \(2013\)](#), le rayon d à considérer est celui qui maximise la relation entre la réponse biologique étudiée et la quantité d'habitat. Il correspond à ce que d'autres travaux ont nommé "l'échelle de l'effet" ou *scale of effect* ([Martin, 2018a](#) ; [Miguet et al., 2016](#)). La *scale of effect*, sujet d'étude récurrent de l'écologie du paysage ([Jackson et Fahrig, 2015a](#) ; [Moraga et al., 2019](#) ; [Stuber et Gruber, 2020](#)), dépend entre autres de la capacité des espèces étudiées à se disperser autour du site échantillonné ([Jackson et Fahrig, 2012](#) ; [Laroche et al., 2020](#)). Or, la capacité d'une espèce à se déplacer dans un paysage dépend de l'hétérogénéité de la matrice paysagère mais également de la configuration spatiale de l'habitat. En optimisant la *scale of effect* pour mesurer la quantité d'habitat, l'HAH prend ainsi indirectement en compte la configuration spatiale des habitats et les mouvements des individus. Les travaux de [Doak et al. \(1992\)](#) avaient déjà démontré l'importance de l'échelle spatiale dans l'étude de l'influence de la configuration des habitats et appelé à considérer les mouvements des individus de façon plus explicite.

1.4 Connectivité des habitats et biodiversité

1.4.1 Origines et évolutions de la définition de la connectivité des habitats

La connectivité du paysage : un concept basé sur les mouvements entre taches d'habitat

Le lien entre la structure du paysage et les mouvements des individus est étudié depuis de nombreuses années. En 1992, [Dunning et al. \(1992\)](#) ont souligné l'importance des phénomènes de complémentation et supplémentation paysagères et des effets sources-puits et de voisinage pour expliquer les patrons de biodiversité dans des paysages hétérogènes. Peu après, [Taylor et al. \(1993\)](#) ont remarqué que le rôle des mouvements des individus dans ces phénomènes n'était pas suffisamment pris en compte. Ils ont alors défini la notion de connectivité du paysage⁷ comme "le degré auquel le paysage facilite ou limite les mouvements d'organismes entre des taches de ressources". Compte tenu du contexte dans lequel cette notion a été définie, la connectivité du paysage a par la suite été utilisée comme une propriété générale d'un paysage. D'autres travaux ont complété cette définition initiale. Le modèle tache-matrice-corridor ([Forman, 1995](#)) a contribué à cela en représentant le paysage comme un ensemble de taches d'habitat reliées par des corridors au sein d'une matrice hétérogène. Le fait que la matrice exerce un effet hétérogène sur le déplacement a également été démontré ([Pither et Taylor, 1998](#) ; [Ricketts, 2001](#)).

De la connectivité structurelle du paysage à la connectivité fonctionnelle des habitats

En 2006, [Taylor et al. \(2006\)](#) ont rappelé la définition du concept de connectivité du paysage qu'ils avaient défini treize ans plus tôt et qui a souvent été mal utilisé par la suite selon eux. Ils insistaient sur la dépendance des mesures de connectivité du paysage aux espèces considérées, rendant le terme de connectivité des habitats plus approprié. Ce focus sur une espèce justifie également que l'on passe d'études de la connectivité dite structurelle à des études de la connectivité fonctionnelle. Dans le premier cas, la connectivité dépend essentiellement de la continuité physique des zones d'habitat et des éléments paysagers ayant une structure similaire aux habitats. Dans le second, la connectivité fonctionnelle intègre les capacités de dispersion propres à chaque espèce et tient compte du fait qu'une espèce peut traverser une zone ne correspondant pas à son habitat ([Calabrese et Fagan, 2004](#)).

De la connectivité des habitats à la disponibilité des habitats : une question de définition et d'échelle

Mesurer la connectivité des habitats pose là encore la question de la distinction entre la quantité d'habitat et sa configuration spatiale. Imaginons un paysage contenant une tache d'habitat de 10000 m² et un second paysage contenant deux taches d'habitat distinctes, de 5000 m² chacune. [Tischendorf et Fahrig \(2000\)](#) ont observé qu'en assimilant la connectivité au taux de migration entre les taches, on attribue une connectivité nulle au premier paysage qui ne contient qu'une tache, et une connectivité non nulle au second paysage car il en contient deux. Or, l'inverse serait plus en accord avec la notion de connectivité. En effet, si à surface totale équivalente les deux paysages comptent le même nombre d'individus, l'ensemble de ces individus ont une probabilité plus faible de se rencontrer en se déplaçant dans le second paysage que dans le premier. Ainsi, [Tischendorf et Fahrig \(2000\)](#) ont suggéré de considérer les mouvements ayant lieu à l'intérieur des taches pour ne pas attribuer une connectivité plus grande aux paysages les plus fragmentés. Ils ont indiqué comment le faire en mesurant la connectivité au niveau de zones de surface équivalente, sans se focaliser sur la délimitation des taches d'habitat. [Pascual-Hortal et Saura \(2006\)](#) et plus récemment [Spanowicz et Jaeger](#)

7. Bien que [Merriam \(1984\)](#), co-auteur de cet article, ait déjà utilisé le terme de "connectivité".

(2019) ont développé des métriques permettant de mesurer la connectivité des habitats sans abandonner la notion de taches tout en tenant compte de leur subdivision et des déplacements intra-taches.

Par ailleurs, mesurer la connectivité des habitats pour évaluer son rôle écologique implique de choisir une échelle spatiale. Selon [Tischendorf et Fahrig \(2000\)](#), la connectivité des habitats est une propriété du paysage. Tester son influence sur la biodiversité requiert donc de mesurer la connectivité et une réponse biologique conjointement et dans plusieurs paysages, et non pas de mesurer ces variables dans plusieurs taches d'un même paysage⁸. Ils ont mentionné une situation "hybride" permettant de tester cette influence à l'échelle de taches d'habitat en mesurant la connectivité au sein du paysage local qui les entoure. La différence entre ces approches repose essentiellement sur la définition du paysage et de son échelle spatiale. En effet, lorsque la connectivité des habitats est mesurée au niveau de l'ensemble des taches d'habitat d'une zone considérée comme étant le paysage, la délimitation de cette zone est arbitraire et ne dépend pas de paramètres écologiques. À l'inverse, le paysage peut être défini en fonction de la perception des individus d'une espèce depuis chaque point de la zone d'étude. À l'instar des approches d'écologie comportementale, le paysage équivaut alors à la zone qu'un individu peut parcourir durant sa vie ("*lifetime track*" sensu [Nathan et al. \(2008\)](#)) ([Baguette et al., 2013](#)). Mesurer la connectivité à l'échelle d'une tache d'habitat en considérant le paysage environnant revient à adopter cette seconde définition du paysage. Cela reflète directement les métriques utilisées dans les modèles de métapopulations pour prédire les taux d'immigration au niveau des taches ([Moilanen et Nieminen, 2002](#)). De telles mesures de connectivité peuvent être agrégées pour mesurer la connectivité à l'échelle d'un paysage plus large, rendant cette définition compatible avec des analyses répliquées dans plusieurs paysages.

Une des principales leçons de l'écologie du paysage est qu'il faut multiplier les échelles d'analyse pour comprendre l'influence des patrons paysagers sur les patrons et processus écologiques ([Holland et Yang, 2016](#) ; [Risser et al., 1984](#)). Bien que les individus perçoivent le paysage à une échelle spatiale limitée par leurs capacités de dispersion, il ne faut pas négliger l'analyse de la connectivité à une échelle plus large. L'approche multi-échelle se justifie par les différentes échelles spatiales et temporelles auxquelles se produisent les processus écologiques influencés par la connectivité ([Saura et Rubio, 2010](#)). En effet, au cours de sa vie, un individu se déplace quotidiennement pour se nourrir ou pour trouver des zones d'abris. Ces déplacements ont essentiellement lieu à l'intérieur du domaine vital de l'individu, c'est-à-dire dans la tache d'habitat qu'il occupe ou dans son voisinage proche. Par ailleurs, certaines espèces effectuent des migrations saisonnières. Ces déplacements peuvent être aussi bien régionaux qu'intercontinentaux et leur permettent en particulier de s'adapter à la disponibilité des ressources alimentaires à l'échelle d'une année. Enfin, au sein d'un réseau de taches d'habitat, un individu peut se déplacer vers une tache différente de celle dans laquelle il est né pour s'y reproduire, donnant lieu à un flux génétique. En écologie⁹, on qualifie ce déplacement de mouvement de **dispersion** ([Roncè, 2007](#)). Il a lieu à l'échelle d'une génération et est limité spatialement par les capacités de dispersion de l'individu. Cependant, le patrimoine génétique de cet individu pourra être transmis dans d'autres populations suite à la dispersion de ses descendants. Pour cette raison, le flux génétique est un processus ayant lieu à une échelle spatiale et temporelle supérieure à celle de la dispersion.

8. De façon identique à la fragmentation des habitats (voir précédemment).

9. En génétique des populations, le terme "migration" est utilisé pour qualifier ce type de mouvement, alors que le même terme est réservé aux déplacements périodiques tels que les migrations saisonnières en écologie.

L'organisation spatiale du réseau de taches d'habitat détermine la fréquence des déplacements s'effectuant d'une partie à l'autre du réseau en plusieurs générations. Certaines taches du réseau jouent le rôle de ponts ou *stepping-stones* en permettant les mouvements entre des ensembles de taches bien connectés (Saura *et al.*, 2014). Ces propriétés doivent donc être prises en compte en mesurant la connectivité des habitats liée à des déplacements à large échelle si l'objectif est d'étudier son influence sur des processus écologiques tels que les flux génétiques.

Compte tenu des éléments précédents, **nous définissons la connectivité des habitats d'une espèce comme la surface d'habitat atteignable lors du déplacement des individus au sein de leurs taches d'habitat et vers les autres taches. Elle se mesure à partir de chaque tache d'habitat de la zone d'étude à plusieurs échelles spatiales définies par les capacités de dispersion de l'espèce à l'échelle temporelle d'une ou plusieurs générations. Aux échelles spatiales les plus larges, la connectivité d'une tache d'habitat tient compte de son rôle dans la connectivité du réseau d'habitat dans son ensemble.** De telles mesures locales de connectivité peuvent être résumées, en calculant leur somme ou leur moyenne par exemple, pour évaluer la connectivité à l'échelle d'un paysage. Cette définition de la connectivité des habitats est donc compatible avec des analyses répliquées dans plusieurs paysages en considérant chacun d'eux comme une observation indépendante. Elle est également proche de celle introduite par Saura (2018) lors de sa présentation d'une hypothèse alternative à l'HAH tenant compte de façon plus explicite du mouvement des individus dans le paysage. Elle est également au cœur des travaux de Santiago Saura et ses collègues ayant donné naissance à des métriques de disponibilité de l'habitat (*habitat availability*). L'avantage de cette définition est qu'elle permet aussi de considérer l'hétérogénéité de la matrice paysagère dans le calcul des surfaces atteignables lors des déplacements¹⁰. Sa pertinence écologique a été validée empiriquement dans les études d'Archaux *et al.* (2018), Awade *et al.* (2012) et Mony *et al.* (2018), entre autres. Nous verrons dans la section 2.2.2 que les métriques basées sur cette définition de la connectivité ont l'avantage (i) de distinguer les situations dans lesquelles les taches sont subdivisées des situations dans lesquelles les zones d'habitat sont contiguës, et (ii) de contrôler le poids accordé à cette subdivision ainsi qu'aux surfaces des taches en elles-mêmes dans le calcul.

1.4.2 Influence de la connectivité des habitats sur la biodiversité

Attendus théoriques

La connectivité des habitats influence directement les processus écologiques de colonisation et de dispersion et les flux génétiques. Aussi, bien que la définition de la connectivité des habitats ait varié au cours du temps, de nombreux travaux théoriques et empiriques fournissent des éléments concernant son influence sur la biodiversité. Les attendus théoriques de cette influence découlent notamment de la TBI en ce qui concerne les communautés d'espèces, des modèles de métapopulations concernant la dynamique des populations d'une espèce et de la **génétique des populations** pour la structure génétique intra-spécifique. En effet, selon ces théories, la connectivité des habitats devrait favoriser la diversité spécifique des communautés (Brown et Kodric-Brown, 1977), la présence et l'abondance

10. Certes, cette définition ne distingue pas la quantité d'habitat de sa configuration spatiale mais nous avons pu voir qu'une telle distinction était difficile à faire à une échelle supérieure à celle de la tache, seule échelle à laquelle elle aurait un intérêt, compte tenu de l'interdépendance de ces deux variables.

des populations d'une espèce (Fahrig et Paloheimo, 1988 ; Hanski, 1989), et maintenir la diversité génétique de ces populations (Frankham, 2006 ; Gilpin, 1991).

Preuves empiriques

Ces attendus théoriques ont été testés en raison de l'intérêt scientifique porté à la connectivité (Crooks et Sanjayan, 2006 ; Fletcher *et al.*, 2016) et grâce à un meilleur accès aux données géographiques, à l'amélioration des méthodes de cartographie des habitats et de la matrice paysagère et à la mise en place de protocoles expérimentaux dédiés (Haddad *et al.*, 2015 ; Tewksbury *et al.*, 2002). Nous disposons donc aujourd'hui de plusieurs synthèses et méta-analyses concernant l'influence de la connectivité sur la biodiversité (Baguette *et al.*, 2013 ; Kool *et al.*, 2013 ; Fletcher *et al.*, 2016 ; Resasco, 2019). 46 % des travaux sur la connectivité publiés entre 2005 et 2015 ont testé ses effets sur des réponses biologiques (Fletcher *et al.*, 2016). Leurs résultats étaient significatifs dans 91 % des cas¹¹. Bien que ces auteurs n'aient pas indiqué la proportion exacte des résultats significatifs qui étaient en accord avec les hypothèses formulées par les chercheurs les ayant obtenus, les études empiriques ont souvent confirmé les attendus théoriques. Par exemple, la connectivité mesurée au niveau des mares d'un réseau de mares avait un effet positif sur la richesse spécifique des communautés d'amphibiens occupant ces habitats (Ribeiro *et al.*, 2011). Mony *et al.* (2018) ont également montré que les surfaces forestières accessibles lors des déplacements autour des forêts échantillonnées influençaient les communautés d'espèces végétales forestières dispersées par les animaux. En faisant l'hypothèse que les distances de dispersion étaient courtes, plus les surfaces forestières accessibles étaient grandes, plus la richesse spécifique de ces espèces ainsi que le couvert forestier qu'elles représentaient étaient importants¹². Diekötter *et al.* (2008) ont également montré que la composition des communautés d'espèces d'insectes occupant des taches d'habitat plus ou moins connectées d'un paysage agricole dépendait des capacités de dispersion contrastées de ces espèces. Au niveau de l'espèce, des observations ont confirmé que la connectivité des habitats favorisait les mouvements des individus entre les taches (Gilbert-Norton *et al.*, 2010 ; Tewksbury *et al.*, 2002). De plus, la synthèse de Bowne et Bowers (2004) a montré qu'une augmentation de la fréquence des mouvements entre taches faisait généralement augmenter la taille des populations. L'influence positive de la connectivité sur l'abondance d'une espèce prédatrice peut également faire augmenter la pression de prédation sur sa proie (voir par exemple Koh *et al.* (2013)). Enfin, en accord avec la théorie, des effets positifs de la connectivité sur la diversité génétique mesurée dans une population (voir Keyghobadi *et al.* (2005)) et négatifs sur les différences génétiques entre populations (voir Cushman et Lewis (2010)) ont été observés. Ces derniers résultats sont issus des travaux de **génétique du paysage** (*landscape genetics*), une discipline récente dans laquelle s'inscrit le travail réalisé durant cette thèse et qui fera l'objet de la section 3.3.

À cet ensemble d'effets de la connectivité des habitats que l'on pourrait qualifier de "positifs" pour les espèces et la diversité des communautés étudiées, s'ajoutent des effets "négatifs" (Taylor *et al.*, 2006). En effet, une connectivité importante peut favoriser la propagation d'espèces exotiques envahissantes (Bullock *et al.*, 2018), la dispersion de maladies (Jules *et al.*, 2002) ou encore freiner l'adaptation génétique de certaines populations à leur environnement local en homogénéisant leurs patrimoines génétiques (Crispo *et al.*, 2011 ; Lenormand, 2002). Pour cette raison, certains auteurs estiment que des niveaux de connectivité intermédiaires sont les plus bénéfiques à long terme pour les

11. On ne peut pas exclure l'éventualité qu'un biais lié à la publication plus aisée de résultats significatifs soit la cause d'un chiffre si élevé.

12. Le premier effet étant surtout observé pour les espèces épi-zoochores et le second pour les endo-zoochores

populations (Minor et Urban, 2008). Par ailleurs, les effets positifs de la connectivité pour la survie d'une espèce dépassent souvent ses effets négatifs (McCallum et Dobson, 2002).

1.4.3 Implications pour la conservation de la biodiversité

L'influence de la connectivité des habitats sur les patrons de biodiversité et les processus écologiques a été comprise progressivement et continue de faire l'objet de travaux. Peu à peu, les travaux de **biologie de la conservation** ont intégré les résultats de l'écologie scientifique concernant la connectivité pour les traduire en préconisations pour la conservation de la biodiversité. Ces dernières décennies, les politiques de conservation se sont ensuite inspirées de ces préconisations pour adapter leurs objectifs et leurs plans d'action. La notion de connectivité des habitats a ainsi influencé tour à tour l'écologie scientifique, la biologie de la conservation et les politiques de conservation de la biodiversité¹³. Cette thèse porte sur des méthodes d'analyse de la connectivité des habitats pouvant contribuer d'une part à des études à l'interface entre écologie scientifique et biologie de la conservation, et d'autre part à la mise en œuvre opérationnelle des politiques de conservation de la connectivité (Figure 7). Pour cette raison, nous revenons dans la section suivante sur l'historique de la prise en compte de la connectivité des habitats dans ces différents domaines pour comprendre les enjeux actuels et en quoi cette thèse pourrait y répondre.

Les premières zones protégées

Les sites sacrés et sanctuaires sont les plus anciens exemples de zones protégées, dans ce cas précis pour des motifs religieux ou spirituels (Ramakrishnan, 1996 ; Verschuuren *et al.*, 2010). Ce sont des raisons essentiellement économiques qui ont motivé ensuite la protection des écosystèmes pour les ressources qu'ils fournissent à la société. En France par exemple, Louis XIV a signé en 1669 une ordonnance royale rédigée par Colbert et encadrant la gestion des forêts françaises afin de pérenniser les constructions de navires militaires et le gibier des chasses royales (Ford, 2004). Cette ordonnance n'a pas interdit l'exploitation de la forêt, elle l'a rationalisée. Les travaux scientifiques sur la sylviculture qu'elle a initiés ont ensuite été repris par l'administration des Eaux et Forêts pour optimiser la production forestière. Mais les conséquences de cette exploitation des forêts ont inquiété la société. Dans les années 1850, les peintres de l'École de Barbizon se sont alarmés de l'impact qu'avaient les coupes de feuillus et les plantations de résineux dans la forêt de Fontainebleau sur les paysages qu'ils peignaient. Dès 1853, des parcelles forestières ont été protégées de toute exploitation et en 1861, un décret impérial a créé la série artistique de Fontainebleau, première réserve naturelle de l'histoire. À la même époque, Victor-Emmanuel II, futur roi d'Italie, a créé en 1856 une réserve royale de chasse pour protéger les effectifs de bouquetins du massif du Gran Paradiso, menacés par la démocratisation des armes à feu parmi les chasseurs.

Ces exemples illustrent le fait que la rationalisation et les excès de l'exploitation des écosystèmes sont en grande partie à l'origine des premières mesures de conservation. Outre-Atlantique, les menaces pesant sur la *wilderness* des grands espaces ont motivé leur protection aux États-Unis. En 1864, le *Yosemite Valley Grant Act* est signé par le président Lincoln pour protéger la vallée de Yosemite en

13. À l'instar d'autres facteurs tels que le changement climatique, la pollution ou les espèces exotiques envahissantes, entre autres, dont les écologues ont révélé l'impact sur la biodiversité dans le cadre des changements globaux actuels, avant qu'ils ne soient mis à l'agenda des politiques publiques après un certain délai de latence.

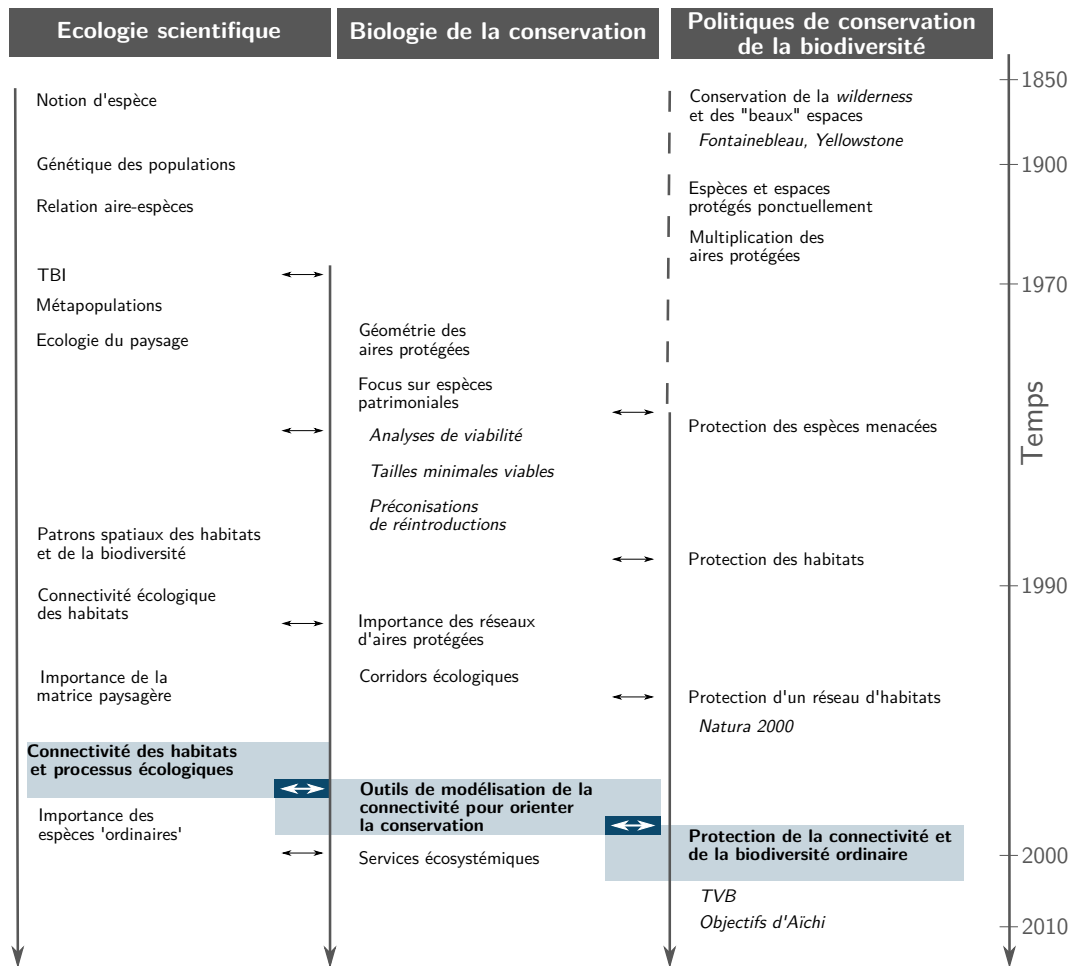


FIGURE 7 – Positionnement disciplinaire de la thèse. L'évolution parallèle de l'écologie scientifique, de la biologie de la conservation et des politiques de conservation de la biodiversité est représentée schématiquement pour mettre en évidence (i) la temporalité des échanges entre ces disciplines ayant mené à la considération de la connectivité écologique des habitats pour conserver la biodiversité ordinaire et (ii) la façon dont cette thèse contribue aux échanges disciplinaires destinés *in fine* à améliorer la mise en œuvre des politiques de conservation. Les flèches blanches sur fond bleu foncé indiquent les échanges auxquels cette thèse contribue.

Californie. Puis, en 1872, le premier parc national est créé à Yellowstone¹⁴. Ces premières mesures de conservation consistaient à "mettre sous cloche" de grands espaces pour protéger leur caractère grandiose et l'esthétique des paysages qu'ils offrent au regard. En plus de ces raisons d'ordre sociétal¹⁵, les premières mesures de conservation avaient souvent un intérêt économique. Par exemple, les premiers parcs nationaux canadiens ont été créés à l'initiative des compagnies de chemins de fer pour protéger les paysages traversés par leurs lignes et ainsi accroître leurs revenus suscités par le tourisme (Leader-Williams *et al.*, 1990).

La protection des espèces et espaces remarquables

Le progrès des connaissances naturalistes et la pression grandissante exercée par l'être humain sur les écosystèmes ont fait changer les raisons pour lesquelles les politiques environnementales ont été mises en œuvre ainsi que leurs moyens d'application. Les premières mesures de protection des espèces ont vu le jour au début du XX^e siècle. La "Convention pour la protection des oiseaux utiles

14. La vallée de Yosemite est devenue un parc national en 1890 suite aux efforts de John Muir notamment ; l'État de Californie n'assurant pas suffisamment la conservation du site.

15. Le Parc National de Yellowstone a été qualifié initialement de "*public park or pleasuring ground for the benefit and enjoyment of the people*".

à l'agriculture" signée par douze États européens à Paris en 1902 en est un exemple. Tout au long de ce siècle, de nouveaux outils réglementaires sont apparus pour encadrer la protection de certaines espèces menacées et des sites "remarquables" dans lesquels elles subsistaient. En plus des parcs nationaux, différents types d'aires protégées ont été créés, tels que les réserves naturelles ou les arrêtés de protection de biotope en France par exemple. Néanmoins, bien que les motivations à l'origine de la création des aires protégées aient peu à peu pris en considération des enjeux écologiques, ce sont le plus souvent des critères socio-économiques qui ont déterminé leur localisation spatiale (Chassé *et al.*, 2021 ; Simberloff, 1988).

Les apports de la biologie de la conservation

Dans la seconde moitié du XX^e siècle, la **biologie de la conservation** a constitué une interface entre l'écologie scientifique et les politiques de conservation (Soulé, 1985, 1986). Cette nouvelle discipline scientifique a contribué à l'élaboration de "règles" permettant d'optimiser et rationaliser la création des aires protégées et les mesures de conservation de manière générale. Des revues scientifiques lui ont alors été dédiées (*Biological Conservation* en 1968 et *Conservation Biology* en 1987). La biologie de la conservation s'est tout d'abord basée sur la théorie de la biogéographie insulaire. Diamond (1975) a directement appliqué cette théorie pour suggérer des "principes géométriques" de répartition spatiale optimale des aires protégées. Selon ces principes, il faudrait préférer la protection d'une grande réserve continue plutôt que plusieurs réserves, à superficie totale équivalente. Il faudrait également minimiser la distance entre les réserves. Ces principes ont rapidement été critiqués (Simberloff et Abele, 1976) et ont donné lieu au débat *SLOSS* (*Single Large Or Several Small*), encore en cours aujourd'hui (Fahrig, 2020). Ils ont néanmoins été appliqués assez largement suite à leur intégration à la stratégie mondiale de conservation de l'Union Internationale pour la Conservation de la Nature (UICN) en 1980. À l'époque, les concepts scientifiques servaient essentiellement à maximiser la probabilité de survie des espèces tout en allouant une surface minimale à la conservation (Sarkar *et al.*, 2006). Les aires protégées étaient encore considérées comme des îles au sein d'un océan anthropisé et hostile.

La conservation a ainsi permis de protéger des patrons de biodiversité et a également contribué aux inventaires de biodiversité. Ces efforts ont permis d'évaluer le succès des mesures de conservation. Les organismes en charge de la gestion des aires protégées ont aujourd'hui encore un rôle clé dans ce suivi. Les données qu'ils acquièrent orientent la gestion, tout en alimentant les travaux d'écologie scientifique et de biologie de la conservation. Par ailleurs, les échanges entre scientifiques et gestionnaires contribuent à l'application concrète des concepts de biologie de la conservation. On peut noter qu'aujourd'hui la plupart des aires protégées françaises comportent un Conseil Scientifique et que certaines aires protégées mènent leurs propres programmes de recherche, à l'instar de Parks Canada ou U.S National Parks Services en Amérique du Nord. Les aires protégées elles-mêmes contribuent ainsi à la forte dimension scientifique des stratégies de conservation.

Au cours du temps, les préconisations formulées pour orienter la conservation se sont complexifiées en intégrant l'ensemble des travaux sur les métapopulations, la génétique des populations et la TBI (Simberloff, 1988 ; Soulé, 1985)¹⁶. Cette intégration a été cruciale pour l'identification des processus

16. On peut d'ailleurs remarquer que les modèles de métapopulations intègrent à la fois la TBI et des concepts de génétique des populations développés par Sewal Wright ou Motoo Kimura, entre autres. Dans ses articles considérés comme étant les fondements de ces modèles (Levins, 1969, 1970), Richard Levins cite ces deux généticiens et s'inspire

démographiques et génétiques responsables de ce que Gilpin et Soulé (1986) ont nommé la spirale d'extinction (*extinction vortex*). Lorsqu'une espèce ou une population est engagée dans une spirale d'extinction, la diminution de ses effectifs perturbe sa reproduction, ce qui fait diminuer davantage ses effectifs. Cela conduit à la diminution de sa diversité génétique, puis à l'accumulation de variations génétiques délétères qui à leur tour réduisent les effectifs et le succès de la reproduction. La consanguinité génétique a également une influence négative sur les populations dans ce cas (Frankel et Soulé, 1986 ; Frankham, 1996a ; Nieminen *et al.*, 2001 ; Saccheri *et al.*, 1998 ; Soulé et Simberloff, 1986). Enfin, ces processus démographiques et génétiques qui s'auto-amplifient conduisent à l'extinction.

En somme, le maintien des patrons de biodiversité dépend d'un ensemble de processus écologiques. Les travaux de biologie de la conservation ont caractérisé les paramètres nécessaires à la réalisation de ces processus ainsi que les solutions concrètes permettant d'éviter l'extinction d'une population lorsqu'ils n'ont plus lieu. Pour cela, la discipline a pris une dimension relativement "technique" (Godet, 2010). Par exemple, la taille minimale des populations viables ou le temps qui sépare une population de son extinction sont calculés à partir d'analyses de viabilité (Boyce, 1992). Des modèles basés sur la génétique des populations identifient les effectifs à maintenir pour limiter les risques de consanguinité (Franklin, 1980). Ils se basent sur le concept de taille de population "efficace" (N_e) développé pour étudier le rythme et l'ampleur des variations de diversité génétique (cf. section 3.1.2). Selon ces travaux, une population doit compter au moins 500 individus pour être viable génétiquement (Franklin, 1980), bien que ce chiffre suscite des débats (Frankham, 1996a ; Frankham *et al.*, 2014). Lorsque tous ces paramètres ne sont pas réunis, des mesures de conservation des espèces en captivité (*ex-situ*) ou des réintroductions ponctuelles (*in-situ*) sont envisagées (Frankham *et al.*, 2017 ; Ralls et Ballou, 1986). Une telle situation a par exemple motivé la réintroduction d'ours bruns slovènes dans les Pyrénées françaises à partir de 1996.

En parallèle de ces approches menant à des mesures de conservation ponctuelles, l'échelle spatiale à laquelle les mesures de conservation *in-situ* sont pensées s'est élargie. Les travaux récents visant à quantifier la surface totale et la connectivité du réseau mondial des aires protégées (Saura *et al.*, 2017, 2018, 2019) illustrent le fait que ce changement d'échelle permet aujourd'hui d'évaluer les efforts de conservation à une échelle continentale voire mondiale. Par ailleurs, la conservation a peu à peu pris en compte l'influence des espaces non protégés et de l'hétérogénéité de leurs paysages sur le succès des mesures mises en œuvre. La protection de corridors entre les aires protégées avait déjà été préconisée pour favoriser les processus écologiques liés au déplacement des individus (Diamond, 1975). Néanmoins, l'absence de preuves empiriques de l'efficacité des corridors a initialement suscité des interrogations (Hobbs, 1992). Les travaux d'écologie du paysage ont ensuite amélioré la compréhension de leur rôle (cf. section 1.4.1) et attesté de leur efficacité du point de vue de la conservation (Resasco, 2019 ; Tewksbury *et al.*, 2002). Des guides méthodologiques ont alors été élaborés pour guider les aménageurs du territoire et gestionnaires dans leur identification et leur gestion (Bennett, 1999).

Bien que les mesures de conservation aient tenu compte de processus écologiques tels que le déplacement des espèces à une large échelle et de leur distribution spatiale dans le paysage, elles sont longtemps restées focalisées sur les aires protégées et les espèces remarquables, patrimoniales ou me-

largement de leurs découvertes sur le rôle des tailles de populations et des réseaux de dispersion de type *stepping-stones* sur la structure génétique des populations. Le lien entre les travaux de R. Levins et la TBI est pourtant mentionné plus fréquemment comme étant à l'origine des modèles de métapopulations.

nacées (Gaston et Fuller, 2008). Or, le fonctionnement des écosystèmes dépend en grande partie des espèces communes, qui sont les plus abondantes, et que l'on rencontre également dans des paysages considérés comme ordinaires (Godet, 2010). Un léger déclin de leur abondance perturbe de façon disproportionnée ce fonctionnement (Gaston et Fuller, 2008). Néanmoins, le rôle de la biodiversité ordinaire a été reconnu assez tardivement dans l'histoire de la biologie de la conservation (Godet, 2010 ; Hanski, 2005). À partir du début des années 2000, la quantification des services écosystémiques rendus par la biodiversité aux sociétés humaines (Costanza *et al.*, 1997), critiquée à de nombreux égards (Maris, 2014), a eu le mérite de faire reconnaître l'importance de la biodiversité ordinaire pour le fonctionnement des écosystèmes (Godet, 2010). La protection de la biodiversité ordinaire implique de penser la conservation à des échelles larges et dans le cadre de l'aménagement des territoires. Le caractère pluri-disciplinaire de la biologie de la conservation mais aussi de l'écologie du paysage ont favorisé ces dernières décennies la formulation de préconisations de conservation tenant compte de l'ensemble des éléments du paysage et insistant particulièrement sur la connectivité des habitats qu'on y trouve (Burel et Baudry, 2003).

En résumé, l'évolution des mesures mises en œuvre ou préconisées pour la conservation de la biodiversité s'est caractérisée par un passage :

- de préoccupations sociétales ou économiques initialement (anthropo-centrées), à des préoccupations écologiques par la suite (éco-centrées) ;
- de décisions de conservation ponctuelles (réserves naturelles, parcs nationaux) à des préconisations issues de travaux scientifiques et concernant de larges échelles spatiales (réseau d'aires protégées connectées par des corridors) ;
- de la protection de patrons de biodiversité (richesse spécifique, diversité génétique) à la protection de processus écologiques (colonisation, flux génétique) ;
- de la protection d'espèces remarquables à la protection de la biodiversité ordinaire à l'échelle des paysages.

La prise en compte de la connectivité des habitats est en quelque sorte l'aboutissement de ces évolutions.

Les politiques de conservation de la connectivité écologique des habitats

Ces dernières décennies, les évolutions des politiques de conservation de la biodiversité ont été à l'image de celles observées en écologie scientifique et en biologie de la conservation. En particulier, ces politiques ont pris en compte le rôle de la connectivité des habitats pour la conservation de l'ensemble des composantes de la biodiversité à l'échelle des paysages. Des objectifs chiffrés, des textes réglementaires et des plans d'action dédiés à la connectivité ont été déclinés aux niveaux local, national et international.

Une étape importante dans l'intégration de la connectivité des habitats aux politiques environnementales est l'adoption de la Directive Habitats par les États membres de l'Union Européenne en 1992 (CE, 1992). Elle a donné naissance à Natura 2000, un réseau européen d'aires protégées basé sur la distribution des habitats et des espèces d'intérêt communautaire inclus dans les annexes de

la Directive Habitats¹⁷. Par ailleurs, l'article 10 de la Directive Habitat stipule explicitement (CE, 1992) :

"les États membres s'efforcent d'encourager la gestion d'éléments du paysage qui revêtent une importance majeure pour la faune et la flore sauvages. Ces éléments sont ceux qui, de par leur structure linéaire et continue (tels que les rivières avec leurs berges ou les systèmes traditionnels de délimitation des champs) ou leur rôle de relais (tels que les étangs ou les petits bois), sont essentiels à la migration, à la distribution géographique et à l'échange génétique d'espèces sauvages."

Cet article incite donc les États membres à intégrer dans leur législation la protection de la connectivité d'un réseau d'habitat. Le réseau Natura 2000 y contribue mais certains états ont parallèlement mis en œuvre des politiques spécifiques à l'échelle nationale.

En France par exemple, le constat de l'importance de la biodiversité ordinaire dans les paysages est à l'origine de la politique de Trame Verte et Bleue (TVB) suite au Grenelle de l'environnement de 2007¹⁸. Cette politique est encadrée par les lois Grenelle I et II¹⁹ qui imposent aux services de l'État et aux collectivités territoriales de cartographier les espaces contribuant à la connectivité écologique des milieux terrestres et aquatiques, à savoir les "réservoirs de biodiversité" et les "corridors écologiques". Les trames cartographiées sont intégrées aux documents de planification territoriale et d'urbanisme à l'échelle régionale (SRADDET²⁰) et locale (SCoT²¹, PLU(I)²²). Ces documents de planification doivent prévenir l'impact de l'aménagement du territoire sur les zones constituant la TVB. Par ailleurs, depuis la Loi pour la reconquête de la biodiversité, de la nature et des paysages du 8 août 2016, les communes françaises peuvent désigner des espaces de connectivité écologique dans les PLU. Néanmoins, ces outils réglementaires ne sont pas opposables, ce qui signifie que les communes peuvent y déroger et autoriser des aménagements impactant potentiellement les espaces de connectivité. Des objectifs de conservation de la connectivité des habitats ont également été intégrés aux plans d'action nationaux, tels que la Stratégie Nationale pour la Biodiversité 2011-2012 (MEDDE, 2011) en France.

Au niveau international, des objectifs de connectivité écologique sont fixés lors des conférences des parties (CoP) à la Convention sur la Diversité Biologique de 2010 (CBD, 2010). Par exemple, l'objectif 11 des objectifs d'Aichi constituant le Plan Stratégique pour la Biodiversité 2011-2020 prévoyait que 17 % des aires terrestres et 10 % des zones marines côtières forment un réseau d'aires protégées inter-connectées à horizon 2020. Cet objectif n'a pas été atteint par tous les États, malgré des efforts en termes de création d'aires protégées durant la période (Carrasco *et al.*, 2021 ; Saura *et al.*, 2019).

17. Le réseau Natura 2000 est composé de Zones Spéciales de Conservation (ZSC) et de Zones de Protection Spéciale (ZPS). Les premières incluent les habitats d'intérêt communautaire définis par la Directive Habitats tandis que les secondes sont définies par la Directive Oiseaux de 1979.

18. Cette politique ayant également pour objectif de protéger la qualité écologique des masses d'eau, la Directive Cadre sur l'Eau a contribué à la prise de conscience ayant favorisé sa mise en œuvre.

19. Et conséquemment par les articles L.371-1 II et R.371-19 III du code de l'environnement

20. Schémas Régionaux d'Aménagement, de Développement Durable et d'Égalité des Territoires, qui intègrent les anciens Schémas Régionaux de Cohérence Écologique (SRCE) depuis 2016 et la loi NOTRE.

21. Schémas de Cohérence Territoriale

22. Plans Locaux d'Urbanisme (Intercommunaux).

Enjeux actuels associés à la connectivité des habitats

Malgré la multiplication des études scientifiques, des préconisations de gestion et des politiques de conservation, la biodiversité poursuit son érosion (Díaz *et al.*, 2019). Au-delà du fait que les principales causes sont d'ordre socio-économique et difficiles à infléchir, certains auteurs considèrent que la pratique actuelle de la biologie de la conservation limite l'identification de solutions à ce problème. Ils reprochent notamment aux scientifiques d'insister de façon contre-productive sur l'impact négatif des activités humaines et aux gestionnaires et décideurs de se focaliser essentiellement sur les aires protégées aux dépens des enjeux présents dans les zones les plus anthropisées (Godet et Devictor, 2018). Certains estiment également que la conservation de la biodiversité ne peut pas être efficace si ce sont les intérêts de l'être humain, et non la biodiversité en tant que telle, qui sont la motivation finale (Doak *et al.*, 2015). Par ailleurs, d'autres considèrent que les résultats et outils permettant d'orienter efficacement les politiques de conservation sont déjà disponibles et que le principal enjeu est celui de leur utilisation pertinente par les gestionnaires et décideurs (Chassé *et al.*, 2020 ; Garnett et Lindenmayer, 2011 ; Mair *et al.*, 2018).

Cependant, les mesures de conservation se sont révélées efficaces à de nombreuses reprises et la biologie de la conservation a su adopter de nouvelles approches (Godet et Devictor, 2018), remettant en question le changement radical de paradigme appelé par certains. Ainsi, les enjeux scientifiques actuels consistent à adapter les préconisations à la conservation de l'ensemble des composantes de la biodiversité et des processus responsables de leur dynamique, y compris dans des paysages considérés comme ordinaires (Godet, 2010). Par exemple, pour atteindre les objectifs visés par les mesures de conservation de la connectivité des habitats, il est nécessaire de comprendre son influence précise dans une large gamme de contextes sur des réponses biologiques telles que la **structure génétique des populations**. En effet, bien que le maintien de la variabilité génétique soit au cœur des programmes de conservation, les études mettant en lien des réponses génétiques avec la connectivité des habitats sont encore récentes. Dans ce contexte, cette thèse a pour objectif de développer des méthodes facilitant la mise en œuvre de ces études tout en apportant de nouveaux résultats sur le sujet (Figure 7).

En ce qui concerne le transfert de connaissances de la sphère scientifique à la sphère opérationnelle, il y a d'une part un besoin de résultats scientifiques plus directement interprétables par les acteurs de la conservation (Keller *et al.*, 2015). D'autre part, ces acteurs ont besoin d'outils validés scientifiquement leur permettant de localiser dans leurs territoires les mesures contribuant à la connectivité des habitats. Ainsi, cette thèse a également pour objectif de valider empiriquement des méthodes trouvant des applications opérationnelles pour la conservation de la connectivité des habitats (Figure 7).

Que cela soit pour approfondir les connaissances scientifiques ou pour localiser les mesures de conservation à mettre en œuvre, la modélisation spatiale joue un rôle clé. Dans la section suivante, nous présentons les méthodes de modélisation spatiale de la connectivité des habitats. Nous nous focalisons sur les **graphes paysagers**, qui peuvent être utilisés à la fois pour des études contribuant à la compréhension des processus écologiques et dans un contexte opérationnel.

2 Les graphes paysagers, outils de modélisation de la connectivité des habitats

2.1 Modélisation spatiale de la connectivité des habitats

Évaluer la connectivité des habitats implique de considérer le paysage dans son ensemble pour cartographier précisément les taches d'habitat et les zones contribuant aux déplacements entre ces taches. Mener une telle évaluation à grande échelle par le biais d'un inventaire systématique sur le terrain n'est pas envisageable compte tenu du coût et du temps à investir. Pour cette raison, la modélisation spatiale revêt un grand intérêt. Elle s'appuie sur les Systèmes d'Information Géographique (SIG), qui permettent de manipuler des données géographiques aujourd'hui très variées et de plus en plus accessibles. Nous allons voir que différentes méthodes de modélisation de la connectivité des habitats existent. Elles diffèrent principalement par leur réalisme écologique, leur besoin en connaissances préalables sur les espèces étudiées et par la durée des calculs qu'elles impliquent.

Connectivité structurelle

La méthode la plus simple pour cartographier la connectivité des habitats consiste à localiser sur une carte les zones qui correspondent à l'habitat de l'espèce. Cela peut se faire à partir d'un SIG et de couches d'information géographique concernant l'occupation du sol. Certaines bases de données d'occupation du sol sont disponibles gratuitement et couvrent toute l'Europe (les données Corine Land Cover par exemple). Mesurer les distances géodésiques²³ entre les taches d'habitat permet d'estimer leur connectivité. Des méthodes telles que la "dilatation-érosion" (Vogt *et al.*, 2007) (Figure 8) ont l'avantage de localiser les corridors permettant de se déplacer entre les taches d'habitat dès lors qu'elles sont suffisamment proches les unes des autres.

Ces méthodes cartographient essentiellement la connectivité structurelle (Calabrese et Fagan, 2004) ou la continuité écologique. À l'image du modèle tache-matrice-corridor (Forman, 1995) ou encore des premiers textes de lois sur la conservation de la connectivité des habitats (CE, 1992), elles ne tiennent pas compte du comportement de l'espèce étudiée lors de son déplacement à travers une matrice paysagère hétérogène. Bien que cette limite soit moins problématique lorsque les espèces étudiées ont de très faibles capacités de dispersion en dehors de leurs taches d'habitat, elle confère à ces méthodes un faible réalisme écologique lorsque les espèces peuvent traverser la matrice, en particulier si cette matrice contient des éléments paysagers fortement hétérogènes.

Connectivité fonctionnelle

Des méthodes de modélisation tiennent compte des capacités de dispersion et du comportement des espèces pour analyser la connectivité fonctionnelle de leurs habitats. Dans ce cas, la **résistance** que chaque élément de la matrice exerce sur les déplacements des individus est prise en compte. Cette modélisation intègre le plus souvent des données relatives à l'occupation du sol ainsi qu'à des éléments paysagers linéaires tels que des infrastructures de transport ou des cours d'eau. À chacun de ces éléments paysagers (type d'occupation du sol, route, rivière, etc.) est attribuée une valeur de résistance,

23. Le terme distance "géodésique" désigne la distance la plus courte entre deux points à la surface de la Terre. Le terme "distance géographique" est souvent utilisé pour désigner cela, tout comme celui de distance euclidienne dès lors que la distance est calculée à partir d'un système de coordonnées projeté et non polaire. Dans le premier système de coordonnées, les coordonnées sont exprimées en mètres par rapport à un point de référence. Dans le second, elles sont exprimées en degrés par rapport à l'Équateur et à un méridien de référence (latitude et longitude, respectivement).

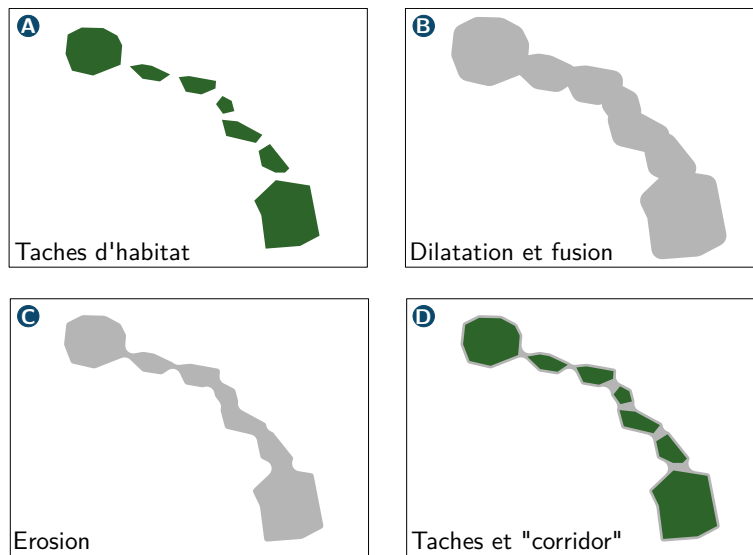


FIGURE 8 – Modélisation de la connectivité structurelle à l'aide de la méthode de dilatation-érosion. (A) Les taches d'habitat sont cartographiées. (B) Les taches d'habitat sont dilatées en les élargissant jusqu'à une distance d à partir de leurs limites. L'ensemble obtenu est fusionné. (C) Les taches dilatées et fusionnées sont érodées en les faisant rétrécir d'une distance d vers l'intérieur à partir de leurs limites. (D) Le résultat final contient les taches d'habitat (en vert) occupant la même surface qu'initialement. Lorsque les taches sont proches les unes des autres, la méthode permet de créer de fins corridors entre elles (en gris). Ils représentent les zones potentiellement traversées par les individus pour aller d'une tache à une autre en parcourant une distance minimale.

un coût, représentant l'effet qu'il exerce sur le déplacement des espèces (Figure 9). La matrice paysagère est alors représentée par une surface de résistance qui prend la forme d'une grille de pixels, c'est-à-dire une couche raster dans le vocabulaire des SIG. Sur cette surface de résistance, on attribue à chaque pixel le coût correspondant au type d'élément paysager qu'il contient. L'hétérogénéité de la matrice paysagère est donc explicitement prise en compte.

La méthode la plus couramment utilisée ensuite consiste à calculer des **chemins de moindre coût** (Adriaensen *et al.*, 2003) à l'aide de l'algorithme de Dijkstra (1959). Par construction, le tracé de ces chemins minimise le coût de déplacement entre les taches d'habitat à travers la matrice paysagère compte tenu des traits fonctionnels de l'espèce (Figure 10). Le résultat obtenu est un chemin, c'est-à-dire une entité spatiale ayant la forme d'une ligne dans un SIG. Chaque chemin a une longueur qui s'exprime en unités métriques et qui diffère de la distance géodésique entre deux points dès lors que la matrice paysagère est hétérogène. Cette distance peut également s'exprimer comme le coût total accumulé le long du chemin. On parle alors de **distance-coût** et nous verrons que cette valeur est utilisée dans plusieurs types d'analyse.

Pour l'attribution des coûts à chaque type d'occupation du sol, il est courant de se baser sur des dires d'expert ou sur des résultats issus de la littérature scientifique concernant les capacités de déplacement de l'espèce étudiée dans différents milieux (Gonzales et Gergel, 2007 ; Stevenson-Holt *et al.*, 2014) (Figures 11A et 11B). Bien que les experts aient parfois une excellente connaissance des espèces étudiées, leurs indications restent nécessairement subjectives. De plus, même si le modélisateur et les experts interrogés connaissent parfois les résultats issus de la littérature, il n'est pas garanti que des valeurs de coût estimées dans un contexte particulier soient applicables dans un autre site d'étude. Aussi, lorsque des données biologiques issues du terrain sont disponibles, Beier *et al.* (2008) conseillent de les utiliser pour calibrer les coûts. L'utilisation de ces données

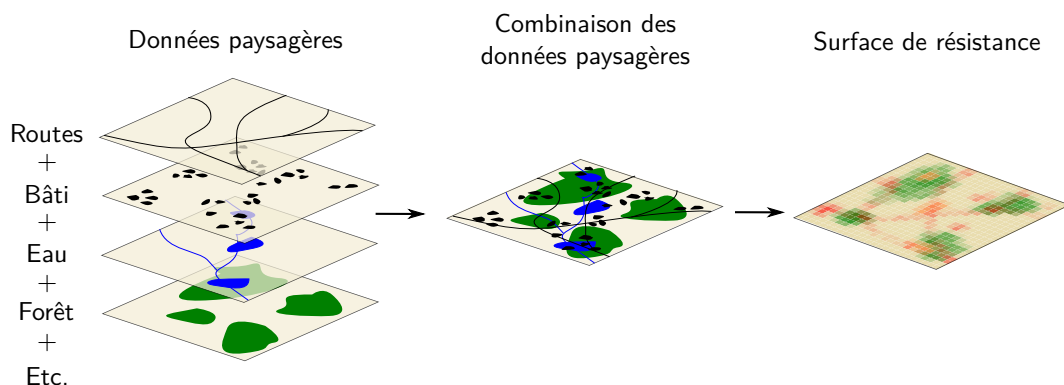


FIGURE 9 – Création d’une surface de résistance à partir de données géographiques. De façon générale, une carte d’occupation du sol couvrant l’ensemble du territoire est complétée par les éléments qu’elle n’inclut pas ; notamment des éléments linéaires tels que les cours d’eau ou les infrastructures de transport, ou encore des données plus précises concernant la localisation de parcelles agricoles ou du bâti.

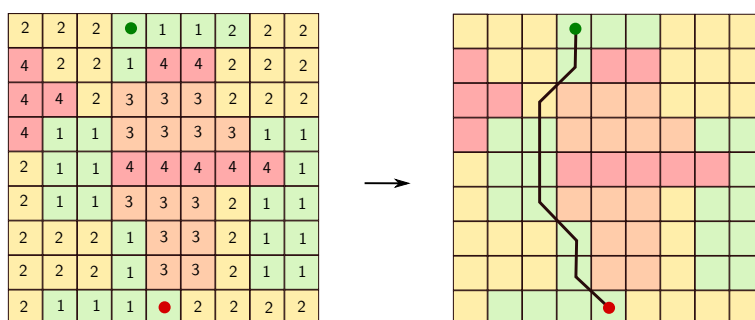


FIGURE 10 – Calcul des chemins de moindre coût. Chaque couleur de pixel correspond à un type d’occupation du sol et les chiffres correspondent aux valeurs de coût qui leur sont attribuées (de 1 à 4 en allant du vert au rouge). Le chemin noir relie les points vert et rouge en minimisant la somme des coûts des pixels traversés ; c’est le chemin de moindre coût.

est supposée augmenter le réalisme écologique de la modélisation (Beier *et al.*, 2008 ; Zeller *et al.*, 2012, 2018). Par exemple, des observations de la présence de l’espèce étudiée permettent de modéliser sa distribution (Figure 11C). Un **modèle de distribution** attribue à chaque pixel de la zone d’étude la probabilité que l’espèce étudiée y soit présente en se basant sur une combinaison de données environnementales (cf. Annexe B3). Ces probabilités de présence sont ensuite converties en coûts de déplacement pour calculer les chemins de moindre coût. La formule utilisée pour la conversion tient habituellement compte du fait qu’une espèce est plus exigeante dans le choix de l’habitat dans lequel elle s’établit que dans ses choix de déplacement dans le paysage (Keeley *et al.*, 2016, 2017). Cependant, observer la présence d’un individu ne signifie pas qu’il est en train de se déplacer, ni qu’il survivra ou s’établira dans un nouvel habitat s’il se déplace. De façon similaire, ne pas observer d’individus de l’espèce étudiée dans un milieu donné ne signifie pas qu’ils ne peuvent pas le traverser. Aussi, les données de présence ne sont pas les plus appropriées pour modéliser la connectivité des habitats. Une alternative supposée augmenter le réalisme des modèles de connectivité consiste à utiliser des traces GPS ou des données génétiques (Beier *et al.*, 2008 ; Zeller *et al.*, 2018). En effet, elles reflètent plus directement le mouvement des individus entre les taches d’habitat. Les traces GPS permettent de modéliser la préférence d’une espèce pour certains types d’occupation du sol lors de ses déplacements. Les coûts sont fixés à partir de ces préférences relatives, pour modéliser ensuite la connectivité des habitats à l’échelle du paysage (Zeller *et al.*, 2018) (Figure 11D). Enfin, lorsque des données génétiques sont disponibles, il est possible d’identifier les valeurs de coût qui expliquent au mieux la structure génétique des populations occupant

un ensemble de taches d'habitat (Figure 11E). En effet, la résistance de la matrice paysagère détermine la fréquence des flux génétiques et les données génétiques reflètent cette influence (cf. section 3.3.2).

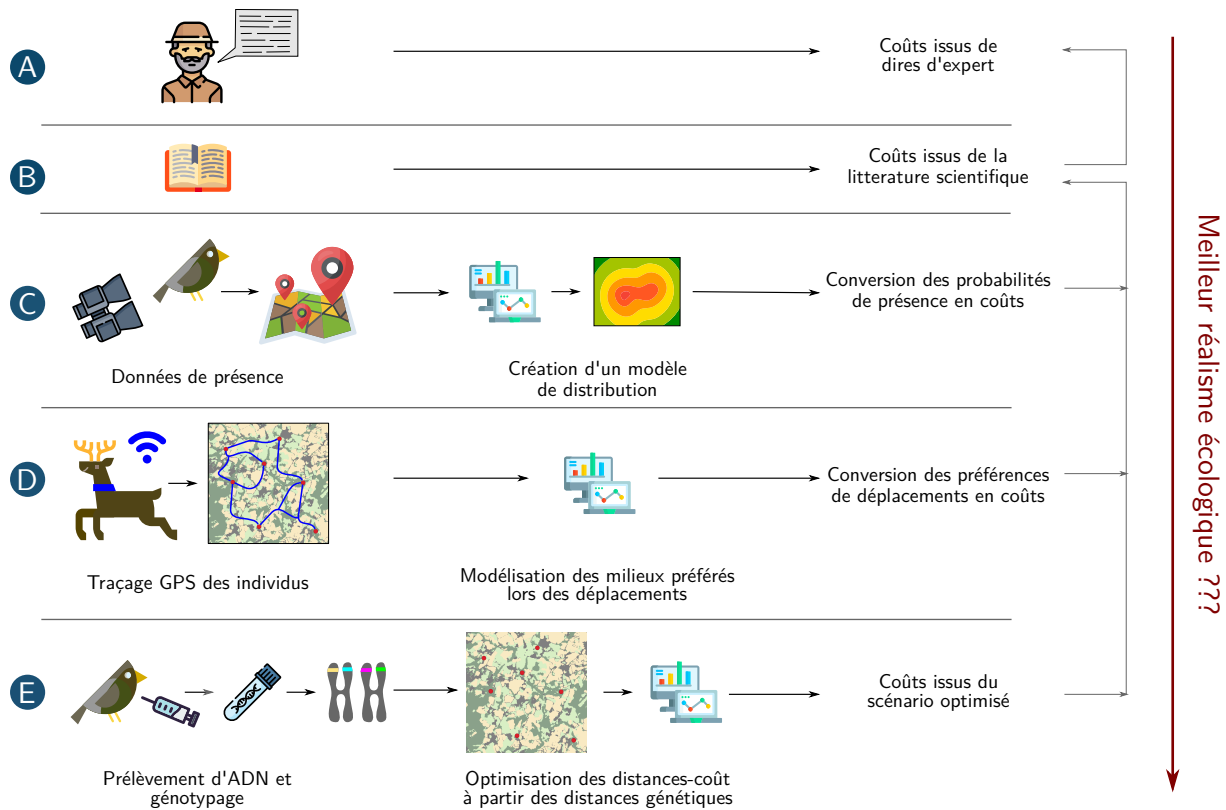


FIGURE 11 – Différentes méthodes de choix des valeurs de coût associées à chaque type d'occupation du sol. La flèche rouge à droite de la figure illustre le fait que d'après [Beier et al. \(2008\)](#) le réalisme écologique de la modélisation est censé augmenter lorsque plutôt qu'attribuer les coûts en fonction de dires d'experts (A), on utilise des résultats issus de la littérature scientifique (B), des données de présence (C), des données de télémétrie (D) ou des données génétiques (E). Les flèches grises ascendantes indiquent que les résultats des analyses de données empiriques complètent la littérature scientifique, qui peut elle-même étendre les connaissances des experts naturalistes.

Considérer que les chemins de moindre coût sont les chemins qu'emprunte une espèce pour se disperser revient à supposer que les individus ont une connaissance préalable du paysage dans lequel ils vivent, ce qui paraît peu réaliste ([Sawyer et al., 2011](#)). Des modèles plus complexes prennent en compte le fait qu'un individu peut se disperser dans toutes les directions en suivant des chemins alternatifs relativement favorables. Ces modèles se basent sur la théorie des circuits ([McRae, 2006](#)) (mais voir aussi [Panzacchi et al. \(2016\)](#)). Ils assimilent le paysage à un circuit électrique sur lequel la résistance entre deux points dépend de circuits en parallèle et peut être calculée en appliquant la loi d'Ohm. Néanmoins, ce calcul est nettement plus long que le calcul de chemins de moindre coût. Chacune des méthodes continue d'être utilisée car (i) leurs résultats sont relativement semblables dans de nombreuses situations ([Balbi et al., 2018](#) ; [Simpkins et al., 2018](#)) et (ii) bien que l'utilisation de la théorie des circuits soit la plus pertinente dans certains cas ([Khimoun et al., 2017](#) ; [McRae, 2006](#)), elle ne donne pas systématiquement de meilleurs résultats que l'utilisation des chemins de moindre coût pour expliquer des réponses biologiques ([Zeller et al., 2018](#)).

Enfin, des méthodes plus complexes sont fondées sur la simulation de déplacements d'individus au sein du paysage sur plusieurs générations en fonction d'un grand nombre de paramètres ([Palmer](#)

et al., 2011 ; Pe'er *et al.*, 2011). Ces modèles individus-centrés sont en théorie les plus réalistes mais leur paramétrage demande une connaissance précise de l'espèce étudiée et de son comportement. Ils impliquent par ailleurs des calculs très longs. On en retrouve peu d'applications dans la littérature scientifique sur la connectivité (Coulon *et al.*, 2015 ; Lookingbill *et al.*, 2010) ; leur utilisation dans un cadre opérationnel est encore plus rare.

Connectivité réalisée

Outre la connectivité structurelle ou fonctionnelle, les zones de connectivité réalisée sont des zones au niveau desquelles des déplacements d'individus ont été observés sur le terrain (Calabrese et Fagan, 2004). Les traces GPS, ou encore les données de capture-marquage-recapture, apportent directement ce genre d'informations. Elles sont précieuses pour estimer la distance à laquelle les individus se déplacent et la proportion des individus concernés (voir Riley *et al.* (2006) pour un exemple et Bowne et Bowers (2004) pour une synthèse de ces études). Néanmoins, elles permettent difficilement d'estimer la connectivité des habitats à l'échelle d'un paysage, par manque d'exhaustivité des relevés. Par ailleurs, les données génétiques apportent indirectement des informations sur les épisodes de dispersion ayant eu lieu entre deux populations durant les générations précédentes. La façon dont on peut utiliser ces données pour calibrer des coûts de déplacement est au cœur de cette thèse et sera abordée dans la section 3.3.2 de cette introduction.

Compromis entre réalisme écologique et complexité des modèles

Le degré de réalisme écologique des différentes méthodes de modélisation de la connectivité des habitats augmente théoriquement avec leur besoin en informations préalables et les temps de calcul (Calabrese et Fagan, 2004). On peut néanmoins considérer que (i) ce degré de réalisme n'augmente pas de façon linéaire avec la complexité de la méthode et (ii) que le modèle ne sera jamais en adéquation parfaite avec la réalité (Moulherat, 2019) (Figure 12). Il est ainsi nécessaire de trouver un compromis entre le réalisme de l'analyse de la connectivité, et la complexité et le coût de la méthode utilisée.

2.2 Les graphes paysagers

La connectivité fonctionnelle des habitats peut être modélisée à partir d'une quantité réduite d'informations préalables en créant des graphes paysagers. Ils constituent un bon compromis entre qualité et complexité de mise en œuvre de l'analyse (Calabrese et Fagan, 2004). Nous allons dans un premier temps présenter le principe de cette méthode basée sur la théorie des graphes. Puis, nous évaluerons ses principaux avantages et inconvénients, qui justifient que l'on s'y intéresse dans le cadre de cette thèse.

2.2.1 La théorie des graphes en écologie

La **théorie des graphes** est la branche des mathématiques qui s'intéresse à l'analyse des réseaux (Harary, 1969 ; Kolaczyk et Csárdi, 2014) (Encadré 3). Dans la suite de cette thèse, le terme "réseau" (*network*) sera utilisé pour désigner les objets étudiés, et le terme "**graphe**" (*graph*) pour désigner la représentation que l'on en fait dans ce cadre sous la forme d'un ensemble de nœuds et de liens. La phrase "*Le tout est plus que la somme de ses parties*" illustre l'intérêt de l'analyse de l'ensemble des éléments d'un réseau sous la forme d'un graphe. En se focalisant sur les relations entre chaque paire de

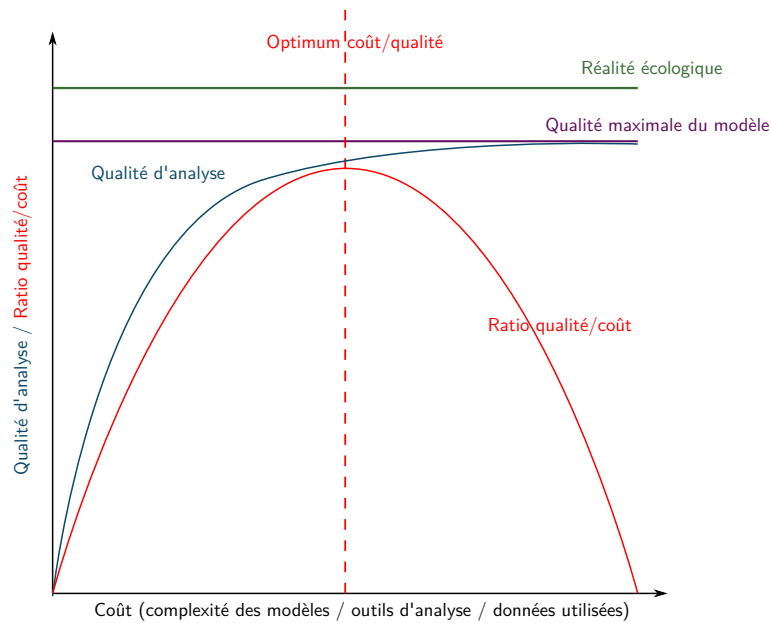


FIGURE 12 – Compromis entre la qualité et la complexité des modèles de connectivité. D’après Moulherat (2019). Cette représentation schématique se base sur l’hypothèse rarement testée que l’utilisation d’une plus grande quantité de données et de modèles plus complexes améliore le réalisme écologique de la modélisation.

nœuds sans tenir compte de la relation de ces nœuds avec tous les autres, on ne peut comprendre certaines des propriétés du système. Les graphes permettent d’étudier ces propriétés dites "émergentes", dues par exemple au rôle clé d’un nœud ou d’un lien pour la cohérence de l’ensemble du réseau ou à des connexions plus fréquentes entre les nœuds d’un sous-ensemble de nœuds.

Encadré 3 : Théorie des graphes : origines et vocabulaire

La théorie des graphes a été introduite pour la première fois par Leonhard Euler en 1736 alors qu’il cherchait à fournir la preuve qu’il était impossible de traverser chacun des sept ponts de Königsberg (actuelle Kaliningrad en Russie) une seule et unique fois lors d’une promenade revenant à son point de départ. Pour résoudre le problème, Euler avait assimilé les îles et les berges à des nœuds et les ponts à des liens, bâtissant ainsi les bases de la théorie des graphes. Un graphe $G = (V, E)$ est composé d’un ensemble $V = \{1, \dots, n\}$ de n nœuds connectés les uns aux autres par un ensemble E de liens.

Le lien $e = \{i, j\} \in E$ relie les nœuds i et j . Lorsque chacun des nœuds du graphe est relié à chacun des autres nœuds, on dit que le graphe est complet (Figure 13). Il contient un nombre maximal de liens. À l’inverse, lorsque certains liens ne sont pas observés, le graphe est incomplet. On dit que le graphe est dirigé si le lien $\{i, j\}$ et le lien $\{j, i\}$ sont différents, et non dirigé s’ils sont identiques (Figure 13). Dans un graphe dit simple, on ne considère pas les liens connectant les nœuds à eux-mêmes, aussi appelés boucles (*self-loop*). Ainsi, un graphe simple complet à n nœuds contiendra $n(n - 1)$ liens s’il est dirigé et $n(n - 1)/2$ liens s’il est non dirigé.

On distingue les graphes unipartites des graphes bipartites (Figure 13). Dans le premier cas, un lien peut exister entre n’importe quelle paire de nœuds du graphe. Dans le second, les nœuds

V forment deux ensembles distincts V_1 et V_2 tel que $V_1 \cap V_2 = \emptyset$ et tout lien $e = \{u, v\} \in E$ est tel que $u \in V_1, v \in V_2$. Il est ainsi impossible qu'un lien connecte deux nœuds appartenant au même sous-ensemble V_1 ou V_2 . Par exemple, si on étudie le réseau formé par les interactions entre des plantes (V_1) et leurs insectes pollinisateurs (V_2), il ne contiendra jamais de liens entre deux plantes ni entre deux pollinisateurs. Parfois, des sous-ensembles de nœuds d'un graphe sont totalement déconnectés les uns des autres. On dit que le graphe est déconnecté et qu'il comporte plusieurs composantes connexes (Figure 13). À l'inverse, si on ne distingue pas de sous-ensembles de nœuds déconnectés les uns des autres, le graphe est dit connecté ou connexe.

L'ensemble des liens d'un graphe définit sa topologie. Deux nœuds i et j d'un graphe peuvent être connectés indirectement par le biais de leurs liens avec d'autres nœuds. On appelle chemin entre i et j l'ensemble des liens $e_1, \dots, e_k \in E$ permettant de relier ces deux nœuds. Il est tel que e_1 est issu de i , e_k est issu de j et pour tout $1 \leq t \leq k - 1$, les liens e_t et e_{t+1} ont un nœud en commun. Un chemin composé de plusieurs liens et reliant un nœud à lui-même est un cycle. Un graphe ne comportant pas de cycle est un arbre (Figure 13), en référence au fait que les extrémités des branches d'un arbre ne se connectent jamais les unes aux autres pour former des cycles. Il existe parfois plusieurs chemins possibles entre deux nœuds. On appelle alors chemin le plus court celui qui compte le nombre minimum de liens. On peut également associer des poids aux nœuds et aux liens d'un graphe, qui devient alors un graphe pondéré (Barrat *et al.*, 2004), par opposition aux graphes binaires (Figure 13). Dans un tel graphe, le chemin le plus court entre deux nœuds peut se calculer en minimisant la somme des poids des liens qui les séparent. Enfin, lorsque les nœuds d'un graphe ont des coordonnées spatiales, on parle de graphes spatiaux (Dale et Fortin, 2010) (Figure 13).

Un graphe se note habituellement sous la forme d'une matrice d'adjacence A . Elle est carrée et comporte autant de lignes et de colonnes qu'il y a de nœuds dans le graphe ($n \times n$). Une matrice d'adjacence binaire indique par des 0 et des 1 l'absence ou la présence de liens entre les nœuds, respectivement. Ainsi, $A = (A_{ij})_{1 \leq i, j \leq n}$ est telle que :

$$A_{ij} = \begin{cases} 1 & \text{si } \{i, j\} \in E, \\ 0 & \text{sinon} \end{cases}$$

Si le graphe est non dirigé, alors la matrice d'adjacence A est symétrique : $A_{ij} = A_{ji}$. Si le graphe est simple, alors tout élément diagonal de A est nul : $A_{ii} = 0$. Enfin, si le graphe est pondéré, on peut remplacer les 1 de la matrice A par le poids associé à chaque lien, tel que :

$$A_{ij} = \begin{cases} w_{ij} & \text{si } \{i, j\} \text{ existe et a pour poids } w_{ij}, \\ 0 & \text{sinon} \end{cases}$$

La théorie des graphes offre une gamme de méthodes d'analyse et de métriques pour caractériser chacun des nœuds, des liens ou l'ensemble du graphe. Par exemple, le nombre de liens connectant chaque nœud avec les autres nœuds est son degré d , tel que $d_i = \sum_{j=1}^n A_{ij}$ pour un graphe binaire. Le degré est un exemple de métrique locale de centralité et nous verrons dans la section 2.2.2 qu'il

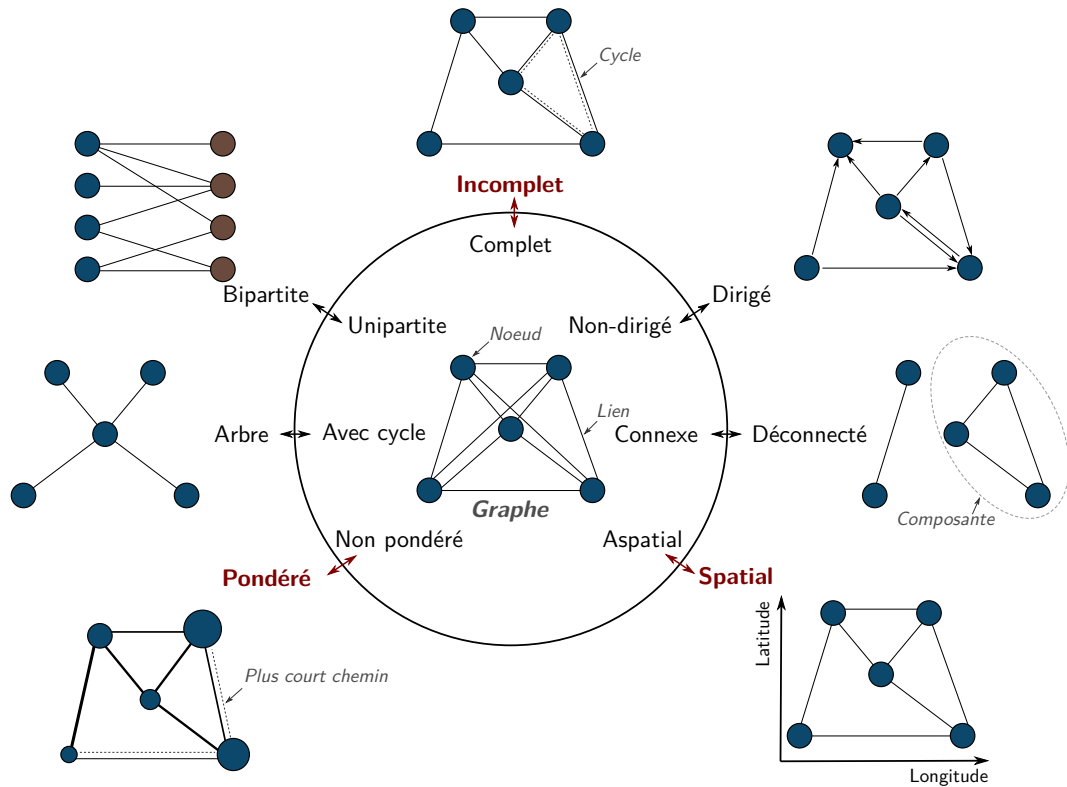


FIGURE 13 – Différents types de graphes. Les graphes dont les caractéristiques sont notées en gras et en rouge sont ceux que nous utiliserons le plus dans cette thèse.

en existe d'autres, plus adaptées aux analyses de connectivité des habitats notamment. À l'échelle de l'ensemble du graphe, la distribution des degrés des nœuds distingue les graphes dont la connectivité repose sur quelques nœuds très connectés (les *hubs* d'un réseau de transport par exemple) de ceux dans lesquels tous les nœuds sont connectés de façon équivalente (Watts et Strogatz, 1998).

Décrire la topologie d'un graphe consiste à étudier ses liens en se basant sur leur nombre, sur la présence de cycles, de plusieurs composantes connexes ou encore en calculant les chemins les plus courts entre les paires de nœuds. Ce calcul permet en outre de transformer une matrice d'adjacence binaire contenant des 0 et des 1 en une matrice de distance sur le graphe contenant les nombres de liens séparant chaque paire de nœuds lorsqu'on suit le chemin le plus court. Ce type de matrice a un intérêt particulier lorsque les graphes sont spatiaux et que les liens sont pondérés par des distances.

Des algorithmes permettent également de partitionner le graphe en identifiant des sous-ensembles de nœuds plus fréquemment reliés les uns aux autres qu'aux nœuds des autres sous-ensembles (Newman, 2006) (Figure 14). Ils forment ainsi des communautés, aussi appelées modules et identifiées à l'aide d'indices de modularité. Bien qu'il existe plusieurs formules pour le calcul de tels indices, ils prennent généralement une valeur forte lorsqu'il y a beaucoup de liens entre les nœuds d'un même module et une valeur faible lorsque les liens relient des nœuds appartenant à des modules différents (Figure 14). On les calcule pour un grand nombre de partitions alternatives et la partition la plus cohérente est celle qui maximise l'indice. Par exemple, la formule de Newman et Girvan (2004) appliquée à la partition \mathcal{C} des nœuds d'un graphe en K modules donne l'indice de modularité $\mathcal{Q}(\mathcal{C})$ tel que :

$$Q(\mathcal{C}) = \frac{1}{2m} \sum_{i,j;\mathcal{C}(i)=\mathcal{C}(j)} (A_{ij} - \frac{d_i d_j}{2m})$$

avec m le nombre de liens du graphe, A_{ij} la valeur de la matrice d'adjacence associée à la paire de nœuds i, j et d_i le degré de i . $\sum_{i,j;\mathcal{C}(i)=\mathcal{C}(j)}$ signifie que l'on considère toutes les paires de nœuds appartenant au même module selon la partition dont on calcule l'indice de modularité. Le terme de droite dans la différence permet de standardiser la valeur en tenant compte du nombre total de liens m et du degré des nœuds.

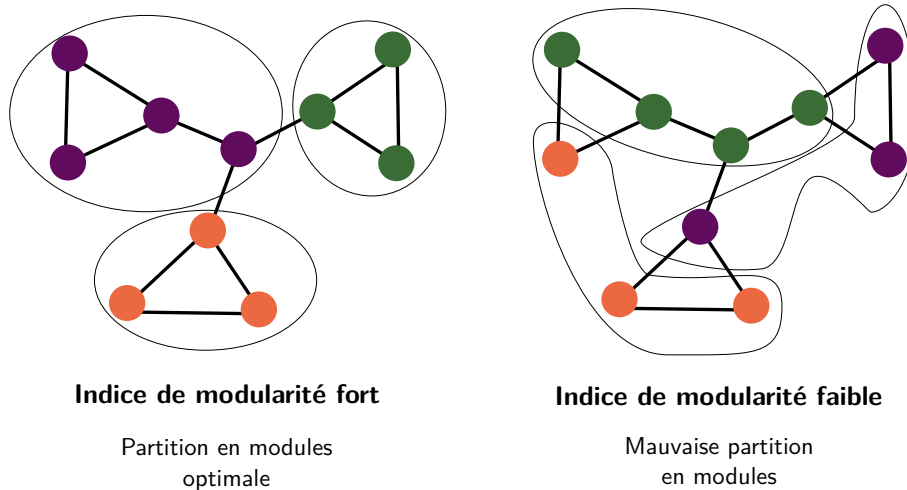


FIGURE 14 – Principe de la partition d'un graphe à l'aide du calcul de la modularité

Enfin, la visualisation du graphe en elle-même aide à interpréter sa structure. Elle implique de déterminer les coordonnées des nœuds pour représenter la structure des interactions dans un plan à deux dimensions de façon optimale. Un pan entier des travaux sur la théorie des graphes est dédié au développement d'algorithmes pour obtenir ces coordonnées, tels que les algorithmes de [Fruchterman et Reingold \(1991\)](#) ou de [Kamada *et al.* \(1989\)](#). Dans le cas des graphes spatiaux, les coordonnées spatiales des nœuds sont le plus souvent utilisées pour visualiser le graphe, même si l'utilisation d'autres algorithmes peut révéler des caractéristiques du graphe non visibles lors de la visualisation spatiale.

Applications de la théorie des graphes

La théorie des graphes a trouvé des applications dans un grand nombre de domaines en raison de l'omniprésence des réseaux dans les systèmes étudiés et de la diversité des méthodes qu'elle offre pour l'analyse de leur structure ([Miele *et al.*, 2019](#)). Ainsi, elle a été utilisée pour analyser des réseaux sociaux ([Latouche *et al.*, 2011](#)), de communication, de transport ([Lagesse *et al.*, 2015](#)), de transmission de maladies ([Auerbach *et al.*, 1984](#)), de régulation de l'expression génétique ou de voies métaboliques ([Magwene et Kim, 2004](#)) ou pour optimiser des calculs en informatique, entre autres exemples. En écologie, les premières analyses de réseaux se trouvent dans des travaux de phytosociologie portant sur la structure des communautés d'espèces de plantes ([Dale, 1977](#)). De nombreux graphes d'interaction ont été analysés depuis pour décrire des communautés d'espèces ([Bascompte, 2007](#)), mais aussi les relations sociales entre des animaux de la même espèce ([Farine et Whitehead, 2015](#)), entre autres exemples ([Dale, 2017](#) ; [Proulx *et al.*, 2005](#)). Les graphes bipartites ont également un franc succès car ils permettent de répondre aux questions que se posent actuellement les écologues concernant les réseaux

trophiques ou les réseaux d'interaction entre plantes et pollinisateurs (Thébault et Fontaine, 2010) ou entre hôtes et parasites (Pilosof *et al.*, 2014).

L'intérêt des graphes spatiaux en écologie

Les réseaux étudiés en écologie sont fréquemment constitués d'entités de même nature localisées à différents endroits dans l'espace (des taches d'habitat ou des populations échangeant des individus et leurs allèles par exemple). Les graphes spatiaux unipartites trouvent donc de nombreuses applications dans ce domaine (Dale et Fortin, 2010). Leurs liens sont habituellement pondérés par des distances ou par des indices de (dis)similarité représentant la fréquence des échanges potentiels ou réalisés entre leurs nœuds (par exemple la fréquence des flux de dispersion ou des flux génétiques). Il est d'ailleurs fréquent que la pondération des liens de ces graphes soit entièrement basée sur les données disponibles au niveau de leurs nœuds. Par exemple, lorsque les liens sont pondérés par des distances géographiques, ces distances sont calculées à partir des coordonnées spatiales des nœuds. De même, lorsqu'ils sont pondérés par des indices de similarité, ce sont les propriétés respectives des nœuds de chaque paire qui permettent de calculer ces indices (voir Economo et Keitt (2008) pour une application en écologie des communautés). L'utilisation de ce type de graphes en génétique des populations sera abordée en détail dans la suite de cette thèse.

La matrice d'adjacence pondérée d'un graphe spatial complet est équivalente à une matrice de distance ou à une matrice de similarité. La plupart des graphes spatiaux diffèrent ainsi des graphes dont les liens sont construits à partir de l'observation directe d'une interaction. Dans ce contexte, les analyses de graphes spatiaux pondérés partagent des points communs avec les analyses de statistiques spatiales, notamment avec celles dédiées aux relations entre entités spatiales (Dale et Fortin, 2014).

Prenons un exemple. On parle d'autocorrélation spatiale positive lorsqu'une variable prend des valeurs similaires au niveau d'entités proches spatialement (Tobler, 1970), ce qui est très couramment observé en écologie (Nekola et White, 1999). Les analyses d'autocorrélation spatiale consistent alors à évaluer la corrélation entre les valeurs d'une même variable mesurée au niveau d'entités voisines spatialement. L'indice I de Moran est fréquemment utilisé pour estimer le degré d'autocorrélation d'une variable quantitative z . Si la variable z est mesurée en n points distincts spatialement, l'indice se calcule ainsi :

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2},$$

avec \bar{z} la valeur moyenne de z . La matrice $\mathbf{W} = (w_{ij})_{1 \leq i, j \leq n}$ définit les relations de voisinage entre chaque paire d'observations i et j de telle sorte que $w_{ij} = 1$ si i et j sont voisines spatialement et $w_{ij} = 0$ sinon. Cet indice est souvent calculé en considérant plusieurs distances d pour définir le voisinage des observations i et j . Le graphique représentant l'évolution de $I(d)$ en fonction de d , appelé corrélogramme permet d'analyser l'échelle de la dépendance spatiale de la variable étudiée. Chaque matrice $\mathbf{W}(\mathbf{d})$ est équivalente à la matrice d'adjacence binaire d'un graphe spatial dont les nœuds correspondent aux n sites d'observation et dont les liens ont été supprimés dès lors qu'ils correspondaient à une distance géodésique supérieure à d (Dale et Fortin, 2014). Lorsqu'on supprime les liens d'un graphe, on dit que le graphe est "élagué" (*pruned*). Cet exemple illustre l'intérêt que

peut avoir l'**élagage** d'un graphe spatial pour analyser la dépendance spatiale de certaines variables ²⁴.

L'élagage présenté dans l'exemple précédent est simplement basé sur des distances géodésiques mais d'autres critères plus directement basés sur la topologie d'un graphe peuvent être utilisés. Par exemple, on peut considérer les connexions entre un nœud et ses k voisins les plus proches seulement, ou connecter les nœuds entre eux de telle sorte que les liens ne se croisent pas. On parle dans ce dernier cas de graphes planaires construits en réalisant une triangulation de Delaunay. Un autre exemple de graphe connectant uniquement les nœuds voisins spatialement est le graphe de Gabriel ([Gabriel et Sokal, 1969](#)), dans lequel deux nœuds i et j sont connectés dès lors que le cercle dont le diamètre est le segment reliant i et j ne contient aucun autre nœud. La grande variété des méthodes d'élagage permet de tester différentes hypothèses sur les déterminants de la dépendance spatiale des phénomènes étudiés. Elles ont donc un réel intérêt en écologie où elles permettent de comprendre les processus spatiaux responsables des réponses écologiques. Le choix des méthodes d'élagage à appliquer à des graphes spatiaux est au cœur du premier chapitre de cette thèse (voir aussi section 4.1).

La similitude entre certaines méthodes d'analyse spatiale et les analyses de graphes spatiaux ²⁵ s'observe également pour les métriques calculées au niveau des nœuds des graphes. Par exemple, l'indice $K(r)$ de [Ripley \(1977\)](#) mesure le degré d'agrégation spatiale d'un semis de points en faisant la somme du nombre de points situés à une distance inférieure à r de chacun des points et en la divisant par la densité globale du semis de points. Si l'ensemble des points du semis de points correspondaient aux nœuds d'un graphe, alors cet indice serait égal à la somme des degrés des nœuds d'un graphe spatial élagué en considérant une distance r , divisé par la densité spatiale des nœuds du graphe. En considérant différentes méthodes pour élaguer ce graphe mais aussi en associant des pondérations aux nœuds et aux liens connectant les nœuds, de tels indices peuvent être complexifiés. Nous verrons que ce genre de calcul est à la base de certaines mesures de connectivité basées sur des graphes spatiaux.

Origines de l'utilisation des graphes paysagers

Un des objectifs de l'écologie du paysage est de comprendre comment les individus se déplacent au sein de réseaux de taches d'habitat à l'échelle de paysages hétérogènes. Il n'est donc pas étonnant que [Risser et al. \(1984\)](#) aient tôt identifié le potentiel de la théorie des graphes dans ce domaine. Il a fallu attendre les travaux de [Cantwell et Forman \(1993\)](#) pour que les taches formées par les différents types d'occupation du sol au sein d'une mosaïque paysagère soient représentées comme les nœuds d'un graphe spatial. L'objectif de ces premiers graphes paysagers (*landscape graph*) était alors d'identifier une typologie de paysages à partir des propriétés de ces réseaux ²⁶. Puis, [Keitt et al. \(1997\)](#) ont représenté le réseau formé par les taches d'un habitat spécifique sous la forme d'un graphe. Ils ont montré que cette représentation permet de quantifier la connectivité de cet habitat à différentes échelles spatiales en se focalisant sur des liens et chemins de plus en plus longs à partir de chacun des

24. Notons que dans le cas de l'analyse de l'autocorrélation spatiale à différentes distances d tout comme lorsqu'on analyse des graphes élagués, il est important de comparer les résultats obtenus à ceux obtenus à partir d'une matrice ou d'un graphe complets pour pouvoir estimer la significativité de la dépendance spatiale mise en évidence.

25. [Dale et al. \(2002\)](#) ont largement décrit les relations mathématiques et conceptuelles qui existent entre les méthodes d'analyse spatiale et ont mentionné à ce titre celles qui sont basées sur des graphes spatiaux.

26. À titre d'exemple, [Cantwell et Forman \(1993\)](#) identifient le motif dit en "candélabre" caractérisé par une grande tache d'occupation du sol reliée d'un côté à une autre grande tache et à un grand nombre de petites taches de l'autre. Il est fréquemment observé dans les plaines alluviales ou les marais salants. Dans le cas de la plaine alluviale, les grandes taches sont la rivière et sa ripisylve tandis que les petites sont par exemple des parcelles de prairies inondables.

nœuds. Ces premiers travaux sont à l'origine de l'utilisation des graphes paysagers pour la modélisation et l'analyse des réseaux d'habitat.

2.2.2 Modéliser les réseaux d'habitat sous la forme de graphes paysagers

Définition des graphes paysagers

L'article fondateur d'Urban et Keitt (2001) définit de façon précise ce que sont les graphes paysagers²⁷ et la façon dont ils peuvent être utilisés en biologie de la conservation et en écologie du paysage. Les **graphes paysagers** sont des graphes spatiaux simples dont les nœuds correspondent aux taches d'habitat des espèces considérées et les liens aux chemins de dispersion potentiels entre ces nœuds. Bien qu'ils soient rarement dirigés, leurs liens sont souvent pondérés par des distances tenant compte de la résistance du paysage entre les taches d'habitat. Pour cette raison, les graphes paysagers peuvent se représenter de façon réaliste en visualisant la localisation spatiale exacte des taches et des liens, mais aussi de façon topologique en représentant les nœuds et les liens par de simples points connectés par des lignes droites. De même, les taches sont généralement pondérées par une valeur de "capacité" correspondant à leur surface, leur qualité ou leur productivité, en tant qu'estimations de leur capacité de charge (Urban et Keitt, 2001).

Urban et Keitt (2001) ont présenté l'intérêt des graphes paysagers en adoptant deux démarches complémentaires. D'une part, des simulations basées sur la théorie des métapopulations montrent que les populations persistent théoriquement dès lors que la topologie du réseau reste connectée. D'autre part, l'intérêt concret des graphes paysagers pour orienter les mesures de conservation est illustré par une application au programme de conservation de la Chouette tachetée (*Strix occidentalis lucida*) en Utah, Colorado, Nouveau Mexique et Arizona. Suite à ce premier article fondamental, l'utilisation des graphes paysagers s'est popularisée à partir du début des années 2000 et ils sont aujourd'hui largement utilisés, aussi bien par des chercheurs que par des gestionnaires de la biodiversité (Galpern *et al.*, 2011). Leur popularisation vient en partie du progrès des SIG, de la disponibilité des données géographiques et des capacités de calcul informatique. Elle vient également du développement de logiciels gratuits permettant leur construction et analyse. Parmi eux²⁸, le logiciel Graphab a été créé en 2012 à l'UMR ThéMA et utilisé à de nombreuses reprises depuis (Foltête *et al.*, 2012b, 2021).

Construction des graphes paysagers

Nous présentons maintenant la façon dont les graphes paysagers sont construits et analysés lorsque l'objectif est d'analyser la connectivité des habitats d'une espèce dont les traits écologiques sont connus, mais que le modélisateur ne dispose pas de données provenant de la zone d'étude sur la présence et les déplacements de cette espèce. Ce cas reflète la majorité des études à finalité scientifique ou opérationnelle dans lesquelles les graphes paysagers sont utilisés (Foltête *et al.*, 2020).

Pour construire un graphe paysager dans ce contexte, il faut disposer de données géographiques concernant les zones d'habitat et les éléments paysagers qui favorisent ou limitent les déplacements

27. Le terme "graphe paysager" sera utilisé dans toute cette thèse en tant que traduction de "*landscape graph*", terme initialement utilisé par Urban et Keitt (2001). Les termes "*habitat graph*", "*habitat network*", "*patch-based graph*", entre autres, se rencontrent également dans la littérature pour désigner ces outils (Galpern *et al.*, 2011 ; Rayfield *et al.*, 2010 ; Saura *et al.*, 2014).

28. On peut également citer le logiciel Conefor Sensinode (Saura et Torne, 2009).

de l'espèce étudiée (Figure 15A). Ces données se basent principalement sur l'occupation du sol de la zone d'étude, comme présenté dans la section 2.1 (cf. Figure 9).

Ainsi, à partir de ces données géographiques, on délimite les types d'occupation du sol correspondant aux zones d'habitat de l'espèce. Des ensembles de pixels contigus correspondant à ces zones forment les taches d'habitat (Figure 15B). Une taille minimale des taches peut être spécifiée dès lors que l'on connaît la surface minimale du domaine vital de l'espèce étudiée. Ces taches constitueront les nœuds du graphe paysager, auxquels pourra être attribuée une capacité.

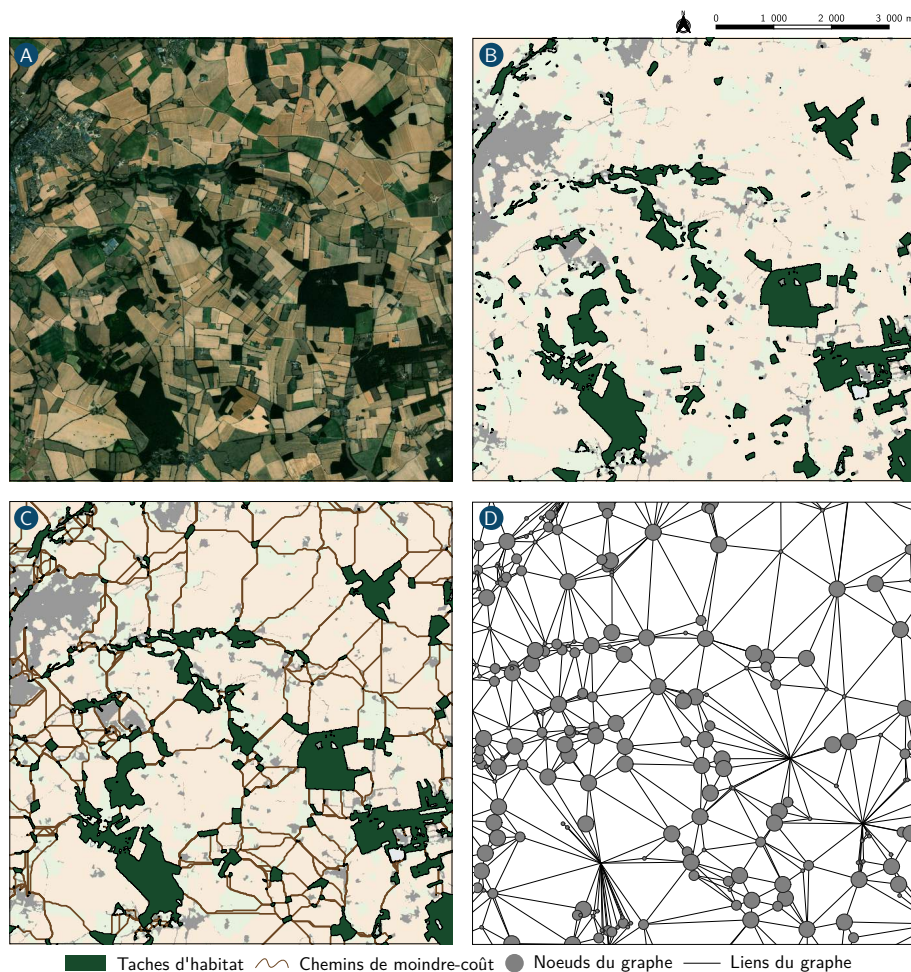


FIGURE 15 – Étapes de construction d'un graphe paysager. (A) Photographie aérienne de la zone choisie pour l'exemple. (B) Carte d'occupation du sol de la zone. Les taches vertes correspondent aux taches d'habitat (ici forestier) choisies pour l'exemple. (C) Représentation réaliste du graphe paysager formé par les taches et les chemins de moindre coût calculés entre elles selon une topologie planaire. (D) Représentation topologique du graphe paysager.

Ensuite, on calcule les chemins de moindre coût permettant de relier les taches d'habitat sur la surface de résistance (Figure 15C). Cette étape implique d'attribuer à chaque type d'occupation du sol présent sur la surface de résistance un coût en fonction des connaissances dont on dispose sur les déplacements de l'espèce étudiée (cf. Figure 11). Néanmoins, il n'est pas rare que de simples distances géodésiques euclidiennes soient calculées entre les taches, en particulier lorsqu'il est difficile d'estimer les coûts de déplacement des espèces étudiées ou qu'ils dépendent peu de l'occupation du sol (pour des espèces végétales par exemple, voir Moran-Lopez *et al.* (2016) ou Schoville *et al.* (2018)). Dans tous les cas, les distances associées à ces chemins sont censées refléter la façon dont l'espèce étudiée

perçoit le paysage, et nous les désignerons par le terme de "distances paysagères" par la suite.

Le graphe à proprement parler est créé en reliant les nœuds par des liens correspondant à ces chemins de moindre coût, pondérés par les distances paysagères qui leur sont associées (Figure 15D). Comme souvent lorsque les liens d'un graphe spatial ne sont pas basés sur des observations directes, le modélisateur contrôle la topologie du graphe. Ainsi, le graphe paysager peut être élagué pour ne conserver qu'un sous-ensemble des liens reliant les taches d'habitat. Cet élagage a parfois des justifications écologiques (Bunn *et al.*, 2000), mais facilite également la visualisation du graphe (Fall *et al.*, 2007) et les calculs (Foltête *et al.*, 2012b). Un élagage basé sur un seuil de distance permettra de ne conserver que les liens pouvant être parcourus par l'espèce étudiée si l'on connaît ses capacités de dispersion maximales (Bunn *et al.*, 2000). La création d'un arbre couvrant minimal, c'est-à-dire le graphe connexe sans cycle dont le poids total des liens est minimal, permet quant à elle de se focaliser sur la "colonne vertébrale" du réseau d'habitat selon Bunn *et al.* (2000). Les topologies planaires sont également utilisées pour la construction des graphes paysagers dans la mesure où (i) leur visualisation spatiale fournit une image synthétique de la connectivité du réseau (Fall *et al.*, 2007) et (ii) les métriques calculées à partir de graphes planaires sont calculées plus rapidement tout en étant de bonnes approximations des métriques calculées à partir du graphe complet (Foltête *et al.*, 2012b)²⁹. Par ailleurs, les graphes paysagers dont la topologie conserve les connexions entre les taches les plus voisines sont probablement à l'image de la façon dont s'effectuent les déplacements des espèces, compte tenu de la perception de leur environnement immédiat et de la fréquence des déplacements de type *stepping-stones* chez les espèces animales (Saura *et al.*, 2014). Enfin, l'analyse de graphes paysagers construits avec différents types d'élagage apporte en elle-même une information sur le patron spatial de la connectivité des habitats (Brooks, 2006, 2003 ; Bunn *et al.*, 2000 ; Keitt *et al.*, 1997)³⁰.

Comme on l'a vu, l'élagage est une manière de simplifier la topologie du graphe paysager, en ne conservant que les liens correspondant potentiellement aux connexions les plus fortes (distances paysagères les plus faibles ou liens empruntés pour relier des taches voisines). La pondération des liens est aussi un moyen de leur accorder une importance qui reflète la capacité des espèces à les parcourir. De plus, les poids des liens sont fréquemment transformés en probabilités de dispersion (Urban et Keitt, 2001). Dans ce cas, l'élagage a une influence moindre sur les analyses dans la mesure où des probabilités de dispersion extrêmement faibles sont attribuées aux liens que la plupart des méthodes d'élagage suppriment.

Pour convertir une distance paysagère en probabilité de dispersion, on suppose fréquemment que la probabilité de dispersion entre deux taches i et j décroît de façon exponentielle lorsque la distance paysagère d_{ij} augmente, de sorte que $p(d_{ij}) = e^{-\alpha d_{ij}}$ (Figure 16). Pour ce calcul, la valeur de la constante α est fixée en définissant la distance $d_{ij0.05}$ à laquelle la probabilité de dispersion est égale à 0.05, si bien que $\alpha = -\frac{\log(0.05)}{d_{ij0.05}}$. Avec les critères de délimitation des taches et les valeurs de coût, le choix de cette distance dépend des connaissances du modélisateur sur l'espèce étudiée. Ce calcul de la probabilité de dispersion entre les taches d'un graphe paysager a été introduit par Urban

29. La distance géodésique maximale associée au chemin entre deux nœuds sur un graphe planaire spatial est au maximum 2.42 fois plus grande que la distance géodésique directe entre ces nœuds (Keil et Gutwin, 1992).

30. On peut considérer de façon très schématique qu'une telle analyse basée sur des élagages successifs consiste à analyser la connectivité des habitats comme l'autocorrélation spatiale de la distribution de l'habitat lorsqu'on considère une matrice d'adjacence pondérée et intégrant les capacités de dispersion de l'espèce étudiée.

et Keitt (2001). Bien que ces auteurs aient mentionné d'autres fonctions de transformation d'une distance paysagère en probabilité de dispersion, ce calcul a souvent été réalisé tel qu'ils l'avaient présenté. Cela vient en partie du fait qu'il est directement inspiré des travaux menés sur la théorie des métapopulations (Hanski, 1994), pour lesquels on dispose de validations empiriques (Moilanen et Nieminen, 2002). En particulier, Hanski (1994) a défini une fonction d'incidence (*Incidence Function Model*, IFM) permettant d'estimer le nombre de migrants S_i au niveau de chaque tache d'habitat i en fonction de leur degré d'isolement par rapport aux autres taches et de la surface A_j de ces taches. Elle est telle que :

$$S_i = \sum_{j=1}^n p_j e^{-\alpha d_{ij}} A_j,$$

avec $p_j = 1$ si la tache j est occupée par une population et $p_j = 0$ sinon.

Cet héritage de la théorie des métapopulations se retrouve également dans les analyses de connectivité réalisées à partir des nœuds et des liens des graphes paysagers (Urban *et al.*, 2009) ; comme nous allons à présent le voir.

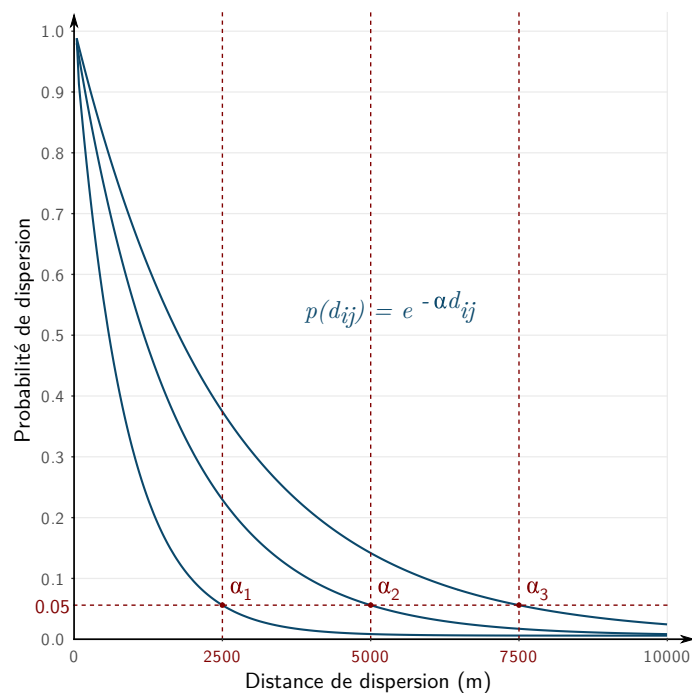


FIGURE 16 – Fonction de transformation d'une distance de dispersion en probabilité de dispersion. La fonction utilisée est telle que la probabilité de dispersion décroît de façon exponentielle à mesure que la distance de dispersion augmente : $p(d_{ij}) = e^{-\alpha d_{ij}}$. La valeur α est paramétrée de façon à ce que la probabilité de dispersion soit égale à 0.05 pour une distance de 2500, 5000 ou 7500 m (α_1 , α_2 et α_3 , respectivement).

Analyse des graphes paysagers

Le graphe paysager permet d'évaluer la connectivité du réseau d'habitat via le calcul d'une variété de métriques. D'une part, les métriques locales évaluent la contribution de chacun des nœuds à la connectivité des habitats. D'autre part, les métriques globales évaluent la connectivité de l'ensemble du réseau d'habitat (Galpern *et al.*, 2011) et permettent de comparer les différentes composantes du graphe paysager lorsqu'il n'est pas connexe. Les métriques diffèrent également par le poids qu'elles accordent aux nœuds et aux liens et par leur dépendance à la topologie du graphe. En effet, Rayfield

et al. (2011) distinguent les métriques essentiellement topologiques, prenant en compte les connexions entre nœuds indépendamment de leur force relative, des métriques qui tiennent compte de la connectivité en intégrant la pondération des liens. Certaines prennent en compte ces deux aspects.

Outre ces différences liées à l'échelle à laquelle les métriques sont calculées et à leur prise en compte des éléments du graphe, les métriques mesurent des aspects de la connectivité des habitats n'influençant pas exactement les mêmes processus écologiques (Bunn *et al.*, 2000). Tout d'abord, la capacité des taches est censée refléter les processus de natalité et de mortalité au niveau de chaque tache (Figure 17A). En effet, Urban *et Keitt* (2001) se basent sur la capacité des taches pour estimer le potentiel de "recrutement" des taches. Elle est souvent calculée à partir de leur surface faite de variables reflétant de façon plus précise le potentiel démographique de la tache. De façon similaire, dans le modèle des métapopulations, la probabilité d'extinction locale est estimée à partir de la surface des taches (Hanski, 1994). En reflétant la quantité d'habitat atteignable au niveau d'une tache, la capacité est censée influencer la démographie en favorisant l'accès aux ressources vitales pour l'espèce, mais aussi le déplacement local et l'appariement des individus lors de la reproduction.

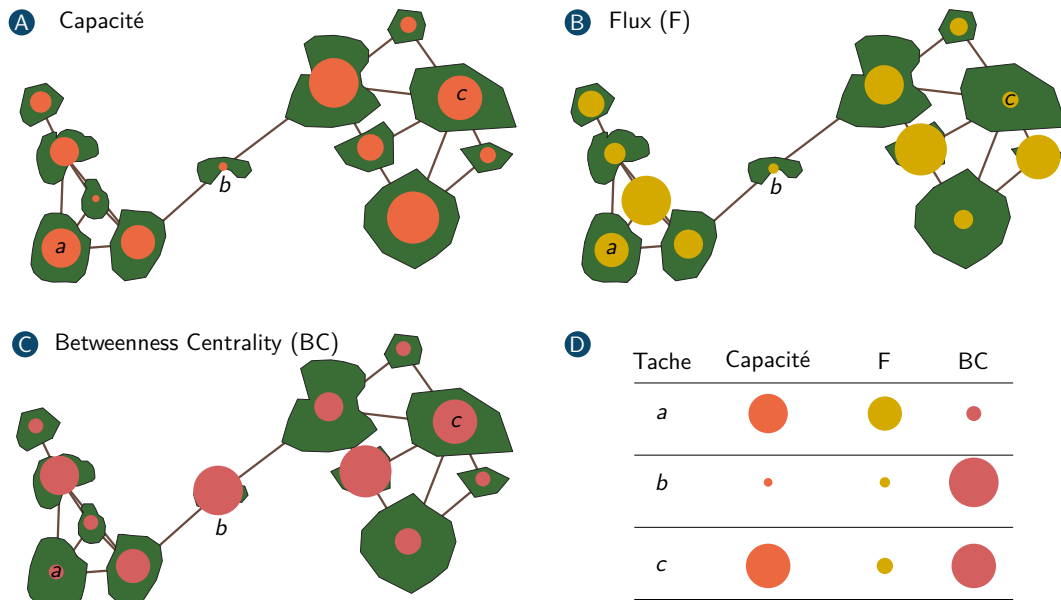


FIGURE 17 – Métriques de connectivité des habitats issues d'un graphe paysager. Les trois métriques de (A) capacité, (B) F et (C) BC reflètent la quantité d'habitat atteignable par les individus occupant les taches de différentes manières et à différentes échelles spatiales. Le graphe créé pour l'illustration est élagué en considérant une distance seuil de 4000 m. La métrique F a été calculée en considérant une probabilité de dispersion égale à 0.05 à 1000 m tandis que cette probabilité était de 0.5 à 2000 m pour le calcul de la métrique BC . Les tailles des nœuds sont proportionnelles aux valeurs des métriques (échelle standardisée pour les trois métriques). Le tableau (D) présente les différences de valeurs relatives des métriques pour trois taches (a , b , c) aux propriétés particulières.

Ensuite, le fait qu'une tache soit connectée à d'autres taches va influencer le nombre de migrants qu'elle reçoit ou qu'elle peut émettre. Pour cela, des métriques locales de connectivité basées sur la fonction d'incidence de la théorie des métapopulations, telle que la métrique de Flux (F) vont quantifier les flux potentiels de migrants s'effectuant depuis ou vers chacune des taches. Pour une tache i , la métrique F se calcule ainsi :

$$F_i = \sum_{j=1, j \neq i}^n a_j^\beta e^{-\alpha d_{ij}},$$

avec n le nombre de taches dans le graphe, a_j la capacité de la tache j , β l'exposant déterminant le poids accordé à la capacité dans le calcul et $e^{-\alpha d_{ij}}$ la probabilité de dispersion entre les taches i et j . Cette métrique, aussi appelée *Area Weighted Flux* dans certains travaux correspond à la somme des capacités des taches pondérées par la probabilité de dispersion entre la tache focale et chacune de ces taches (Figure 17B). Elle reflète d'une part la quantité d'habitat atteignable par les migrants depuis une tache donnée compte tenu de leurs capacités de dispersion. D'autre part, si les taches ayant une valeur de Flux élevée ont également une forte capacité, on peut s'attendre à ce qu'elles jouent le rôle de sources dans les modèles "source-puits" issus de la théorie des métapopulations (Pulliam, 1988), comme l'ont fait remarquer Urban et Keitt (2001) et Bunn *et al.* (2000). Cette métrique peut être calculée en faisant varier la valeur de α pour modifier le poids des connexions selon l'échelle spatiale à laquelle elles sont considérées (Figure 16). Cela permet d'adapter l'analyse aux capacités de dispersion des espèces et d'étudier la dépendance de la connectivité des habitats à ces capacités. Si $\beta = 0$, la métrique F mesure la probabilité de dispersion entre la tache i et les autres taches du graphe indépendamment de leur capacité. Cela permet de se focaliser sur le degré d'isolement des taches compte tenu de la topologie du réseau d'habitat indépendamment des capacités des taches voisines.

Certaines taches jouent un rôle important pour la connectivité de l'ensemble du réseau d'habitat à l'échelle de plusieurs générations. Elles contribuent à ce qu'Urban et Keitt (2001) ont appelé la traversabilité du réseau d'habitat de par leur rôle de *stepping-stones*. Ces taches sont clés dans les phénomènes de *spreading of risk* ou de *long-distance rescue* (Den Boer, 1968 ; Levins, 1969). Au niveau du graphe paysager, elles se trouvent souvent sur les chemins les plus courts permettant de relier des taches éloignées les unes des autres (Figure 17C). La prise en compte de la topologie du réseau d'habitat à l'aide de la théorie des graphes est essentielle pour évaluer cette propriété. Pour cela, la métrique de centralité intermédiaire (*Betweenness Centrality*, BC), couramment utilisée dans les analyses de graphes, a été appliquée aux graphes paysagers (Bodin et Norberg, 2007). Elle mesure le nombre de fois où une tache se trouve sur les chemins les plus courts entre chacune des paires de taches du graphe. Dans le cas des graphes paysagers, chaque paire de taches comptabilisée dans le calcul de la métrique BC est pondérée par le produit des capacités des deux taches et de la probabilité de dispersion entre elles. Pour une tache i , la métrique BC se calcule ainsi :

$$BC_i = \sum_{j=1}^n \sum_{k=1}^n a_j^\beta a_k^\beta e^{-\alpha d_{jk}}$$

$$j, k \in \{1, \dots, n\}; k < j; i \in P_{jk},$$

avec P_{jk} l'ensemble des taches traversées par le plus court chemin entre les taches j et k . La métrique BC mesure donc à quel point une tache contribue à la quantité d'habitat atteignable à l'échelle du réseau d'habitat par des flux de dispersion s'effectuant sur plusieurs générations. Selon Zetterberg *et al.* (2010), cette propriété des taches d'habitat est censée favoriser le brassage génétique. Elle permet aussi de localiser les taches d'habitat dont la suppression aurait le plus de conséquences sur la connectivité du réseau d'habitat (Thompson *et al.*, 2014).

Les métriques locales calculées au niveau des nœuds d'un graphe paysager permettent de mesurer la connectivité des habitats comme une quantité d'habitat atteignable par les individus depuis les taches d'habitat à différentes échelles spatiales et temporelles, en accord avec la définition de la connectivité employée dans cette thèse. Elles reflètent différents processus écologiques, nommés recrutement, flux

et traversabilité par [Urban et Keitt \(2001\)](#). Ils sont théoriquement liés aux processus d'extinction, de colonisation et de *long-distance rescue* sur lesquels se focalisent la théorie des métapopulations. Nous verrons dans la section 3 qu'ils influencent également la structure génétique des populations.

Les métriques globales se calculent quant à elles à l'échelle de l'ensemble du graphe. La difficulté dans ce cas est de considérer dans une seule métrique les différentes façons par lesquelles la capacité des taches et leurs connexions contribuent à la connectivité du réseau d'habitat. Le "comportement" des métriques globales a donc été étudié pour identifier celles qui sont les plus sensibles aux variations de connectivité ([Laita et al., 2011](#) ; [Pascual-Hortal et Saura, 2007](#)). Lors de ces travaux, les métriques tenant compte de la connectivité intra-tache et de la probabilité de dispersion entre chacune des paires de taches du réseau, y compris le long de chemins comportant plusieurs liens, se sont révélées les plus pertinentes ([Laita et al., 2011](#)). Parmi elles, la métrique *PC* (*Probability of Connectivity*) mesure la probabilité que deux individus tirés au hasard dans les taches d'habitat puissent interagir ([Saura et Pascual-Hortal, 2007](#)). Elle se calcule ainsi :

$$PC = \frac{1}{A^2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j e^{-\alpha d_{ij}},$$

avec a_i la surface de la tache i et A la somme des surfaces des taches de la zone d'étude. On remarque que s'il n'y a qu'une seule tache de 10000 m² dans la zone d'étude, alors $PC = 1$. En revanche, s'il y a deux taches de 5000 m² séparées d'un kilomètre et que la probabilité de dispersion à cette distance est égale à 0.05, alors $PC = 0.5125$. Cette métrique tient donc compte de la connectivité intra-tache, tout en étant sensible à la fragmentation de l'habitat. Elle est directement transformable en une quantité d'habitat atteignable à l'échelle de l'ensemble du réseau d'habitat. C'est pourquoi on utilise aussi fréquemment la métrique *EC* (*Equivalent Connectivity*) correspondant à la surface de la tache unique d'habitat qui aurait une probabilité de connectivité *PC* identique à celle du réseau d'habitat étudié ([Saura et al., 2011](#)). Elle se calcule ainsi :

$$EC = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_i a_j e^{-\alpha d_{ij}}} = \sqrt{A^2 \times PC}$$

À partir de ces métriques globales, il est possible d'évaluer à quel point chaque tache d'habitat contribue à la quantité d'habitat atteignable à l'échelle du réseau d'habitat. Cette contribution se calcule à l'aide de "delta-métriques". Elles mesurent la variation d'une métrique globale suite à la suppression d'un des nœuds ou des liens du graphe. Soit M une métrique globale et i une tache d'habitat, la contribution dM_i de la tache i à la métrique M est :

$$dM_i = \frac{M - M_{-i}}{M},$$

avec M_{-i} la valeur de M calculée dans le graphe paysager ne contenant pas la tache i . Ce type de calcul permet d'identifier les taches du réseau d'habitat qui contribuent le plus à sa connectivité et trouve donc des applications pour la mise en œuvre de mesures de conservation de la connectivité. Néanmoins, les delta-métriques calculées à partir des métriques globales identifient fréquemment les plus grandes taches comme les taches les plus importantes pour la connectivité du réseau. En d'autres termes, elles ont tendance à négliger l'importance des taches pour les flux de dispersion ou la tra-

versabilité du réseau d’habitat. Pour remédier à ce problème, [Saura et Rubio \(2010\)](#) ont proposé de décomposer la valeur des delta-métriques pour isoler les différentes contributions des taches d’habitat à la quantité d’habitat atteignable. Ils se sont basés sur la métrique PC et ont défini la décomposition suivante de la contribution dPC_i de la tache i à la métrique PC ³¹ :

$$dPC_i = dPC_{i.area} + dPC_{i.flux} + dPC_{i.connector}$$

En effet, lorsqu’une tache est supprimée, la métrique PC diminue car :

- La perte de la tache i supprime la contribution de cette tache à la connectivité intra-tache ($a_i a_i \times e^0 = a_i^2$).
- La perte des liens impliquant la tache i supprime les connexions entre cette tache et les autres ($\sum_{j=1, j \neq i}^n a_i a_j e^{-\alpha d_{ij}}$).
- La perte des liens impliquant la tache i supprime sa contribution aux connexions indirectes entre taches éloignées se faisant via des chemins passant par la tache i ($\sum_{j=1}^n \sum_{k=1}^n a_j a_k e^{-\alpha d_{jk}}$ avec $j, k \in \{1, \dots, n\}$, $k < j$ et $i \in P_{jk}$ tel que P_{jk} est l’ensemble des taches traversées par le chemin le plus court entre j et k).

On retrouve ici les trois composantes de la quantité d’habitat atteignable pouvant influencer les processus écologiques (recrutement, flux, traversabilité). En effet, alors que la composante dPC_{area} est fortement liée à la capacité des taches, les composantes dPC_{flux} et $dPC_{connector}$ sont proches des métriques F et BC respectivement³². Les valeurs des delta-métriques sont par construction dépendantes du degré de connectivité de l’ensemble du réseau d’habitat et exprimées comme des fractions d’une métrique globale. De plus, elles sont calculées en considérant une échelle spatiale unique alors que ces différentes composantes sont censées refléter le rôle d’une tache dans des processus écologiques s’effectuant à des échelles spatiales et temporelles différentes. Cela limite leur utilisation pour l’explication de réponses biologiques. L’utilisation des métriques de capacité, F et BC , calculées indépendamment, semble donc plus appropriée dans ce dernier cas.

Nous n’avons présenté ici qu’un faible nombre de métriques calculées à partir d’un graphe paysager au regard de la variété des métriques disponibles. Des travaux ont été menés pour guider les utilisateurs des graphes paysagers dans le choix des métriques à considérer dans leurs analyses ([Baranyi et al., 2011](#) ; [Rayfield et al., 2011](#)). Ces travaux ont conclu qu’il était inutile de considérer un grand nombre de métriques différentes, compte tenu de leurs redondances. Ils conseillent au contraire de se focaliser sur trois métriques reflétant de façon complémentaire les différentes contributions des taches à la quantité d’habitat atteignable identifiées précédemment ([Baranyi et al., 2011](#)). Cela permet d’interpréter les valeurs des métriques au regard des processus écologiques que le réseau d’habitat rend possibles.

En parallèle des métriques calculées au niveau des nœuds ou de l’ensemble du graphe paysager, il est possible de calculer des matrices de distances supposant que les déplacements s’effectuent uniquement en suivant les liens du graphe. On les calcule en additionnant pour chaque paire de nœuds les distances paysagères associées à chacun des liens formant le chemin le plus court entre deux nœuds.

31. La notation suivante est celle du manuel d’utilisation du logiciel Graphab, disponible en suivant ce [lien](#), et ne reflète pas exactement celle de l’article original de [Saura et Rubio \(2010\)](#).

32. Ces métriques ne sont pas exactement équivalentes néanmoins.

Enfin, au sein des réseaux d'habitat, on peut identifier des sous-ensembles de taches davantage connectées les unes aux autres qu'avec les taches d'habitat d'autres sous-ensembles. Ce type de topologie est mis en évidence par une partition du graphe basée sur un indice de modularité. Les indices de modularité calculés à partir des graphes paysagers ont été adaptés pour prendre en compte les probabilités de dispersion entre les taches ainsi que leurs capacités. C'est en utilisant ce type d'indices que [Foltête et Vuidel \(2017\)](#) ont identifié les modules de taches d'habitat prairial expliquant en partie la synchronicité des pullulations de Campagnol terrestre (*Arvicola terrestris*) en Franche-Comté.

Maintenant que nous avons vu comment construire et analyser les graphes paysagers, nous allons voir quels sont les principaux intérêts et limites liés à leur utilisation.

2.2.3 Intérêts et limites des graphes paysagers

Intérêts de l'utilisation des graphes paysagers

Les graphes paysagers sont un outil d'analyse de la connectivité fonctionnelle des habitats relativement facile à mettre en œuvre ([Calabrese et Fagan, 2004](#)). Ils offrent aux gestionnaires de la biodiversité un moyen efficace d'obtenir des réponses aux enjeux de conservation dont ils ont la charge, à partir de peu de données et connaissances initiales. De plus, l'utilisation de données spatiales de plus en plus précises permet de tenir compte de l'intégralité des taches d'habitat potentielles de la zone d'étude et de l'ensemble de la matrice paysagère dans l'analyse.

[Foltête et al. \(2014\)](#) ont présenté la façon dont un graphe paysager pouvait être mis en œuvre dans une logique opérationnelle. Ils ont illustré ces usages en montrant comment (i) identifier les étangs et mares les plus importants pour la connectivité des habitats d'espèces d'amphibiens, (ii) localiser les zones dans lesquelles il serait le plus favorable de restaurer des étangs ou des mares et (iii) évaluer l'impact d'une infrastructure de transport sur la connectivité des habitats³³. Cet exemple illustre concrètement comment les graphes paysagers peuvent être utilisés pour la conservation de la biodiversité.

Par ailleurs, les gestionnaires interrogés par [Bergsten et Zetterberg \(2013\)](#) ont souligné l'intérêt des graphes paysagers pour cartographier de façon explicite les enjeux de connectivité. Cela facilite la sensibilisation des décideurs à l'origine des mesures de conservation de la connectivité. L'intérêt "cartographique" des graphes paysagers explique probablement leur intégration dans les analyses de connectivité des documents de planification territoriale dans le cadre de la politique française de Trame Verte et Bleue ([Clauzel et Bonneville, 2019](#) ; [Foltête et al., 2012a](#)). Parallèlement, les graphes paysagers peuvent faciliter la mise en œuvre de la séquence Éviter-Réduire-Compenser (ERC) visant à limiter l'impact écologique de l'aménagement du territoire français ([Bergès et al., 2019](#) ; [Tarabon et al., 2020a](#)).

[Galpern et al. \(2011\)](#) et [Foltête et al. \(2021\)](#) listent un ensemble de questions traitées à l'aide de graphes paysagers, ainsi que les exemples d'application correspondants. On y retrouve à la fois des questions liées à la conservation de la biodiversité que des questions d'écologie plus théoriques. C'est probablement la flexibilité de la construction et de l'analyse des graphes paysagers qui justifie

33. Cette étude concernait la Rainette verte (*Hyla arborea*), le Crapaud calamite (*Bufo calamita*) et le Sonneur à ventre jaune (*Bombina variegata*) dans le contexte de la construction d'une ligne ferroviaire à grande vitesse à proximité de la ville de Belfort (LGV Rhin-Rhône).

leur utilisation dans une si vaste gamme de situations. [Fletcher et al. \(2016\)](#) attribuent l'utilisation fréquente des graphes paysagers ces dernières années à leur prise en compte de la surface des taches et des flux indirects dans des métriques de connectivité basées sur la théorie des métapopulations et décomposables en plusieurs composantes fonctionnelles. Néanmoins, ces outils de modélisation possèdent également un certain nombre de limites, que nous abordons à présent.

Limites de l'utilisation des graphes paysagers

Bien que présentant beaucoup d'intérêts, l'utilisation des graphes paysagers ne se justifie pas toujours ([Urban et Keitt, 2001](#)). Certaines espèces occupent des micro-habitats qu'il n'est pas possible de cartographier de façon exhaustive pour la réalisation d'un graphe paysager (par exemple : les arbres morts constituant l'habitat de la faune saproxylique). À l'inverse, des espèces s'établissent dans des habitats si vastes et de natures si variées qu'il est impossible de discrétiser leurs taches d'habitat. Pour ce type d'espèces souvent généralistes, il est plus pertinent de considérer le paysage comme un "continuum" au sein duquel la qualité de l'habitat varie ([Fischer et Lindenmayer, 2006](#)). Enfin, quand bien même l'espèce étudiée est spécialiste d'un habitat que l'on peut cartographier précisément, si cet habitat occupe la majeure partie du paysage, l'analyse de connectivité basée sur les graphes paysagers sera peu informative. L'utilisation des graphes paysagers se limite donc aux espèces formant des populations au sein de taches d'habitat subdivisées.

Lorsque l'utilisation des graphes est adaptée à l'espèce étudiée, les choix arbitraires réalisés par le modélisateur limitent potentiellement la validité de l'analyse ([Correa Ayram et al., 2016](#) ; [Galpern et al., 2011](#)). La flexibilité des méthodes de construction et d'analyse des graphes devient alors potentiellement une faiblesse. En effet, si les choix sont faits sur la base d'une vision erronée de la réalité écologique, l'analyse ne sera pas pertinente. De même, [Moilanen \(2011\)](#) liste les limites de l'utilisation des graphes paysagers en ciblant :

1. l'intérêt incertain et potentiellement limité des métriques issues des graphes en comparaison à des métriques basées sur la théorie des métapopulations utilisées antérieurement,
2. l'incertitude associée à la délimitation des taches et à l'élagage des liens,
3. les temps de calcul importants dès lors que la zone d'étude est grande et sa résolution fine,
4. l'importance démesurée accordée selon lui à la connectivité des habitats.

La troisième limite mentionnée par cet auteur est aujourd'hui moins problématique compte tenu de l'augmentation des capacités de calcul et de leur optimisation. Quant à la dernière, elle n'est pas liée aux graphes paysagers en eux-mêmes mais davantage à leur interprétation ou à leur utilisation exclusive. En revanche, il n'est pas possible de déterminer à quel point les métriques utilisées et les choix de construction du graphe sont pertinents sans les confronter à la réalité des réponses biologiques mesurées sur le terrain ([Correa Ayram et al., 2016](#) ; [Galpern et al., 2011](#) ; [Kadoya, 2009](#)). En effet, il est nécessaire de valider les méthodes de construction et d'analyse des graphes paysagers à l'aide de données biologiques. Depuis la publication de l'article de [Moilanen \(2011\)](#), le nombre d'études utilisant en même temps des graphes paysagers et des données biologiques a augmenté. Il convient de s'y intéresser pour comprendre comment elles ont renforcé la validité écologique de ces outils tout en identifiant les questions de recherche auxquelles les réponses manquent encore.

2.3 Modalités d'utilisation conjointe de données biologiques et de graphes paysagers

Comme nous l'avons vu, les graphes paysagers sont utilisés à la fois (i) pour améliorer nos connaissances théoriques en écologie et (ii) pour accompagner la mise en œuvre de mesures de conservation de la biodiversité. Les méthodes de construction et d'analyse utilisées dans ces deux cas sont souvent différentes. Ces différences devraient donc se retrouver lorsque des données biologiques sont intégrées à l'utilisation des graphes paysagers. Dans ce dernier cas, on peut se demander s'il y a un lien entre l'objectif principal des études menées à l'aide de graphes paysagers, la façon dont les graphes sont construits et analysés, le type de données biologiques utilisées et la façon dont ces données sont intégrées à l'analyse. Dans le cadre d'un travail collectif mené par l'équipe de recherche rassemblée autour du logiciel Graphab et piloté par Jean-Christophe Foltête, nous avons réalisé une synthèse bibliographique systématique pour répondre à cette question (Annexe A1 ³⁴). L'objectif de cette synthèse était d'identifier les raisons pour lesquelles les données biologiques étaient intégrées aux analyses afin de dégager une typologie des utilisations de graphes paysagers. L'hypothèse formulée en amont de cet état de l'art était la suivante :

L'objectif des études détermine la façon dont les données biologiques sont considérées, créant un contraste entre les applications opérationnelles lors desquelles les graphes paysagers sont construits de façon simple (à partir d'une carte d'occupation du sol) par souci d'efficacité et les applications scientifiques lors desquelles les chercheurs intègrent les données biologiques aux analyses par une méthodologie visant à maximiser l'ajustement du modèle à la réalité écologique.

Cette étude a révélé la vaste gamme de méthodes de construction et d'analyse des graphes paysagers mises en œuvre ces dernières années, tout en questionnant leurs limites. Nous résumons à présent la méthodologie employée pour cette étude et ses principaux résultats.

Méthodologie de l'analyse bibliographique systématique

La première étape de ce travail a consisté à identifier le corpus d'articles scientifiques dans lesquels les auteurs avaient utilisé à la fois des graphes paysagers et des données biologiques. La sélection a été faite au sein de la base de données en ligne Scopus rassemblant les articles de la littérature scientifique anglophone. La requête utilisée assurait que les articles satisfaisaient les quatre critères suivants :

1. Présence d'un terme apparenté à la notion de "réseau écologique" dans le titre, l'abstract ou les mots clés associés à l'article (*landscape* ou *habitat* ou *ecological* ou *patch* et *graph* ou *network*)
2. Présence d'un terme associé aux différents types de données biologiques utilisables dans de telles études (par exemple *presence* ou *genetic*, parmi d'autres)
3. Étude concernant la biodiversité terrestre
4. Citation d'au moins un des six articles de référence relatifs aux graphes paysagers, à savoir : [Urban et Keitt \(2001\)](#), [Saura et Pascual-Hortal \(2007\)](#), [Urban et al. \(2009\)](#), [Bunn et al. \(2000\)](#), [Minor et Urban \(2008\)](#) et [Pascual-Hortal et Saura \(2006\)](#)

34. Ce travail a donné lieu à un article publié dans la revue *Landscape Ecology* en mars 2020 : Foltête, J. C., Savary, P., Clauzel, C., Bourgeois, M., Girardet, X., Sahraoui, Y., Vuidel, G. & Garnier, S. 2020. Coupling landscape graph modeling and biological data : a review. *Landscape Ecology*, 35(5), 1035-1052. Cette étude s'est déroulée de façon concomitante à ce travail de thèse et en justifie les objectifs. La version intégrale de cet article constitue l'Annexe A1.

Ensuite, chacun des articles remplissant ces critères a été analysé indépendamment par au moins deux des auteurs de l'étude, afin de les décrire en suivant une grille d'analyse prédéfinie (Figure 18). Les principaux critères retenus étaient :

- Le type de données biologiques
- Le contexte géographique et biologique
- L'objectif de l'étude et de l'utilisation des données biologiques
- L'implication de gestionnaires de la biodiversité
- Les méthodes de construction du graphe
- Le type d'analyses de connectivité réalisées et les métriques utilisées (le cas échéant)

<p>Données biologiques</p> <ul style="list-style-type: none"> - Présence - Div. spécifique - Div. génétique - Abondance - Télémétrie - Autres 	<p>Objectif</p> <ul style="list-style-type: none"> - Compréhension théorique - Priorisation pour la conservation - Evaluation d'impact - "Amélioration" de la fonctionnalité du réseau écologique - Développement méthodologique <p><i>Lien entre connectivité et données biologiques</i></p> <ul style="list-style-type: none"> - Pas de mise en relation - Corrélation, description $\text{cor}(Y_{\text{bio}}, X_{\text{connect}})$ - Modélisation $Y_{\text{bio}} \sim X_{\text{connect}}$ <p><i>Utilisation des données biologiques</i></p> <ul style="list-style-type: none"> - Pour créer le graphe (<i>a priori</i>) - Après création (<i>a posteriori</i>) <p><i>Temporalité de l'analyse</i></p> <ul style="list-style-type: none"> - Analyse statique (présent) - Analyse rétrospective - Analyse prospective 	<p>Construction du graphe</p> <p><i>Création des noeuds</i></p> <ul style="list-style-type: none"> - A partir de l'OS - A partir d'un SDM - A partir d'aires protégées <p><i>Pondération des noeuds</i></p> <ul style="list-style-type: none"> - Poids uniforme - Surface de la tache - Qualité de la tache (SDM) <p><i>Pondération des liens</i></p> <ul style="list-style-type: none"> - Distance géodésique euclidienne - Distances-coût basée sur l'OS - Distances-coût basées sur le SDM <p>Analyses de connectivité</p> <ul style="list-style-type: none"> - Au niveau du graphe global - Au niveau des composantes - Au niveau des noeuds - Au niveau des liens <p><i>Utilisation de métriques</i></p> <ul style="list-style-type: none"> - PC, IIC, dPC, dIIC ? - BC, F, Dg, ECS ?
--	--	---

FIGURE 18 – Grille utilisée pour analyser les articles couplant graphes paysagers et données biologiques. Chaque catégorie de critères est notée en gras. Les critères à choisir pour chacune d'elles sont précédés d'un tiret. SDM : modèle de distribution d'espèce (*Species Distribution Model*), OS : occupation du sol. Dg : degré d'un nœud, ECS : taille moyenne d'une composante du graphe paysager (*Expected Cluster Size*).

Les données obtenues à l'issue de cette analyse ont été soumises à trois Analyses des Correspondances Multiples (ACM) (Tenenhaus et Young, 1985) étant donnée la nature qualitative des critères étudiés. Ces différentes ACM avaient pour objectifs d'étudier le lien entre le type de données biologiques (dans chaque ACM) et, successivement et indépendamment :

1. Le continent, le milieu, le taxon et le type de mouvements concernés par l'étude
2. L'objectif de l'étude, l'implication des gestionnaires et la méthode de construction du graphe
3. Les analyses de connectivité réalisées

La seconde ACM étant celle qui impliquait le plus de variables différentes, elle a été suivie d'une Classification Ascendante Hiérarchique (CAH) pour identifier une typologie d'articles.

Résultats de l'analyse systématique des articles satisfaisant les critères de sélection

L'application des critères de sélection et une relecture finale ont permis d'identifier 71 articles publiés entre 2005 et 2018 pour cette analyse systématique. La majorité d'entre eux ont été publiés après 2012. Les données de présence étaient les plus utilisées dans ces articles, suivies des données

d'abondance et des données génétiques ou de **télémetrie** (Figure 19A). Les études réalisées en Europe (52.1 %) et Amérique du Nord (23.9 %) dominaient, tout comme celles concernant les mammifères (35.2 %) et les oiseaux (23.9 %). De même, les études concernant les milieux forestiers (57.7 %) étaient majoritaires. L'ACM concernant le lien entre les données biologiques, le continent, le milieu, le taxon et le type de mouvements (dispersion, mouvements quotidiens) n'a pas révélé d'associations très fortes entre les modalités de ces catégories.

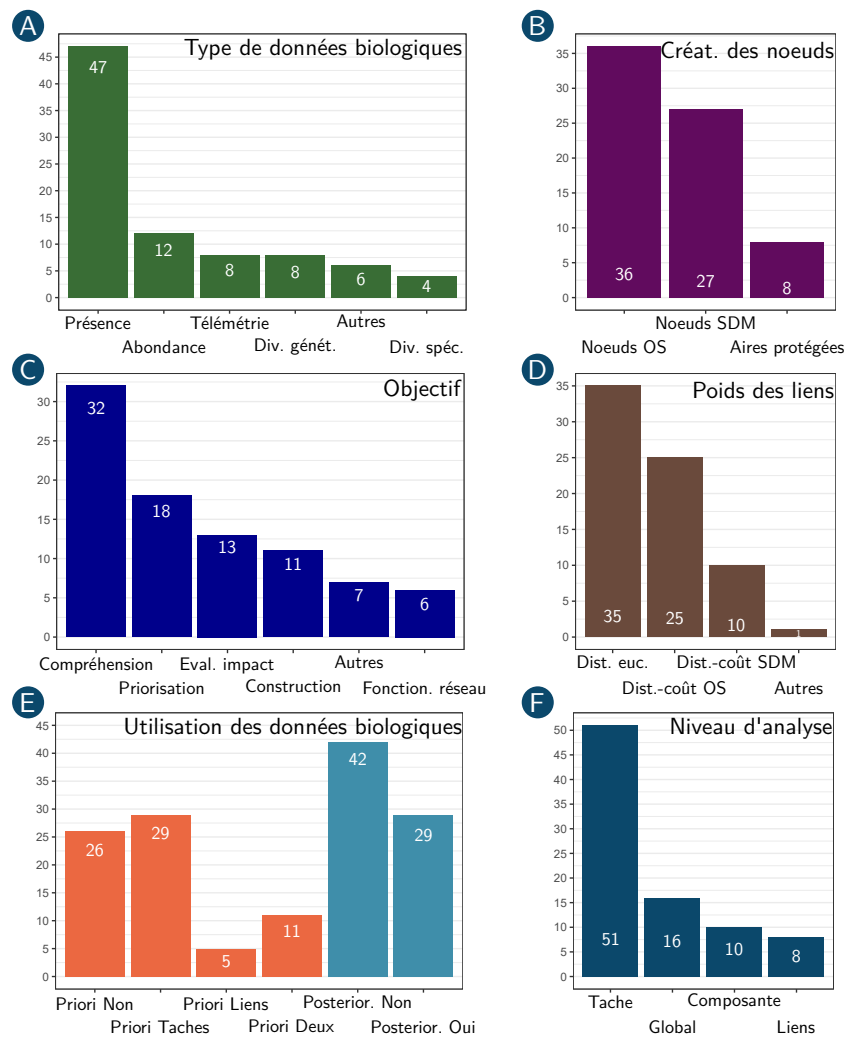


FIGURE 19 – Fréquence de quelques caractéristiques des études couplant graphes paysagers et données biologiques analysées par Foltête *et al.* (2020). (A) Type de données biologiques utilisées. (B) Données utilisées pour définir les nœuds du graphe. OS : occupation du sol. SDM : modèle de distribution d'espèce. 'Aires protégées' signifie que les nœuds du graphe sont basés sur les limites d'aires protégées. (C) Objectifs de l'analyse. (D) Type de pondération des liens du graphe. 'Dist-coût OS' et 'Dist-coût SDM' signifient respectivement que les coûts utilisés pour calculer les distances-coûts sont basés sur l'occupation du sol ou sur les probabilités de présence issues d'un SDM. (E) Moment de l'analyse auquel les données biologiques sont utilisées. 'Priori' signifie que les données biologiques sont utilisées avant ou pendant la construction du graphe. 'Priori Taches', 'Priori Liens' et 'Priori Deux' signifient que les données biologiques sont utilisées pour créer ou pondérer les taches, les liens ou les deux. 'Posteriori' signifie que les données biologiques sont confrontées aux métriques ou autres variables issues du graphe paysager après la construction du graphe. (F) Niveau auquel l'analyse du graphe est réalisée.

Près de la moitié des études avaient pour objectif la compréhension de l'écologie des espèces ou des communautés (Figure 19C). Un quart visaient à identifier les éléments du réseau écologique prioritaires pour la conservation, tandis que les autres cherchaient à évaluer un impact, améliorer la fonctionnalité du réseau écologique ou développer de nouvelles méthodes. Environ 20 % des études n'étaient pas basées uniquement sur le réseau actuel et cherchaient à en prédire l'évolution (12.7 %) ou s'intéres-

saient à ses évolutions passées (7.0 %). Même si les gestionnaires de la biodiversité étaient peu souvent co-auteurs des études (16.9 %), ils étaient intégrés à la démarche scientifique dans plus de la moitié des cas (57.7 %).

Les données biologiques étaient utilisées lors de la construction du graphe (*a priori*, avant l'analyse de connectivité) dans 45 articles (Figure 19E). Le plus souvent, il s'agissait de données de présence utilisées pour définir et/ou pondérer les nœuds du graphe suite à la réalisation d'un SDM (modèle de distribution d'espèce, *Species Distribution Model*) (Figure 19B). Les données biologiques servaient également dans ce cas à pondérer les liens (Figure 19D), en calculant les valeurs de coût à partir du SDM (15.5 %) ou plus rarement à partir de traces GPS (télémétrie) ou de mesures de **différenciation génétique** (8.4 %). À l'inverse, les données biologiques étaient utilisées après la construction du graphe (*a posteriori*) dans 42.2 % des articles. Dans ce cas, leur corrélation avec les métriques de connectivité était évaluée (14.1 %) ou, plus fréquemment (28.2 %), elles étaient les variables réponses de modèles dont les variables explicatives étaient issues du graphe paysager.

L'ACM analysant les associations entre données biologiques, objectifs des études, implications des gestionnaires et méthodes de construction et d'analyse des graphes a permis de distinguer des approches contrastées. Le premier facteur séparait les études dans lesquelles les données étaient intégrées à la construction du graphe de celles dans lesquelles elles étaient utilisées *a posteriori* (Figure 20). Dans le premier cas, les données étaient plus fréquemment des données de présence et les nœuds et liens du graphe étaient définis et/ou pondérés à l'aide d'un SDM. Cette construction se justifiait essentiellement par le caractère opérationnel des objectifs poursuivis. En revanche, dans le second cas, les éléments du graphe étaient principalement définis à partir de cartes d'occupation du sol. Des données de diversité spécifique ou d'abondance étaient confrontées au modèle de connectivité obtenu, dans un objectif de compréhension de leurs relations. Au-delà de ce premier contraste, le second facteur mettait en évidence un gradient lié à l'implication des gestionnaires de la biodiversité, qui semble plus fréquente dans les études utilisant des données de télémétrie ou génétiques (Figure 20). À l'inverse, ces acteurs étaient moins souvent impliqués dans les études d'impact.

La CAH appliquée à ces deux axes factoriels a identifié une typologie des études à quatre classes (Figure 21), que nous décrivons ainsi :

- Classe 1 : "Approche déductive, liens euclidiens" : les études visent à comprendre les relations entre la connectivité des habitats et les populations ou communautés d'espèces en étudiant la corrélation entre métriques de connectivité et réponses biologiques. Les données utilisées sont principalement des données de présence ou d'abondance. Les taches sont définies à partir de l'occupation du sol (90 % des cas) et les liens pondérés par des distances euclidiennes (80 %). Les gestionnaires sont peu souvent impliqués.
- Classe 2 : "Approche déductive, liens fonctionnels" : les objectifs sont les mêmes que pour la classe 1 mais les données biologiques sont plus variées. Les taches et les liens sont tous deux fréquemment définis et/ou pondérés à partir de l'occupation du sol. Les gestionnaires sont impliqués.
- Classe 3 : "Approche opérationnelle, évaluation d'impact" : les études visent à évaluer l'impact d'un changement paysager passé ou futur. Les taches sont souvent définies à partir d'un

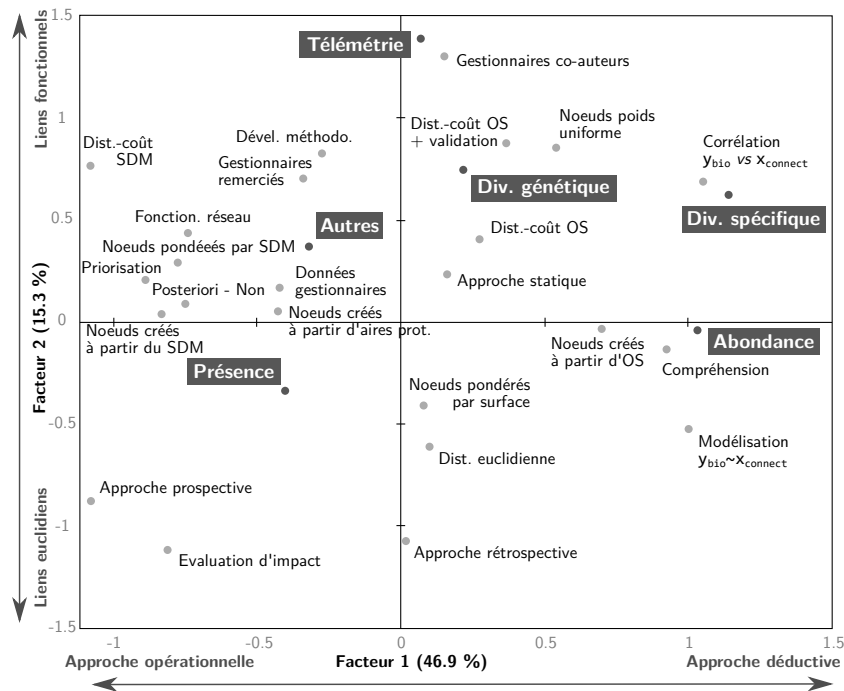


FIGURE 20 – Résultat de l'Analyse des Correspondances Multiples (ACM) des liens entre données biologiques, objectifs des études, implications des questionnaires et méthodes de construction et d'analyse des graphes paysagers. SDM : modèle de distribution d'espèces (*Species Distribution Model*), OS : occupation du sol.

SDM tandis que les liens sont pondérés par une distance euclidienne. Les questionnaires sont peu impliqués.

- Classe 4 : "Approche opérationnelle, fonctionnalité du réseau" : les études ont un objectif opérationnel et les questionnaires sont souvent impliqués. Les nœuds sont définis et les liens pondérés à partir d'un SDM.

Les analyses de connectivité réalisées dans ces différents contextes reposent le plus souvent sur le calcul de métriques de connectivité au niveau des nœuds (Figure 19F). Les métriques dPC , F , BC et le degré sont fréquemment utilisés dans ce cadre. Une analyse combinant les données biologiques et des variables issues du graphe est réalisée dans 42.2 % des cas. On distingue alors (i) les approches basées sur une modélisation statistique dans lesquelles les métriques de connectivité sont des variables explicatives de la réponse biologique (28.1 %) et (ii) les approches descriptives dans lesquelles la corrélation entre métriques de connectivité et réponses biologiques est simplement estimée ou visualisée (14.1 %).

Principaux enseignements de cet état de l'art

Malgré l'utilisation fréquente des graphes paysagers ces dernières années, déjà constatée par [Correa Ayram et al. \(2016\)](#) et [Fletcher et al. \(2016\)](#), le nombre d'études couplant graphes paysagers et données biologiques reste faible (71 articles pour plus de 500 articles utilisant des graphes paysagers selon notre analyse bibliographique). Lorsque ce couplage est réalisé, deux types de raisonnement scientifique sont adoptés (Figure 22) :

- Approche inductive : les données spatiales et les données biologiques sont utilisées simultanément pour construire le graphe paysager en supposant que ces observations particulières reflètent une réalité écologique plus générale (induction). Cette utilisation de données biologiques lors de la

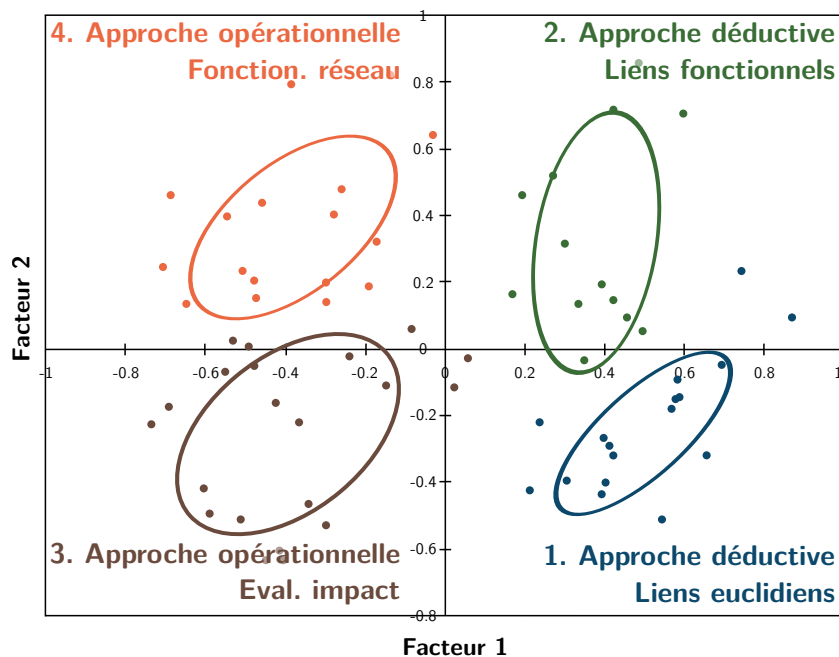


FIGURE 21 – Classification Ascendante Hiérarchique (CAH) réalisée à partir de l'ACM des liens entre données biologiques, objectifs, implications des questionnaires et méthodes de construction et d'analyse des graphes paysagers. Chaque point correspond à une étude couplant graphes paysagers et données biologiques ($n = 71$). Leur couleur dépend de la classe à laquelle chaque étude est associée lors de la CAH réalisée à partir des deux premiers axes factoriels de l'ACM. Les noms des classes ont été attribués en fonction des caractéristiques des études, visibles sur la figure 20.

construction est censée donner sa validité au graphe. On part du particulier (données) pour généraliser (graphe modélisant la connectivité) (Figure 22A).

- Approche hypothético-déductive : il s'agit d'une approche scientifique "classique". Un graphe paysager est réalisé selon des hypothèses concernant la réalité écologique (par déduction), par exemple des dires d'experts sur les types d'occupation du sol utilisés par l'espèce et ses capacités de dispersion. L'adéquation de ce graphe avec les données biologiques est vérifiée en les confrontant *a posteriori* aux métriques de connectivité (Figure 22B). Cela permet de confronter les hypothèses de départ à des observations particulières pour tester leur validité.

Les approches inductives sont plus souvent employées lorsque l'étude a des visées opérationnelles (classes 3 et 4, Figure 21). L'intégration des données biologiques à la construction du graphe se fait alors de façon complexe, en se basant le plus souvent sur un SDM créé à partir de données de présence (Duflot *et al.*, 2018). À l'inverse, les approches hypothético-déductives sont davantage employées à des fins de compréhension théorique (classes 1 et 2, Figure 21). Dans ce cas, la construction des graphes se fait de façon assez simple, à partir de dires d'expert et de données d'occupation du sol. Les métriques de connectivité obtenues servent alors à expliquer les réponses biologiques (voir Martín-Queller et Saura (2013) et Mony *et al.* (2018) par exemple). Ainsi, en contraste avec notre hypothèse de départ, les méthodes de construction des graphes les plus complexes ne sont pas employées à des fins de connaissance scientifique mais au contraire à des fins opérationnelles.

Selon les objectifs, on remarque que les postulats concernant la connectivité des habitats et sa modélisation sont différents. Dans le cas de l'approche inductive, on part du principe que les réponses biologiques sont influencées par la connectivité, si bien que leur utilisation permet de modéliser correctement cette connectivité. L'utilisation du modèle obtenu vise à orienter des mesures de conservation dont l'objectif est de conserver la connectivité en raison de son influence sur les réponses biologiques.

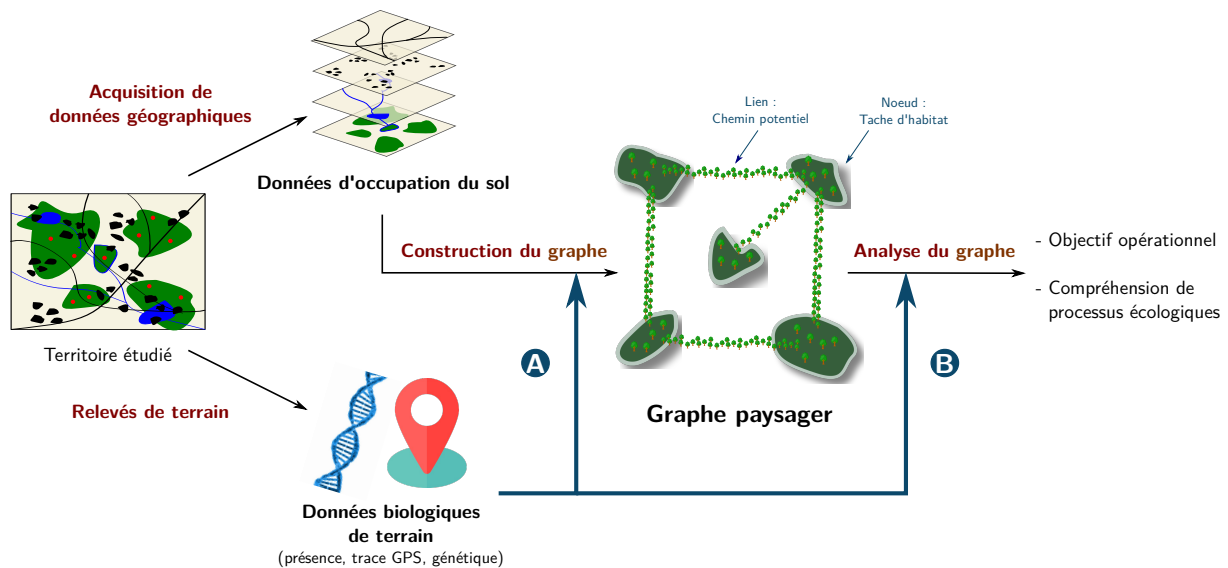


FIGURE 22 – Utilisation conjointe de données biologiques et de graphes paysagers. Les données biologiques peuvent être intégrées à l’analyse pour guider la construction du graphe (A, *a priori*) ou confrontées aux variables issues du graphe après sa construction (B, *a posteriori*). Valider le graphe construit en suivant l’approche A implique d’utiliser un jeu de données indépendant et de le confronter au graphe via l’approche B.

Dans le cas de l’approche hypothético-déductive, on suppose que les graphes paysagers reflètent correctement la connectivité des habitats, auquel cas on peut tester l’influence de cette connectivité sur les réponses biologiques. La validité de ces approches suppose toutefois que l’on réponde aux questions suivantes : (i) dans quelle mesure la connectivité des habitats influence les processus écologiques à l’origine des réponses biologiques mesurées ?, (ii) est-ce que les graphes paysagers reflètent correctement la connectivité des habitats ? Pour y répondre, il serait nécessaire d’utiliser deux types de données biologiques et d’adopter tour à tour ces deux approches. Seules quelques rares études fournissent de tels exemples (Bergerot *et al.*, 2013 ; Galpern *et al.*, 2012 ; O’Brien *et al.*, 2006 ; Ribeiro *et al.*, 2011). Par exemple, Galpern *et al.* (2012) ont défini les coûts de déplacements à partir de données de télémétrie, avant de confronter leur modèle de connectivité à la structure génétique de populations de Caribou (*Rangifer tarandus caribou*).

Les études se distinguent également par la façon dont les liens des graphes sont définis et pondérés. D’une part, les liens sont pondérés par de simples distances euclidiennes (classes 1 et 3, Figure 21) quand d’autre part ils sont pondérés en assignant des valeurs de résistance à la matrice paysagère en fonction des types d’occupation du sol ou des probabilités de présence issues d’un SDM (classes 2 et 4, Figure 21). C’est dans la seconde situation que se trouvent tous les exemples d’utilisation de données de télémétrie. Elles sont souvent fournies par des gestionnaires d’aires protégées d’Amérique du Nord, expliquant le lien entre implication des gestionnaires et utilisation des données biologiques *a priori*.

Bien que ce ne soit pas la prise en compte de la connectivité fonctionnelle *vs* structurelle qui distingue les études à visée scientifique des études à visée opérationnelle, on perçoit une différence dans la façon dont la connectivité fonctionnelle est modélisée dans ces deux situations. Les études à visée opérationnelle intègrent fréquemment les coûts de déplacement des espèces en utilisant un SDM créé à partir de données de présence. À l’inverse, les données reflétant plus directement les mouvements des espèces (télémétrie, génétique) sont davantage utilisées pour la compréhension théorique de l’influence de la connectivité. Alors que la littérature scientifique indique que ce dernier cas permet de modéliser

la connectivité de façon relativement correcte (Zeller *et al.*, 2018), le premier cas pose néanmoins question. En effet, des études ont montré que les données de présence ne reflétaient pas toujours la façon dont les espèces se déplacent (Keeley *et al.*, 2017). Leur utilisation pour la définition des taches semble plus appropriée. En effet, l'objectif premier de la réalisation d'un modèle de distribution d'espèce est de localiser les zones d'habitat de l'espèce.

Certains types de données biologiques, telles que les données génétiques et les données concernant la composition des communautés d'espèces, peuvent être confrontées aux variables issues du graphe paysager à plusieurs niveaux. En effet, la connectivité écologique des habitats est censée influencer la diversité spécifique et la diversité génétique mesurées au niveau des taches d'habitat ainsi que les différences génétiques et de composition des communautés entre les taches. Néanmoins, ce sont surtout les liens entre (i) diversité spécifique locale et métriques de connectivité ou entre (ii) différenciation génétique et distances paysagères qui sont étudiés. Cela limite ainsi notre compréhension de l'influence de la connectivité des habitats sur la diversité génétique locale ou sur les différences de composition des communautés d'espèces (diversité β). La similitude des mécanismes à l'origine de la structure génétique des populations (**flux génétique, dérive génétique**) et de la composition spécifique des communautés (extinction, colonisation) (Vellend et Geber, 2005) et leurs liens avec la connectivité des habitats devraient faire l'objet de nouvelles analyses.

Notre analyse indique donc que les couplages de graphes paysagers et de données biologiques sont encore rares. Quelle que soit l'approche utilisée pour réaliser ce couplage, les postulats sur lesquels elle repose nécessitent d'être validés. Pour cela, il faudrait par exemple vérifier que l'approche courante consistant à construire un graphe paysager à partir de données de présence est valide d'un point de vue écologique. Tester cela impliquerait de confronter le graphe à un jeu de données biologiques indépendant reflétant plus directement les mouvements des espèces, telles que des données génétiques ou de télémétrie. Construire le graphe à partir de ce type de données pourrait également être une option valide. Par ailleurs, lorsque les données biologiques sont confrontées aux graphes paysagers pour améliorer notre compréhension de l'influence de la connectivité des habitats sur les réponses biologiques, il semble nécessaire de mener l'analyse à différents niveaux. Cela semble particulièrement important lorsqu'on étudie la composition des communautés d'espèces ou la structure génétique des populations. En effet, on peut étudier leur diversité au niveau de chaque communauté ou population, mais aussi étudier les différences spécifiques ou génétiques entre communautés et entre populations, respectivement. Or, comprendre les déterminants de ces deux types de réponse biologique à ces différents niveaux est crucial du point de vue de la conservation de la biodiversité.

L'utilisation conjointe de données génétiques et de graphes paysagers permettrait d'une part de valider les méthodes de construction et d'analyse des graphes paysagers, et d'autre part d'approfondir notre compréhension de l'influence de la connectivité des habitats sur la structure génétique des populations. En effet, ce type de données a l'avantage de refléter la dispersion d'individus qui sont parvenus à se reproduire. De plus, contrairement aux données de télémétrie qui dépendent fortement du comportement particulier des individus au moment précis de leur suivi, la structure génétique évaluée à partir de prélèvements génétiques individuels est le résultat des mouvements de plusieurs individus au cours de plusieurs générations. Même si ce dernier point empêche parfois d'estimer la connectivité contemporaine, l'utilisation de ces données reste malgré tout un moyen particulièrement efficace pour évaluer la connectivité (Zeller *et al.*, 2018). En outre, les données génétiques sont de plus en plus accessibles en raison du développement des techniques de biologie moléculaire. Ces techniques permettent aujourd'hui d'extraire de l'ADN de matériaux biologiques ne nécessitant pas la capture des individus (poils, fèces, plumes par exemple, voir Afonso *et al.* (2016) ou Lehnen *et al.* (2021)), les distinguant des suivis par télémétrie nécessitant de capturer et d'équiper les individus.

Dans ce contexte, cette thèse a pour objectif principal d'identifier des méthodes permettant l'utilisation conjointe de données génétiques et de graphes paysagers. Nous allons désormais présenter comment la connectivité des habitats est censée influencer la structure génétique des populations et comment analyser cette influence.

3 La structure génétique des populations, une réponse biologique influencée par la connectivité des habitats

3.1 La structure génétique des populations

3.1.1 Diversité génétique locale et différenciation génétique

Il existe plusieurs façons de quantifier l'ampleur des variations génétiques ([Encadré 4](#)) au sein d'une espèce. Comptabiliser le nombre total d'allèles différents existant pour chacun des différents *loci* informe sur la richesse allélique intra-spécifique totale, et donc sur un des niveaux d'organisation de la biodiversité définis précédemment (cf. section 1.1). Néanmoins, cette mesure exhaustive de richesse allélique est très coûteuse à obtenir et on ne peut généralement considérer qu'un échantillon d'individus chez lesquels on étudie quelques *loci*. En génétique des populations, on considère des individus issus d'une ou plusieurs populations. Bien qu'il existe un grand nombre de définitions du terme "population" dans cette discipline ([Waples et Gaggiotti, 2006](#)), nous définissons ici une **population** comme un groupe d'individus vivant suffisamment proches les uns des autres pour que chaque individu puisse potentiellement se reproduire avec un individu de sexe opposé de ce même groupe, à l'instar de [Frankham et al. \(2017\)](#). La diversité génétique locale, dans une population donnée, est estimée par le nombre d'allèles différents qu'on y observe (richesse allélique) et par la proportion d'individus hétérozygotes ([Figure 23](#)). Lorsque le nombre d'individus échantillonnés n'est pas le même dans chaque population, des méthodes basées sur le principe de raréfaction permettent d'estimer la richesse allélique qu'on devrait observer dans chacune des populations si on n'y avait échantillonné partout autant d'individus ([Szpiech et al., 2008](#)). Dans une population, chaque allèle i est présent avec une certaine fréquence f_i . Une autre façon de mesurer la diversité consiste alors à estimer la probabilité H_e qu'un individu soit hétérozygote compte tenu des fréquences alléliques ([Nei, 1978](#)) et en supposant que l'équilibre de Hardy-Weinberg est respecté ([Encadré 5](#)) :

$$H_e = 1 - \sum_{i=1}^k f_i^2,$$

avec k le nombre d'allèles au *locus* considéré ([Hamilton, 2011](#)). Notons que lorsque l'équilibre de Hardy-Weinberg est vérifié, l'hétérozygotie attendue H_e et l'hétérozygotie observée H_o sont censées être égales.

Outre la diversité génétique présente dans une population, les généticiens s'intéressent aux différences de patrimoine génétique observées entre plusieurs populations. Aussi, dans la suite de cette thèse, nous utiliserons le terme "structure génétique" pour désigner la façon dont la variabilité génétique se répartit dans un ensemble de populations. Pour désigner les deux composantes de la **structure génétique des populations**, nous emploierons les termes de (i) "**diversité génétique**" pour désigner la diversité génétique locale, c'est-à-dire au sein d'une population et (ii) celui de "**différenciation génétique**" pour désigner les différences génétiques entre populations ([Figure 23](#))³⁵.

35. Une telle définition de la structure génétique ne fait pas consensus dans la littérature. Parfois, structure génétique y est synonyme de différenciation génétique. Nous faisons ce choix terminologique de façon à englober dans un même terme diversité et différenciation génétiques, de la même façon que dans les sections précédentes, la structure spatiale des habitats incluait leur quantité et leur configuration.

Tout comme il existe plusieurs façons de quantifier la diversité génétique locale, il existe plusieurs mesures de différenciation génétique entre populations (Jost, 2008 ; Meirmans et Hedrick, 2011 ; Verity et Nichols, 2014). Nous nous sommes particulièrement focalisés sur deux d'entre elles dans le cadre de cette thèse, à savoir le F_{ST} et le D_{PS} , couramment utilisées dans notre discipline. Le F_{ST} est un des indices de fixation développés par Sewall Wright pour analyser le partitionnement de la variabilité génétique entre populations (Wright, 1931). Cet indice se base sur l'écart entre les taux d'hétérozygotie observés dans deux populations et celui qu'on mesurerait si ces deux populations n'en formaient qu'une. Il quantifie ainsi le degré de divergence de leurs fréquences alléliques. Le D_{PS} (Bowcock *et al.*, 1994) se base quant à lui sur la proportion d'allèles qui sont partagés par des individus. Initialement conçu pour mesurer la différenciation génétique entre deux individus, le D_{PS} a été adapté au calcul de la différenciation entre deux populations i et j (Waits et Storfer, 2015) :

$$D_{PS} = 1 - \frac{\sum_d^D \sum_k^K \min(f_{a_{kd,i}}, f_{a_{kd,j}})}{D},$$

tel que a_{kd} est l'allèle k au *locus* d , $f_{a_{kd,i}}$ est la fréquence de a_{kd} dans la population i , D est le nombre total de *loci* et K est le nombre d'allèles à chaque *locus*. Nous verrons que cette mesure se révèle utile pour comprendre l'influence de la connectivité des habitats sur la différenciation génétique entre populations.

Enfin, d'autres indices génétiques constituent des intermédiaires entre les mesures de diversité et de différenciation. Par exemple, la richesse allélique privée mesure le nombre d'allèles présents dans une population mais absents de toutes les autres (Szpiech *et al.*, 2008) (Figure 23). De façon similaire, la richesse allélique manquante correspond au nombre d'allèles absents d'une population alors qu'ils sont présents partout ailleurs.

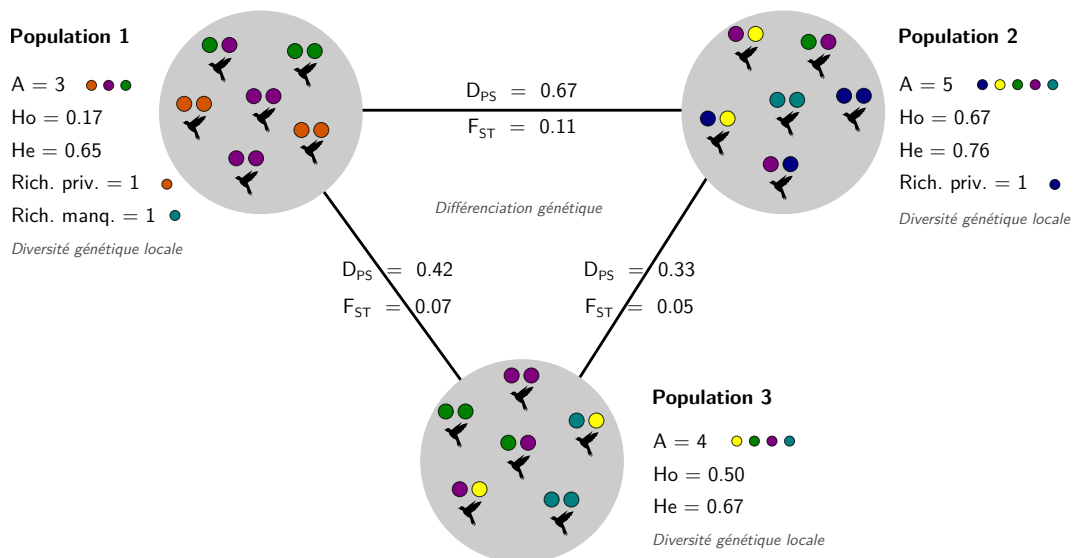


FIGURE 23 – Diversité génétique locale et différenciation génétique. Chaque disque gris correspond à une population de six individus. Les points colorés représentent les versions que possède chaque individu pour un *locus* et les couleurs correspondent donc chacune à un allèle différent. A : richesse allélique, H_o et H_e : hétérozygotie observée et attendue, respectivement, Rich. priv. : richesse privée, Rich. manq. : richesse manquante, F_{ST} et D_{PS} sont deux mesures de différenciation génétique.

Encadré 4 : L'ADN, support de l'information génétique

Chacune des cellules d'un être vivant contient des molécules d'ADN (acide désoxyribonucléique)^a constituées d'une séquence de nucléotides reliés chacun à un sucre (désoxyribose) et à un groupe phosphate. Il n'existe que quatre nucléotides (A, T, C, G pour Adénine, Thymines, Cytosine et Guanine, respectivement) et ce sont les séquences formées par leurs combinaisons qui définissent l'information génétique. Chaque brin d'ADN est apparié à son brin complémentaire et l'ensemble s'enroule pour former la double-hélice d'ADN. Le patrimoine génétique d'un individu, ou génotype, est constitué de plusieurs double-hélices d'ADN formant des chromosomes dont le nombre est propre à chaque espèce. Chez les organismes diploïdes, les chromosomes sont présents par paires homologues^b. Chaque chromosome d'une paire est transmis par l'un des deux parents par le biais des gamètes lors de la reproduction sexuée.

Chaque segment de la molécule ADN, c'est-à-dire chaque portion de chromosome, est appelé un *locus* et correspond à un gène^c. Même si au sein d'une même espèce, les séquences d'ADN de deux individus sont fortement similaires, elles ne sont pas exactement identiques. Les différentes versions de la séquence d'un gène sont appelées des allèles. Un individu diploïde possède deux copies d'un même gène à un *locus* donné. Il peut hériter de deux allèles différents pour un même gène de la part de ses deux parents. Il sera dit hétérozygote au *locus* considéré ; et au contraire homozygote s'il hérite d'allèles identiques.

a. À l'exception des érythrocytes ("globules rouges").

b. Il y a 23 paires de chromosomes chez l'espèce humaine.

c. Nous ne restreignons pas la définition de "gène" aux portions codantes de l'ADN, transcrites en ARNm puis traduites en protéines, compte tenu de la focalisation sur les variations génétiques neutres dans le cadre de cette thèse.

Encadré 5 : Le principe de Hardy-Weinberg

La structure génétique des populations est un patron observable à un instant donné. L'évolution de ce patron au cours du temps se fait théoriquement selon le principe suivant. Considérons un *locus* à deux allèles A et a de fréquences respectives p et q dans une population. Considérons également que :

- la population compte un nombre infini d'individus,
- les individus ne quittent pas la population et aucun individu extérieur ne peut la rejoindre,
- les individus se reproduisent tous de façon équivalente (pas de sélection naturelle) et leur patrimoine génétique ne subit pas de mutations,
- les accouplements entre individus et l'association de leurs gamètes se font de manière aléatoire (panmixie et pangamie, respectivement),
- les générations ne se chevauchent pas.

Lorsque ces conditions sont vérifiées, la probabilité qu'un individu porte le génotype AA à la génération suivante suite à l'appariement aléatoire des gamètes est égale à p^2 . Pour le génotype aa ,

cette probabilité est égale à q^2 et pour le génotype Aa (ou aA) à $2pq$. La somme de ces probabilités vaut alors :

$$p^2 + 2pq + q^2 = (p + q)^2 = 1$$

La fréquence de l'allèle A dans la génération formée est égale à :

$$p^2 + \frac{1}{2}2pq = p(p + q) = p(p + 1 - p) = p$$

Celle de l'allèle a est de la même manière égale à q . Ainsi, au cours des générations, les fréquences alléliques et génotypiques restent théoriquement constantes quelles que soient les valeurs initiales de p et q . Dans ce cadre, les taux d'hétérozygotie observés et attendus sont égaux. Ce constat a été fait indépendamment par Godfrey Hardy et Wilhelm Weinberg au début du XX^e siècle et nommé principe d'Hardy-Weinberg. Il réconcilie à l'époque les théories de l'évolution darwinienne et de l'hérédité mendélienne, donnant naissance à la génétique des populations.

3.1.2 Dérive et flux génétique : principaux déterminants des variations génétiques neutres

Les conditions permettant d'observer l'équilibre prédit par le principe d'Hardy-Weinberg ([Encadré 5](#)) ne sont jamais réunies dans la réalité. En effet, la structure génétique des populations réelles évolue au cours du temps en raison de processus que l'on nomme forces évolutives. La **génétique des populations** s'intéresse précisément à la façon dont les quatre forces évolutives que sont les mutations, la sélection naturelle, la dérive et le flux génétique entraînent le non-respect de l'équilibre d'Hardy-Weinberg ([Gillespie, 2010](#)) et déterminent la structure génétique ([Figure 24](#)).

En effet, la structure génétique des populations évolue du fait de mutations qui modifient aléatoirement le matériel génétique des individus, pouvant ainsi faire apparaître de nouveaux allèles. Par ailleurs, les allèles confèrent parfois un avantage ou un désavantage reproductif aux individus qui les possèdent. Dans ce cas, la sélection naturelle fait augmenter ou diminuer, respectivement, leurs fréquences dans la population au cours des générations. Néanmoins, la plupart des allèles formés par mutation ne confèrent aucun avantage ou désavantage reproductif aux individus ([Kimura *et al.*, 1968](#)). Ils sont dits neutres et se trouvent essentiellement au niveau de séquences non codantes. L'évolution de leurs fréquences dépend de la dérive génétique et du flux génétique, deux forces évolutives ayant des déterminants démographiques. En raison du lien entre la connectivité écologique des habitats et la démographie des populations, nous nous focalisons dans le cadre de cette thèse sur l'influence de ces deux forces sur la variabilité génétique neutre.

Les populations réelles comptent un nombre fini d'individus et lors de la reproduction sexuée un individu diploïde ne transmet à chacun de ses descendants qu'une moitié de son patrimoine génétique portée par un seul de ses gamètes. À chaque génération, le tirage aléatoire et sans remise d'un nombre fini d'allèles forme le patrimoine génétique de la génération suivante. Il est alors possible que certains des allèles présents à une génération ne soient pas transmis à la suivante. C'est ce qu'on appelle la **dérive génétique**. Elle est aléatoire et responsable de la disparition de certains allèles au cours du temps. À l'issue de ce processus, il ne reste plus qu'un seul allèle dans une population. La fréquence allélique de l'allèle "fixé" est égale à 1 ([Figure 25](#)).

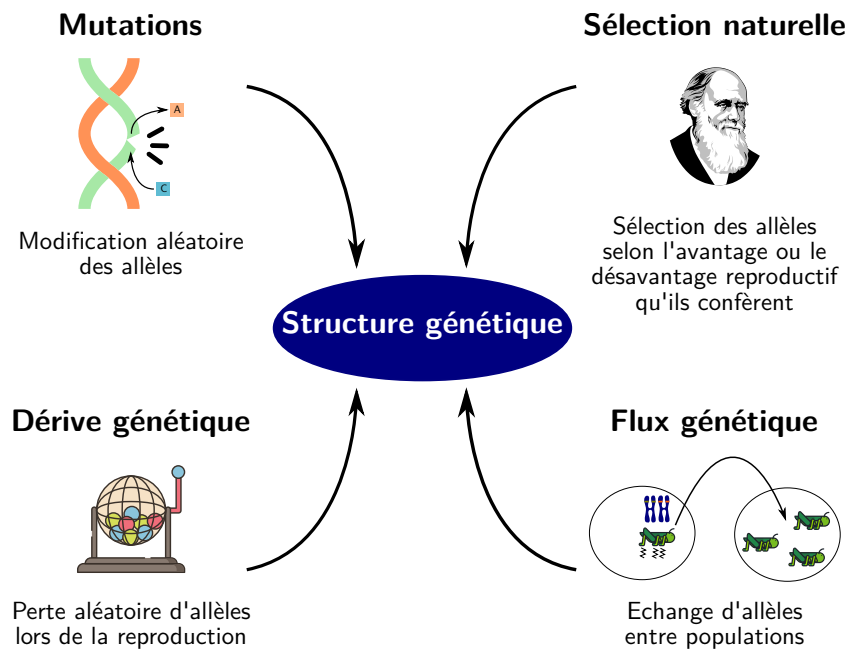


FIGURE 24 – Forces évolutives à l'origine des variations génétiques

Ce processus d'échantillonnage se modélise à l'aide d'une loi de probabilité binomiale et des démonstrations mathématiques³⁶ montrent qu'il réduit le taux d'hétérozygotes d'une proportion égale à $\frac{1}{2N}$ à chaque génération, avec N le nombre d'individus dans la population. L'effet de la dérive étant inversement proportionnel à la taille de la population qu'elle affecte, c'est dans les petites populations que la diversité génétique diminue le plus vite. Sur la figure 25, on remarque que dans la population comptant 2000 individus, un seul des huit allèles présents au départ a disparu au bout des 500 générations simulées. Il n'en subsistait plus qu'un lorsque cette population ne comptait que 200 individus.

L'évolution de la diversité génétique en réponse à la dérive est à l'origine du concept de taille de population efficace, notée N_e . Elle correspond à la taille de la population idéale dans laquelle on observerait une perte de diversité génétique équivalente à celle que l'on mesure dans la population réelle (Frankham *et al.*, 2004). La taille réelle d'une population ("*census size*"), que nous noterons N , serait de l'ordre de dix fois supérieure à la taille de population efficace selon certaines estimations (Frankham, 1996a). Cela provient entre autres du fait que tous les individus d'une population ne participent pas à la reproduction. Ainsi, la dérive peut avoir des conséquences importantes sur la diversité génétique d'une population, quand bien même le nombre d'individus y est relativement élevé.

La diminution de la diversité génétique d'une population cause des problèmes de consanguinité qui réduisent le succès de la reproduction de cette population et augmentent sa probabilité d'extinction (Frankham *et al.*, 2017). Par exemple, Nieminen *et al.* (2001) et Saccheri *et al.* (1998) ont établi un lien entre la consanguinité de populations de papillons et le risque d'extinction locale de ces populations. Par ailleurs, une moindre diversité génétique réduit le potentiel d'adaptation d'une population face à des changements environnementaux, augmentant alors son risque d'extinction à long terme. Il a également été montré que lorsqu'une espèce est menacée d'extinction, elle subit déjà une érosion de sa diversité génétique (Spielman *et al.*, 2004). Selon certains auteurs, le lien entre diversité génétique

36. Ces démonstrations sont présentées par Hamilton (2011), entre autres.

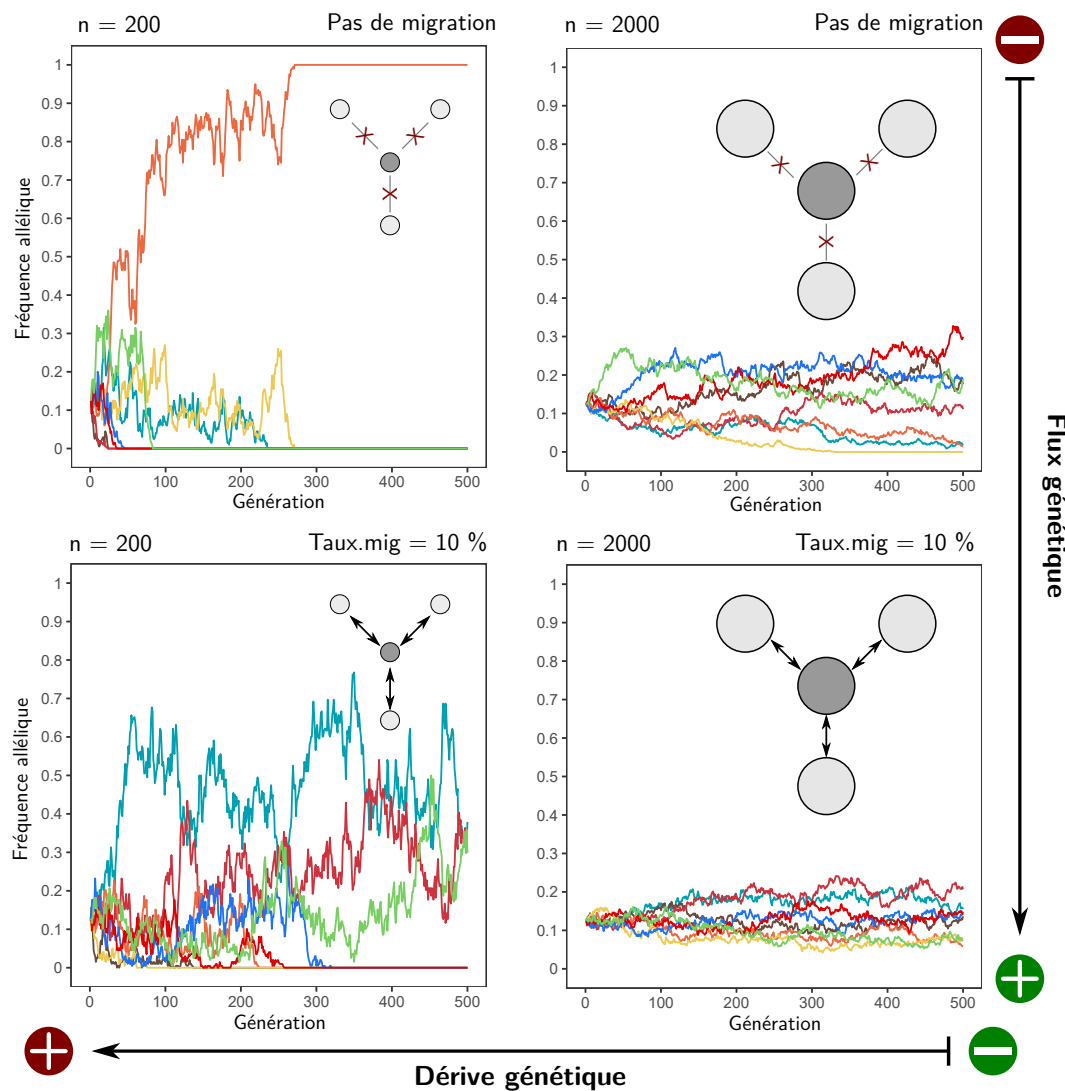


FIGURE 25 – Influence de la dérive et du flux génétique sur la diversité génétique locale. Évolution simulée de la fréquence allélique des huit allèles d'un *locus* dans la population représentée par un disque gris foncé au cours de 500 générations non chevauchantes. Cette population compte 200 (à gauche) ou 2000 (à droite) individus diploïdes se reproduisant aléatoirement à chaque génération pour former autant d'individus. Elle n'échange aucun individu avec les autres populations grises (en haut) ou échange 10 % de ses individus à chaque génération avec chacune des trois autres populations équitablement. La dérive génétique a un effet plus important dans les populations de 200 individus que dans celles de 2000 individus, tandis que le flux génétique est plus important lorsqu'il y a des échanges d'individus que lorsqu'il n'y en a pas.

et survie et adaptation des populations serait toutefois moins fort dès lors que l'on s'intéresse à la variabilité génétique neutre (Teixeira et Huber, 2021).

La perte de diversité génétique dans une population de taille finie n'en est pas pour autant inexorable dès lors que cette population n'est pas totalement isolée. En effet, l'échange d'individus provenant d'autres populations et participant à la reproduction assure un brassage génétique susceptible de restaurer la diversité perdue à cause de la dérive ("*genetic rescue*", Ingvarsson (2001)). On parle de **flux génétique**. Sur la figure 25, on voit que lorsque 10 % des individus d'une population en comptant 200 sont issus d'autres populations à chaque génération, quatre allèles persistent encore dans la population à l'issue de la simulation. L'effet du flux génétique sur la diversité génétique est parfois contre-intuitif. En théorie, l'arrivée d'un seul migrant³⁷ par génération dans une population

37. Le terme "migrant" employé ici est synonyme de celui de "disperseur" utilisé dans les sections précédentes. Nous le conservons pour évoquer le principe "*One Migrant Per Generation*".

est suffisante pour contrebalancer l'effet de la dérive sur la diversité génétique, et cela quelle que soit la taille de cette population. En effet, dans une grande population, l'effet de la dérive est faible donc un faible effet du flux génétique le contrebalance, tandis que ces deux effets seront forts dans une petite population (Mills et Allendorf, 1996 ; Wang, 2004). Néanmoins, compte tenu de la différence entre N et N_e , plusieurs migrants sont généralement requis pour qu'un flux génétique ait lieu. Par ailleurs, un seul migrant par génération ne suffit pas à limiter les conséquences de la consanguinité dans une population (Frankham *et al.*, 2017), et le rétablissement de la diversité génétique est lent avec un flux génétique si faible (Varvio *et al.*, 1986).

Le flux génétique maintient donc la diversité génétique d'une population. En assurant un brassage génétique, il tend aussi à limiter la différenciation génétique entre populations. Dans une population, la possession de certains génotypes peut augmenter le succès reproductif (*fitness*) compte tenu des conditions locales. L'homogénéisation du patrimoine génétique liée aux flux génétiques peut alors être un frein à la sélection naturelle (Crispo *et al.*, 2011 ; Lenormand, 2002). Néanmoins, bien qu'il existe d'autres cas particuliers dans lesquels l'augmentation du flux génétique au sein d'une espèce peut avoir un effet négatif sur la survie de ses populations (Frankham *et al.*, 2017), dans la majorité des cas ces échanges sont bénéfiques selon la méta-analyse de Frankham (2015).

Les premiers travaux de génétique des populations étaient essentiellement théoriques et basés sur des démonstrations mathématiques³⁸. Ils représentaient les populations de façon schématique, comme dans le modèle îles-continent de Wright-Fisher qui étudie la structure génétique de populations de taille finie (îles) échangeant toutes autant de migrants avec une population centrale de taille infinie (continent) (Wright, 1931). Plus tard, Kimura et Weiss (1964) ont développé le modèle "stepping-stones" dans lequel chaque population d'un ensemble de populations de taille finie échange des migrants avec ses populations voisines uniquement. En intégrant plus explicitement l'influence de la distribution spatiale des populations, les théories initiales ont été approfondies (Malécot, 1948). La disponibilité des données génétiques empiriques a permis de confirmer et d'affiner ces théories à partir de la fin du XX^e siècle. Elles ont pris en compte le fait que le paysage au sein duquel se trouvent les populations réelles est hétérogène, ce qui affecte à la fois la dérive et le flux génétique. C'est à partir de ces résultats concernant la distribution spatiale de la variabilité génétique que l'influence de la connectivité des habitats sur la structure génétique des populations a peu à peu été comprise. Ce sujet étant au cœur de cette thèse, nous allons désormais présenter le cadre théorique et méthodologique dans lequel ces travaux s'inscrivent, ainsi que leurs limites actuelles.

3.2 Distribution spatiale des variations génétiques

Lorsque la structure génétique d'un ensemble de populations subdivisées est essentiellement influencée par la dérive et le flux génétique, la distribution spatiale de la diversité et de la différenciation génétique des populations est censée reproduire certains attendus théoriques.

38. On notera par exemple que Ronald Fisher, pionnier de la génétique des populations du début du XX^e siècle est aussi un pionnier de la statistique moderne.

3.2.1 Influence de la subdivision des populations et de l'hétérogénéité de leurs tailles

Subdivision des populations

Tout d'abord, la subdivision des populations détermine l'effet de la dérive sur la diversité génétique. Cet effet est inversement proportionnel à la taille efficace de la population N_e . Ainsi, le taux d'hétérozygotes dans une population isolée de 100 individus diminue d'une proportion égale à $\frac{1}{200}$ à chaque génération (cf. section 3.1.2), tandis que la diminution de cette proportion d'hétérozygotes à chaque génération vaut $\frac{1}{20}$ dans 10 populations isolées de 10 individus. On démontre donc mathématiquement qu'à nombre total d'individus équivalent, la subdivision des populations a un effet négatif sur la diversité génétique en l'absence de flux génétique entre les populations (Frankham *et al.*, 2017). Par conséquent, on s'attend à ce que la configuration des taches d'habitat dans le paysage influence le niveau de diversité génétique des populations qui les occupent. D'autre part, des variables paysagères affectant potentiellement la taille des populations, telles que la capacité des taches d'habitat, devraient expliquer la structure génétique des populations, en tant qu'elles reflètent l'influence de la dérive.

D'autre part, lorsque les populations sont subdivisées et isolées, les évolutions indépendantes et aléatoires de leurs fréquences alléliques sous l'effet de la dérive devraient les faire diverger d'un point de vue génétique, c'est-à-dire faire augmenter la différenciation génétique entre elles. Si ces populations ne sont pas totalement isolées, leur différenciation génétique dépend généralement de la distance qui les sépare (cf. section 3.2.2).

Hétérogénéité spatiale de la taille des populations

Lorsque la taille efficace d'une population fluctue au cours du temps, la diversité génétique subit la plus forte érosion lorsque la taille de la population est la plus faible (Allendorf *et al.*, 2012). La taille de population efficace N_e qui reflète l'influence de la dérive dans une population au cours de t générations est égale, non pas à la moyenne arithmétique des tailles de populations N_1, N_2, \dots, N_t , mais à leur moyenne harmonique (Hartl *et al.*, 1997) :

$$\frac{1}{N_e} = \frac{1}{t} \left(\frac{1}{N_1} + \frac{1}{N_2} + \dots + \frac{1}{N_t} \right)$$

On a donc :

$$N_e = \frac{t}{\sum_{i=1}^t \left(\frac{1}{N_i} \right)}$$

Cette moyenne est fortement influencée par les plus faibles valeurs, ce qui reflète bien le fait qu'une forte réduction de la taille d'une population pendant une génération peut faire diminuer sa diversité génétique de façon considérable. Ce dernier cas correspond à ce que les généticiens des populations appellent un goulot d'étranglement (Allendorf et Luikart, 2009). Pour des raisons similaires, lorsqu'une population est créée par un petit nombre d'individus, la diversité génétique y est fortement réduite par rapport à celle de la population d'origine. On parle alors d'effet fondateur.

Par ailleurs, si les populations réelles comptent rarement un nombre constant d'individus au cours du temps, toutes les populations ne comptent pas non plus le même nombre d'individus à un instant t . L'effet de la dérive varie donc non seulement dans le temps mais aussi dans l'espace. Des variables paysagères affectant potentiellement la taille des populations, telles que la capacité des taches d'habi-

tat, devraient donc expliquer la structure génétique des populations. En effet, l'hétérogénéité spatiale de l'intensité de la dérive influence la distribution spatiale de la diversité génétique au sein des populations, mais aussi leur différenciation génétique (Jaquière *et al.*, 2011 ; Richardson *et al.*, 2016 ; Whitlock et McCauley, 1999). En effet, considérons deux paires de populations issues d'une même population ancestrale non subdivisée. La dérive génétique aura fait diverger de façon plus importante le patrimoine génétique de la paire de populations dont la moyenne harmonique des tailles de populations est la plus faible. À flux génétique équivalent, la différenciation génétique de cette paire de populations devrait être la plus forte. On peut alors s'attendre à ce que des variables paysagères reflétant l'influence de la dérive dans les populations expliquent aussi le niveau de différenciation génétique entre populations.

3.2.2 L'isolement par la distance

Nous avons vu précédemment que la perte de diversité génétique causée par la dérive pouvait être contrebalancée par les flux génétiques (Figure 25). L'intensité des flux génétiques est fortement liée à la distribution spatiale des populations (Slatkin, 1987). En effet, elle dépend du nombre d'individus qui se dispersent entre les taches d'habitat qu'occupent les populations. Théoriquement, plus la distance géographique entre des populations est importante, plus la différenciation génétique entre ces populations est forte car peu d'individus peuvent se déplacer entre elles. Initialement établi par Wright (1943), ce phénomène d'**isolement par la distance** (noté IBD pour *Isolation By Distance*) a fait l'objet de travaux théoriques (Slatkin, 1993), de validations empiriques (Garnier *et al.* (2004)) et sa modélisation s'est complexifiée au fil du temps (Bradbury et Bentzen, 2007 ; Castric et Bernatchez, 2003). Ces travaux ont permis d'étendre le modèle considérant initialement des individus distribués de façon homogène dans l'espace à des populations échangeant des individus de proche en proche (Kimura, 1953)³⁹.

Hutchison et Templeton (1999) ont défini une typologie des patrons d'IBD basée sur l'influence respective de la dérive et du flux génétique sur la différenciation génétique (Figure 26). La différenciation génétique entre populations atteint un état d'équilibre lorsqu'il y a autant d'allèles qui disparaissent à chaque génération à cause de la dérive génétique, que d'allèles qui arrivent dans une population grâce aux flux génétiques. Cette situation d'équilibre (*migration-drift equilibrium*) est atteinte progressivement lorsque des populations colonisent une nouvelle zone ou suite à une modification des conditions dans lesquelles les flux génétiques ont lieu. Une fois la situation d'équilibre de la différenciation génétique atteinte à l'échelle de toutes les populations étudiées, le patron d'IBD se matérialise par une augmentation linéaire et continue de la distance génétique entre populations à mesure que la distance géographique entre ces mêmes populations augmente (patron d'IBD de type I, Figure 26). La variabilité de la distance génétique autour de sa valeur attendue compte tenu de la distance géographique augmente également avec la distance géographique. En effet, lorsqu'une grande distance sépare les populations, l'influence du flux génétique sur la différenciation génétique est plus faible que celle de la dérive aléatoire, ce qui crée une forte variabilité. L'inverse s'observe lorsque cette distance est faible.

En revanche, lorsque la situation d'équilibre n'est pas atteinte à l'échelle de toute la zone d'étude, la distance génétique augmente de façon linéaire jusqu'à une certaine distance géographique, au-delà de laquelle elle atteint un plateau. Ce patron d'IBD de type IV, selon la typologie d'Hutchison et Tem-

39. Guillot *et al.* (2009) décrivent en détail l'historique des travaux sur l'IBD.

pleton (1999) (Figure 26), s’observe lorsque les flux génétiques entre les populations les plus éloignées, y-compris multi-générationnels, sont trop faibles pour que la distance géographique soit corrélée à la distance génétique. Aux échelles les plus larges, la dérive détermine donc la différenciation génétique entre populations indépendamment du flux génétique, et donc de la distance géographique entre populations. Enfin, les patrons d’IBD de type II et III correspondent respectivement aux situations dans lesquelles le flux génétique entre populations est très important (proche d’une situation de panmixie) ou au contraire inexistant (isolement total) (Figure 26).

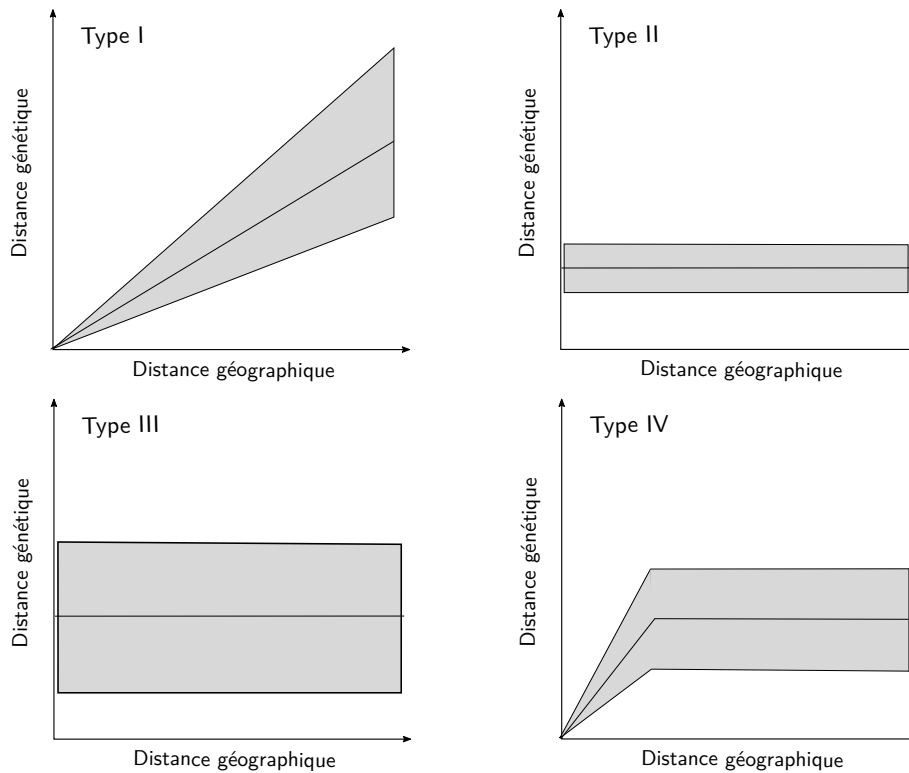


FIGURE 26 – Typologie des patrons d’isolement par la distance définie par Hutchison et Templeton (1999)

La présence du patron d’IBD de type I se teste en évaluant le degré de corrélation entre le F_{ST} linéarisé ($F_{ST}/(1-F_{ST})$) et la distance géographique entre populations ou entre le logarithme du F_{ST} et celui de la distance géographique (Rousset, 1997)⁴⁰. Cette corrélation est estimée par le coefficient de corrélation de Mantel entre les deux matrices de distances considérées (Mantel, 1967). Sa significativité est testée à l’aide de permutations aléatoires (test de Mantel) car la dépendance statistique des valeurs de distances empêche l’utilisation de tests paramétriques classiques. En revanche, la présence d’un patron d’IBD de type IV est plus difficile à tester statistiquement. On peut assimiler l’IBD à l’autocorrélation spatiale des fréquences alléliques des populations étudiées. Les méthodes de statistique spatiale et issues de la théorie des graphes ont donc parfois été utilisées pour étudier les patrons d’IBD (Brooks, 2006 ; Smouse et Peakall, 1999 ; Wagner *et al.*, 2005). Nous étudierons justement dans le cadre de cette thèse l’intérêt de la théorie des graphes pour étudier les patrons spatiaux de différenciation génétique.

40. D’après les travaux théoriques de Rousset (1997), la première option est à utiliser lorsque les populations forment un réseau uni-dimensionnel (linéaire) et la seconde quand ce réseau est bi-dimensionnel. Lorsque le patron d’IBD de type I est détecté, ce type d’analyse permet également d’estimer des paramètres démographiques des populations étudiées.

Étant donné que la localisation spatiale des populations détermine la distribution statistique des valeurs de distances entre populations, le patron d'IBD est théoriquement influencé par la connectivité des taches d'habitat. Ainsi, des travaux mobilisant les acquis de l'écologie du paysage en génétique des populations, et réciproquement, ont été menés pour comprendre cette influence. Cela a contribué à l'émergence de la génétique du paysage (*landscape genetics*), discipline à laquelle nous nous intéressons à présent.

3.3 L'apport de la génétique du paysage pour analyser l'influence de la connectivité sur la structure génétique

3.3.1 Origines et objectifs de la génétique du paysage

Créée au début des années 2000, la **génétique du paysage** est la discipline qui vise à tester l'influence de la structure du paysage sur la structure génétique des populations (Balkenhol *et al.*, 2016 ; Manel *et al.*, 2003 ; Storfer *et al.*, 2007). Contrairement aux approches phylogéographiques fréquemment adoptées par le passé en génétique des populations, la génétique du paysage s'intéresse à la façon dont le paysage contemporain influence d'une part la sélection naturelle et d'autre part la variabilité génétique neutre (Manel *et al.*, 2003). Dans les deux cas, la dimension spatiale est explicitement prise en compte (Wagner et Fortin, 2013). Néanmoins, lorsqu'on vise à identifier les variables environnementales locales à l'origine des processus adaptatifs, l'autocorrélation spatiale des variables considérées limite la robustesse des analyses statistiques (Diniz-Filho *et al.*, 2009). On cherche alors à l'estimer pour séparer ce "bruit" du signal étudié. À l'inverse, dans les travaux concernant la variabilité génétique neutre, le patron de variation spatiale des réponses génétiques et des variables paysagères est le signal étudié, et non pas un bruit dont il faudrait s'affranchir (Dale et Fortin, 2014). La modélisation spatiale revêt donc de l'importance pour atteindre ce second objectif et c'est dans ce cadre que nos travaux s'inscrivent.

Les méthodes de modélisation spatiale que nous utilisons sont adaptées aux situations dans lesquelles les taches d'habitat des espèces étudiées sont subdivisées (cf. section 2.2). Nous considérons donc les populations occupant ces taches d'habitat comme les unités élémentaires de nos analyses, bien que les analyses centrées sur les individus puissent aussi être pertinentes (Balkenhol et Fortin, 2016 ; Prunier *et al.*, 2013). Ce niveau d'analyse est le plus fréquent en génétique du paysage (Milligan *et al.*, 2018 ; Waits et Storfer, 2015), ce qui justifie davantage notre choix compte tenu des objectifs de développement méthodologique poursuivis dans le cadre de cette thèse.

3.3.2 Prise en compte de l'hétérogénéité du paysage

Bien que certains modèles théoriques de génétique des populations considèrent l'espace de façon explicite, la prise en compte de l'hétérogénéité du paysage est relativement récente dans l'analyse de la structure génétique des populations. Elle constitue un des apports majeurs de la génétique du paysage (Dyer, 2015a), dans la mesure où elle permet de spatialiser l'influence de la dérive et du flux génétique sur la structure génétique en la mettant notamment en lien avec la connectivité écologique des habitats (Figure 27).

Des travaux de génétique du paysage ont confirmé empiriquement le lien existant entre la surface des taches d'habitat, ou d'autres variables mesurées localement et influençant potentiellement

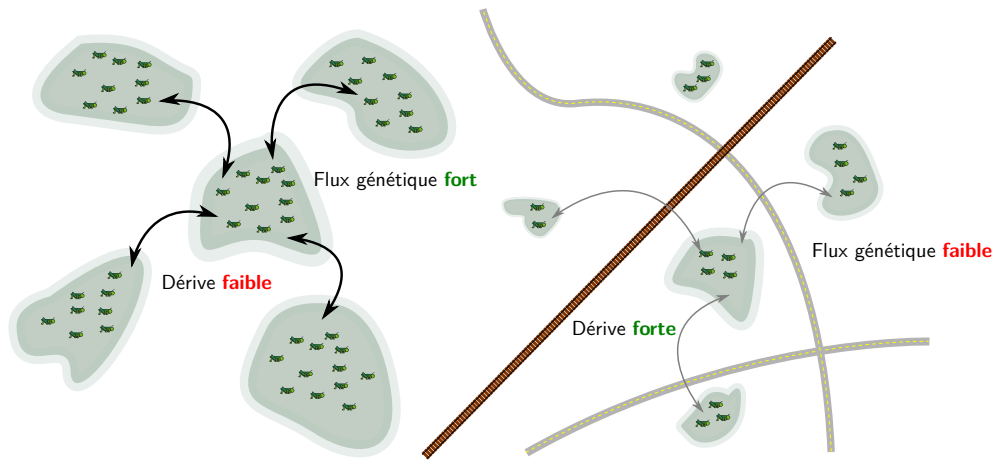


FIGURE 27 – Influence de la connectivité des habitats sur la dérive et le flux génétique. Si on considère que la taille des populations dépend de la surface des taches d’habitat et que la distance et la présence d’infrastructures anthropiques limitent la dispersion entre les taches, alors (i) la dérive devrait être plus forte dans les populations situées dans de petites taches d’habitat et (ii) le flux génétique devrait être plus faible entre les populations séparées par de grandes distances ou par des barrières naturelles (reliefs, cours d’eau) ou anthropiques (infrastructures linéaires de transport). Ces deux effets influencent conjointement la diversité et la différenciation génétique des populations.

la dérive génétique, et la diversité génétique des populations occupant ces taches (Barr *et al.*, 2015 ; Taylor et Hoffman, 2014 ; Toma *et al.*, 2015). Néanmoins, c’est l’influence de l’hétérogénéité de la matrice séparant les taches d’habitat qui a été la plus étudiée dans ce domaine (DiLeo et Wagner, 2016). Le patron classique d’IBD a été observé fréquemment à partir de données empiriques et il est basé sur un cadre mathématique permettant d’inférer des paramètres démographiques. Toutefois, la distance géodésique euclidienne entre des populations explique souvent moins bien leur différenciation génétique que des distances paysagères intégrant l’hétérogénéité de la matrice. Par exemple, Arnaud (2003) a montré que la différenciation génétique entre les populations d’escargots (*Helix aspersa*) d’un paysage agricole de Bretagne était mieux expliquée par la distance euclidienne mesurée le long des fossés de drainage ou de bord de routes et des haies que par la distance géodésique euclidienne. De la même façon, Coulon *et al.* (2004) ont montré que la différenciation génétique entre des chevreuils du sud-ouest de la France était mieux expliquée par une distance paysagère tenant compte du fait que ces individus se déplaçaient préférentiellement en forêt que par la distance géodésique euclidienne.

Par la suite, McRae (2006) a introduit une méthode basée sur la théorie des circuits pour calculer une distance de résistance tenant compte de l’hétérogénéité de la matrice et de chemins de dispersion alternatifs dans le paysage. Il a montré par des simulations les raisons pour lesquelles cette distance, tout comme la distance-coût associée au chemin de moindre coût entre les populations, devrait mieux expliquer la différenciation génétique que la distance géodésique euclidienne dans un grand nombre de situations. Ce même auteur a démontré empiriquement l’intérêt de son approche (McRae et Beier, 2007) et développé le logiciel Circuitscape pour la mettre en œuvre. Ce travail a donné naissance au concept d’**isolement par la résistance du paysage** (IBLR, *Isolation By Landscape Resistance sensu Cushman et al.* (2006)), extension sémantique du patron d’IBD dans le cas où la matrice est hétérogène (Guillot *et al.*, 2009)⁴¹. Les distances-coût sont également utilisées fréquemment pour expliquer la dif-

41. Nous utilisons le terme IBLR utilisé par Cushman *et al.* (2006) pour englober les situations dans lesquelles des distances-coût ou des distances issues de la théorie des circuits sont utilisées pour expliquer la différenciation génétique. Ces deux situations ont en commun de considérer l’hétérogénéité de la matrice paysagère. Le terme IBR popularisé par McRae (2006) désigne spécifiquement l’utilisation de la théorie des circuits et nous voulons éviter toute ambiguïté avec l’utilisation des distances-coût.

férenciation génétique. Parfois, le degré de corrélation entre distances-coût et différenciation génétique est plus fort que celui entre la distance de résistance obtenue à l'aide de la théorie des circuits et la différenciation génétique (voir [Balbi *et al.* \(2018\)](#) ou [Zeller *et al.* \(2018\)](#)). D'autre fois, c'est l'inverse (voir par exemple [Khimoun *et al.* \(2017\)](#)), et certains auteurs considèrent que ces deux approches sont tout aussi performantes la plupart du temps ([Simpkins *et al.*, 2018](#)).

L'analyse d'un patron d'IBLR nécessite d'attribuer des coûts aux différents pixels d'une surface de résistance, le plus souvent basée sur l'occupation du sol ([Spear *et al.*, 2010](#)) (cf. section 2.1). Habituellement, plusieurs scénarios de coût sont comparés jusqu'à identifier celui qui maximise l'ajustement entre les indices de différenciation génétique (ou distance génétique) et les distances paysagères. Bien que cette opération puisse se faire "manuellement", en définissant chacun des scénarios testés (voir [Mergey *et al.* \(2017\)](#) ou [Ruiz-Gonzalez *et al.* \(2015\)](#)), elle peut aussi être automatisée via l'utilisation d'algorithmes d'optimisation ([Graves *et al.*, 2013](#) ; [Peterman, 2018](#)). Parmi eux, l'algorithme génétique⁴² développé par [Peterman \(2018\)](#) identifie un scénario de coût optimal dans un vaste champ des possibles compte tenu de la combinatoire presque infinie des scénarios de coût. Le résultat d'une telle optimisation basée sur le principe d'IBLR permet d'une part d'identifier les éléments de la matrice influençant la différenciation génétique entre populations. Il permet potentiellement de comprendre quels sont les facteurs favorisant ou limitant le flux génétique et donc la connectivité écologique des habitats. D'autre part, les valeurs de coût optimisées peuvent être utilisées pour modéliser la connectivité écologique des taches d'habitat. Cela permet de localiser les chemins de dispersion potentiels entre les taches, de leur associer des distances-coût et de calculer des métriques de connectivité (cf. section 2.2.2).

Comme indiqué précédemment, la différenciation génétique dépend à la fois du flux génétique et de la dérive. Tout comme le flux génétique entre populations est hétérogène compte tenu des distances qui les séparent et de l'hétérogénéité de la matrice, l'intensité de la dérive est hétérogène lorsque les populations ne comptent pas le même nombre d'individus. [Prunier *et al.* \(2017\)](#), [Serrouya *et al.* \(2012\)](#) et [Weckworth *et al.* \(2013\)](#) ont mis en évidence l'influence de l'hétérogénéité des tailles de populations sur la différenciation génétique entre populations. [Prunier *et al.* \(2017\)](#) ont développé des indicateurs basés sur la moyenne harmonique des tailles de populations pour mesurer cette hétérogénéité. Ils ont ainsi montré que l'influence de cette hétérogénéité sur la différenciation génétique était d'autant plus forte que le degré d'hétérogénéité était fort et que le flux génétique était faible. Selon ces auteurs, des variables paysagères liées à la taille des populations peuvent être utilisées pour évaluer cette influence lorsque les tailles des populations sont difficilement estimables directement, ce qui est fréquemment le cas ([Wang, 2005](#)).

Enfin, le patron spatial de différenciation génétique se traduit parfois par l'existence d'ensembles de populations fortement différenciés les uns des autres en raison d'une quasi-absence de flux génétiques. Ces situations sont décrites depuis longtemps et le développement du logiciel STRUCTURE ([Pritchard](#)

42. Un algorithme génétique permet d'optimiser un résultat à l'aide d'une fonction d'objectif et de règles inspirées de la sélection naturelle lorsque le nombre de résultats possibles est quasiment infini. Pour cela, plusieurs résultats sont initialement générés aléatoirement. La performance de chacun d'eux est évaluée à l'aide de la fonction d'objectif. Les meilleurs résultats sont conservés. Ils subissent des mutations, c'est-à-dire de légères modifications aléatoires. Les nouveaux résultats créés par mutation sont évalués. Les meilleurs parmi eux sont conservés et soumis eux-aussi à la mutation. Le processus continue jusqu'à ce que l'amélioration de leurs performances s'arrête ou soit négligeable. Le meilleur résultat est alors conservé. C'est un optimum local néanmoins.

et al., 2000), basé sur un algorithme de *clustering* bayésien, y a largement contribué. Le fait qu’elles puissent être confondues avec un patron d’IBD (Bradburd *et al.*, 2018 ; Safner *et al.*, 2011), tout comme la possible existence d’un patron d’IBD au sein de chacun de ces ensembles de populations (Garnier *et al.*, 2004), ont également été étudiés. Dans ce contexte, la génétique du paysage vise à identifier quelles sont les barrières paysagères responsables de ces fortes discontinuités dans le gradient spatial de différenciation génétique (voir Manel *et al.* (2007)).

3.3.3 Plusieurs niveaux de mise en relation des réponses génétiques et des variables paysagères

Nous avons vu que l’hétérogénéité du paysage est intégrée aux analyses de génétique du paysage de différentes façons. Wagner et Fortin (2013) ont formalisé les différents niveaux d’analyse auxquels les réponses génétiques et les variables paysagères sont mises en relation (Figure 28). Tout d’abord, les réponses génétiques peuvent être étudiées au niveau des populations (en mesurant leur diversité génétique par exemple) et mises en relation avec des variables paysagères mesurées à ce même niveau (par exemple avec la surface des taches d’habitat) dans le cadre d’analyses de type *node level* selon leur typologie (Figure 28A). Le lien entre les distances génétiques entre populations et les distances paysagères fait quant à lui l’objet des analyses de type *link level* (Figure 28B). De façon intermédiaire, la mise en relation de réponses génétiques mesurées au niveau des populations avec des variables paysagères tenant compte de la position d’une population vis-à-vis des autres, telles que des métriques de connectivité écologique, constitue les analyses de type *neighbourhood level* (Figure 28C). Enfin, la recherche des éléments paysagers constituant les barrières au flux génétique responsables d’un patron de différenciation génétique caractérisé par des *clusters* est l’objet des analyses de type *boundary level* (Figure 28D).

Le vocabulaire utilisé dans cette typologie rappelle celui de la théorie des graphes et nous verrons dans la section 5.1 qu’elle permet de distinguer les différentes utilisations de cette théorie en génétique du paysage. Nous verrons également dans la section 5.2 et dans le chapitre 6 que des approches hybrides consistant à mettre simultanément en relation des variables mesurées au niveau des populations et entre les populations peuvent également être utiles.

3.3.4 La boîte à outils de la génétique du paysage

La mise en œuvre de ces différentes analyses nécessite (i) de décrire le contexte paysager à l’aide de données spatiales, (ii) de décrire la structure génétique des populations à l’aide de données génétiques et (iii) de les mettre en relation à l’aide de méthodes statistiques. Les données paysagères utilisées et leur traitement sont relativement similaires à ce qui est fait habituellement en écologie du paysage. Pour étudier la variabilité génétique neutre, les principaux marqueurs génétiques utilisés sont les **microsatellites** (Storfer *et al.*, 2010)⁴³. Il s’agit de *loci* au niveau desquels des motifs composés de deux ou trois nucléotides sont répétés plusieurs dizaines de fois. Ce qui définit un allèle est alors le nombre de fois où le motif est répété (Figure 29). En raison d’un taux de mutations plus élevé qu’à d’autres portions du génome, ils ont une variabilité importante. En outre, ils se situent le plus souvent au sein de portions non codantes du génome. Ils permettent donc d’étudier assez finement l’influence du paysage contemporain sur les processus démographiques déterminant la structure génétique des

43. Et dans une moindre mesure les AFLPs (*Amplified Fragment Length Polymorphism*).

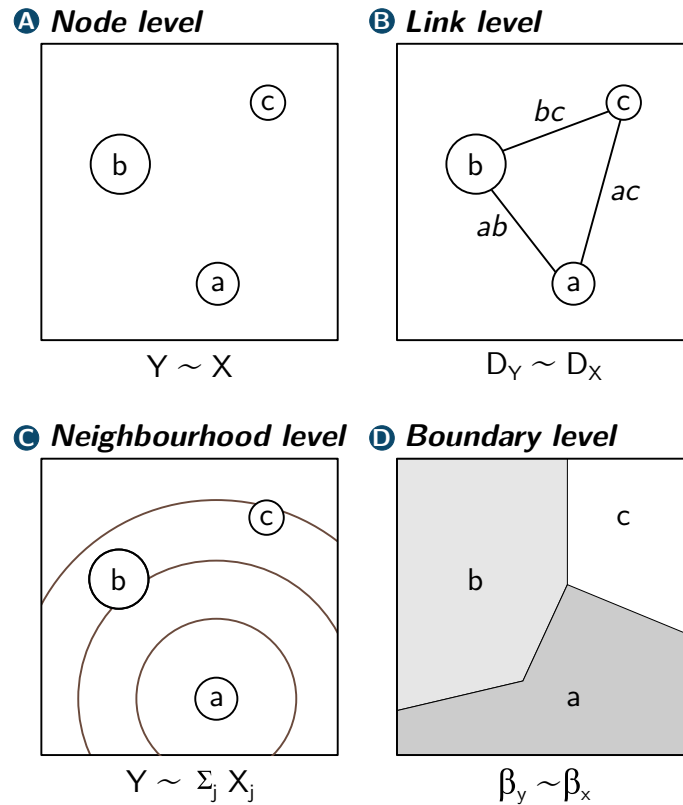


FIGURE 28 – Quatre niveaux de mise en relation des données génétiques et paysagères en génétique du paysage (d'après la figure 2 de [Wagner et Fortin \(2013\)](#)). (A) Au niveau des "nœuds", le patrimoine génétique (Y) des populations des sites (a, b, c) est relié aux variables paysagères locales (X) observées au niveau de ces sites. (B) Au niveau des "liens", les distances génétiques (D_Y) entre les populations des différents sites (ab, ac, bc) sont mises en relation avec les distances paysagères (D_X) entre ces sites. (C) Au niveau du voisinage, le patrimoine génétique (Y) des populations (a, b, c) est mis en relation avec des métriques de connectivité calculées au niveau des taches d'habitat qui tiennent compte du contexte local et des connexions avec les autres taches j ($\sum_j X_j$). (D) Au niveau des "frontières", la distinction entre des populations ou des ensembles de populations faite à partir des données génétiques (obtenue par *clustering* par exemple) est mise en relation avec des variables paysagères pour identifier les éléments paysagers constituant des barrières au flux génétique.

populations. Par ailleurs, les marqueurs SNPs (*Single Nucleotide Polymorphism*) sont de plus en plus utilisés dans ce domaine. Dans ce cas, c'est l'identité des nucléotides présents à des positions précises dans le génome qui permet de décrire le polymorphisme génétique (Figure 29). Tandis que moins de vingt microsatellites sont généralement étudiés pour caractériser la structure génétique, ce sont plusieurs milliers de SNPs qui sont habituellement considérés simultanément. Un marqueur SNP ne comporte que quatre allèles au maximum alors qu'un microsatellite peut en comporter un nombre bien plus élevé, ce qui est potentiellement compensé par le nombre important de *loci* considérés dans le cas des SNPs. Davantage d'analyses de bio-informatique sont nécessaires pour traiter ce type de données. Il est également nécessaire de distinguer les *loci* qui sont soumis à la sélection naturelle (*outlier loci*) de ceux qui ne le sont pas ([Cushman et al., 2018](#) ; [Foll et Gaggiotti, 2008](#)), pour utiliser l'un ou l'autre type de *loci* selon l'objectif des analyses réalisées.

Les approches basées sur des simulations (*virtual ecologist approach sensu* [Zurell et al. \(2010\)](#)) permettent quant à elles de simuler la structure génétique des populations qui résulterait de l'existence de contraintes démographiques liées à la structure du paysage ([Landguth et al., 2016](#)). Des logiciels tels que CDPOP ont précisément été développés pour simuler l'effet de la taille et de la distribution spatiale des populations, et de l'hétérogénéité de la matrice sur les variations génétiques temporelles

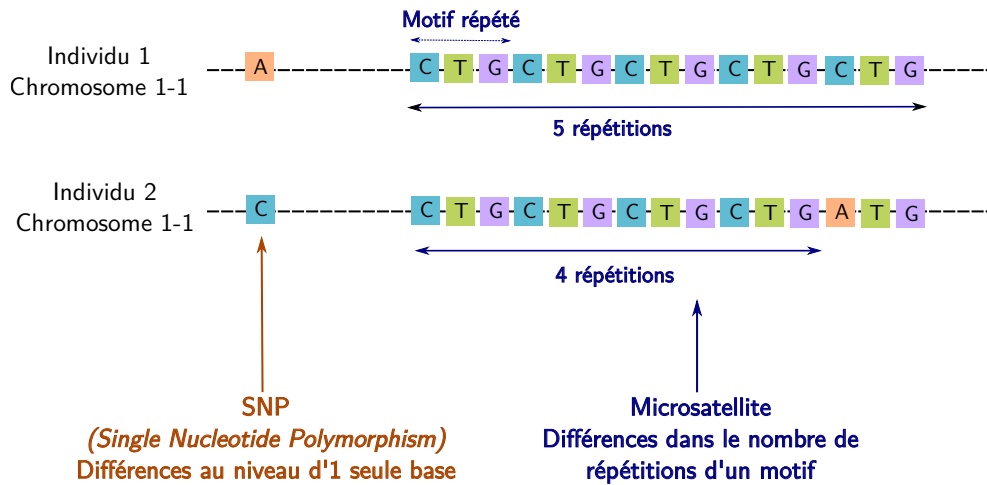


FIGURE 29 – Principaux marqueurs génétiques utilisés en génétique du paysage. Chaque segment d’ADN représenté appartient à l’un des deux brins d’ADN formant un des deux chromosomes homologues de la même paire de chromosomes chez deux individus diploïdes. Tandis que l’étude d’un marqueur SNP se focalise sur une seule base (nucléotide), celle des microsatellites consiste à comparer le nombre de répétitions d’un motif formé de deux ou trois nucléotides (CTG ici).

et spatiales (Landguth et Cushman, 2010). Cela permet d’identifier les principaux facteurs influençant la structure génétique (voir Cushman *et al.* (2012) ou Jackson et Fahrig (2015b)), mais aussi de tester la validité de la myriade de méthodes statistiques existantes (voir Graves *et al.* (2012)).

En effet, les méthodes d’analyse statistique utilisées en génétique des populations ont évolué à un rythme soutenu durant les deux dernières décennies (Balkenhol *et al.*, 2009b ; Guillot *et al.*, 2009). Lorsque les variables sont mises en relation au niveau des populations (*node-* ou *neighbourhood-level*), des méthodes classiques basées sur des coefficients de corrélation, des modèles linéaires ou des modèles linéaires mixtes sont le plus souvent utilisées (Storfer *et al.*, 2010), bien que des méthodes de régression intégrant l’autocorrélation spatiale des résidus (*SAR models*) ou intégrant l’espace comme un ensemble de co-variables à part entière (*Moran’s Eigenvector Maps*, Dray *et al.* (2006)) aient également été employées (Diniz-Filho *et al.*, 2009 ; Taylor et Hoffman, 2014).

L’approche statistique la plus fréquente en génétique du paysage repose sur le coefficient de corrélation de Mantel et son test de significativité (Manel et Holderegger, 2013 ; Storfer *et al.*, 2010). Elle est utilisée pour évaluer la relation entre des distances génétiques et des distances paysagères (*link level*). Cette approche a pourtant fait l’objet de nombreuses critiques, en raison notamment de son risque élevé d’erreur de type I (Balkenhol *et al.*, 2009b ; Legendre et Fortin, 2010). Des modèles linéaires mixtes permettent aussi d’évaluer la relation entre distances génétiques et paysagères⁴⁴. Considérer l’identité des populations impliquées dans chaque paire de populations comme un effet aléatoire dans ces modèles permet de modéliser la covariance des résidus liée au fait qu’on considère plusieurs distances associées à la même population. Cette méthode nommée MLPE (*Maximum Likelihood Population Effect*) a été développée par Clarke *et al.* (2002) et introduite en génétique du paysage par Van Strien *et al.* (2012). Shirk *et al.* (2017b) ont montré sa supériorité par rapport à d’autres méthodes. Enfin, les **modèles gravitaires** permettent d’intégrer à la fois des variables relatives aux liens entre populations et à chacune des populations dans les analyses (cf. Annexe B5). D’abord utilisés en

44. Nous n’évoquons pas de façon exhaustive les méthodes d’analyse utilisables dans ce contexte. Il faudrait pour cela mentionner la régression de matrices de distances (Legendre et Legendre, 2012) et la méthode de *causal modelling* développée par Cushman *et al.* (2006) et basée sur le test partiel de Mantel, entre autres.

économie et en géographie (Anderson, 1979 ; Fotheringham et O’Kelly, 1989), ils ont été introduits en génétique du paysage par Murphy *et al.* (2010a).

Les analyses de *clustering* bayésien basées sur le logiciel STRUCTURE sont également très fréquentes dans ce domaine (Storfer *et al.*, 2010). Les barrières paysagères entre les *clusters* peuvent alors être identifiées à l’aide de méthodes telles que l’algorithme de Monmonier ou l’algorithme bilinéaire de Womble (*boundary level*, voir Manel *et al.* (2007) ou Mani *et al.* (2004)). Enfin, les méthodes issues de la théorie des graphes sont parfois utilisées en génétique du paysage (Murphy *et al.*, 2016). Elles sont compatibles avec l’ensemble des niveaux d’analyse caractérisés par Wagner et Fortin (2013) et avec la plupart des analyses statistiques décrites précédemment, et constituent le sujet principal de cette thèse. Nous reviendrons donc par la suite (cf. section 5) sur la façon dont les méthodes statistiques décrites précédemment peuvent s’intégrer dans le cadre des approches développées dans la thèse. Il convient néanmoins de lister quelques-unes des limites actuelles des études de génétique du paysage avant d’identifier comment l’utilisation des graphes dans ce domaine pourrait y remédier.

3.4 Limites actuelles des études de génétique du paysage

La génétique du paysage a fait l’objet de plusieurs synthèses bibliographiques ayant identifié ses principales limites et des pistes de recherches futures (Balkenhol *et al.*, 2009a ; DiLeo et Wagner, 2016 ; Dyer, 2015a ; Keller *et al.*, 2015 ; Manel et Holderegger, 2013 ; Richardson *et al.*, 2016 ; Segelbacher *et al.*, 2010). Les limites mentionnées dans ces articles sont de plusieurs ordres. Elles concernent :

- Les questions de recherche et les variables étudiées pour y répondre
- Les régions et taxons étudiés
- Les analyses statistiques mises en œuvre
- La traduction des résultats en préconisations pour la conservation de la biodiversité

Limites des questions de recherche et inadéquations avec les variables étudiées

Tandis que le traitement des données spatiales s’est nettement complexifié, le traitement des données génétiques reste le plus souvent assez basique d’après la synthèse de Dyer (2015a). Par exemple, il est préférable d’utiliser la plupart des mesures de différenciation génétique basées sur des indices de fixation (le F_{ST} par exemple) lorsque l’équilibre entre dérive et flux génétique est atteint à l’échelle de l’ensemble de la zone d’étude (Whitlock et McCauley, 1999). Pourtant, ces mesures sont souvent utilisées en génétique du paysage dans des contextes paysagers ayant changé récemment et dans lesquels cet équilibre n’est probablement pas atteint (Balkenhol *et al.*, 2009a ; Bruggeman *et al.*, 2010 ; Segelbacher *et al.*, 2010 ; Storfer *et al.*, 2010). Au-delà de la nature de la distance génétique, ce point rappelle également l’importance des échelles spatiales et temporelles considérées (Anderson *et al.*, 2010 ; Balkenhol *et al.*, 2009a). Par exemple, Van Strien (2017) et Wagner et Fortin (2013) ont mentionné le fait qu’il pourrait parfois être plus robuste de se focaliser sur un sous-ensemble de paires de populations dans les analyses.

En ce qui concerne l’étude de l’influence du patron spatial des habitats sur la structure génétique des populations, DiLeo et Wagner (2016) ont mis en évidence le fait que la plupart des études constatent un effet prépondérant de la configuration des habitats. Cela viendrait en partie du fait

que la plupart des analyses réalisées mettent en relation la différenciation génétique avec des distances paysagères, et plus rarement avec des quantités d'habitat. On notera également les conclusions contrastées de [Cushman *et al.* \(2012\)](#) et [Jackson et Fahrig \(2015b\)](#) à ce sujet. Alors que les premiers ont conclu que la configuration de l'habitat est le principal facteur de différenciation génétique, les seconds ont montré que c'est la quantité d'habitat à l'échelle du paysage qui influence le plus le degré de différenciation génétique. [Jackson et Fahrig \(2015b\)](#) ont également conclu que ce sont les distances entre taches d'habitat et non leur nombre qui influencent le plus la structure génétique. Leur forte corrélation avec la quantité d'habitat expliquerait ce résultat. [DiLeo et Wagner \(2016\)](#) ont insisté sur l'importance de la prise en compte de la matrice pour mieux comprendre l'influence du patron spatial des habitats sur la structure génétique des populations. Les mesures de connectivité écologique des habitats issues des graphes paysagers et basées sur la quantité d'habitat atteignable ont l'avantage de prendre en compte de façon intégrée la quantité et la configuration de l'habitat ainsi que la résistance de la matrice. Pourtant, elles ont rarement été utilisées en génétique du paysage et pourraient donc permettre d'éclaircir ces questions.

Les variables étudiées en génétique du paysage sont la diversité génétique et la différenciation génétique. Quand elles caractérisent le patron de variation génétique neutre, elles sont toutes deux influencées par deux processus, à savoir la dérive et le flux génétique. Pourtant, le lien entre la diversité génétique et les variables paysagères locales est parfois interprété uniquement vis-à-vis de la dérive tandis que les variables influençant la différenciation génétique sont censées refléter uniquement le flux génétique ([DiLeo et Wagner, 2016](#) ; [Richardson *et al.*, 2016](#)). Il serait donc nécessaire de reconsidérer la complexité de cette relation patron-processus ([Bruggeman *et al.*, 2010](#)), en étudiant simultanément l'effet de différentes variables paysagères sur la diversité et la différenciation génétique. Réaliser ce type d'analyse au même niveau (*node-* et *neighbourhood-level*) serait souhaitable pour rendre les résultats comparables. Ce type d'analyse est rare dans la littérature de génétique du paysage, tout comme les analyses concernant la diversité génétique plus généralement ([DiLeo et Wagner, 2016](#)).

Compte tenu de l'influence importante des variables locales sur la dérive mais aussi sur la dispersion des individus, il serait souhaitable de les considérer davantage dans les analyses, même dans l'analyse de distances génétiques mesurées entre des paires de populations ([Pffüger et Balkenhol, 2014](#)). De même, malgré le fait que l'hétérogénéité spatiale des tailles de populations influence de façon significative la différenciation génétique, son interaction avec d'autres variables déterminant cette réponse génétique a rarement été prise en compte dans les analyses. Cela pourrait permettre de séparer ces différents effets pour mieux comprendre à quel point la matrice paysagère en elle-même affecte le flux génétique.

Biais dans les régions et taxons étudiés

Les études de génétique du paysage sont majoritairement menées en milieux tempérés ([Manel et Holderegger, 2013](#) ; [Storfer *et al.*, 2010](#)) et concernent moins souvent les milieux tropicaux alors que les enjeux liés à la biodiversité y sont très forts ([Myers *et al.*, 2000](#)). Or, les espèces tropicales ont parfois des traits d'histoire de vie très différents de ceux des espèces de milieux tempérés ([Stutchbury *et al.*, 2001](#)). Aussi, les résultats concernant l'influence du paysage sur les espèces de milieux tempérés ne sont peut-être pas transposables en milieu tropical, justifiant que des études soient menées dans ce contexte. De plus, les espèces de vertébrés font l'objet de près des deux tiers de ces études selon

Storfer *et al.* (2010). Multiplier le nombre d'études concernant les invertébrés ou les plantes serait donc souhaitable, d'autant plus que leur rapport à la structure du paysage est potentiellement différent de celui des vertébrés.

Besoin d'amélioration des méthodes statistiques

Les nombreuses analyses statistiques utilisées dans ce domaine mériteraient d'être validées et améliorées si nécessaire (Balkenhol *et al.*, 2009b). C'est en particulier le cas des méthodes permettant les analyses au niveau des paires de populations (*link level*). Compte-tenu des différents niveaux auxquels les analyses sont mises en œuvre, de nouvelles utilisations de méthodes multivariées au niveau des populations (*node-* et *neighbourhood-level*) pourraient également être utiles. Les modèles gravitaires introduits par Murphy *et al.* (2010a) offrent également un potentiel intéressant pour la prise en compte simultanée des variables mesurées au niveau des populations et entre elles (DiLeo et Wagner, 2016). Pour évaluer ces méthodes, les approches basées sur des simulations seraient particulièrement utiles (Balkenhol *et al.*, 2009b ; Landguth *et al.*, 2016).

Manque de traduction des résultats en préconisations pour la conservation

Enfin, une des principales limites de la génétique du paysage est le fait que seuls 4 % des études publiées dans ce domaine sont menées à des fins de conservation de la biodiversité (Bowman *et al.*, 2016). Essentiellement descriptive, la discipline devrait envisager davantage d'approches prédictives selon Manel et Holderegger (2013). Les modèles prédictifs sont en effet très rares dans ce domaine (mais voir Van Strien *et al.* (2014)). Au-delà de l'aspect prédictif, l'analyse de variables facilement interprétables par les gestionnaires encouragerait leur prise en compte de ces résultats (Keller *et al.*, 2015). Par exemple, l'identification de la surface minimale d'habitat en deçà de laquelle la diversité génétique diminue fortement est plus utile aux gestionnaires que les résultats d'une analyse statistique complexe mettant en relation des composantes factorielles paysagères et génétiques. Enfin, les méthodes et outils dédiés à la planification des mesures de conservation gagneraient à être validés à partir de données empiriques pour renforcer la confiance des gestionnaires en leur utilisation.

Dans ce contexte, la thèse vise à trouver des solutions aux limites de la génétique du paysage mentionnées précédemment en utilisant d'une part des graphes paysagers et d'autre part des **graphes génétiques**. Ce second type de graphe est particulièrement utile en génétique du paysage dans la mesure où il permet d'analyser la structure génétique à différents niveaux. Avant d'aborder la façon dont ces outils peuvent être utilisés conjointement aux graphes paysagers en génétique du paysage, nous présentons pourquoi et comment ils sont habituellement construits et analysés.

4 Les graphes génétiques, des outils pour l'analyse de la structure génétique des populations

Les graphes génétiques sont des graphes dont les nœuds représentent des populations⁴⁵, c'est-à-dire des groupes d'individus de la même espèce pour lesquels on dispose de données génétiques. Leurs liens sont quant à eux pondérés par les distances génétiques entre ces populations (Murphy *et al.*, 2016). Comme les populations sont géolocalisées, ces graphes pondérés sont le plus souvent représentés sous la forme de graphes spatiaux, même quand l'information géographique n'est pas utilisée pour définir leur topologie ou pour la pondération des liens. Ils permettent ainsi d'étudier la distribution spatiale de la variabilité génétique en tirant profit des méthodes issues de la théorie des graphes.

Les réseaux de populations ont été représentés tôt sous la forme de graphes en génétique des populations (voir Excoffier *et al.* (1992)), et cette représentation se trouvait dans les travaux pionniers même si le terme "graphe" n'y était pas employé (voir Kimura et Weiss (1964)). La première utilisation de la théorie des graphes pour mettre en relation des distances génétiques et des distances paysagères se trouve dans le travail d'Arnaud (2003). Néanmoins, ce sont Dyer et Nason (2004) qui ont défini explicitement les *population graphs* et mis en évidence leur intérêt en génétique du paysage. Alors que les modèles théoriques tels que le modèle îles-continent ou *stepping-stones* reposent sur une topologie des connexions entre populations relativement schématique, la flexibilité des approches issues de la théorie des graphes permet de dépasser ces topologies afin d'identifier celle qui reflète au mieux les processus génétiques ayant lieu (Greenbaum et Fefferman, 2017). Nous allons voir quels sont les différents objectifs poursuivis lorsqu'on utilise ces graphes et comment ils sont habituellement construits et analysés.

4.1 Objectifs de l'utilisation des graphes génétiques

Tout d'abord, le graphe génétique peut être construit pour localiser les paires de populations échangeant directement des individus lors d'épisodes de dispersion (Figure 30). Ces individus sont habituellement des juvéniles ou propagules dont le déplacement d'une population à une autre engendre un flux génétique. Certaines méthodes d'élagage des graphes génétiques sont censées sélectionner les paires de populations entre lesquelles ces individus se déplacent. La méthode introduite par Dyer et Nason (2004) en est un exemple (cf. section 4.2.1). L'identification de ces liens permet de formuler des hypothèses sur les facteurs empêchant ou facilitant la dispersion des individus. Elle permet également de réaliser des analyses complémentaires basées sur la topologie du graphe pour identifier les populations dont le rôle est clé pour la connectivité de l'ensemble (voir Cross *et al.* (2018)) ou des ensembles de populations échangeant beaucoup d'individus (modules, voir Fortuna *et al.* (2009)).

Tandis que les flux de dispersion des individus se déroulent à l'échelle d'une génération ("directs"), les flux génétiques peuvent eux être multi-générationnels ("indirects") (Figure 30). L'utilisation de graphes génétiques permet ainsi de déterminer le nombre minimum de générations nécessaires pour qu'un allèle traverse le réseau dans son ensemble (Boulanger *et al.*, 2020). Elle permet également de tenir compte de la topologie du réseau de populations pour sélectionner les paires de populations

45. Bien qu'il existe des travaux dans lesquels chaque nœud correspond à un individu (voir Castillo *et al.* (2016)), nous nous focalisons dans le cadre de cette thèse sur les graphes dont les nœuds correspondent à des populations. Ce choix est justifié par notre focalisation sur la structure génétique des populations occupant des taches d'habitat discrètes.

à considérer dans les analyses de type *link level* (Van Strien, 2017). En particulier, des méthodes d'élagage permettent d'ajuster l'échelle spatiale et la topologie des connexions à considérer pour évaluer l'influence du paysage sur la structure génétique (voir Keller *et al.* (2013)).

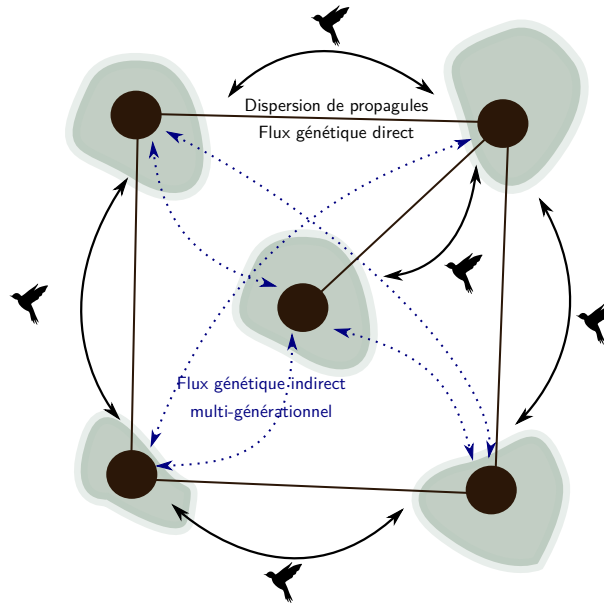


FIGURE 30 – Graphe génétique, flux de dispersion direct et flux génétiques multi-générationnels

4.2 Construction et analyse des graphes génétiques

4.2.1 Méthodes de construction des graphes génétiques

D'une façon similaire à ce que nous avons présenté pour les graphes paysagers (cf. section 2.2.2), la construction d'un graphe génétique implique de définir des nœuds, de pondérer des liens et de choisir une topologie avant d'analyser le graphe.

Données nécessaires et définition des nœuds

Les données nécessaires à la construction d'un graphe génétique sont les génotypes d'individus pour plusieurs *loci*. Les nœuds correspondent le plus souvent aux sites d'échantillonnage des populations d'individus.

Pondération des liens

Les nœuds du graphe génétique sont ensuite reliés par des liens pondérés par une distance génétique. Ainsi, le graphe complet initialement formé n'est autre que la représentation graphique d'une matrice symétrique de distance génétique. De nombreuses distances génétiques ont été utilisées pour cette pondération : le F_{ST} (Creech *et al.*, 2014 ; Munwes *et al.*, 2010), le D_{PS} (Naujokaitis-Lewis *et al.*, 2013), des distances génétiques dites euclidiennes auparavant utilisées en génétique des populations (Dyer et Nason, 2004 ; Fortuna *et al.*, 2009), entre autres (Bruvo *et al.*, 2004 ; Draheim *et al.*, 2016 ; Koen *et al.*, 2012 ; Rozenfeld *et al.*, 2008). Le graphe complet est ensuite élagué et plusieurs méthodes existent pour ce faire.

Élagage du graphe génétique

De la même manière que pour les graphes paysagers, l'élagage d'un graphe génétique complet peut reposer sur des critères biologiques ou topologiques. Par exemple, on peut supprimer tous les liens associés à des distances paysagères supérieures à la distance de dispersion maximale de l'espèce étudiée (Castillo *et al.*, 2016 ; Creech *et al.*, 2014). Des critères topologiques ont quant à eux été utilisés pour construire des graphes de Gabriel (Arnaud, 2003 ; Keller *et al.*, 2013) ou des arbres couvrants minimaux (Naujokaitis-Lewis *et al.*, 2013)⁴⁶. Une autre méthode consiste à supprimer les liens associés aux poids les plus forts jusqu'à ce que le graphe se déconnecte (Figure 31). Si le dernier lien dont la suppression cause la déconnexion est conservé, alors on obtient le graphe connexe dont le poids total des liens est minimal. Le poids de ce lien est appelé "seuil de percolation". Cette méthode a notamment été utilisée pour simplifier un graphe génétique (Munwes *et al.*, 2010 ; Rozenfeld *et al.*, 2008), mais on en trouve également des applications dans les premiers travaux menés sur les graphes paysagers (Keitt *et al.*, 1997 ; Urban *et al.*, 2009). Dans le cas des critères biologiques, il existe généralement une incertitude sur la distance de dispersion maximale ; alors que dans celui des critères topologiques, l'élagage n'a pas nécessairement de signification écologique.

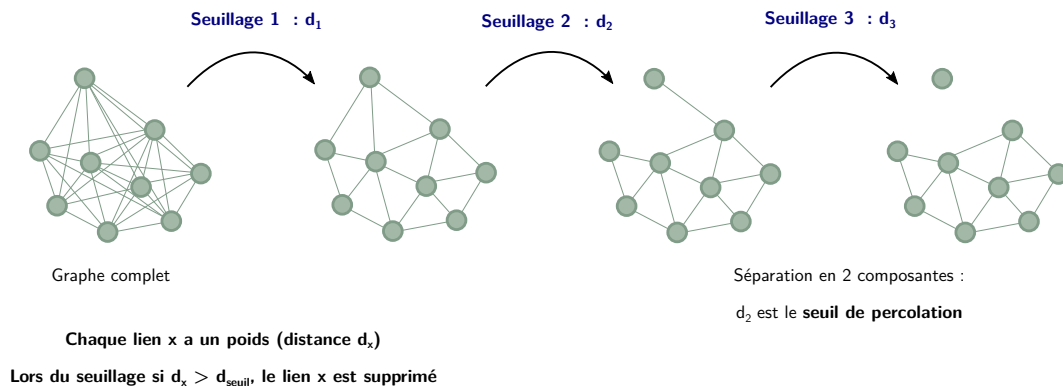


FIGURE 31 – Élagage d'un graphe à l'aide d'un seuil de percolation

La particularité des graphes génétiques est que l'élagage peut aussi être réalisé sur la base d'un critère statistique. Cette méthode, introduite par Dyer et Nason (2004), vise à sélectionner les liens reliant des populations qui échangent directement des migrants. Pour cela, on évalue à quel point les fréquences alléliques mesurées dans deux populations sont corrélées entre elles, tout en tenant compte de leurs éventuelles corrélations avec celles mesurées dans d'autres populations. On évalue ainsi la corrélation partielle entre deux populations situées dans un réseau de populations et un lien est conservé lorsque cette corrélation partielle est significative au sens statistique du terme (cf. Annexe B1). On parle alors de dépendance et d'indépendance conditionnelles entre populations lorsqu'il y a respectivement présence ou absence de lien direct entre elles. Imaginons trois populations A, B et C tel que A et B sont séparées par une route infranchissable par les individus qui les constituent (Dyer, 2015b) (Figure 32). Chacune de ces deux populations échange des migrants avec la population C. Cependant, comme elles n'échangent pas de migrants directement, il y aurait peu de sens à les connecter directement sur un graphe génétique censé représenter les flux de dispersion. Les fréquences alléliques des individus de A et B devraient être similaires en raison de l'échange d'individus entre ces deux populations et la population C. En revanche, la corrélation partielle entre les fréquences

46. Cf. section 2.2.2 pour un rappel de ces termes liés à l'élagage.

alléliques de A et B, c'est-à-dire après prise en compte de leur corrélation avec celles de C, sera nulle. Dans un graphe construit selon le principe d'indépendance conditionnelle, il n'y aura donc pas de lien entre A et B mais des liens entre A et C et entre B et C (Figure 32). Cette méthode a été utilisée à de nombreuses reprises en génétique du paysage et Dyer (2015b) a fait une synthèse de ces applications. Nous y reviendrons dans le premier chapitre.

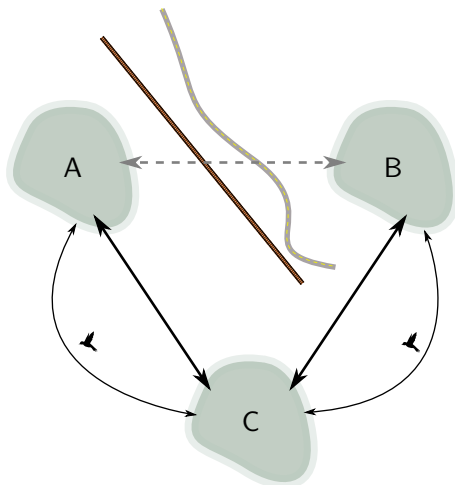


FIGURE 32 – Élagage d'un graphe à l'aide du principe d'indépendance conditionnelle

4.2.2 Méthodes d'analyse des graphes génétiques

Une fois le graphe génétique construit, il peut être analysé à différents niveaux, tout comme nous l'avons vu pour les graphes paysagers.

Analyses au niveau des nœuds

Des métriques issues de la théorie des graphes peuvent être calculées pour l'ensemble du graphe, mais aussi au niveau de chacun des nœuds. Parmi les métriques calculées pour caractériser le rôle de chaque population dans le réseau, on retrouve des indices de centralité tels que le degré (Cross *et al.*, 2018 ; Dyer, 2007), l'indice de centralité intermédiaire (*betweenness centrality*) (Rodger *et al.*, 2018) ou de proximité (*closeness*) (Creech *et al.*, 2014 ; Herrera-Arroyo *et al.*, 2013 ; Pérez-Méndez *et al.*, 2018 ; Schoville *et al.*, 2018). Par exemple, Pérez-Méndez *et al.* (2018) ont calculé ces métriques à partir de graphes génétiques construits pour une espèce de plante (*Neochamaelea pulverulenta*) sur trois îles des Canaries dans lesquelles les communautés d'espèces de lézards responsables de la dispersion des graines sont différentes. Ils ont ainsi mis en évidence des différences de connectivité de ces graphes selon le contexte écologique de chacune des îles.

Cross *et al.* (2018) ont établi une typologie des métriques locales issues des graphes génétiques, distinguant les métriques indiquant les populations jouant le rôle de relais dans le réseau, basées notamment sur le principe de centralité intermédiaire, de celles indiquant l'importance d'une population pour les connexions plus locales. Cette distinction rappelle celle mentionnée au sujet des métriques issues des graphes paysagers (Baranyi *et al.*, 2011). Notons néanmoins que le calcul de métriques telles que l'indice de centralité intermédiaire est basé sur un calcul des chemins les plus courts sur le graphe et implique donc potentiellement des comparaisons entre des sommes de distances génétiques.

La signification d'une telle mesure est donc sujette à précaution. Par ailleurs, [Koen et al. \(2016\)](#) ont également comparé des métriques sur la base de simulations de flux génétiques d'une part et en analysant des données empiriques issues d'une espèce de Martre (*Martes americana*) d'autre part. Ils ont observé une corrélation forte entre le nombre de migrants passant par un nœud lors des simulations et la moyenne des poids de ses liens ou de l'inverse des poids de ses liens. Cette moyenne de l'inverse des poids des liens, que nous appellerons MIW (*Mean Inverse Weight*) dans la suite de cette thèse (Figure 33), était également corrélée positivement à la qualité de l'habitat des sites occupés par les populations de Martre. Dans un graphe génétique, l'inverse du poids d'un lien correspond à l'inverse d'une différenciation génétique et est donc d'autant plus fort que la différenciation est faible entre deux populations⁴⁷. À l'inverse, le degré, la *betweenness* et la *closeness* d'un nœud étaient faiblement corrélés au nombre de migrants dans l'étude de [Koen et al. \(2016\)](#).

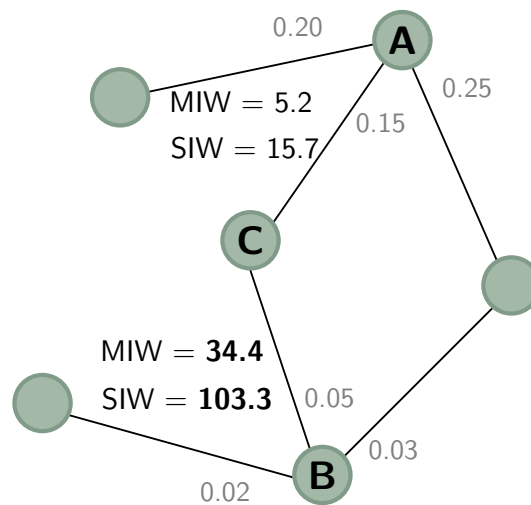


FIGURE 33 – Signification écologique de la moyenne de l'inverse des poids des liens d'un graphe génétique. Les valeurs en gris au niveau des liens correspondent aux poids des liens, c'est-à-dire à des indices de différenciation génétique (par exemple le D_{PS}). MIW (*Mean Inverse Weight*) : moyenne de l'inverse des poids des liens. SIW (*Sum of Inverse Weight*) : somme de l'inverse des poids des liens. La population B est moins différente d'un point de vue génétique des populations auxquelles elle est connectée que ne l'est la population A. Cela se traduit par des indices MIW et SIW plus élevés.

Analyses au niveau des liens

Les graphes génétiques étant construits à partir de matrices de distances génétiques, ils sont également utiles pour les analyses de type *link level* en génétique du paysage ([Dyer et al., 2010](#) ; [Garroway et al., 2011](#) ; [Van Strien, 2017](#)). Ils permettent notamment de sélectionner les paires de populations à considérer dans les analyses en se basant sur un critère d'élagage topologique (voir [Keller et al. \(2013\)](#)). Des études dans lesquelles un sous-ensemble de paires de populations est considéré sur la base d'un critère de distance et sans qu'il soit fait mention de la théorie des graphes existent également dans la littérature (voir [Angelone et al. \(2011\)](#)). Elles montrent indirectement l'intérêt potentiel de l'élagage d'un graphe dans ce type d'analyse.

Lorsqu'un sous-ensemble de liens est sélectionné, il est courant de considérer la matrice de distance incomplète ainsi obtenue. [Dyer et al. \(2010\)](#) ont toutefois introduit une approche alternative consistant

47. Cette corrélation n'est pas sans rappeler le lien de proportionnalité inverse entre le F_{ST} et le nombre de migrants du modèle des îles de [Wright \(1931\)](#). Il est tel que : $F_{ST} = \frac{1}{1+4Nm}$ avec N la taille de chaque population et m le taux de migration.

à recréer une matrice de distance complète à partir de cette matrice "élaguée". Pour ce faire, on calcule la somme des distances génétiques entre les populations connectées le long du chemin le plus court sur le graphe génétique élagué. Le poids obtenu est associé au lien entre deux populations lorsque le lien direct entre ces populations n'existe pas (Figure 34). Il est nécessaire que le graphe soit connexe pour que la matrice de distance complète puisse être créée (Dyer, 2015b). Telle que présentée par Dyer *et al.* (2010), cette méthode s'applique aux graphes élagués à l'aide du principe d'indépendance conditionnelle (*population graphs sensu* Dyer et Nason (2004)). Les distances génétiques obtenues s'appellent donc distances génétiques conditionnelles (*cGD* pour *conditional genetic distance*). Dyer *et al.* (2010) ont montré à l'aide de simulations que ce type de distance reflète mieux la réalité des processus simulés qu'une matrice complète de F_{ST} . Cette méthode a notamment été utilisée par Garroway *et al.* (2011) pour calibrer une surface de résistance. Néanmoins, elle pourrait théoriquement s'appliquer à tous les types d'élagage et à tous les types de distances génétiques, comme nous le verrons dans le chapitre 1.

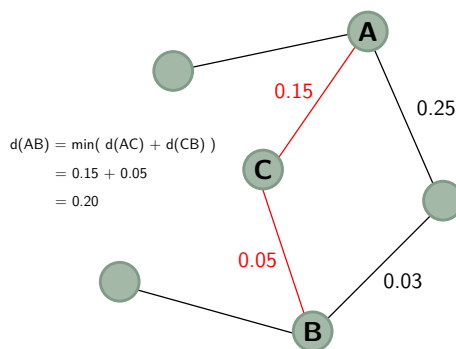


FIGURE 34 – Calcul de la distance génétique conditionnelle

Analyses au niveau des modules

Au sein des réseaux de populations représentés sous la forme de graphes génétiques, on distingue parfois des ensembles de populations qui sont nettement plus proches génétiquement les unes des autres au sein d'un même ensemble qu'elles ne le sont des populations d'autres ensembles. Elles forment des modules sur le graphe qu'on peut identifier à partir d'algorithmes de modularité (Albert *et al.*, 2013 ; Fletcher *et al.*, 2013 ; Fortuna *et al.*, 2009 ; Peterman *et al.*, 2016 ; Oppen *et al.*, 2011), comme pour les graphes paysagers (cf. section 2.2.2). Dans le cas des graphes génétiques, les algorithmes de modularité prennent uniquement en compte la présence et le poids des liens entre les populations. Ils constituent donc des alternatives basées sur les distances génétiques aux méthodes de *clustering* communément utilisées en génétique du paysage et reposant habituellement sur des postulats d'équilibre (équilibre de Hardy-Weinberg notamment) (Greenbaum *et al.*, 2016).

En somme, les graphes génétiques peuvent être construits de différentes façons et analysés à différents niveaux. Cette flexibilité comporte des avantages pour les analyses de génétique du paysage mais également des limites et des parts d'ombre, que nous allons aborder à présent.

4.3 Limites de l'utilisation des graphes génétiques

Quelle adéquation entre les méthodes utilisées et le contexte des recherches ?

La distinction faite à la section 4.1 entre les différentes questions auxquelles les graphes génétiques peuvent apporter des réponses (Figure 30) est rarement faite en ces termes dans la littérature relative

à ces outils. Elle nous paraît néanmoins cruciale dans la mesure où elle devrait déterminer les méthodes de construction et d'analyse de ces graphes. Parmi la multitude de méthodes existantes, il est encore difficile pour l'utilisateur de choisir alors que ce choix a une influence (cf. Keller *et al.* (2013) ; Van Strien (2017)). Nous pensons que ce choix doit être déterminé par les objectifs poursuivis mais aussi par le contexte dans lequel les graphes génétiques sont utilisés (taxon, zone d'étude, histoire évolutive et contemporaine des populations étudiées). Par exemple, les capacités de dispersion de l'espèce devraient d'une part déterminer la méthode d'élagage choisie pour mettre en évidence les flux de dispersion. D'autre part, le patron de différenciation génétique et son adéquation avec l'équilibre migration-dérive devraient déterminer l'élagage mis en œuvre pour évaluer l'influence du paysage sur le flux génétique.

Par ailleurs, les distances génétiques utilisées pour pondérer les liens ont potentiellement une influence sur les résultats obtenus, tout comme les différentes façons dont les matrices de distance incomplètes sont intégrées aux analyses. Compte tenu de tous ces éléments, une comparaison des différentes méthodes de construction et de d'analyse des graphes génétiques semble nécessaire. Elle constitue l'objet du chapitre 1 de cette thèse.

Comment mettre en relation un graphe génétique avec des variables paysagères ?

Les graphes génétiques revêtent une grande diversité de formes. Néanmoins, il n'existe pas de logiciel unique permettant de créer ces différents graphes génétiques et d'évaluer l'importance des paramètres de construction, à l'instar du logiciel Graphab dans le cas des graphes paysagers. Nous avons donc développé un outil de ce type, il fait l'objet du chapitre 2 de cette thèse.

Une couverture partielle du réseau de populations ?

En génétique du paysage, les analyses sont limitées lorsqu'un faible nombre de populations a été échantillonné et que le protocole d'échantillonnage n'assure pas un maillage suffisamment dense et uniforme de la zone d'étude. Cette limite affecte également les graphes génétiques dont les nœuds ne correspondent pas à l'ensemble exhaustif des populations de la zone d'étude, contrairement aux graphes paysagers qui considèrent l'ensemble des taches d'habitat potentielles de cette zone. Des travaux ont montré que les métriques issues des graphes génétiques étaient sensibles au nombre de populations composant le graphe (Koen *et al.*, 2013 ; Naujokaitis-Lewis *et al.*, 2013). Il serait donc nécessaire d'étudier à quel point les résultats obtenus à partir de ces graphes sont, d'une part, valides et d'autre part généralisables à un plus large ensemble de populations à l'échelle de la zone d'étude.

5 Enjeux de l'utilisation conjointe des graphes génétiques et paysagers

Comme nous l'avons vu, les graphes paysagers sont des modèles de la connectivité écologique des habitats qui trouvent leur origine dans la théorie des métapopulations et les concepts de l'écologie du paysage. Ils représentent l'ensemble exhaustif des taches d'habitat et chemins de dispersion potentiellement utilisés par les espèces étudiées. Bien que le réalisme écologique de ces outils manque d'une validation à partir de données biologiques, le patron de connectivité qu'ils modélisent est censé déterminer les flux génétiques entre populations. D'autre part, les graphes génétiques, développés dans le cadre de la génétique des populations et de la génétique du paysage, ne considèrent le plus souvent qu'un sous-ensemble de populations mais permettent d'étudier leur structure génétique à partir de données empiriques. Ils donnent ainsi un aperçu du patron de connectivité réalisée. Dans la mesure où les graphes génétiques et paysagers s'appuient tous deux sur la large gamme de méthodes d'analyse offertes par la théorie des graphes, il serait opportun de combiner ces approches. En effet, graphes génétiques et paysagers ont été utilisés à de nombreuses reprises en écologie du paysage, génétique des populations et génétique du paysage (Galpern *et al.*, 2011 ; Greenbaum et Fefferman, 2017) (Annexe A2⁴⁸), mais rarement de façon conjointe. De précédents travaux ont suggéré d'utiliser des données génétiques dans le cadre de la modélisation de la connectivité écologique à l'aide de graphes paysagers (Foltête et Vuidel, 2017 ; Luque *et al.*, 2012). Toutefois, ils ne mentionnaient pas explicitement la complémentarité entre les graphes génétiques et paysagers. De même, Garroway *et al.* (2011), Manel et Holderegger (2013) et Murphy *et al.* (2016) ont mis en évidence l'intérêt potentiel de la comparaison ou de l'intégration de ces graphes mais n'ont pas détaillé précisément les modalités de ces utilisations conjointes car ce n'était pas l'objectif premier de leurs travaux. Nous pensons qu'il est nécessaire de le faire pour étendre la contribution de ces approches basées sur des graphes en génétique du paysage.

Dans ce contexte, cette thèse vise à étudier les modalités d'utilisation conjointe des graphes génétiques et paysagers dans le but de dépasser les limites inhérentes à ces deux types d'outil, tout en répondant aux questions pour lesquelles on manque encore de réponses en génétique du paysage. En somme, elle vise à répondre à la question suivante :

- **Comment utiliser conjointement des graphes génétiques et paysagers pour modéliser la connectivité écologique des habitats et comprendre son influence sur la structure génétique des populations ?**

Nous présentons désormais les enjeux de cette utilisation conjointe (cf. Annexe A2⁴⁹).

48. En complément de l'article de positionnement mentionné ci-dessous, nous avons réalisé un état de l'art concernant les méthodes de construction et d'analyse de graphes génétiques et paysagers construits dans le cadre de 107 articles scientifiques.

49. Cette section a donné lieu à un article de positionnement soumis à la publication dans la revue *Landscape Ecology* en novembre 2020 : Savary, P., Foltête, J. C., Savary, P., Moal, H. & Garnier, S. Combining landscape and genetic graphs to address key issues in landscape genetics. Soumis à *Landscape Ecology*. La version intégrale de cet article constitue l'Annexe A2.

La génétique du paysage hérite une grande partie de ses méthodes et de ses questions de recherche à la fois de l'écologie du paysage et de la génétique des populations. Quand un écologue du paysage se demande si la quantité d'habitat importe plus que sa configuration pour les réponses biologiques (Lindenmayer *et al.*, 2020), un généticien des populations se demande si la structure du paysage influence davantage la diversité génétique locale ou la différenciation génétique (Keyghobadi *et al.*, 2005). De même, les études d'écologie du paysage visant à identifier l'échelle à laquelle le paysage influence les réponses biologiques ("*scale of effect*") (Jackson et Fahrig, 2012 ; Miguet *et al.*, 2016) s'apparentent aux études de génétique des populations dont le but est d'identifier la taille de voisinage ou l'échelle à laquelle à la fois la dérive et le flux génétique influencent la différenciation génétique (Hardy et Vekemans, 1999 ; Van Strien *et al.*, 2015). D'autres questions parallèles existent entre ces domaines et nous pensons que chacune d'entre elles pourrait être traitée dans des analyses impliquant les propriétés des nœuds et des liens de graphes génétiques et paysagers. L'utilisation conjointe des graphes génétiques et paysagers serait un pas de plus vers la véritable interdisciplinarité recherchée en génétique du paysage (Dyer, 2015a).

Nous présentons dans un premier temps comment la simple comparaison de ces graphes permettrait de répondre à certaines questions théoriques. Puis, nous discutons de l'intérêt potentiel de l'intégration de ces deux outils dans la mesure où elle permettrait de tirer profit de leur complémentarité. Ces deux approches constituent les deux parties principales de cette thèse.

5.1 Comparer graphes génétiques et paysagers pour traiter des problématiques de génétique du paysage

Associer des prédicteurs paysagers et des réponses génétiques aux éléments de graphes spatiaux similaires ouvre des opportunités d'analyses au niveau des nœuds, de leur voisinage, des liens et des limites entre modules (*node-*, *neighbourhood-*, *link-* et *boundary-level* respectivement, *sensu* Wagner et Fortin (2013)). En génétique du paysage, comparer des graphes génétiques et paysagers à ces différents niveaux permettrait de traiter des thématiques telles que (i) les relations entre les composantes de la structure du paysage et de la structure génétique, (ii) l'échelle spatiale à laquelle ces relations s'établissent et (iii) l'identification de barrières à la dispersion (Figure 35).

5.1.1 Analyser les relations entre les différentes composantes de la structure du paysage et de la structure génétique

Des variables paysagères et des réponses génétiques peuvent toutes deux être mesurées (i) au niveau local pour décrire chaque tache d'habitat ou population (*node level*) ou (ii) en tenant compte des connexions qu'établit chaque tache ou population avec d'autres taches ou populations du réseau (*neighbourhood level*) (Wagner et Fortin, 2013). Les graphes plus que n'importe quels autres outils d'analyse permettent de calculer ces deux types de variables pour caractériser les différentes composantes de la structure du paysage et de la structure génétique des populations (Figure 35).

La génétique des populations et l'écologie du paysage considèrent que la structure génétique des populations dépend à la fois des caractéristiques des taches d'habitat et de leur position dans le réseau d'habitat. En effet, diversité et différenciation génétiques sont toutes deux influencées à la fois par la dérive et le flux génétique, qui sont eux-mêmes influencés par la connectivité des habitats. Nous avons

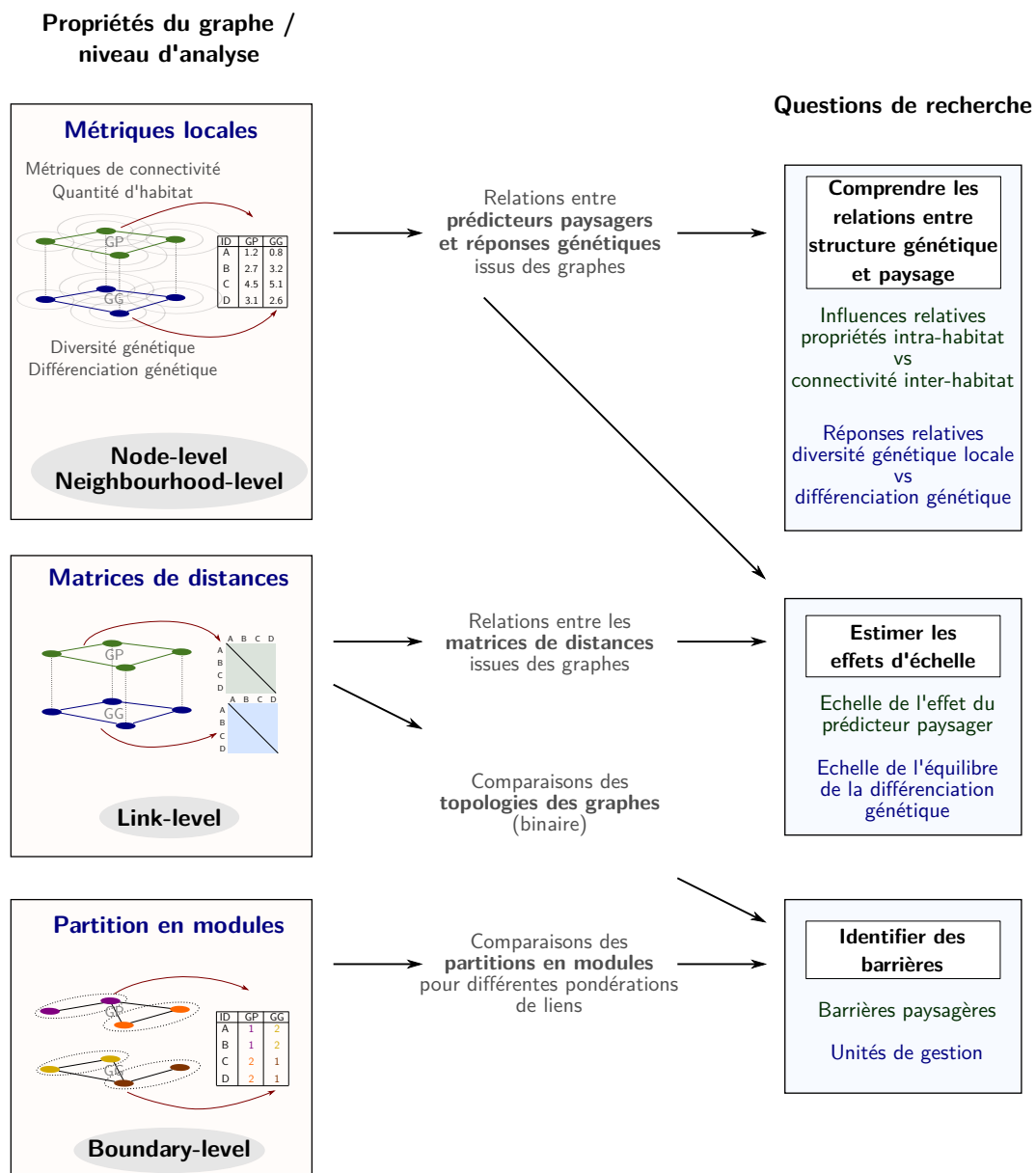


FIGURE 35 – Enjeux de l'utilisation conjointe des graphes génétiques et paysagers

vu dans la section 2 que les graphes paysagers fournissent un cadre flexible pour calculer des métriques mesurant de façon complémentaire la quantité d'habitat atteignable dans le réseau d'habitat. On peut donc s'attendre à ce que ces métriques soient de bons prédicteurs des différentes composantes de la structure génétique des populations.

Les graphes génétiques permettent quant à eux de calculer des indices génétiques au niveau des populations en tenant compte de leurs caractéristiques locales (par exemple la richesse allélique, *node level*) ou de leur différenciation génétique avec chacune des autres populations (par exemple l'indice MIW, *neighbourhood level*, figure 33).

Lorsque des métriques sont calculées à partir de ces deux types de graphes, leurs relations peuvent être étudiées par le biais d'analyses statistiques. Pour que les relations mises en évidence entre la structure du paysage et des réponses biologiques soient comparables et leurs différences interprétables, il est préférable qu'elles soient évaluées avec le même type d'analyse statistique (Leroux *et al.*, 2017). Les

métriques locales de type "*neighbourhood level*" issues de graphes génétiques et paysagers ont l'avantage d'être calculées en intégrant les connexions et les distances entre les taches ou les populations, qu'il s'agisse de distances paysagères ou génétiques, respectivement. Cela rend possible l'utilisation de modèles dans lesquels les individus statistiques sont des taches ou des populations, que l'on vise à expliquer la diversité ou la différenciation génétique et que l'on utilise la quantité d'habitat locale ou atteignable en dehors des taches comme variable explicative. Une telle approche permet également d'intégrer des distances dans l'analyse autrement qu'en utilisant les tests de Mantel, dont les limites sont connues (Balkenhol *et al.*, 2009b ; Legendre et Fortin, 2010).

5.1.2 Évaluer les effets d'échelle

Les résultats obtenus en comparant des variables calculées au niveau des nœuds de graphes génétiques et paysagers dépendront probablement de la taille du voisinage considéré pour ces calculs, et donc de l'échelle à laquelle les liens sont pris en compte ou pondérés de façon significative dans les calculs. En effet, les effets d'échelle sont omniprésents en écologie du paysage (Brooks, 2003 ; Jackson et Fahrig, 2012 ; Miguet *et al.*, 2016) et en génétique des populations (Hutchison et Templeton, 1999).

Dans les graphes paysagers, la pondération des liens dans le calcul des métriques détermine en quelque sorte la taille du voisinage considéré (cf. section 2.2.2). Lorsque l'utilisateur convertit des distances paysagères en probabilités de dispersion avec une fonction exponentielle négative, il choisit une constante déterminant la distance à laquelle la probabilité de dispersion devient négligeable (Figure 16). Une métrique de connectivité est fréquemment calculée en attribuant à cette constante une gamme de valeurs différentes. En évaluant à chaque fois la corrélation de cette métrique avec les réponses génétiques issues du graphe génétique et en identifiant la valeur de la constante qui maximise cette corrélation, l'utilisateur peut évaluer l'échelle à laquelle la connectivité influence la structure génétique. Comme les variables calculées au niveau des nœuds du graphe génétique reflètent la diversité génétique locale ou la différenciation génétique, cette échelle peut être évaluée pour chacune de ces réponses et la comparaison est possible. Cette approche a également l'avantage d'intégrer l'hétérogénéité de la matrice paysagère via l'utilisation de distances-coût dans le calcul des métriques. Cela conduit à considérer un voisinage anisotrope pour évaluer l'échelle à laquelle les effets s'exercent tandis que l'utilisation de distances géodésiques euclidiennes impliquant un voisinage isotrope est la norme jusqu'à présent (mais voir Miguet *et al.* (2017)).

Par ailleurs, la façon dont les graphes ont été élagués détermine les connexions prises en compte lors du calcul des métriques au niveau des nœuds (*neighbourhood level*). Comparer la force de la relation entre les réponses génétiques et les métriques paysagères issues de graphes génétiques et paysagers élagués de différentes manières peut donc fournir des informations sur la topologie la plus probable des événements de dispersion.

Les graphes génétiques et paysagers peuvent aussi être comparés en évaluant la relation entre des matrices de distances calculées à partir de leurs liens entre des nœuds équivalents, c'est-à-dire les populations du graphe génétique et les taches d'habitat du graphe paysager occupées par ces populations (Figure 35). Différentes méthodes et modèles sont couramment utilisés pour ce type d'analyse (*link level*) en génétique du paysage (corrélations de Mantel, modèles linéaires mixtes, entre autres). L'utilisation de plusieurs méthodes d'élagage des graphes permet d'identifier l'échelle à laquelle cette relation

se met en place. Sur un graphe paysager élagué, les distances entre taches sont calculées comme la somme des distances paysagères associées aux liens du chemin le plus court. Elles permettent de tester la vraisemblance des modèles de dispersion de type "stepping-stones" (Saura *et al.*, 2014). En revanche, l'élagage d'un graphe génétique élimine de l'analyse les valeurs de distances associées à certaines paires de populations, ce qui pourrait améliorer la puissance de ce type d'analyse selon Wagner et Fortin (2013). Nous pensons également que l'élagage du graphe génétique qui maximise l'ajustement de la relation entre les distances génétiques et paysagères fournit une information sur l'échelle à laquelle la différenciation génétique dépend à la fois de la dérive et du flux génétique. Cette question se pose depuis longtemps en génétique des populations (Hardy et Vekemans, 1999 ; Hutchison et Templeton, 1999).

5.1.3 Identifier des barrières dans le paysage

Les études de génétique du paysage ont parfois pour objectif d'identifier des barrières paysagères isolant des groupes de populations les uns des autres (Manel *et al.*, 2003 ; Segelbacher *et al.*, 2010 ; Storfer *et al.*, 2007). Si un graphe paysager représente correctement le réseau d'habitat, il ne devrait pas inclure de liens entre les populations qui n'échangent pas d'individus, mettant ainsi en évidence l'absence de connectivité entre leurs taches d'habitat. Quand l'élagage des graphes génétiques est basé sur un critère purement génétique, par exemple lors de la création de "population graphs" selon le principe d'indépendance conditionnelle (Dyer et Nason, 2004), leur topologie est indépendante de toute hypothèse concernant la structure du paysage. Elle peut alors être comparée à celle du graphe paysager pour évaluer sa validité écologique. Cette analyse ("topological congruence" *sensu* Dyer (2015b)) est un moyen d'identifier les barrières à la dispersion ainsi que les capacités de dispersion maximales de l'espèce étudiée. Quand les graphes partagent exactement les mêmes nœuds, elle consiste à comparer deux classifications binaires (présence ou absence des liens) et peut se faire à l'aide d'indices de comparaison de classifications tels que l'indice de corrélation de Matthews (1975) (voir aussi Fletcher *et al.* (2011)).

Les modules calculés au niveau de graphes génétiques et paysagers peuvent aussi être comparés (*boundary level*) (Figure 35). Deux partitions en modules seront similaires si deux nœuds du même module d'un graphe sont aussi dans le même module dans l'autre graphe. La similarité des partitions peut s'évaluer avec des indices de comparaison de partitions tels que l'*Adjusted Rand Index* (Hubert et Arabie, 1985) ou l'indice de *Normalized Mutual Information* (Danon *et al.*, 2005 ; Reichert *et al.*, 2016). Une forte similarité indiquerait que la structure spatiale de la variation génétique est due à l'existence de barrières à la dispersion. Elle validerait aussi l'identification d'unités de gestion des populations réalisée à partir de graphes paysagers (Foltête et Vuidel, 2017).

Même si la comparaison des graphes génétiques et paysagers offre donc des perspectives prometteuses, elle a peu été exploitée jusqu'à présent (mais voir Castillo *et al.* (2016), Creech *et al.* (2014), Draheim *et al.* (2016) et Schoville *et al.* (2018)). Par ailleurs, l'intérêt de l'utilisation conjointe de ces deux types de graphes ne se réduit pas au fait qu'elle permet de valider la pertinence écologique des graphes paysagers. Comme nous l'avons vu précédemment, alors que les graphes génétiques ne considèrent généralement qu'un nombre réduit de populations, les graphes paysagers représentent l'ensemble exhaustif des taches d'habitat potentielles (Figure 36). Les graphes génétiques et paysagers sont

donc complémentaires et la modélisation de la connectivité bénéficierait d'une véritable interaction entre ces outils.

5.2 Intégrer graphes génétiques et paysages pour bénéficier de leur complémentarité

Au-delà de la comparaison de ces deux types de graphes, une autre manière de les utiliser conjointement consisterait à les intégrer. Cela consisterait à combiner les données et les méthodes habituellement utilisées séparément pour construire ces graphes dans la même analyse. Une telle approche permettrait de développer des modèles prédictifs, dont l'utilisation en génétique du paysage est actuellement rare malgré l'intérêt qu'ils auraient du point de vue de la conservation (mais voir [Van Strien *et al.* \(2014\)](#)). Si des modèles robustes liant les propriétés de graphes génétiques et paysagers étaient calibrés, leur extrapolation permettrait d'identifier le rôle potentiel pour la connectivité et les flux génétiques d'un grand nombre de taches d'habitat, échantillonnées aussi bien que non échantillonnées. Par exemple, [Creech *et al.* \(2014\)](#) ont extrapolé un modèle liant des distances paysagères à des flux génétiques en utilisant une approche basée sur des graphes. Ils ont ainsi modélisé les différences de topologie des réseaux de dispersion et de flux génétique liées au fait que la dispersion s'effectue différemment selon le sexe des individus chez une espèce d'ongulé sauvage (*Ovis canadensis nelsoni*).

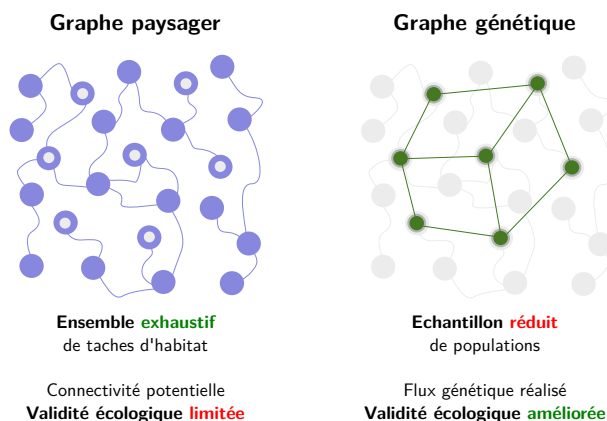


FIGURE 36 – Complémentarité des graphes génétiques et paysagers

L'utilisation de modèles gravitaires ([Murphy *et al.*, 2010a](#)) mènerait plus avant l'intégration de ces deux types de graphes. Dans ces modèles, des variables prédictives calculées au niveau des taches d'habitat ainsi qu'entre les taches sont incluses dans la même analyse. L'évaluation de l'ajustement de ces modèles indiquerait leur pertinence en tant que modèles prédictifs. De plus, inclure des estimations de la taille des populations en tant que variable explicative pourrait permettre de s'affranchir de l'influence de l'hétérogénéité spatiale des tailles de populations sur l'intensité de la dérive, et potentiellement d'expliquer de façon plus fiable l'influence de la matrice paysagère sur la différenciation génétique ([Prunier *et al.*, 2017](#)). Les modèles gravitaires ont déjà été utilisés en génétique du paysage ([Murphy *et al.*, 2010a](#) ; [Robertson *et al.*, 2018a](#) ; [Zero *et al.*, 2017](#)), mais ils n'ont pas été combinés à des graphes génétiques et paysagers pour modéliser voire prédire la différenciation génétique. Nous pensons que cette approche présente un intérêt pour la compréhension des processus en jeu ainsi que dans une perspective de conservation.

Organisation du manuscrit de thèse

La première partie de ce manuscrit est dédiée à la comparaison des graphes génétiques et paysagers. Dans un premier temps, nous présentons une analyse des méthodes de construction et d'analyse des graphes génétiques à mettre en œuvre selon l'objectif poursuivi et le contexte des recherches. Nous présentons ensuite le package R `graph4lg`, développé pour permettre la construction et l'analyse des graphes génétiques et paysagers de façon séparée ou conjointe et à différents niveaux.

Suite à ces deux chapitres, nous comparons ces deux types de graphes dans deux contextes différents. Premièrement, nous évaluons à quel point des métriques issues d'un graphe paysager et évaluant de façon complémentaire la quantité d'habitat atteignable expliquent les différentes composantes de la structure génétique des populations décrites à l'aide d'un graphe génétique. Ce travail se base sur des données génétiques issues des populations de Criquet ensanglanté (*Stethophyma grossum*) d'un paysage agricole suisse. Il vise à valider empiriquement la définition de la connectivité écologique des habitats présentée en introduction. De plus, il illustre l'intérêt de la comparaison des graphes pour comprendre les processus écologiques reliant la connectivité à la structure génétique des populations.

Ensuite, nous évaluons la capacité de différentes méthodes de construction de graphes paysagers à expliquer la structure génétique des populations à partir d'analyses basées sur les nœuds et les liens de graphes génétiques et paysagers. Nous nous basons pour cela à la fois sur des observations de présence/absence et sur des données génétiques concernant la Paruline caféïette (*Setophaga plumbea*), passereau forestier de l'île de Guadeloupe. Ce chapitre évalue la pertinence des graphes paysagers dans l'optique d'une mise en œuvre opérationnelle.

La seconde partie est dédiée à l'intégration des graphes génétiques et paysagers. Elle considère en particulier l'intérêt de ces graphes pour la calibration d'une surface de résistance à partir de données génétiques. Une telle analyse implique d'identifier le scénario de coût qui explique au mieux la structure génétique des populations. Nous nous intéressons donc d'abord à la sensibilité des chemins de moindre coût et des distances-coût aux scénarios de coût, dans la mesure où cette sensibilité détermine la qualité de l'inférence de la surface de résistance à partir de données génétiques.

Le dernier chapitre évalue l'intérêt de l'utilisation de modèles gravitaires intégrant à la fois les propriétés des liens et des nœuds de graphes génétiques et paysagers pour inférer la résistance du paysage au flux génétique. Il porte en particulier sur la prise en compte des propriétés des nœuds des graphes pour intégrer l'hétérogénéité spatiale des tailles de populations dans l'analyse.

Enfin, une discussion générale clôture ce manuscrit. Nous y revenons sur l'intérêt des graphes génétiques et paysagers pour la compréhension de la relation entre connectivité écologique des habitats et structure génétique ainsi que pour la conservation de la connectivité. Nous identifions également les perspectives de recherche auxquelles cette thèse ouvre la voie.

Partie I

Comparaison de graphes génétiques et paysagers

Chapitre 1

Comment construire et analyser des graphes génétiques pour étudier l'effet du paysage sur les réseaux de dispersion et de flux génétique ?

Résumé

Quand une espèce forme des populations dans des taches d'habitat discrètes, les graphes génétiques peuvent être utilisés pour (i) identifier les chemins de dispersion directs suivis par les individus ou (ii) quantifier les effets du paysage sur le flux génétique multi-générationnel. L'influence des paramètres de construction de ces graphes sur ces analyses restait néanmoins à déterminer. Suite à des simulations de flux génétiques, nous avons construit des graphes génétiques en utilisant plusieurs méthodes d'élagage (distances géodésiques seuils, contraintes topologiques, inférence statistique) et plusieurs distances génétiques pour pondérer les liens (F_{ST} , D_{PS} , distances génétiques euclidiennes). Nous avons ensuite comparé la capacité de ces différents graphes à (i) identifier la topologie précise du réseau de dispersion et (ii) inférer la résistance du paysage au flux génétique à partir de la relation entre distances-coût et distances génétiques. Même s'ils n'étaient pas toujours très tranchés, nos résultats montrent que les méthodes basées sur des distances géographiques seuils sont celles qui identifient le mieux les réseaux de dispersion. Bien que cela soit contre-intuitif, ils démontrent aussi que la sélection d'un sous-ensemble de distances entre populations via l'élagage du graphe peut améliorer l'inférence des effets du paysage sur le flux génétique. Finalement, nous montrons que des distances génétiques telles que le D_{PS} ou des distances génétiques euclidiennes devraient être préférées au F_{ST} pour l'inférence des effets du paysage vu qu'elles répondent plus rapidement aux changements paysagers.

Cet article a été publié dans la revue *Molecular Ecology Resources* en janvier 2021 :

Savary, P., Foltête, J. C., Moal, H., Vuidel, G. & Garnier, S. 2021. Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, 21(4), 1167-1185

*I invoke the first law of geography :
everything is related to everything else,
but near things are more related than distant things.*

Waldo R. Tobler

1 Introduction

Les études de génétique du paysage permettent d'évaluer l'effet du paysage sur la dispersion et le flux génétique car les données génétiques reflètent les mouvements de dispersion qui sont suivis d'une reproduction, tandis que les données GPS ou de capture-marquage-recapture reflètent tous les types de mouvements indifféremment (Mateo-Sánchez *et al.*, 2015 ; Zeller *et al.*, 2018). Néanmoins, l'analyse et l'interprétation des données génétiques posent encore des questions théoriques et méthodologiques (Balkenhol *et al.*, 2009a,b ; Dyer, 2015a).

Les approches basées sur des graphes sont particulièrement pertinentes pour étudier la dispersion au sein d'un réseau de populations (Greenbaum et Fefferman, 2017). Dans ce cas, un graphe génétique est construit à partir d'une matrice d'adjacence contenant des 0 et des 1 traduisant respectivement l'absence et la présence de liens entre les populations. Ces liens sont généralement pondérés par des distances génétiques, c'est-à-dire des mesures de différenciation génétique. Il est souvent recommandé d'élaguer le graphe complet pour rendre sa topologie plus directement interprétable visuellement et pour ne garder que les liens les plus pertinents au regard de l'objectif de l'étude.

Le flux génétique entre deux populations peut résulter d'un échange direct d'individus et/ou d'échanges indirects via des populations intermédiaires sur plusieurs générations. Une simple mesure de différenciation génétique entre populations ne permet pas de distinguer ces deux situations. En revanche, l'utilisation du principe d'indépendance conditionnelle (Magwene, 2001 ; Whittaker, 2009) est supposée le permettre. En effet, les fréquences alléliques de deux populations sont censées être indépendantes conditionnellement à celles des autres populations si elles n'échangent des allèles que par le biais de flux génétiques indirects (Dyer, 2015b). Ainsi, l'utilisation de cette méthode pour élaguer un graphe génétique devrait identifier la topologie précise du réseau de dispersion, c'est-à-dire le jeu de liens suivis par les individus entre les populations.

Par ailleurs, un graphe génétique peut être utilisé pour quantifier la résistance des éléments du paysage au flux génétique à l'aide d'analyses basées sur les distances (*link level*) (Garroway *et al.*, 2011). Évaluer la corrélation entre distances génétiques et paysagères permet d'identifier l'hypothèse qui s'ajuste le mieux aux données génétiques, reflétant ainsi l'influence du paysage sur le flux génétique (Cushman *et al.*, 2006 ; Khimoun *et al.*, 2017 ; Peterman, 2018 ; Ruiz-Gonzalez *et al.*, 2015). Bien que ce type d'inférence soit habituellement basé sur des matrices complètes de distances, plusieurs auteurs ont suggéré que réduire ces matrices à un sous-ensemble de populations pourrait améliorer la puissance de l'analyse (Van Strien *et al.*, 2015 ; Van Strien, 2017 ; Wagner et Fortin, 2013 ; Zeller *et al.*, 2016). L'élagage du graphe génétique consiste précisément à restreindre l'analyse à un sous-ensemble de paires de populations, ce qui rend les graphes génétiques potentiellement utiles dans ce contexte.

Réduire un jeu de données en éliminant des paires de populations pour améliorer l'inférence de la résistance du paysage peut paraître contre-intuitif. Le principe sous-jacent est le suivant. Étant donné que la dispersion est généralement limitée spatialement, les modèles théoriques de génétique des populations prédisent une corrélation linéaire et continue entre la différenciation génétique et la distance géodésique entre populations, à condition que suffisamment de temps se soit écoulé pour qu'un équilibre se soit établi (Guillot *et al.*, 2009 ; Kimura et Weiss, 1964 ; Slatkin, 1993 ; Wright, 1943). Les modèles prévoient également que l'échelle spatiale à laquelle ce patron d'IBD (*Isolation By Distance*)

a atteint son état stationnaire s'élargisse suite à l'établissement des populations (Slatkin, 1993). En d'autres termes, avant que l'équilibre complet ne soit atteint, le patron d'IBD est uniquement observé entre les populations les plus proches spatialement mais pas entre les plus éloignées.

Ces modèles théoriques supposent un effet homogène du paysage (Guillot *et al.*, 2009). Or, la plupart des paysages étant hétérogènes, les distances géodésiques sont souvent remplacées par des distances paysagères (des distances-coût ou des distances de résistance par exemple) (Balkenhol *et al.*, 2016 ; Coulon *et al.*, 2004 ; Peterman, 2018). Bien que le patron d'IBLR (*Isolation By Landscape Resistance*) ait fait l'objet de moins d'études théoriques que le patron d'IBD, les travaux de McRae (2006) indiquent qu'il devrait évoluer de la même manière que le patron d'IBD au cours du temps. Par conséquent, on peut s'attendre à ce que de meilleures inférences de la résistance du paysage au flux génétique soient obtenues en considérant un sous-ensemble de paires de populations lorsque l'état d'équilibre n'est pas atteint. Le test de cette hypothèse revêt une certaine importance car les études de génétique du paysage ont fréquemment lieu dans des paysages anthropisés dont les modifications récentes ont pu affecter les flux génétiques entre populations (Manel et Holderegger, 2013 ; Storfer *et al.*, 2010). Néanmoins, l'intérêt de l'élagage des graphes génétiques dans ce contexte a rarement été étudié.

Dans ce chapitre, nous avons adopté une approche basée sur des simulations pour comparer les performances relatives de plusieurs méthodes d'élagage de graphes génétiques, plusieurs distances génétiques et plusieurs méthodes d'analyse de ces graphes par rapport à deux objectifs : (i) identifier la topologie précise du réseau de dispersion, (ii) évaluer correctement la relation entre distances paysagères et distances génétiques. Tout d'abord, nous avons évalué l'efficacité de trois critères d'élagage pour supprimer les liens des graphes : des distances géodésiques seuils, des critères topologiques et l'application du principe d'indépendance conditionnelle à des données génétiques (Dyer et Nason, 2004). Ensuite, nous avons comparé quelques-unes des distances génétiques couramment utilisées pour pondérer les liens de graphes génétiques (Murphy *et al.*, 2016) : le F_{ST} (Keller *et al.*, 2013 ; Munwes *et al.*, 2010), le D_{PS} (Naujokaitis-Lewis *et al.*, 2013 ; Keller *et al.*, 2013) et des distances génétiques euclidiennes (Excoffier *et al.*, 1992). Finalement, nous avons comparé deux pratiques mises en œuvre dans les analyses de distances faites à partir d'un graphe génétique. La première consiste à évaluer la corrélation entre les distances paysagères et génétiques séparant les paires de populations directement connectées sur le graphe génétique. La seconde est basée sur le même type de corrélation mais en considérant l'ensemble des paires de populations, qu'elles soient directement ou indirectement connectées sur le graphe, et en faisant la somme des distances génétiques le long du chemin le plus court entre ces populations. Dyer *et al.* (2010) ont mis en évidence une corrélation plus forte entre de telles distances génétiques calculées à partir d'un graphe génétique d'indépendance (*cGD*) et des distances paysagères qu'entre une matrice complète de F_{ST} et ces mêmes distances paysagères. Néanmoins, ces deux matrices de distances génétiques (*cGD* vs matrice complète de F_{ST}) impliquent à la fois deux distances génétiques différentes (distance génétique euclidienne vs F_{ST}) et deux façons de traiter ces distances avant l'analyse (somme de distances le long des liens conservés vs valeurs brutes des distances associées à tous les liens). Cela introduit un facteur confondant dans la comparaison et l'intérêt de cette approche reste à démontrer.

2 Matériel et méthodes

2.1 Données paysagères

Nous avons tout d’abord simulé 10 paysages de 3600 km² (résolution : 100 m) comportant cinq types d’occupation du sol. Ils avaient une structure similaire à celle de paysages agricoles dominés par les cultures et les prairies (70 % du paysage) dans lesquels subsistent des fragments forestiers (15 %) ¹. Les valeurs de coût attribuées aux différents types d’occupation du sol s’échelonnaient entre 1 (forêt) et 1000 (zones artificielles) et reflétaient la résistance du paysage au déplacement d’une espèce forestière spécialiste (Gurrutxaga *et al.*, 2010 ; Khimoun *et al.*, 2017 ; Pérez-Espona *et al.*, 2008 ; Ruiz-González *et al.*, 2014 ; Schadt *et al.*, 2002 ; Wang *et al.*, 2008). Nous avons ensuite échantillonné aléatoirement 50 populations séparées d’au moins 3 km les unes des autres. Elles correspondaient chacune à une population de 30 individus. Nous avons répété 10 fois cet échantillonnage dans chaque paysage (100 combinaisons paysage/distribution spatiale de populations).

2.2 Simulations de flux génétiques

Nous avons utilisé le logiciel CDPOP (Landguth et Cushman, 2010) pour simuler des flux génétiques entre les populations durant 500 générations non chevauchantes. Les tailles des populations étaient constantes et à chaque génération les individus pouvaient se reproduire dans leur population d’origine ou se disperser pour se reproduire dans une autre population, provoquant ainsi un flux génétique. Chaque couple d’individus était remplacé par sa descendance tout en s’assurant qu’il n’y ait pas d’individus surnuméraires. Chaque individu était caractérisé par son génotype au niveau de 20 *loci* microsatellites comportant initialement 30 allèles.

Les flux génétiques simulés dépendaient de la distance paysagère entre les populations. Le second objectif de ce chapitre consistait à étudier l’intérêt des analyses basées sur les graphes génétiques pour évaluer la relation entre distances paysagères et génétiques. Nous avons donc cherché à simuler des patrons de différenciation génétique contrastés de par l’échelle à laquelle un patron d’IBLR était observé. Ainsi, pour chaque combinaison d’un paysage et d’une distribution spatiale de 50 populations, nous avons construit un graphe paysager dans lequel chaque nœud correspondait à une population. Les distances-coût calculées entre les paires de populations pondéraient les liens de ce graphe. Nous avons calculé la distance-coût correspondant au seuil de percolation de ce graphe et supprimé les liens dont le poids était supérieur à 1.1 fois ce seuil. Ce graphe connexe définissait la topologie du réseau au sein duquel les individus pouvaient se disperser durant les simulations, c’est-à-dire l’ensemble des chemins de dispersion directs. La valeur $1.1 \times \text{seuil.percolation}$ correspondait donc à la distance de dispersion maximale.

Les probabilités de dispersion décroissaient de façon exponentielle à mesure que la distance-coût entre populations augmentait : $p(CD) = e^{-\alpha CD}$. Les valeurs de α étaient fixées de façon à ce que la probabilité de dispersion à une distance égale à 5 % du seuil de percolation soit égale à 0.01. Des tests préliminaires nous avaient montré que ce paramétrage permettait d’obtenir des proportions de migrants similaires à celles calculées par Bowne et Bowers (2004) à partir de données empiriques.

1. La méthodologie détaillée se trouve dans l’annexe A3. Nous avons ici fait le choix de n’en donner que les grandes lignes.

Chaque combinaison d'un paysage et d'une distribution de 50 populations a donné lieu à 10 simulations (1000 au total). Nous avons utilisé les géotypes des individus aux générations 50 et 500 pour construire des graphes génétiques. De plus, pour chaque simulation, nous avons construit un "graphe de dispersion réalisée" dont les liens correspondaient aux liens suivis par au moins un individu au cours de la simulation. Les graphes génétiques construits pour identifier la topologie du réseau de dispersion étaient donc censés reproduire la topologie de ce graphe de dispersion réalisée.

2.3 Graphes génétiques

Nous avons construit les graphes génétiques en utilisant plusieurs méthodes d'élagage et plusieurs distances génétiques pour pondérer les liens (voir Table 1 pour la liste exhaustive des combinaisons).

2.3.1 Méthodes d'élagage

Les graphes génétiques ont été élagués selon trois critères : (i) des distances géodésiques seuils, (ii) des contraintes topologiques, (iii) l'application du principe d'indépendance conditionnelle. Lorsque des distances géodésiques seuils étaient utilisées, nous avons supprimé tous les liens entre les populations séparées par une distance géodésique supérieure à un seuil qui valait tour à tour 10, 15, 20 ou 30 km. Quatre graphes étaient donc créés pour chaque simulation (GEO-10, GEO-15, GEO-20, GEO-30). Deux critères topologiques ont été utilisés pour construire un graphe de Gabriel (GAB) ou un arbre couvrant minimal (MST) en se basant sur les distances géodésiques entre populations pour définir ces topologies (à l'instar d'Arnaud (2003) et Keller *et al.* (2013)). Alors que la topologie d'un graphe de Gabriel s'apparente à celle d'un réseau de dispersion de type *stepping-stones*, celle de l'arbre couvrant minimal correspond à la "colonne vertébrale" du réseau de dispersion selon Bunn *et al.* (2000).

Le troisième critère d'élagage reposait sur le principe d'indépendance conditionnelle (Magwene, 2001 ; Whittaker, 2009). Son application consiste à inférer statistiquement le jeu de liens minimal expliquant le patron de différenciation génétique. Pour cela, nous avons utilisé la méthode originale de Dyer et Nason (2004) (`popgraph`) ainsi qu'une version modifiée de cette méthode (cf. Annexes A3 et B1). L'application du principe d'indépendance conditionnelle implique de calculer une matrice de covariance génétique à partir d'une matrice de distances génétiques aux propriétés euclidiennes (Gower, 1966). Pour cela, nous avons utilisé d'une part une distance génétique euclidienne entre populations issue de l'ACP (Analyse en Composantes Principales) de la table des fréquences alléliques par population, et d'autre part la distance génétique euclidienne calculée par défaut par le package `popgraph`. D'un point de vue strictement mathématique, la formule à utiliser pour calculer la covariance c_{ij} à partir de la distance d_{ij} entre les populations i et j est la suivante : $c_{ij} = -\frac{1}{2} \times (d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2)$ (Everitt et Hothorn, 2011 ; Smouse et Peakall, 1999). Néanmoins, la formule utilisée dans le package `popgraph` est : $c_{ij} = -\frac{1}{2} \times (d_{ij} - d_{i\bullet} - d_{\bullet j} + d_{\bullet\bullet})$ (Dyer et Nason, 2004)². Dans la version modifiée de cette méthode (CI), nous avons utilisé la première formule mais nous avons aussi utilisé la seconde par souci de comparaison (CI2). Nous avons également ajusté les p -valeurs pour limiter les erreurs de type I (Holm, 1979). Ainsi, nous avons construit des graphes génétiques d'indépendance en utilisant soit une version modifiée de la méthode (CI), soit la version originale de Dyer et Nason (2004) (CI2), avec une distance génétique euclidienne issue d'une ACP (PCA) ou la distance génétique euclidienne calculée par le package `popgraph` (PG), et en ajustant (ADJ) les p -valeurs ou non (Table 1). Enfin,

2. $d_{i\bullet}$ et $d_{\bullet j}$ correspondent respectivement à la somme des distances d'une colonne/ligne de la matrice de distance.

nous avons construit des graphes génétiques complets pour avoir une référence permettant d'évaluer l'intérêt de l'élagage.

2.3.2 Distances génétiques

Nous avons calculé quatre distances génétiques pour pondérer les liens des graphes :

- Le F_{ST} linéarisé (égal à $F_{ST}/(1-F_{ST})$) (Rousset, 1997), simplement noté F_{ST} par la suite.
- Le D_{PS} (Bowcock *et al.*, 1994), supposé refléter les flux génétiques contemporains (Murphy *et al.*, 2010b, 2016 ; Naujokaitis-Lewis *et al.*, 2013).
- Une distance génétique euclidienne issue de l'ACP de la table des fréquences alléliques par population (PCA) (Paschou *et al.*, 2014 ; Shirk *et al.*, 2017a).
- La distance génétique euclidienne calculée par défaut par le package `popgraph` (PG).

Les liens des graphes génétiques d'indépendance n'étaient pondérés qu'avec les deux types de distances génétiques euclidiennes et les autres avec le F_{ST} , le D_{PS} et la distance génétique euclidienne issue de l'ACP. Les graphes complets étaient pondérés avec chacune des distances génétiques. Ces combinaisons ont donné lieu à la création de 30 graphes génétiques aux générations 50 et 500 de chaque simulation (Table 1).

2.4 Analyses des graphes

La topologie du graphe de dispersion réalisée reflétait le patron de dispersion simulé et le flux génétique simulé dépendait des valeurs de distances-coût entre populations. Par conséquent, un graphe génétique peut être considéré pertinent pour l'analyse si (i) sa topologie est similaire à celle du graphe de dispersion réalisée ou si (ii) les distances génétiques issues de ses liens sont fortement corrélées aux distances-coût entre populations, selon le type d'analyse.

2.4.1 Analyses de similarité des topologies

Nous avons évalué la similarité des topologies des graphes de dispersion réalisée et des graphes génétiques à l'aide du coefficient de corrélation de Matthews (Matthews, 1975). L'intérêt de ce coefficient est qu'il se base sur toutes les valeurs de la matrice de confusion indiquant à quel point un lien présent/absent d'un graphe est également présent/absent de l'autre graphe (Baldi *et al.*, 2000). Il vaut 0 lorsque cette matrice est équivalente à ce qu'elle serait si les liens étaient choisis aléatoirement, 1 lorsque les topologies sont identiques et -1 lorsqu'elles sont totalement dissimilaires. Dans notre cas, une forte valeur indiquait donc que le graphe génétique identifiait correctement la topologie du réseau de dispersion.

2.4.2 Analyses basées sur les distances

Nous avons calculé le coefficient de corrélation r de Mantel entre distances génétiques et distances-coût (Mantel, 1967). Pour chaque jeu de données génétiques, nous avons considéré trois ensembles de distances génétiques :

1. le sous-ensemble des distances génétiques entre les paires de populations connectées sur le graphe génétique,

2. les distances génétiques calculées sur le graphe pour l'ensemble des paires de populations, connectées ou non, étendant ainsi le calcul des distances de type *cGD* introduites par [Dyer et al. \(2010\)](#) à d'autres types d'élagage de graphes,
3. l'ensemble des distances génétiques entre toutes les paires de populations connectées sur le graphe génétique complet.

Une forte valeur de r indiquait que les distances génétiques considérées permettaient d'évaluer correctement l'effet simulé du paysage sur le flux génétique, compte tenu de la forte relation attendue entre distances-coût et distances génétiques suite aux simulations.

2.4.3 Analyse factorielle des résultats des simulations

La variabilité des paysages et de la distribution spatiale des populations a produit des patrons de différenciation génétique contrastés. Nous avons cherché à caractériser ces contrastes en les situant par rapport à la typologie des patrons d'IBD définie par [Hutchison et Templeton \(1999\)](#), et en particulier vis-à-vis des types I et IV de cette typologie. Alors que le patron d'IBD de type I correspond aux situations dans lesquelles la distance génétique augmente de façon linéaire et continue à mesure que la distance géodésique augmente entre les populations, le patron d'IBD de type IV est caractérisé par l'atteinte d'un plateau dans cette relation à partir d'une certaine distance géodésique. Dans ce dernier cas, le patron de différenciation génétique n'a atteint un état correspondant à l'équilibre entre dérive et flux génétique qu'entre les populations les plus proches spatialement. Cette distance est définie par [Van Strien et al. \(2015\)](#) comme la distance maximale de corrélation (DMC), à savoir la distance géodésique seuil en deçà de laquelle la corrélation entre distances géodésiques et distances génétiques est maximisée. La DMC est donc grande lorsqu'un patron d'IBD de type I est observé et plus courte pour un patron d'IBD de type IV. Nous l'avons calculée pour estimer l'échelle spatiale au-delà de laquelle l'équilibre entre dérive et flux génétique n'est pas atteint. Nous faisons donc le postulat que la typologie d'[Hutchison et Templeton \(1999\)](#) pouvait s'appliquer aux patrons d'IBLR.

Nous avons déterminé la DMC en calculant le coefficient de corrélation de Mantel entre (i) les distance-coût déterminant la dispersion et (ii) le F_{ST} ou le D_{PS} en utilisant itérativement des distances-coût seuils pour sélectionner les paires de populations considérées. Nous avons aussi visualisé des nuages de points pour vérifier qu'ils correspondaient aux différents cas de la typologie d'[Hutchison et Templeton \(1999\)](#). Ces analyses ainsi que les constructions et analyses des graphes ont été réalisées à l'aide du package R `graph4lg` ([Savary et al., 2021b](#)).

Enfin, nous avons réalisé une ACP à partir de huit variables décrivant les résultats des simulations (proportion de migrants par population, distance de dispersion maximale, nombre de liens dans le graphe de dispersion réalisée, distance moyenne parcourue par un individu se dispersant, DMC calculées aux générations 50 et 500 avec le F_{ST} et le D_{PS}). Une analyse de *clustering* hiérarchique appliquée aux axes factoriels de l'ACP nous a permis d'identifier les principales tendances pour nous focaliser ensuite sur quelques cas contrastés le long d'un gradient opposant des situations proches des types I et IV de la typologie d'[Hutchison et Templeton \(1999\)](#).

Nom du graphe	Méthode d'élagage	Distance génétique
COMP-FST	Pas d'élag.	F_{ST}
COMP-DPS	Pas d'élag.	D_{PS}
COMP-PCA	Pas d'élag.	Dist. gén. euc. ACP
COMP-PG	Pas d'élag.	Dist. gén. euc. popgraph
GEO-10-FST	Dist. géo. seuil (10-km)	F_{ST}
GEO-10-DPS	Dist. géo. seuil (10-km)	D_{PS}
GEO-10-PCA	Dist. géo. seuil (10-km)	Dist. gén. euc. ACP
GEO-15-FST	Dist. géo. seuil (15-km)	F_{ST}
GEO-15-DPS	Dist. géo. seuil (15-km)	D_{PS}
GEO-15-PCA	Dist. géo. seuil (15-km)	Dist. gén. euc. ACP
GEO-20-FST	Dist. géo. seuil (20-km)	F_{ST}
GEO-20-DPS	Dist. géo. seuil (20-km)	D_{PS}
GEO-20-PCA	Dist. géo. seuil (20-km)	Dist. gén. euc. ACP
GEO-30-FST	Dist. géo. seuil (30-km)	F_{ST}
GEO-30-DPS	Dist. géo. seuil (30-km)	D_{PS}
GEO-30-PCA	Dist. géo. seuil (30-km)	Dist. gén. euc. ACP
GAB-FST	Topologique (Graphe de Gabriel)	F_{ST}
GAB-DPS	Topologique (Graphe de Gabriel)	D_{PS}
GAB-PCA	Topologique (Graphe de Gabriel)	Dist. gén. euc. ACP
MST-FST	Topologique (MST)	F_{ST}
MST-DPS	Topologique (MST)	D_{PS}
MST-PCA	Topologique (MST)	Dist. gén. euc. ACP
CI-PCA	Indép. condit.	Dist. gén. euc. ACP (covar. depuis dist. au carré)
CI-ADJ-PCA	Indép. condit.	Dist. gén. euc. ACP (covar. depuis dist. au carré) avec ajust. d'Holm-Bonferroni
CI-PG	Indép. condit.	Dist. gén. euc. popgraph (covar. depuis dist. au carré)
CI-ADJ-PG	Indép. condit.	Dist. gén. euc. popgraph (covar. depuis dist. au carré) avec ajust. d'Holm-Bonferroni
CI2-PCA	Indép. condit.	Dist. gén. euc. ACP (covar. depuis dist.)
CI2-ADJ-PCA	Indép. condit.	Dist. gén. euc. ACP (covar. depuis dist.) avec ajust. d'Holm-Bonferroni
CI2-PG	Indép. condit.	Méthode popgraph originale
CI2-ADJ-PG	Indép. condit.	Méthode popgraph originale avec ajust. d'Holm-Bonferroni

TABLE 1 – Paramètres de construction des graphes génétiques. COMP : graphe complet, Dist. gén. euc. : distance génétique euclidienne, Dist. gén. euc. ACP : distance génétique euclidienne calculée à partir des axes factoriels de l'ACP des fréquences alléliques par population, Dist. gén. euc. **popgraph** : distance génétique euclidienne calculée par le package **popgraph**, Dist. géo. seuil : distance géodésique seuil utilisée pour élaguer le graphe, MST (*Minimum Spanning Tree*) : arbre couvrant minimal, CI : *Conditional Independence*, Indép. condit. : principe d'indépendance conditionnelle, covar. depuis dist. (au carré) : covariance entre populations calculée à partir des distances génétiques (au carré) pour appliquer le principe d'indépendance conditionnelle, ajust. d'Holm-Bonferroni : ajustement des p -valeurs à l'aide de la méthode d'Holm (1979).

3 Résultats

3.1 Résultats des simulations

Lors de toutes les simulations, chaque population a échangé des individus avec au moins une autre population et les graphes de dispersion réalisée étaient tous connexes. La proportion d'individus se dispersant durant les simulations variait entre 13.3 % et 24.1 %. Bien que les paysages aient été simulés selon les mêmes paramètres et les populations localisées aléatoirement, les distances maximales de dispersion variaient de façon importante (de 1321 à 3564 unités de coût). De même, le nombre de liens dans les graphes de dispersion réalisée variait de 155 à 858 (Figure 37).

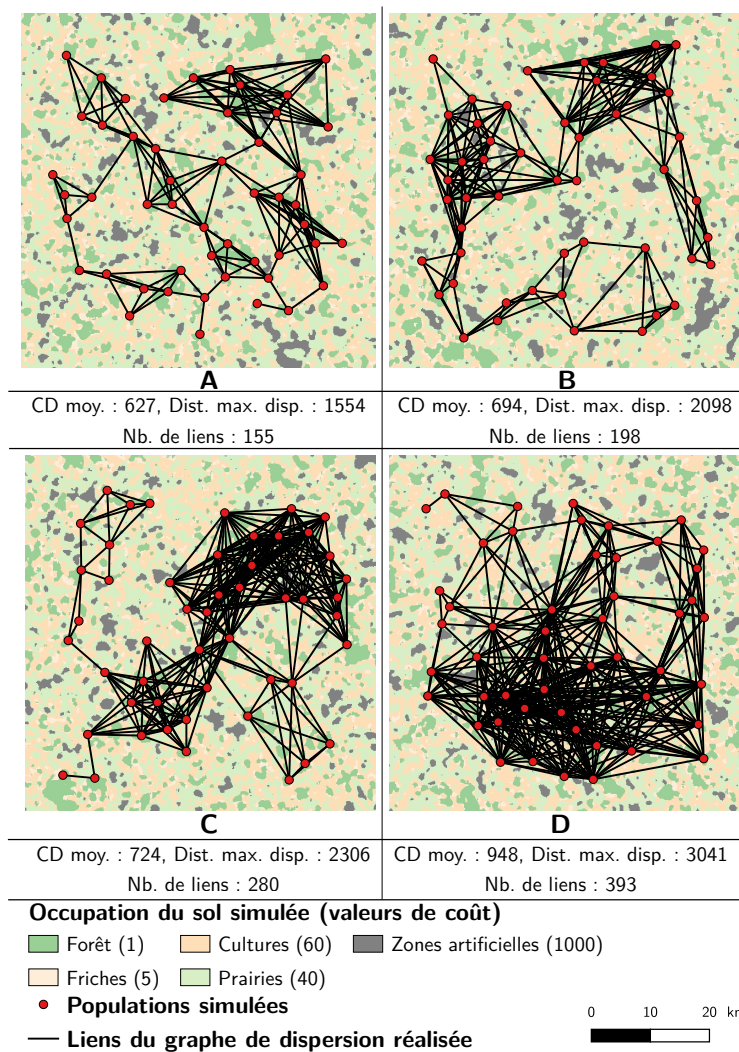


FIGURE 37 – Quatre combinaisons contrastées d'un paysage et d'une distribution spatiale de populations (A, B, C, D) pour lesquelles les nombres de liens des graphes de dispersion réalisée étaient différents. La distance-coût moyenne entre populations, la distance maximale de dispersion en unités de coût et le nombre de liens suivis par les individus lors des simulations sont indiqués pour chacune de ces combinaisons.

Même si un patron d'IBLR de type IV était souvent observé à la génération 50 (Figure 39), les valeurs de DMC augmentaient de la génération 50 à 500, suggérant que le patron de différenciation génétique atteignait son état d'équilibre à une échelle de plus en plus large au cours du temps. Les valeurs de DMC dépassaient souvent la distance maximale de dispersion. L'ACP a mis en évidence les différences entre les simulations faites dans différents paysages et avec différentes distributions spatiales

des populations (Figure 38). En particulier, sa première composante principale (56.6 % de la variance) était corrélée positivement à la DMC, à la distance maximale de dispersion et au nombre de liens dans le graphe de dispersion réalisée. Trois *clusters* de combinaisons d'un paysage et d'une distribution spatiale des populations ont été identifiés (Figure 38). Le premier et le troisième *clusters* s'opposaient sur la première composante principale et incluait des situations correspondant respectivement à des patrons d'IBLR de type IV et I à la génération 50. Dans la suite des analyses, nous nous sommes focalisés sur quatre combinaisons (A, B, C, D, cf. Figures 37, 38 et 39) définissant un gradient entre ces deux patrons le long de la première composante, compte tenu de nos hypothèses concernant l'intérêt potentiel de l'élagage des graphes génétiques dans ces situations. La combinaison A était typique d'un patron d'IBLR de type IV (à la génération 50 en particulier) et la combinaison D d'un patron d'IBLR de type I (Figure 39). Les combinaisons B et C correspondaient à des situations intermédiaires.

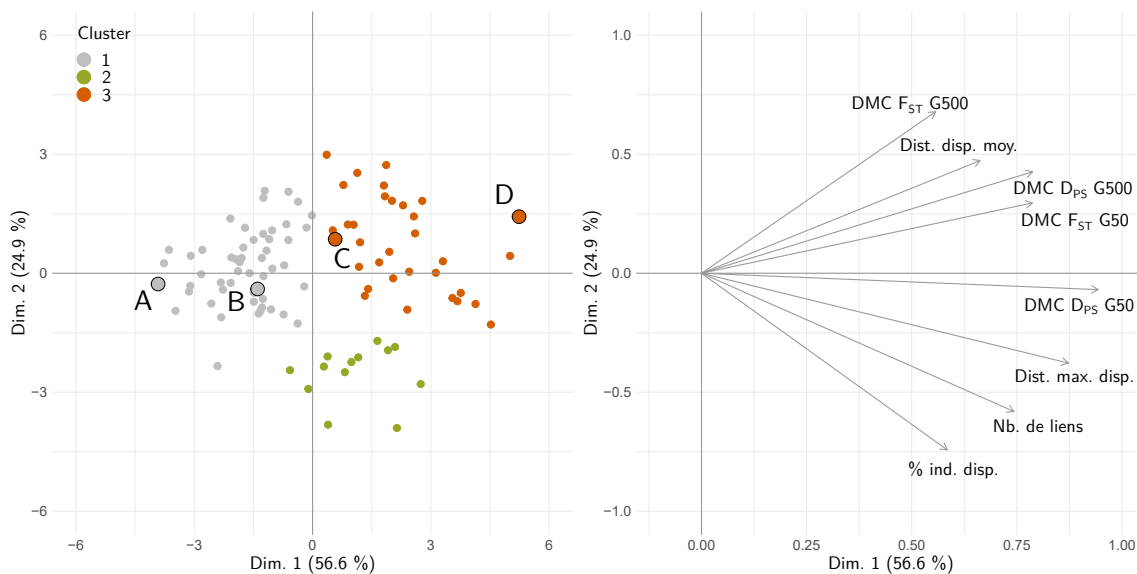


FIGURE 38 – Analyse en Composantes Principales (ACP) des huit variables (valeurs moyennes des 10 simulations pour chaque combinaison, 100 observations) décrivant les résultats des simulations. Les combinaisons A à D sont aussi représentées sur la figure 37.

3.2 Graphes génétiques

3.2.1 Analyses de similarité des topologies

Selon la méthode d'élagage utilisée, le nombre moyen de liens dans les graphes génétiques variait fortement (de 49 (MST) à 802 (GEO-30), Table 2). En revanche, le nombre de liens dans les graphes de dispersion réalisée, dont les graphes génétiques étaient censés reproduire la topologie, était égal à 155, 198, 280 et 393 pour les combinaisons A, B, C et D, respectivement.

Quand les graphes étaient élagués à partir de distances géodésiques seuils ou de contraintes topologiques, leur nombre de liens était stable au cours des générations étant donné que ces critères ne dépendaient pas des données génétiques. Le nombre de liens des arbres couvrants minimaux et des graphes de Gabriel (49 et environ 90, respectivement) était nettement inférieur à celui des graphes de dispersion réalisée (Table 2). Par conséquent, leurs topologies étaient globalement différentes (corrélation de Matthews entre 0.29 et 0.54, Table 2).

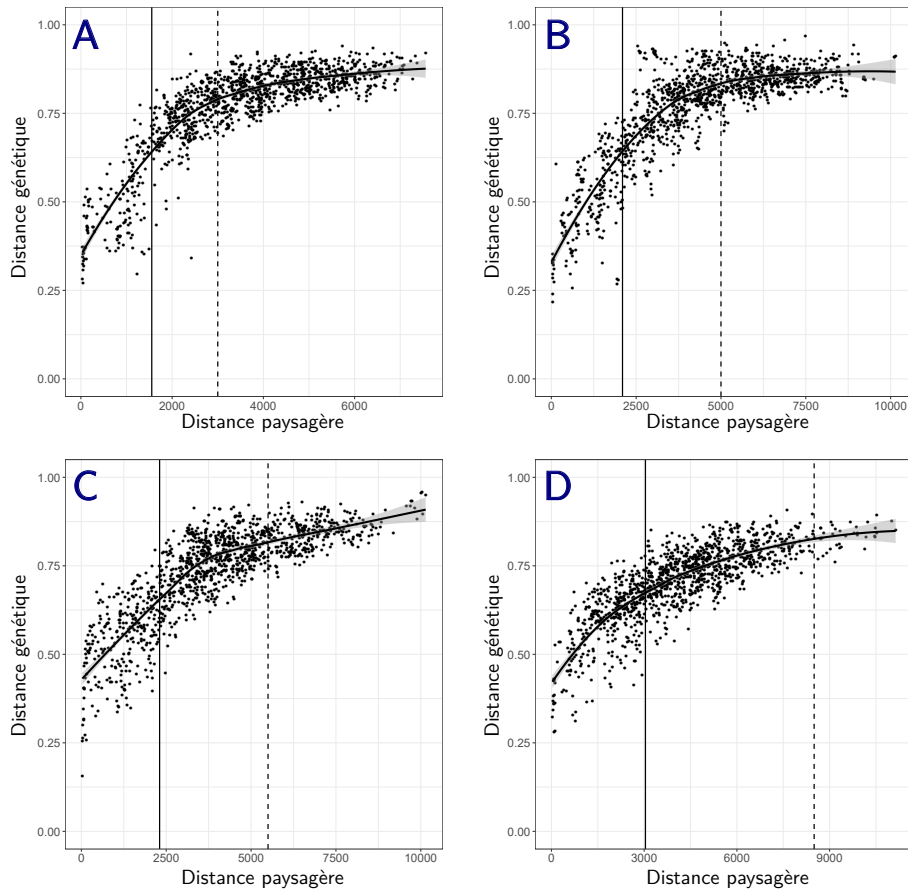


FIGURE 39 – Nuages de points représentant la relation entre les distances génétiques (D_{PS}) et les distances paysagères (distances-coût) à la génération 50. Les cas A à D illustrent le gradient dans les patrons d’IBLR (du type IV au type I). Les lignes continues indiquent la distance de dispersion maximale, les lignes en pointillés la DMC.

Au contraire, vu que l’élagage basé sur le principe d’indépendance conditionnelle repose sur les données génétiques, le nombre de liens des graphes génétiques d’indépendance variait fortement, de 58.5 à 519.7 en moyenne (Table 2). Ce nombre tendait à augmenter de la génération 50 à 500, alors que le nombre total de liens suivis par les individus était stable entre ces générations. Cette tendance était moins marquée en utilisant la méthode originale issue du package `popgraph` (CI2) que la méthode modifiée (CI). De façon générale, avec la méthode originale, le nombre de liens était également plus faible mais parfois les graphes génétiques n’étaient pas connexes. C’était en particulier le cas lorsque les p -valeurs étaient ajustées, ce qui réduisait fortement le nombre de liens.

Les graphes génétiques élagués à l’aide de distances géodésiques seuils présentaient les topologies les plus semblables à celles des graphes de dispersion réalisée (corrélations de Matthews supérieures à 0.6), à l’exception de la combinaison A pour laquelle le graphe de dispersion réalisée comptait peu de liens (Table 2). Plus la distance géodésique (GEO) utilisée pour l’élagage était proche de la conversion de la distance-coût seuil en distance géodésique, et donc d’une estimation de la distance de dispersion maximale en unités métriques (Table 2), plus le graphe génétique reflétait la topologie du graphe de dispersion réalisée. En revanche, la topologie des graphes de dispersion réalisée correspondant à la combinaison A était mieux reproduite par des graphes génétiques d’indépendance. Dans ce cas, des corrélations de Matthews supérieures à 0.6 étaient obtenues à chaque fois que la covariance était calculée à partir de distances génétiques (CI2) et seulement à la génération 50 avec un ajustement des p -valeurs quand elle était calculée à partir de distances génétiques au carré (CI). Pour les combinaisons

B et C, la méthode originale (CI2) donnait aussi des résultats satisfaisants. Notons que les graphes génétiques construits avec la méthode modifiée (CI) incluait parfois des liens entre des populations qui n'étaient pas connectées sur le graphe de dispersion réalisée. L'ajustement des p -valeurs tendait à limiter le nombre de ces "faux" liens.

Globalement, les graphes génétiques qui reproduisaient correctement la topologie des graphes de dispersion réalisée avaient un nombre de liens similaire à ces derniers (Table 2). Toutefois, cela n'était pas suffisant étant donné que des corrélations de Matthews relativement faibles étaient parfois obtenues alors que ce critère était vérifié (Table 2).

3.2.2 Analyses basées sur les distances

Les coefficients de corrélation de Mantel entre les distances génétiques et les distances-coût séparant les paires de populations directement connectées sur les graphes génétiques variaient de 0.47 à 0.86 en moyenne à la génération 50 (Figure 40). Dans tous les cas, les corrélations entre distances génétiques et distances géodésiques étaient plus faibles qu'entre distances génétiques et distances-coût, conformément au résultat attendu compte tenu des paramètres de simulation.

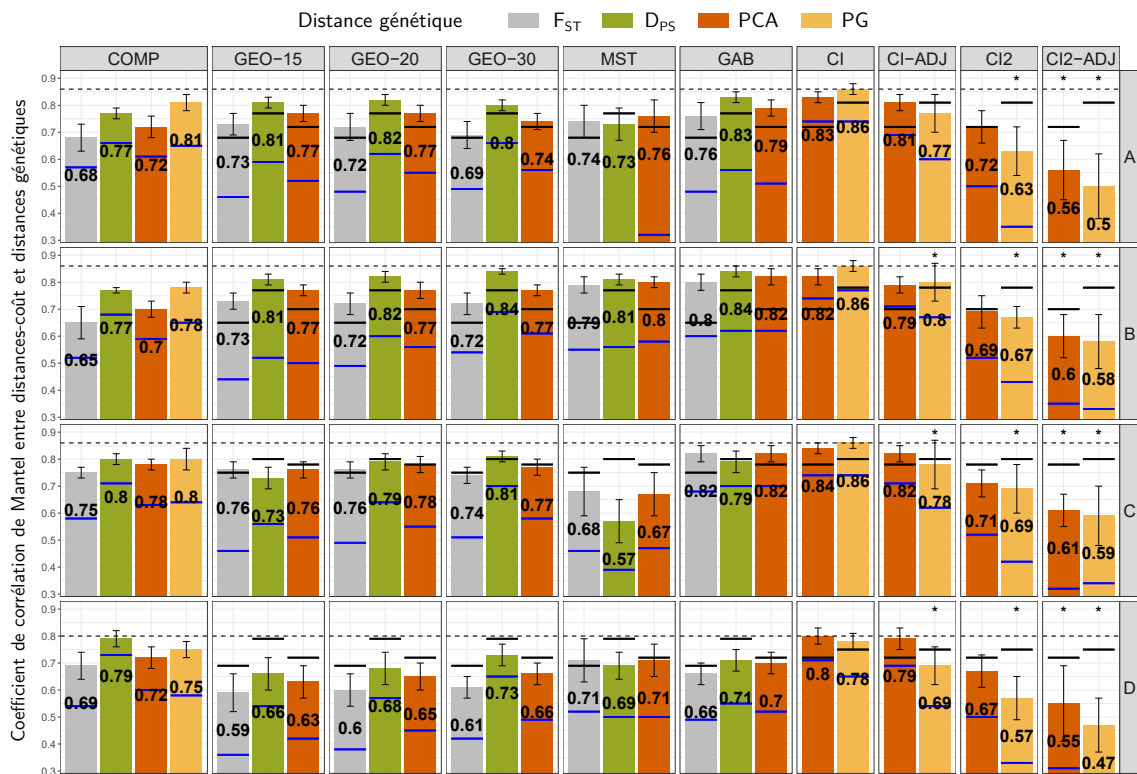


FIGURE 40 – Corrélations de Mantel entre les distances-coût et les distances génétiques séparant les populations directement connectées sur les graphes génétiques, selon le type de distance génétique et la méthode d'élagage, à la génération 50 (cf. Table 1 pour le nom des graphes). Les valeurs moyennes \pm écarts-types sont calculées pour les 10 simulations réalisées pour chaque combinaison. Les barres bleues indiquent le coefficient de corrélation entre distances génétiques et distances géodésiques, s'il est supérieur à 0.3. Les barres noires indiquent le coefficient de corrélation obtenu en considérant toutes les paires de populations pour calculer la corrélation. Quand les barres bleues et noires coïncident, la barre est noire. Les étoiles indiquent les graphes qui ne sont pas connexes. La ligne en pointillés indique la valeur de r maximale obtenue pour chaque combinaison.

Quand un patron d'IBLR de type I était observé à la génération 50 (combinaisons C et D), les coefficients de corrélation les plus élevés étaient toujours ceux obtenus en considérant les distances gé-

nétiques issues des graphes génétiques complets plutôt qu'élagués, à l'exception des graphes génétiques d'indépendance construits avec la méthode modifiée (CI). Au contraire, quand un patron d'IBLR de type IV était observé à la génération 50 (combinaisons A et B), les coefficients de corrélation étaient plus élevés en considérant les distances génétiques associées aux liens des graphes élagués, quelle que soit la méthode d'élagage (Figure 40). À la génération 500, les différences entre les combinaisons étaient moindres car la relation entre distances génétiques et paysagères avait tendance à se linéariser au cours du temps.

Les plus forts coefficients de corrélation étaient toujours obtenus en sélectionnant les valeurs de distance génétique à partir d'un graphe génétique d'indépendance construit en calculant la covariance à partir de distances au carré sans ajustement des p -valeurs (CI). La méthode `popgraph` originale ne renforçait jamais la corrélation obtenue en considérant l'intégralité des valeurs de la matrice de distance génétique équivalente (COMP-PG ou COMP-PCA) (Figure 40). Pour les combinaisons A et B, les coefficients de corrélation étaient légèrement plus élevés lorsque les distances génétiques étaient issues d'un graphe de Gabriel plutôt que d'un arbre couvrant minimal ou d'un graphe élagué selon un critère de distance (Figure 40). Dans la plupart des cas, les valeurs de F_{ST} étaient moins fortement corrélées aux distances-coût que les valeurs de D_{PS} ou de distance génétique euclidienne.

Les coefficients de corrélation entre les distances-coût et les distances génétiques calculées sur les graphes génétiques étaient compris entre 0.57 et 0.93 à la génération 50 et les différences entre combinaisons étaient moins marquées dans ce cas (Figure 41). Ces coefficients prenaient leurs valeurs maximales quand les graphes génétiques étaient élagués selon une distance géodésique seuil ou une contrainte topologique. Néanmoins, lorsque les distances génétiques étaient calculées de cette façon le long des liens des graphes, les coefficients de corrélation entre distances génétiques et distances géodésiques étaient supérieurs à ceux calculés entre distances génétiques et distances-coût. Ainsi, l'hypothèse d'isolement par la distance géodésique semblait plus probable dans ce cas que celle d'isolement par la résistance du paysage, en inadéquation avec le paramétrage des simulations qui constituait ici la "réalité écologique".

Au contraire, lorsque ce type de distance génétique était calculé à partir de graphes génétiques d'indépendance, la corrélation entre distances génétiques et distances-coût était la plus forte (Figure 41). Dans ce cas, les corrélations étaient plus fortes en utilisant la méthode `popgraph` originale (CI2) que sa version modifiée (CI). Néanmoins, la corrélation entre les distances génétiques euclidiennes pondérant les liens de ces graphes génétiques et les distances-coût était plus forte en considérant les valeurs brutes associées aux liens du graphe génétique complet plutôt qu'en calculant les distances génétiques le long des liens de ces graphes génétiques d'indépendance. De plus, nous avons également reproduit le résultat de [Dyer *et al.* \(2010\)](#) qui avaient montré que l'influence du paysage était mieux estimée en utilisant une distance génétique conditionnelle (CI2) qu'en utilisant une matrice complète de F_{ST} .

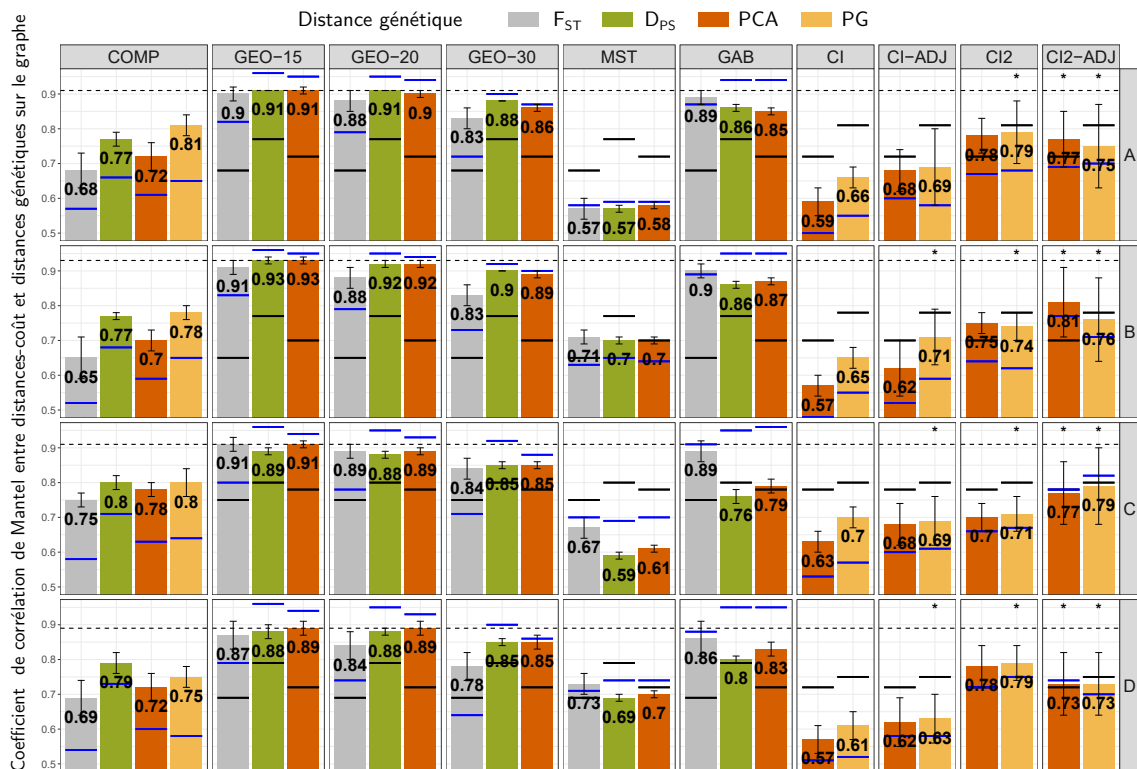


FIGURE 41 – Corrélations de Mantel entre les distances-coût et les distances génétiques calculées sur le graphe génétique le long des chemins les plus courts entre populations, selon le type de distance génétique et la méthode d'élagage, à la génération 50 (cf. Table 1 pour le nom des graphes). Les valeurs moyennes \pm écarts-types sont calculées pour les 10 simulations réalisées pour chaque combinaison. Les barres bleues indiquent le coefficient de corrélation entre distances génétiques et distances géodésiques, s'il est supérieur à 0.5. Les barres noires indiquent le coefficient de corrélation obtenu en considérant toutes les paires de populations pour calculer la corrélation. Quand les barres bleues et noires coïncident, la barre est noire. Les étoiles indiquent les graphes qui ne sont pas connexes. La ligne en pointillés indique la valeur de r maximale obtenue pour chaque combinaison.

Combinaison	A		B		C		D	
a) Graphes de dispersion réalisée								
Dist. max. disp.	1554		2098		2306		3041	
Dist.max.disp.(km)	13.1		15.1		18.2		20.8	
Nb. de liens suivis	155		198		280		393	
b) Nombre de liens								
Génération	G50	G500	G50	G500	G50	G500	G50	G500
GEO-10	127.0	<i>127.0</i>	120.0	<i>120.0</i>	120.0*	<i>120.0*</i>	112.0*	<i>112.0*</i>
GEO-15	274.0	<i>274.0</i>	237.0	<i>237.0</i>	253.0	<i>253.0</i>	246.0	<i>246.0</i>
GEO-20	455.0	<i>455.0</i>	384.0	<i>384.0</i>	408.0	<i>408.0</i>	413.0	<i>413.0</i>
GEO-30	802.0	<i>802.0</i>	693.0	<i>693.0</i>	746.0	<i>746.0</i>	757.0	<i>757.0</i>
GAB	92.0	<i>92.0</i>	89.0	<i>89.0</i>	96.0	<i>96.0</i>	92.0	<i>92.0</i>
MST	49.0	<i>49.0</i>	49.0	<i>49.0</i>	49.0	<i>49.0</i>	49.0	<i>49.0</i>
CI-PCA	274.4	<i>427.7</i>	269.7	<i>412.1</i>	283.8	<i>435.0</i>	305.7	<i>458.8</i>
CI-ADJ-PCA	108.2	<i>246.5</i>	107.7	<i>234.6</i>	109.0	<i>251.5</i>	114.8	<i>275.8</i>
CI-PG	325.5	<i>441.3</i>	330.9	<i>397.9</i>	342.8	<i>447.2</i>	375.3	<i>519.7</i>
CI-ADJ-PG	96.8	<i>152.6</i>	101.2*	<i>152.5</i>	95.7*	<i>149.6*</i>	102.2*	<i>165.3</i>
CI2-PCA	132.6	<i>158.9</i>	134.6	<i>161.2</i>	141.8	<i>176.3*</i>	146.9	<i>204.9</i>
CI2-ADJ-PCA	76.7*	<i>85.8*</i>	71.3*	<i>81.4*</i>	65.0*	<i>81.5*</i>	60.9*	<i>70.0*</i>
CI2-PG	135.4*	<i>148.1</i>	134.3*	<i>149.2*</i>	137.5*	155.4	140.5*	<i>153.9</i>
CI2-ADJ-PG	77.8*	<i>80.5*</i>	72.5*	<i>75.6*</i>	63.7*	<i>72.9*</i>	59.4*	<i>58.5*</i>
c) Coefficient de corrélation de Matthews								
GEO-10	0.64	<i>0.64</i>	0.68	<i>0.68</i>	0.56*	<i>0.56*</i>	0.44*	<i>0.44*</i>
GEO-15	0.59	<i>0.59</i>	0.73	<i>0.73</i>	0.68	<i>0.68</i>	0.62	<i>0.62</i>
GEO-20	0.47	<i>0.47</i>	0.60	<i>0.60</i>	0.68	<i>0.68</i>	0.69	<i>0.69</i>
GEO-30	0.28	<i>0.28</i>	0.38	<i>0.38</i>	0.44	<i>0.44</i>	0.51	<i>0.51</i>
GAB	0.54	<i>0.54</i>	0.54	<i>0.54</i>	0.41	<i>0.41</i>	0.39	<i>0.39</i>
MST	0.49	<i>0.49</i>	0.43	<i>0.43</i>	0.36	<i>0.36</i>	0.29	<i>0.29</i>
CI-PCA	0.42	<i>0.25</i>	0.39	<i>0.24</i>	0.33	<i>0.19</i>	0.26	<i>0.14</i>
CI-ADJ-PCA	0.60	<i>0.34</i>	0.53	<i>0.33</i>	0.43	<i>0.25</i>	0.34	<i>0.18</i>
CI-PG	0.41	<i>0.34</i>	0.39	<i>0.37</i>	0.35	<i>0.32</i>	0.27	<i>0.21</i>
CI-ADJ-PG	0.64	<i>0.52</i>	0.55*	<i>0.48</i>	0.45*	<i>0.41*</i>	0.35*	<i>0.29</i>
CI2-PCA	0.77	<i>0.69</i>	0.69	<i>0.65</i>	0.58	<i>0.55*</i>	0.46	<i>0.40</i>
CI2-ADJ-PCA	0.65*	<i>0.65*</i>	0.55*	<i>0.57*</i>	0.42*	<i>0.46*</i>	0.33*	<i>0.33*</i>
CI2-PG	0.80*	<i>0.75</i>	0.72*	<i>0.69*</i>	0.59*	0.60	0.48*	<i>0.45</i>
CI2-ADJ-PG	0.66*	<i>0.65*</i>	0.56*	<i>0.56*</i>	0.42*	<i>0.44*</i>	0.32*	<i>0.31*</i>

TABLE 2 – Topologies des graphes de dispersion et des graphes génétiques. (a) Topologies des graphes de dispersion. Les distances maximales de dispersion sont données en unités de coût et en kilomètres (conversion faite suite à une régression linéaire des distances-coût en fonction des distances géodésiques). (b) Nombre de liens dans les graphes génétiques. (c) Coefficients de corrélation de Matthews évaluant la similarité des topologies des deux types de graphes (dispersion réalisée et génétique), en fonction du type de distance génétique et de la méthode d'élagage pour les quatre combinaisons d'un paysage et d'une distribution spatiale des populations, à deux générations différentes (cf. Table 1 pour le nom des graphes). Les valeurs moyennes sont calculées pour les 10 simulations réalisées pour chaque combinaison (les écarts-types étaient négligeables). Les coefficients de corrélation de Matthews supérieurs à 0.6 et les nombres de liens correspondants sont notés en gras. Les valeurs relatives à la génération 500 sont notées en italique. Les étoiles indiquent que parmi les dix graphes génétiques considérés, certains n'étaient pas connexes.

4 Discussion

Ce chapitre de la thèse montre que la capacité de différentes méthodes d'élagage à identifier la topologie précise du réseau de dispersion est fortement variable. De plus, nous soulignons ici l'importance de l'élagage des graphes génétiques pour évaluer les effets du paysage sur le flux génétique lorsque le patron de différenciation génétique n'a pas atteint d'équilibre. À partir de ces résultats, que nous discutons dans les sections suivantes, nous avons réalisé la figure 42 qui constitue un guide méthodologique schématique à destination des utilisateurs de graphes génétiques.

4.1 Quand et pourquoi élaguer un graphe génétique ?

Premièrement, l'élagage d'un graphe génétique est difficilement évitable si l'objectif est d'identifier la topologie du réseau de dispersion suivi par les individus. En effet, à l'exception des rares situations de panmixie et/ou lorsque la zone d'étude couvre une superficie très réduite, les épisodes de dispersion n'ont pas lieu entre toutes les paires de populations (Kimura et Weiss, 1964). Deuxièmement, nos résultats montrent que l'intérêt de l'élagage du graphe pour inférer la résistance du paysage au flux génétique dépend de l'échelle à laquelle l'effet du flux génétique sur la différenciation génétique est détectable. Quand un patron d'IBLR est observé à l'échelle de l'ensemble de la zone d'étude, l'élagage du graphe génétique n'améliore pas l'inférence. Au contraire, lorsque ce patron est observé à une échelle restreinte seulement, l'élagage tend à renforcer la corrélation entre distances génétiques et distances-coût, suggérant qu'il est utile pour l'inférence de l'effet du paysage sur le flux génétique.

L'équilibre entre dérive et flux génétique est plus difficilement atteint à l'échelle de toute la zone d'étude lorsque les distances de dispersion de l'espèce considérée sont courtes au regard de la zone d'étude et/ou que le paysage a subi des modifications récentes. De telles situations de non-équilibre correspondent au patron d'IBD de type IV proposé par Hutchison et Templeton (1999). Il a été reproduit dans des études théoriques (Slatkin, 1993) et observé empiriquement (Ciofi *et al.*, 1999 ; Clegg et Phillimore, 2010 ; Hänfling et Weetman, 2006 ; Hutchison et Templeton, 1999 ; Kuehn *et al.*, 2003 ; Méndez *et al.*, 2011). Il est très probablement fréquent dans les études de génétique du paysage menées dans des paysages anthropisés (Manel et Holderegger, 2013 ; Storfer *et al.*, 2010). Ne pas élaguer le graphe génétique dans ces situations peut être problématique si l'objectif est d'inférer la résistance du paysage au flux génétique. En effet, ces inférences intégreraient des distances génétiques qui ne reflètent pas l'effet du paysage sur la différenciation génétique. Nos résultats paraissent alors en accord avec la suggestion de Wagner et Fortin (2013) selon laquelle considérer un sous-ensemble de populations dans les analyses de génétique du paysage basées sur des distances pourrait améliorer leur puissance. Quelques études avaient déjà reporté des relations plus fortes entre différenciation génétique et distances paysagères en faisant une sélection similaire (Angelone *et al.*, 2011 ; Coster *et al.*, 2015 ; Jaquière *et al.*, 2011 ; Keller *et al.*, 2013 ; Van Strien *et al.*, 2015). La plupart considéraient simplement les distances géodésiques entre populations dans cette sélection. Néanmoins, Van Strien (2017) avait déjà signalé que la topologie des populations devrait être davantage prise en compte dans les analyses de génétique du paysage. Dans ce contexte, les méthodes basées sur les graphes paraissent pertinentes et leurs performances relatives n'avaient pas été évaluées jusqu'à présent.

4.2 Comment élaguer un graphe génétique pour identifier la topologie précise du réseau de dispersion ?

Nos résultats montrent que les méthodes d'élagage basées sur des contraintes topologiques sont rarement utiles pour reproduire la topologie du réseau de dispersion. En imposant le nombre de liens à partir du nombre de populations, ce type de critère semble manquer de signification écologique. En revanche, lorsque les capacités de dispersion de l'espèce étudiée sont connues, un élagage basé sur des distances géodésiques seuils permet de reproduire relativement bien la topologie du graphe de dispersion réalisée. Estimer la distance de dispersion d'une espèce est toujours difficile (Schneider, 2003 ; Van Dyck et Baguette, 2005), et lorsqu'on mène une étude empirique, il est impossible de savoir quel graphe génétique reflète au mieux la réalité sans cette information. La similarité entre les graphes génétiques élagués en utilisant des distances seuils et les graphes de dispersion réalisée devrait également dépendre de la force de la relation entre distances-coût et distances géodésiques. Elle est souvent forte (Marrotte et Bowman, 2017) mais ce n'est pas toujours le cas (Balkenhol *et al.*, 2009b), par exemple lorsqu'une barrière paysagère empêche la dispersion entre des populations proches ou lorsqu'au contraire de grandes distances géodésiques sont parcourues par les individus car elles sont associées à un faible coût de déplacement. Ignorer ces rares épisodes de dispersion à longue distance peut être problématique compte tenu de leurs conséquences écologiques et évolutives (Clobert *et al.*, 2012 ; Greenbaum et Fefferman, 2017 ; Nathan *et al.*, 2003).

La construction de graphes génétiques d'indépendance semble être une option convenable pour reproduire la topologie du réseau de dispersion lorsque la distance de dispersion est inconnue, en particulier dans les situations où les connexions sont les plus rares (A et B, Table 2). Dans ces dernières situations, les résultats peuvent être aussi satisfaisants que lorsque la distance de dispersion est connue. La topologie des graphes de dispersion est mieux reproduite quand la covariance est calculée à partir de distances génétiques avec la méthode `popgraph` originale (CI2) qu'à partir de distances génétiques au carré (CI). Dans ce second cas, la présence de liens reliant des populations entre lesquelles les individus ne se dispersent jamais indique que cette méthode n'est pas fiable.

Il peut sembler étonnant d'inférer des épisodes de dispersion directs à partir d'un patron de différenciation génétique formé par plusieurs générations de dispersion. Pourtant, le principe d'indépendance conditionnelle est censé retrouver les chemins directs suivis par les individus (Dyer, 2015b). Bien que nos résultats semblent confirmer cela dans certaines conditions, cette méthode d'élagage nécessiterait davantage de recherches. Il reste par exemple à déterminer pourquoi le nombre de liens restait si stable même quand un nombre important de liens étaient suivis par les individus. On peut également s'attendre à ce que les performances de cette méthode soient limitées lorsque l'échantillonnage est incomplet, ce qui est la norme dans les études empiriques.

4.3 Comment élaguer un graphe génétique pour inférer la résistance du paysage au flux génétique ?

Les graphes génétiques reflétant précisément la topologie du réseau de dispersion ne sont pas nécessairement ceux qui permettent d'évaluer correctement les effets du paysage sur le flux génétique. En effet, la DMC était toujours plus forte que la distance maximale de dispersion (Figure 39), suggérant que les distances génétiques à inclure dans les analyses ne devraient pas être restreintes à celles

qui séparent les populations échangeant directement des individus. L'équilibre entre dérive et flux génétique peut s'établir entre des populations séparées par des distances supérieures aux capacités de dispersion des espèces (Slatkin, 1993), car elles peuvent échanger des allèles sur plusieurs générations sans être directement connectées. Nous pensons qu'il faut inclure ces paires de populations dans les inférences car leur différenciation génétique reflète l'influence du paysage sur le flux génétique. Cette recommandation est contraire à celle qui préconise d'inclure dans les analyses uniquement les paires de populations séparées par une distance inférieure à la distance de dispersion maximale de l'espèce étudiée (Keller *et al.*, 2013 ; Van Strien *et al.*, 2015 ; Van Strien, 2017). Selon nous, une méthode d'élagage permettant d'estimer correctement la résistance du paysage au flux génétique devrait identifier les paires de populations dont la différenciation génétique reflète les flux génétiques à long terme.

Dans ce contexte, nous déconseillons d'utiliser des méthodes d'élagage basées sur des critères fixes (par exemple des distances géodésiques ou des contraintes topologiques), y compris celles qui nous ont permis d'obtenir des corrélations relativement fortes entre distances génétiques et distances-coût. En effet, l'échelle spatiale à laquelle s'établit le patron d'IBLR évolue au cours du temps (McRae, 2006). Les méthodes d'élagage basées sur des données génétiques et une inférence statistique devraient fournir la meilleure inférence de la résistance du paysage car elles tiennent compte de la dynamique du patron d'IBLR. Pour des patrons d'IBLR de type I et IV, la corrélation entre les distances génétiques associées aux liens des graphes génétiques et les distances-coût était maximale lorsque les graphes étaient élagués selon le principe d'indépendance conditionnelle. Mais, ce résultat ne s'observait qu'en calculant la covariance à partir de distances génétiques au carré, en accord avec les démonstrations mathématiques (Everitt et Hothorn, 2011 ; Magwene, 2001 ; Smouse et Peakall, 1999). Alors que la méthode `popgraph` originale reproduisait quant à elle relativement bien le réseau de dispersion, elle produisait les plus faibles corrélations entre distances génétiques et distances-coût. Ces méthodes mériteraient davantage de recherches car certaines paires de populations connectées dans les graphes génétiques d'indépendance (en utilisant la méthode modifiée) étaient séparées par des distances-coût supérieures à la DMC. Même si l'utilisation de la DMC pour déterminer l'échelle spatiale à laquelle la différenciation génétique dépend à la fois du flux génétique et de la dérive mériterait un fondement théorique plus solide, cela suggère que la différenciation génétique entre ces populations n'était pas encore à l'équilibre.

Nous pensons que l'identification des paires de populations vérifiant l'équilibre entre la dérive et le flux génétique est un point essentiel. Ciofi *et al.* (1999) ont développé une approche pour évaluer si le patron de différenciation génétique est mieux expliqué par un modèle d'équilibre entre dérive et flux génétique ou par un modèle basé uniquement sur la dérive. Néanmoins, il semble que cette approche ne détecte pas les patrons d'IBD de type IV convenablement (Hänfling et Weetman, 2006). En supposant que la DMC est une bonne approximation de l'échelle à laquelle l'équilibre migration-dérive s'établit, une approche prometteuse consisterait à élaguer les graphes génétiques avec une distance-coût égale à la DMC. Cette approche nécessiterait une connaissance des valeurs de coût associées aux éléments du paysage pour estimer la distance-coût entre les paires de populations. Cependant, déterminer de telles valeurs de coût est souvent l'objectif des analyses empiriques basées sur les liens (Balkenhol *et al.*, 2016). Une solution alternative consisterait alors à utiliser des graphes génétiques élagués avec des distances-coût en testant plusieurs scénarios de coût de façon itérative via l'utilisation de méthodes d'optimisation des valeurs de coût (Peterman, 2018).

4.4 Quelle distance génétique utiliser et comment ?

Pondérer les liens des graphes génétiques avec le D_{PS} ou des distances génétiques euclidiennes permettrait systématiquement d'obtenir des corrélations avec les distances-coût plus fortes qu'en les pondérant avec le F_{ST} . Bien que le F_{ST} soit une excellente mesure de différenciation génétique, son utilisation pour inférer des paramètres démographiques nécessite que l'équilibre migration-dérive soit atteint (Neigel, 2002 ; Whitlock et McCauley, 1999). Par ailleurs, il a déjà été montré que le D_{PS} reflétait mieux que d'autres distances génétiques l'influence sur la structure génétique de changements paysagers récents (Landguth *et al.*, 2010 ; Robin *et al.*, 2015).

Une fois le graphe génétique élagué, nous déconseillons d'ajouter des distances génétiques le long des chemins les plus courts pour créer une matrice complète de distances génétiques sur le graphe. Cette approche conduisait au résultat erroné selon lequel un phénomène d'isolement par la distance géodésique était à l'origine du patron de différenciation génétique simulé.

4.5 Limites et perspectives

Les simulations produisaient des patrons de connectivité contrastés, mais nous sommes conscients qu'ils ne représentaient que des situations dans lesquelles les populations forment une seule unité fonctionnelle au sein de laquelle toutes les populations peuvent échanger des individus. Les performances des méthodes ici comparées dans des situations où des barrières paysagères isolent complètement des groupes de populations restent à évaluer. De plus, les différences mises en évidence entre les méthodes étaient informatives mais parfois subtiles. Même si les taux de migration ici obtenus étaient similaires à ceux reportés par Bowne et Bowers (2004), ils étaient plus importants que ceux reportés dans d'autres études empiriques (Meirmans, 2014) ou ceux utilisés pour simuler des patrons d'IBD de type IV (Van Strien *et al.*, 2015). Étant donné que les patrons d'IBLR de type IV étaient observés dans les situations où la dispersion était la plus contrainte, reproduire nos simulations en restreignant davantage la dispersion aurait produit des contrastes encore plus marqués dans nos résultats. Cela aurait probablement accentué les différences de performance des graphes complets et élagués pour l'inférence de l'effet du paysage sur le flux génétique et potentiellement aidé à distinguer les méthodes d'élagage les unes des autres. En supposant que les faibles taux de migration sont la norme, cela renforce l'intérêt de l'élagage des graphes génétiques dans les études empiriques.

Nos résultats remettent en question la pratique courante consistant à utiliser la matrice complète de distances génétiques pour inférer la résistance des éléments du paysage. Il faudrait donc étudier à quel point l'élagage des graphes génétiques affecte la calibration des surfaces de résistance. Nous n'avons pas comparé différents scénarios de coût car nous connaissions les "vraies" valeurs de coût déterminant les simulations. Nous avons donc considéré que maximiser la corrélation (r de Mantel) entre les distances génétiques et paysagères entre un sous-ensemble de paires de populations permettrait d'identifier la meilleure méthode de construction d'un graphe, c'est-à-dire celle qui sélectionne le meilleur sous-ensemble de paires de populations pour l'analyse.

Dans nos simulations, nous avons considéré que toutes les populations de la zone d'étude étaient échantillonnées. Une telle intensité d'échantillonnage est rarement possible en pratique, bien qu'elle soit recommandée (Keller *et al.*, 2013 ; Van Strien, 2017). Évaluer à quel point un échantillonnage

partiel affecterait nos conclusions serait nécessaire (voir [Koen *et al.* \(2013\)](#) et [Naujokaitis-Lewis *et al.* \(2013\)](#) pour des approches similaires). Enfin, le flux génétique et la dérive étaient les deux processus affectant la différenciation génétique dans nos simulations. L'intensité de la dérive était homogène et constante au sein des populations. Cette simplification de la réalité néglige le fait que la structure du paysage affecte également les tailles des populations. Il serait donc intéressant d'évaluer à quel point l'hétérogénéité des tailles de populations affecte l'inférence de la résistance du paysage et si l'intégration de variables locales dans les analyses via l'utilisation de modèles gravitaires pourrait améliorer cette inférence.

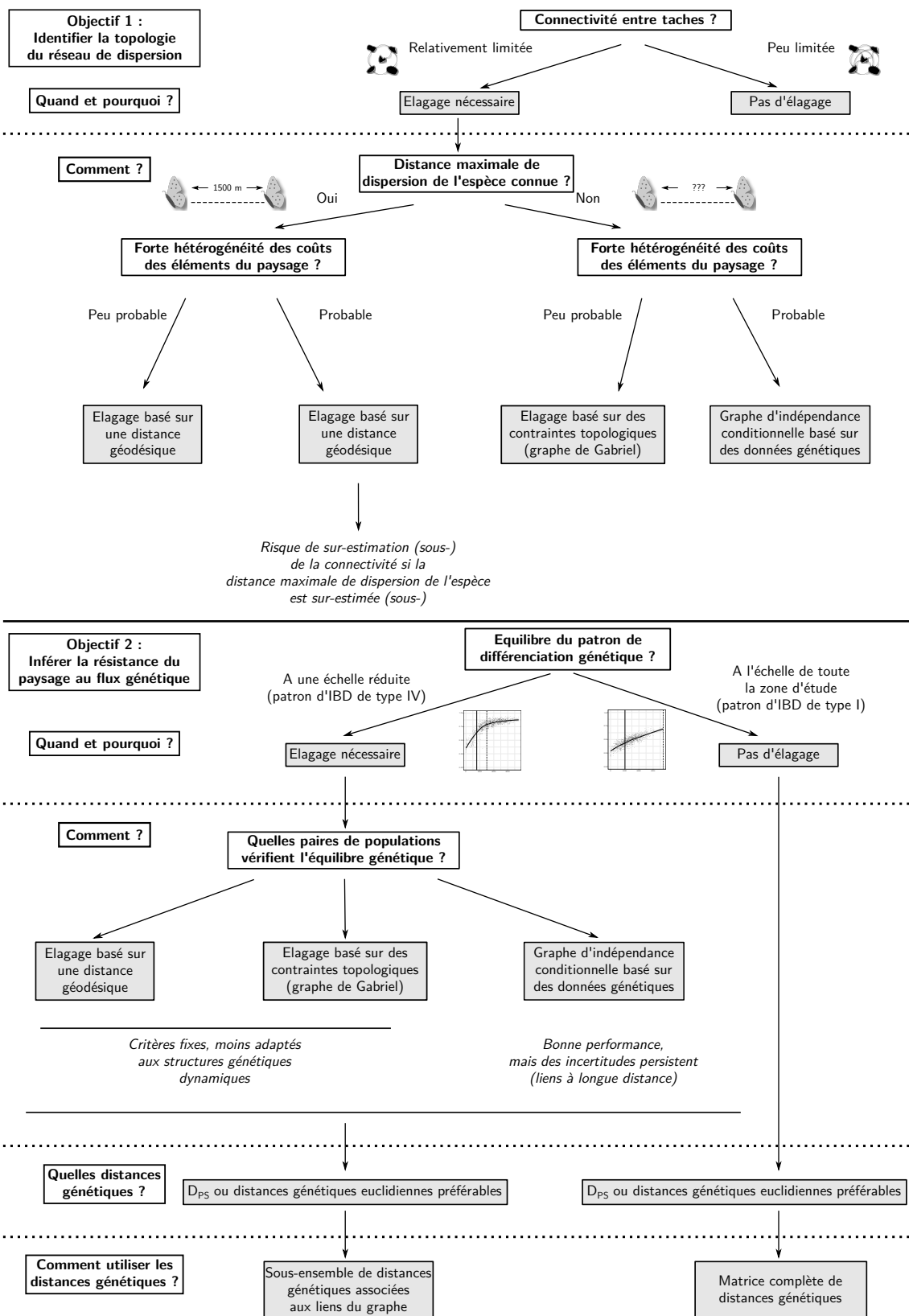


FIGURE 42 – Guide méthodologique pour la construction et l'analyse des graphes génétiques pour (i) identifier la topologie précise du réseau de dispersion ou (ii) inférer la résistance du paysage au flux génétique

Transition

L'analyse de la littérature a montré que les modalités d'utilisation des graphes génétiques ont été moins étudiées que celles des graphes paysagers. Avant de comparer des graphes génétiques et paysagers, il était donc nécessaire de déterminer comment construire et analyser les graphes génétiques selon l'objectif de recherche poursuivi. C'est ce que nous avons fait dans le premier chapitre de cette thèse. Cela nous a permis d'identifier les méthodes d'élagage, de pondération et d'analyse des liens des graphes génétiques permettant d'identifier la topologie d'un réseau de dispersion et d'inférer la résistance du paysage au flux génétique.

Pour comparer des graphes génétiques et paysagers, il est également nécessaire de disposer d'un outil de construction et d'analyse de ces deux types de graphes puisant dans le large éventail des méthodes actuellement disponibles. Le logiciel Graphab, développé au sein du laboratoire ThéMA par Gilles Vuidel dans le langage Java, a facilité l'utilisation des graphes paysagers. Toutefois, il manquait un tel outil dédié aux graphes génétiques. De plus, des méthodes permettant de comparer ces deux graphes à différents niveaux et dans un environnement similaire faisaient également défaut. Pour pallier à ce manque, nous avons développé le package R `graph41g`, que nous présentons dans le chapitre suivant. Comme le précédent, ce chapitre constitue une étape préliminaire à la comparaison des graphes génétiques et paysagers, dédiée au développement de l'outil qui sera mis en œuvre pour la réaliser ensuite.

Chapitre 2

graph4lg : un package R pour construire et analyser des graphes en génétique du paysage

Résumé

Nous avons développé un package R pour faciliter et encourager l'utilisation des graphes génétiques et paysagers pour répondre à des questions de génétique du paysage. Il intègre des fonctions dédiées à la conversion et à l'import de données génétiques et au calcul de distances génétiques. Il permet aussi de calculer des distances géodésiques et des distances-coût à partir de données spatiales. Une importante gamme de paramètres peut être utilisée pour créer des graphes génétiques et paysagers. Plusieurs méthodes d'élagage sont notamment disponibles. Nous avons également rendu accessible l'utilisation du logiciel Graphab aux utilisateurs du logiciel R pour faciliter la construction et l'analyse de graphes paysagers dans cet environnement. Par ailleurs, les fonctions du package permettent des analyses préliminaires ayant pour but d'adapter les choix méthodologiques aux questions de recherche. Les graphes génétiques et paysagers créés peuvent être analysés avec des métriques calculées au niveau des nœuds ou avec des analyses des liens ou des modules de ces graphes. Les utilisateurs peuvent comparer et visualiser les graphes et les exporter sous forme de couches au format *shapefile* pour faciliter l'interprétation et les analyses ultérieures. **graph4lg** contribue ainsi au potentiel des graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats, tout en encourageant de futures recherches sur les aspects méthodologiques relatifs à ces outils.

Cet article a été publié dans la revue *Methods in Ecology and Evolution* en novembre 2020 :

Savary, P., Foltête, J. C., Moal, H., Vuidel, G., & Garnier, S. 2021. **graph4lg** : A package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, 12(3), 539-547

Graph theory investigates all aspects of combinations of nodes with edges joining them; and "all" is no exaggeration.

What is continually impressive about graph theory is the way that it can go from what seems simple and intuitive to very sophisticated (and, yes, difficult) results.

Mark Dale

1 Introduction

Les approches basées sur les graphes se sont révélées pertinentes en écologie du paysage et en génétique des populations de par leur capacité à analyser les interactions entre des taches d'habitat ou des populations formant un réseau (Dale et Fortin, 2010 ; Dale, 2017 ; Fortin *et al.*, 2012 ; Manel et Holderegger, 2013). En écologie du paysage, l'utilisation des graphes paysagers a bénéficié du développement de logiciels gratuits tels que Graphab (Foltête *et al.*, 2012b) ou Conefor (Saura et Torne, 2009). Néanmoins, bien que ces logiciels aient rendu possible l'import de certaines données biologiques (Foltête *et al.*, 2012c), ils ne permettaient pas d'intégrer de données génétiques, dont la mise en forme et le traitement statistique sont particuliers.

Comme nous l'avons vu, les approches basées sur les graphes se sont également développées en génétique des populations (Arnaud, 2003 ; Excoffier *et al.*, 1992) et ont permis d'analyser la structure génétique des populations sous la forme de graphes génétiques (Greenbaum et Fefferman, 2017). Le chapitre précédent illustre le fait que la construction et l'analyse des liens de ces graphes devraient être guidées par la question de recherche, le contexte de l'étude et les connaissances disponibles sur l'espèce étudiée. Par ailleurs, les graphes génétiques peuvent être analysés en calculant des métriques au niveau de leurs nœuds (*node- et neighbourhood-level*) et/ou en localisant les limites de modules composés d'ensembles de nœuds (*boundary-level*) (Wagner et Fortin, 2013). Dans ce cadre, l'absence d'outils permettant de mettre en œuvre l'ensemble des méthodes de construction et d'analyse des graphes génétiques restreignait l'utilisation de ces graphes et la flexibilité de cette utilisation.

Enfin, bien qu'il ait été suggéré à plusieurs reprises que la comparaison de graphes génétiques et paysagers pourrait améliorer la compréhension de l'effet de la distribution spatiale des habitats sur la structure génétique des populations (Foltête et Vuidel, 2017 ; Galpern *et al.*, 2011 ; Manel et Holderegger, 2013 ; Murphy *et al.*, 2016), **un outil permettant de construire, analyser et comparer ces graphes faisait défaut. Nous avons ainsi développé le package R graph4lg pour pallier à ce manque. Cet outil facilite le choix et la mise en œuvre des méthodes de construction et d'analyse des graphes en se basant notamment sur d'autres packages R destinés aux analyses de graphes ou de données génétiques, tels que igraph (Csardi et Nepusz, 2006) ou adegenet (Jombart, 2008).**

2 Organisation des fonctions du package

Les fonctions du package `graph4lg` peuvent être séparées en quatre catégories correspondant aux différentes étapes des analyses de génétique du paysage :

1. Que les données brutes soient des données génétiques ou spatiales, le package permet de les traiter pour préparer la construction des graphes. Pour cela, des fonctions calculent des indices génétiques intra-populationnels et des distances génétiques ou paysagères entre populations. Des analyses préliminaires guidant les choix méthodologiques sont rendues possibles par des fonctions créant des figures "diagnostic".
2. Un ensemble de fonctions est destiné à la construction et à l'analyse des graphes génétiques.
3. En parallèle, des fonctions permettent d'utiliser directement le logiciel Graphab (Foltête *et al.*, 2012b) dans l'environnement de R pour construire et analyser des graphes paysagers.

4. Enfin, les graphes génétiques et paysagers peuvent être comparés, visualisés et exportés dans d'autres formats.

Cette organisation des fonctions est décrite dans les parties suivantes et sur la figure 43.

2.1 Traitements des données brutes

2.1.1 Données génétiques

L'étude des flux génétiques et de l'influence qu'exerce sur eux le paysage se fait principalement à partir de microsatellites supposés neutres ou identifiés comme tels (Storfer *et al.*, 2010). Néanmoins, les marqueurs de type SNP (*Single Nucleotide Polymorphism*) sont de plus en plus utilisés pour réaliser ce type d'études, à condition que les *loci* soumis à des pressions de sélection aient été mis de côté au préalable (Cushman *et al.*, 2018 ; Foll et Gaggiotti, 2008). Pour ces raisons, les fonctions du package `graph4lg` permettent de traiter des données génétiques concernant à la fois des microsatellites ou des SNPs. Ces données sont traitées sous la forme d'un objet de type `genind` dans R, ce qui facilite le recours à des fonctions du package `adegenet` (Jombart, 2008) pour certains calculs. Étant donné que les analyses de génétique du paysage ont le plus souvent pour focus des populations au sein desquelles plusieurs individus ont été échantillonnés, la plupart des fonctions du package nécessitent la définition de populations *a priori* (Milligan *et al.*, 2018 ; Waits et Storfer, 2015).

Pour faciliter la conversion entre les formats de données destinés aux principaux logiciels et packages R utilisés en génétique des populations (`gstudio` (Dyer, 2014), `pegas` (Paradis, 2010), `STRUCTURE` (Pritchard *et al.*, 2000) ou `GENEPOP` (Raymond et Rousset, 1995)) et les objets de type `genind`, nous avons développé plusieurs fonctions de conversion (`gstud_to_genind`, `loci_to_genind`, `structure_to_genind`, `genepop_to_genind`). Nous avons aussi rendu possible l'export des objets `genind` sous la forme de fichiers `.txt` destinés au logiciel `GENEPOP` (fonction `genind_to_genepop`), couramment utilisé dans ce domaine.

Une fois les données génétiques importées sous forme d'objets `genind`, la fonction `mat_gen_dist` calcule huit types différents de distance génétique entre populations (Table 3). Néanmoins, des matrices de distances génétiques calculées différemment par l'utilisateur peuvent également être importées et utilisées par les fonctions décrites par la suite pour construire des graphes génétiques.

2.1.2 Données spatiales

Outre les distances génétiques, l'utilisateur peut calculer deux types de distances paysagères :

- La fonction `mat_geo_dist` calcule une matrice de distances géodésiques euclidiennes à partir des coordonnées d'un ensemble de points ; que ces coordonnées soient exprimées selon un système de coordonnées polaire (en degrés) ou projeté (en mètres).
- La fonction `mat_cost_dist` calcule une matrice de distances-coût à partir d'un ensemble de points, d'une couche raster catégorielle sur laquelle chaque cellule est associée à un type d'occupation du sol et d'une table (`data.frame`) indiquant les valeurs de coût de déplacement associées à chaque type d'occupation du sol. Par défaut, cette fonction utilise des fonctions du package `gdistance` (Van Etten, 2012). Néanmoins, les utilisateurs disposant du logiciel Java peuvent utiliser un programme externe codé dans ce langage pour réduire les temps de calcul de façon significative pour les rasters les plus lourds.

Distance génétique	Description	Eq.	Depend.	Ref.
F_{ST}	Indice de fixation	Oui	diveRsity	Weir et Cockerham (1984)
F_{ST} linéarisé	Indice de fixation linéarisé	Oui	diveRsity	Rousset (1997)
G'_{ST}	Indice de fixation standardisé	Oui	diveRsity	Hedrick (2005)
D_{Jost}	Indice de fixation standardisé	Oui	diveRsity	Jost (2008)
D_{PS}	1 - proportion d'allèles partagés	Non	Aucune	Bowcock <i>et al.</i> (1994), utilisation de la formule du logiciel MSA (Dieringer <i>et Schlotterer</i> , 2003)
Distance génétique euclidienne	Calculée à partir des différences de fréquences alléliques	Non	Aucune	Excoffier <i>et al.</i> (1992)
Distance génétique euclidienne pondérée	Calculée à partir des différences de fréquences alléliques en donnant plus de poids aux allèles rares	Non	Aucune	Fortuna <i>et al.</i> (2009) ; Greenbaum <i>et al.</i> (2016)
Distance génétique euclidienne dérivée d'une ACP	Distance entre populations dans l'espace défini par les composantes principales obtenues par ACP de la table des fréquences alléliques	Non	Aucune	Inspirée par les distances génétiques calculées par Paschou <i>et al.</i> (2014) <i>et Shirk <i>et al.</i> (2017a)</i>
Distance génétique issue du package <code>popgraph</code>	Distance entre populations dans l'espace défini par les composantes principales obtenues par ACP et SVD de la table des fréquences alléliques	Non	Aucune	Dyer <i>et Nason</i> (2004)

TABLE 3 – Distances génétiques entre populations calculées par la fonction `mat_gen_dist`. La colonne 'Eq.' indique si l'utilisation de la distance génétique repose sur le postulat d'équilibre entre migration et dérive. La colonne 'Depend.' indique les packages R dont dépend parfois cette fonction pour le calcul d'une distance génétique.

2.1.3 Analyses préliminaires

L'effet du paysage sur la dispersion peut être inféré à partir de la relation entre les distances génétiques et paysagères associées aux liens d'un graphe. Dans ce cas, visualiser la relation entre ces distances sous la forme d'un nuage de points en considérant l'ensemble des paires de populations (fonction `scatter_dist`) peut être une première étape avant la construction du graphe. Une relation linéaire est attendue lorsque le patron de différenciation génétique déterminé par la dérive et le flux génétique a atteint un équilibre à l'échelle de toute la zone d'étude (Slatkin, 1993). En revanche, une relation non-linéaire caractérisée par un plateau au-delà d'une distance paysagère seuil est attendue lorsque cet équilibre n'a pas été atteint (IBD de type IV selon la typologie d'Hutchison et Templeton (1999)) (Hänfling et Weetman, 2006). Dans ce dernier cas, il faut tenir compte de l'échelle à laquelle ce plateau apparaît pour réaliser l'inférence (Bradbury et Bentzen, 2007 ; Van Strien *et al.*, 2015). Élaguer le graphe génétique pour ne conserver que les paires de populations séparées par une distance paysagère équivalente à la distance à laquelle apparaît le plateau peut être une solution. Par ailleurs, l'atteinte d'un état d'équilibre entre flux génétique et dérive est un pré-requis à l'utilisation d'indices de fixation pour inférer certains paramètres démographiques (Neigel, 2002 ; Whitlock et McCauley, 1999).

Lors d'une étude des déterminants des patrons d'IBD, Van Strien *et al.* (2015) ont estimé la distance seuil entre populations à laquelle la corrélation entre différenciation génétique et distance paysagère est

la plus forte. Cette distance de corrélation maximale (nommée DMC) est une estimation de l'échelle à laquelle la différenciation génétique entre populations atteint un état d'équilibre dépendant à la fois de la dérive et du flux génétique. La fonction `dist_max_corr` calcule cette distance (Figure 44A), ce qui peut éventuellement aider l'utilisateur à choisir la méthode d'élagage d'un graphe génétique (cf. section 2.2.1).

Quand l'objectif est d'identifier des chemins de dispersion directs en considérant la résistance du paysage et les capacités de dispersion de l'espèce étudiée, il peut être nécessaire de convertir une distance de dispersion maximale exprimée en mètres en une distance-coût équivalente exprimée en unités de coût. La fonction `convert_cd` rend cette conversion possible en la basant sur les paramètres de la régression linéaire ou log-log linéaire des distances-coût en fonction des distances géodésiques euclidiennes (Tournant *et al.*, 2013). La valeur convertie peut alors être utilisée comme distance seuil pour l'élagage d'un graphe.

2.2 Construction et analyse de graphes génétiques

2.2.1 Construction de graphes génétiques

Le package `graph4lg` met à disposition de l'utilisateur une large gamme de méthodes pour élaguer des graphes génétiques. Tout d'abord, les graphes peuvent être élagués en supprimant les liens entre des populations séparées par des distances génétiques ou paysagères supérieures ou inférieures à une valeur seuil (`gen_graph_thr`). Cette méthode identifie des chemins de dispersion directs dès lors qu'on dispose d'une estimation précise des capacités de dispersion maximales de l'espèce. Par ailleurs, elle permet de sélectionner les paires de populations incluses dans une analyse selon un critère de distance ; comme l'ont fait Angelone *et al.* (2011) et Keller *et al.* (2013), entre autres.

Quand l'espèce étudiée se disperse essentiellement de proche en proche (*stepping-stone dispersal*), l'élagage peut être défini par des contraintes topologiques à l'aide de la fonction `gen_graph_topo`. Des arbres couvrants minimaux, des graphes planaires, des graphes ne connectant que les k voisins les plus proches ou des graphes de Gabriel, déjà utilisés dans de précédents travaux (Arnaud, 2003 ; Bunn *et al.*, 2000 ; Keller *et al.*, 2013 ; Naujokaitis-Lewis *et al.*, 2013), peuvent alors être construits en appliquant les contraintes topologiques à des matrices de distances génétiques ou paysagères. Une méthode d'élagage basée sur la théorie de la percolation fait également partie de ce panel de méthodes.

Enfin, la fonction `gen_graph_indep` crée un graphe génétique en appliquant le principe d'indépendance conditionnelle (Whittaker, 2009). Elle prolonge et adapte la fonction `popgraph` du package R éponyme en appliquant différentes formules de calcul de covariance (Savary *et al.*, 2021a) et en rendant possible l'ajustement des p -valeurs (Benjamini et Hochberg, 1995 ; Holm, 1979).

2.2.2 Analyses de graphes génétiques

Une fois les graphes génétiques créés, la fonction `compute_node_metric` calcule des métriques de centralité issues de la théorie des graphes telles que le degré, la *closeness* ou la *betweenness* afin d'identifier les nœuds qui sont potentiellement les plus importants du réseau en termes de flux génétique (Cross *et al.*, 2018). Elle calcule également la moyenne ou la somme de l'inverse des distances génétiques pondérant les liens du graphe génétique connectés à chaque nœud (MIW et SIW, respec-

tivement). En parallèle de ces métriques dépendant du voisinage de chaque population défini par la topologie du graphe (*neighbourhood level*), la fonction `pop_gen_index` calcule des indices génétiques intra-populationnels (*node level*) tels que la richesse allélique ou les taux d'hétérozygotie observée et attendue. Toutes ces métriques peuvent être associées aux nœuds du graphe génétique en tant qu'attributs ou poids des nœuds (`add_nodes_attr`).

Les nœuds peuvent être partitionnés en modules en se basant sur le poids des liens qui les relient et en utilisant un algorithme de maximisation de la modularité (`compute_graph_modul`). Des barrières paysagères à la dispersion causant l'apparition de *clusters* de populations d'un point de vue génétique peuvent ainsi être mises en évidence (Fortuna *et al.*, 2009 ; Garroway *et al.*, 2008).

2.3 Construction et analyse de graphes paysagers

Le package `graph4lg` intègre les méthodes de construction et d'analyse de graphes paysagers du logiciel Graphab (Foltête *et al.*, 2012b). Dans un premier temps, la fonction `graphab_project` crée un projet Graphab en délimitant les taches d'habitat à partir d'une couche raster catégorielle représentant l'occupation du sol de la zone d'étude. Ensuite, les chemins de moindre coût entre ces taches d'habitat sont calculés à partir des valeurs de coût de déplacement fournies par l'utilisateur (`graphab_link`). Les taches d'habitat et le jeu de liens créés servent de base à la création de graphes paysagers à la topologie complète ou planaire ou élagués à partir d'une distance seuil (`graphab_graph`).

Une large gamme de métriques de connectivité peuvent être calculées au niveau du graphe dans son ensemble ou de chacun de ses nœuds (`graphab_metric`). Le calcul de "delta-metriques" permet notamment d'identifier les éléments du graphe les plus importants pour la conservation de la connectivité des habitats de l'ensemble de la zone d'étude. Dans la perspective d'analyses confrontant les métriques de connectivité à des réponses biologiques, le package permet d'importer des données ponctuelles dans le projet Graphab et d'obtenir les métriques de connectivité associées aux taches d'habitat les plus proches (`graphab_pointset`); par exemple pour mener ensuite des analyses statistiques dans l'environnement du logiciel R. Enfin, la fonction `graphab_modul` permet de faire une partition des taches d'habitat à l'aide d'un algorithme maximisant la modularité pour identifier de potentielles barrières à la dispersion.

2.4 Comparaisons de graphes génétiques et paysagers

Les graphes génétiques et paysagers ont rarement été comparés directement bien qu'une telle comparaison présente un intérêt pour les études de génétique du paysage (mais voir Draheim *et al.* (2016) et Schoville *et al.* (2018)). Pour faciliter l'interprétation de leurs topologies respectives ainsi que la formulation et le test d'hypothèses concernant leurs similitudes, la fonction `plot_graph_lg` permet aux utilisateurs de visualiser la topologie des graphes (Figure 44D). Cette visualisation se base :

- sur les coordonnées spatiales des nœuds pour les cartographier dans un espace géographique (`mode="spatial"`);
- sur l'algorithme d'attraction-répulsion de Fruchterman et Reingold (1991). Cet algorithme utilise les poids des liens pour calculer les positions des nœuds qui permettent de visualiser au mieux les relations entre les nœuds définies par les poids des liens, indépendamment de leurs localisations spatiales (`mode="aspatial"`).

Les métriques calculées au niveau des nœuds, la partition des nœuds en modules et les poids des liens peuvent être visualisés à l'aide de cette même fonction en utilisant ses paramètres pour définir la taille des nœuds, la couleur des nœuds et la largeur des liens, respectivement (Figure 44D). Un histogramme des poids des liens peut être créé à l'aide de la fonction (`plot_hist_w`, Figure 44C). En effet, la distribution des poids des liens des graphes génétiques et paysagers est un déterminant important du patron spatial des processus de dispersion et de flux génétique qu'ils représentent. De plus, les propriétés des liens conservés suite à l'élagage des graphes peuvent être visualisées sur le nuage de points représentant la relation entre distances génétiques et distances paysagères en y distinguant les paires de populations connectées sur le graphe élagué (`scatter_dist_g`, Figure 44B).

Lorsque les graphes génétiques et paysagers partagent les mêmes nœuds, des hypothèses peuvent être testées pour comprendre la relation entre (i) la connectivité écologique des habitats et (ii) la structure génétique des populations. Pour ce faire, la fonction `graph_node_compar` évalue la corrélation entre les indices génétiques calculés au niveau des populations et les métriques de connectivité associées aux taches d'habitat qu'elles occupent. Par ailleurs, les topologies de deux graphes peuvent être comparées en créant une table de contingence basée sur la présence ou l'absence de chacun des liens dans chacun des deux graphes. La similarité des topologies est alors mesurée en calculant différents indicateurs à partir de la table de contingence (fonction `graph_topo_compar`). La coïncidence de leurs topologies peut également être représentée sur une carte à l'aide de la fonction `graph_plot_compar`. Les différences entre graphes identifiées lors de ces analyses apportent notamment des informations sur la capacité des graphes paysagers à représenter les flux génétiques.

Enfin, compter le nombre de paires de nœuds qui appartiennent au même module dans un graphe tout en appartenant également au même module dans un autre graphe indique à quel point ces deux graphes reflètent la même réalité. Ce type d'analyse est rendu possible par la fonction `graph_modul_compar` qui calcule l'*Adjusted Rand Index* (ARI) (Hubert et Arabie, 1985) pour comparer deux partitions de nœuds en modules.

Les graphes créés à l'aide de `graph4lg` peuvent être exportés sous forme de couches au format *shapefile* pour être intégrés à un SIG (`graph_to_shp`), afin de réaliser des analyses impliquant d'autres types de données géographiques par exemple.

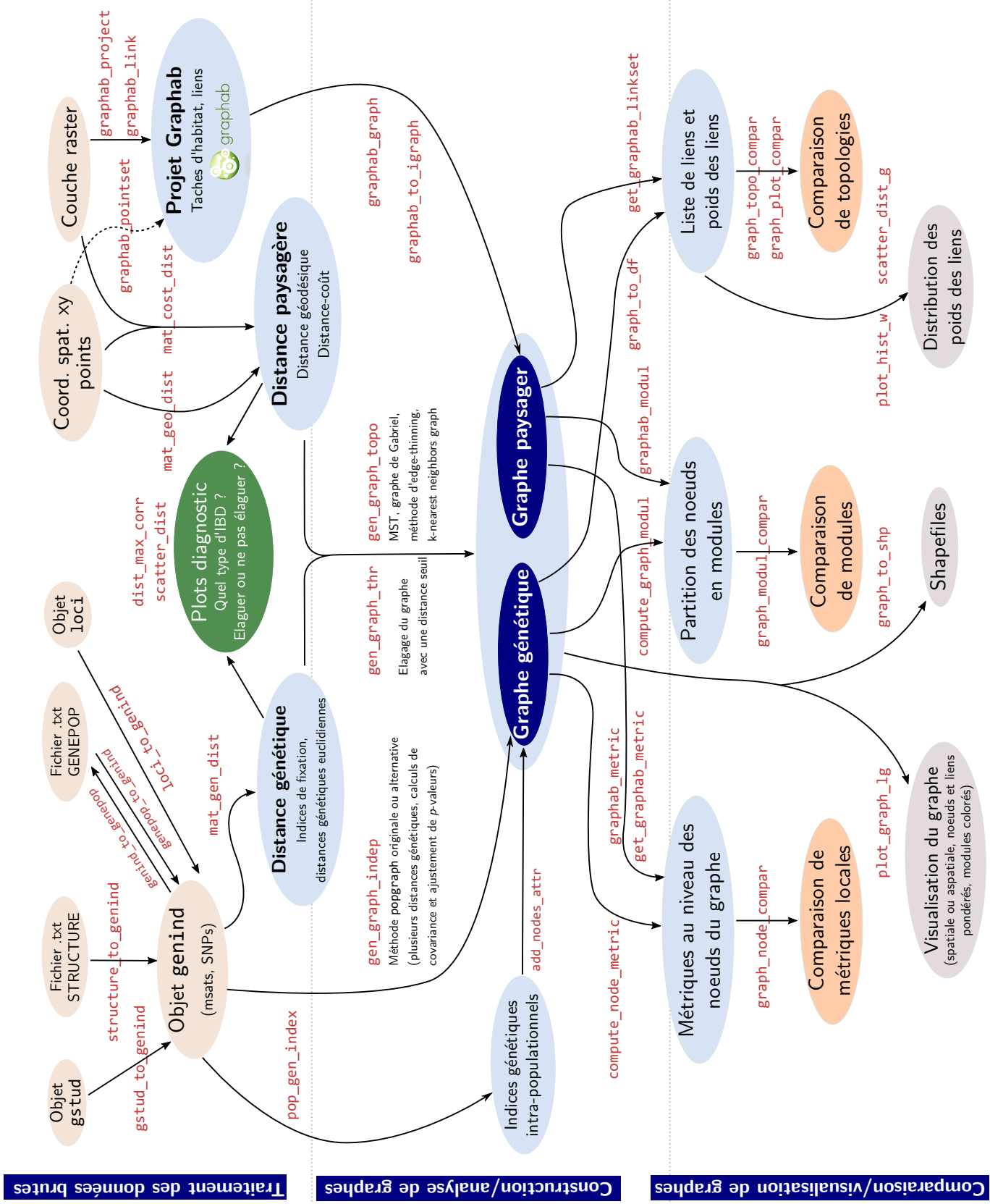


FIGURE 43 – Organisation des fonctions du package `graph41g`. Les fonctions sont représentées par des flèches reliant leurs arguments à leurs résultats. Leurs noms sont écrits en rouge. Les boîtes beige clair correspondent aux données d'entrée tandis que les boîtes grises correspondent aux résultats qui peuvent être exportés sous forme de figures ou de fichiers. Les boîtes bleu clair indiquent les résultats qui sont utilisés pour construire les graphes ou qui sont obtenus à partir des graphes. Les boîtes bleu foncé correspondent aux deux types de graphes qui peuvent être construits avec le package. La boîte verte correspond aux figures de 'diagnostic' dont l'analyse peut guider le choix d'une méthode de construction. Les boîtes orange indiquent les comparaisons de graphes génétiques et paysagers qui peuvent être réalisées à plusieurs niveaux (noeuds, modules).

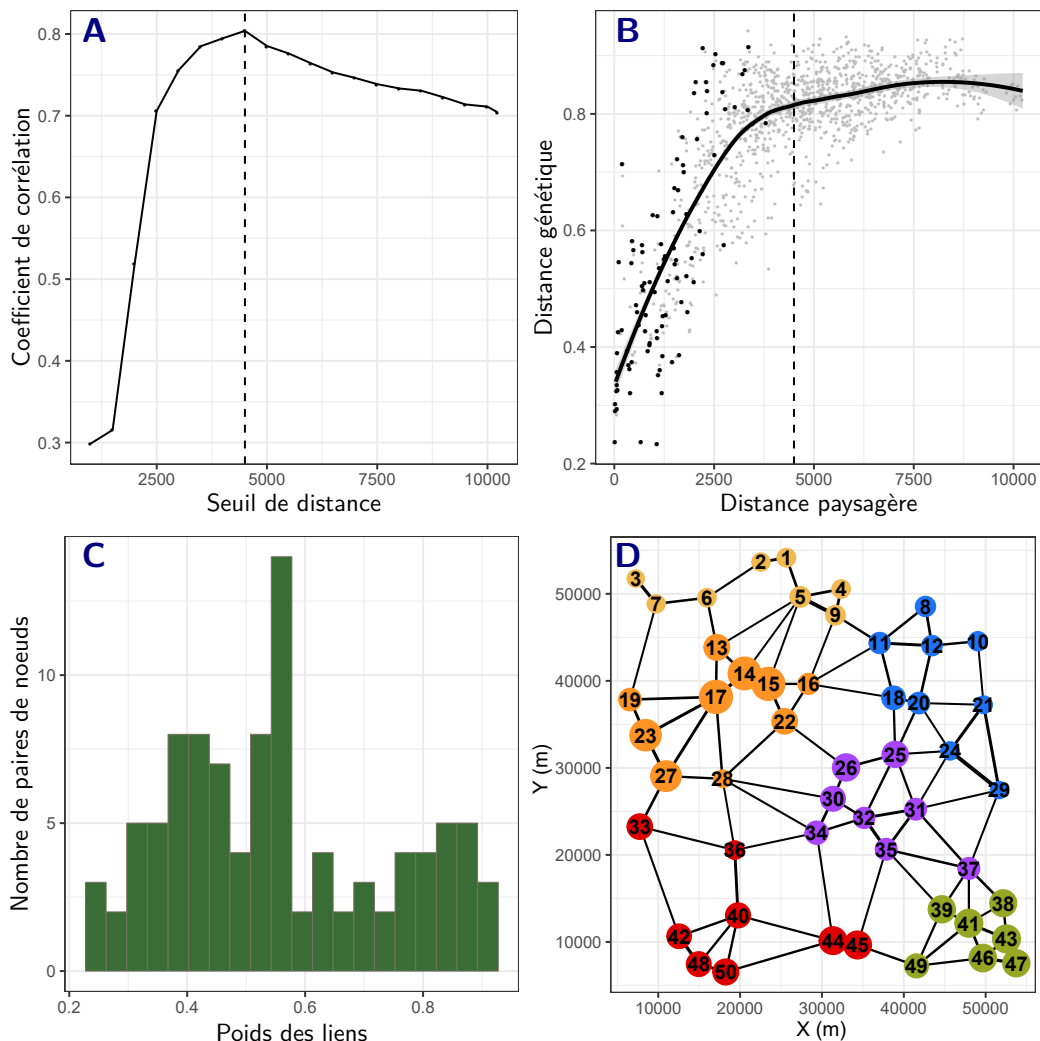


FIGURE 44 – Figures de diagnostic (A, B) et d’analyse de graphes génétiques (C, D) réalisées à partir des jeux de données `data_simul_genind` et `pts_pop_simul` obtenus par simulation à l’aide de CDPOP (Landguth et Cushman, 2010). (A) Identification de la distance de corrélation maximale (DMC) (Van Strien *et al.*, 2015) avec la fonction `dist_max_corr`. Ici, la DMC correspond à la ligne verticale en pointillés sur les figures A et B. (B) Nuage de points créé avec la fonction `scatter_dist_g` et représentant la relation entre distances génétiques (D_{PS}) et distances-coût. La zone grise autour de la courbe d’ajustement lissée correspond à l’intervalle de confiance à 95 % de la fonction de lissage. Les points noirs représentent les paires de populations connectées par un lien sur le graphe de Gabriel. (C) Histogramme des distances génétiques séparant les paires de populations connectées par un lien sur le graphe de Gabriel, obtenu avec la fonction `plot_w_hist`. (D) Graphe de Gabriel cartographié avec la fonction `plot_graph_lg`. La largeur d’un lien est inversement proportionnelle à la distance génétique qui pondère ce lien. La taille d’un nœud est proportionnelle au Flux, une métrique de connectivité calculée à partir du graphe paysager correspondant. La couleur des nœuds indique à quel module appartient chaque nœud suite à l’analyse de modularité faite avec la fonction `compute_graph_modul`.

3 Limites et conclusion

Nous ne prétendons pas ici rassembler l'ensemble exhaustif des méthodes de construction et d'analyse des graphes génétiques et paysagers. D'autres méthodes d'élagage ont été développées (Brooks, 2006 ; Greenbaum *et al.*, 2016 ; Kininmonth *et al.*, 2010 ; Milligan, *in prep* ; Peterson *et al.*, 2019) et pourraient notamment être appliquées à des graphes dirigés. Par ailleurs, une large gamme de métriques inspirées de la théorie des métapopulations ont été développées pour analyser des graphes paysagers et pourraient également inspirer les métriques locales calculées à partir des graphes génétiques. De plus, des tests de significativité statistique basés sur des permutations aléatoires amélioreraient la robustesse de l'analyse des propriétés des graphes. Enfin, bien que le logiciel Graphab autorise le traitement de jeux de données spatiales de taille importante en un temps réduit (Foltête *et al.*, 2012b), la manipulation des jeux de données génétiques les plus lourds (des SNPs par exemple) peut impliquer des temps de calcul plus longs, limitant le champ d'application de ces méthodes. Les prochaines versions du package pourraient introduire des améliorations en ce sens.

Notre but premier lors du développement de `graph4lg` était de rassembler et rendre accessible un large panel de méthodes actuellement utilisées en génétique du paysage pour construire et analyser des graphes. Nous espérons que ce package encouragera l'utilisation des graphes génétiques et paysagers, ainsi que les recherches relatives aux aspects théoriques et méthodologiques associés à leur utilisation.

Transition

Dans les deux premiers chapitres de cette partie de la thèse dédiée à la comparaison de graphes génétiques et paysagers, nous avons présenté successivement la façon dont on pouvait utiliser les graphes génétiques et l'outil informatique permettant d'utiliser ces deux types de graphes séparément ou de façon conjointe. Nous allons à présent mettre en œuvre concrètement la comparaison de ces deux types de graphes en utilisant des données empiriques.

Dans le chapitre suivant, nous mettons en relation des indices génétiques et des métriques de connectivité écologique des habitats calculés au niveau des nœuds des graphes en tenant compte ou non de leurs connexions avec les autres nœuds (*neighbourhood level versus node level*). Les objectifs de cette comparaison de graphes génétiques et paysagers sont multiples. Elle vise premièrement à valider la définition de la connectivité écologique des habitats énoncée en introduction, que nous assimilons à la quantité d'habitat atteignable (ARH, *Amount of Reachable Habitat*). Elle vise ensuite à mettre en évidence les relations complexes qui existent entre les composantes de cette quantité d'habitat atteignable et celles de la structure génétique des populations. Enfin, elle illustre l'intérêt des approches basées sur les graphes pour les analyses de type *neighbourhood level* ou *node level* en génétique du paysage. À ce titre, ce chapitre présente une utilisation des graphes génétiques différente mais complémentaire de celle étudiée dans le chapitre 1 (*link level*).

Chapitre 3

Analyse de l'influence de la quantité d'habitat atteignable sur la structure génétique à l'aide de graphes génétiques et paysagers

Résumé

La structure génétique, c'est-à-dire la diversité génétique de chaque population et la différenciation génétique entre populations, est influencée par la surface et la configuration spatiale des zones d'habitat. Mesurer la quantité d'habitat atteignable (ARH, *Amount of Reachable Habitat*) permet de décrire ces deux composantes du patron spatial des habitats en considérant la connectivité à l'intérieur et entre les taches, les capacités de dispersion et le coût des déplacements à travers la matrice paysagère. Des métriques complémentaires mesurant l'ARH selon différents scénarios de coût sont censées refléter l'influence conjointe de la dérive et du flux génétique sur la structure génétique. En utilisant des données empiriques concernant le Criquet ensanglanté (*Stethophyma grossum*), nous avons évalué la pertinence des métriques d'ARH en tant que variables explicatives de la structure génétique. Pour ce faire, nous avons calculé d'une part des indices de diversité génétique locale et de différenciation génétique au niveau des nœuds de graphes génétiques, et d'autre part plusieurs métriques d'ARH au niveau des nœuds de graphes paysagers, avant d'évaluer leurs relations. Les métriques d'ARH se sont révélées pertinentes pour expliquer les deux composantes de la structure génétique. Alors que la richesse allélique d'une population était expliquée par les trois métriques d'ARH de façon significative, la richesse allélique privée et la différenciation génétique d'une population par rapport aux autres étaient essentiellement liées à l'ARH mesurée en dehors de la tache occupée par une population. La prise en compte de différents scénarios de coût s'est révélée essentielle pour expliquer chacune des réponses génétiques. Nous encourageons l'utilisation future du concept d'ARH et de métriques issues de graphes génétiques et paysagers pour analyser l'influence du patron spatial de l'habitat sur les différentes composantes de la structure génétique.

Cet article est en voie d'être re-soumis après modifications dans la revue *Heredity* :

Savary, P., Foltête, J. C., van Strien, M.J., Moal, H., Vuidel, G. & Garnier, S. Assessing the influence of the amount of reachable habitat on genetic structure using landscape and genetic graphs. Submitted to *Heredity*

*Percevoir un territoire comme un tissu de liens potentiels
lui donne une épaisseur, une texture particulière.*

Alessandro Pignocchi

1 Introduction

Lorsqu'on s'intéresse à des marqueurs génétiques neutres, la dérive et le flux génétique sont les deux principaux processus influençant les composantes de la structure génétique, à savoir la diversité génétique locale et la différenciation génétique entre populations. Dans la mesure où la surface et la configuration des taches d'habitat influencent toutes deux la dérive et le flux génétique, la structure génétique dépend du patron spatial des zones d'habitat (DiLeo et Wagner, 2016 ; Keyghobadi, 2007). Cependant, surface et configuration des taches d'habitat n'influencent pas de la même manière ces deux processus génétiques, donnant lieu à des relations indirectes et complexes entre structure génétique et patron spatial des zones d'habitat. À l'heure où la destruction des habitats menace la biodiversité, il semble crucial de les comprendre précisément.

Pour ce faire, il est nécessaire de caractériser le patron spatial des habitats à l'aide de variables reflétant son influence potentielle sur la dérive et le flux génétique. Elles doivent être calculées à l'échelle des populations sur lesquelles ces processus agissent. Dans cette thèse, nous avons défini la connectivité écologique des habitats comme la quantité d'habitat atteignable, mesurée localement et en tenant compte des perceptions de l'espèce étudiée et de la position des taches au sein du réseau d'habitat. Cette quantité d'habitat atteignable, notée par la suite ARH (*Amount of Reachable Habitat*), a plusieurs composantes. Nous les avons assimilées aux trois propriétés de recrutement, flux et traversabilité conceptualisées par Urban et Keitt (2001).

Chacune des composantes de l'ARH peut se mesurer à partir d'un graphe paysager. Notons d'ailleurs que cet outil d'analyse est à l'origine des principales réflexions sur l'ARH (Bergès *et al.*, 2020 ; Saura et de la Fuente, 2017). Comme nous l'avons vu précédemment, la capacité des taches du graphe reflète la quantité d'habitat atteignable au sein des taches occupées par les populations. Elle devrait déterminer la taille des populations (recrutement), et donc l'intensité de la dérive génétique. Par ailleurs, des métriques telles que le Flux (F) reflètent la quantité d'habitat qu'un individu peut atteindre en se dispersant depuis une tache donnée vers les autres taches du graphe, et réciproquement le nombre d'individus qui peuvent atteindre cette tache. Ainsi, la métrique F reflète potentiellement les flux génétiques s'effectuant au niveau d'une tache. Enfin, des métriques telles que l'indice de centralité intermédiaire (BC, *Betweenness Centrality*) mesurent le rôle d'une tache en tant que *stepping-stone* favorisant la traversabilité du réseau d'habitat. Une métrique telle que le BC est donc censée mesurer le fait qu'une tache puisse être une étape intermédiaire dans des flux génétiques s'effectuant sur plusieurs générations entre des taches d'habitat éloignées les unes des autres. Les trois métriques que sont la capacité, le F et le BC mesurent donc de façon complémentaire l'ARH et son influence potentielle sur la dérive et le flux génétique.

Par ailleurs, dès lors que l'on considère des surfaces d'habitat situées en dehors de la tache focale pour calculer l'ARH, il est préférable de considérer la résistance de la matrice dans le calcul. L'utilisation des graphes paysagers rend cela possible dans la mesure où les liens des graphes sont fréquemment pondérés par des distances-coût. Le calcul des différentes métriques d'ARH s'ajuste également aux capacités de dispersion des espèces étudiées en faisant des hypothèses sur l'échelle à laquelle les individus peuvent se disperser vers d'autres taches. En somme, les graphes paysagers constituent un cadre flexible et unifié pour mesurer l'ARH avec un ensemble de métriques complémentaires.

De telles métriques se sont révélées utiles pour expliquer la composition de communautés d'espèces (Awade *et al.*, 2012 ; Mony *et al.*, 2018) ou pour concevoir des mesures de conservation (Bergès *et al.*, 2020 ; Saura et de la Fuente, 2017). Elles ont plus rarement été utilisées pour expliquer la structure génétique des populations (mais voir Bertin *et al.* (2017), Flavenot *et al.* (2015) et Schoville *et al.* (2018)), malgré l'intérêt potentiel qu'elles présentent dans ce cadre. En effet, la plupart des analyses de génétique du paysage se focalisent sur la relation entre distances génétiques et distances paysagères (analyses *link level*, *sensu* Wagner et Fortin (2013)) pour tester l'effet de la structure du paysage sur la différenciation (DiLeo et Wagner, 2016). En comparaison, il y a peu d'études sur l'influence du paysage sur des indices de diversité génétique ou de différenciation génétique relative mesurés à l'échelle des populations (analyses *node-* ou *neighbourhood level*, *sensu* Wagner et Fortin (2013)). En outre, les quelques études adoptant cette approche se focalisaient principalement sur la diversité génétique locale (Flavenot *et al.*, 2015 ; Toma *et al.*, 2015) et considéraient de simples métriques telles que la quantité d'habitat dans un voisinage circulaire autour des populations ou la distance à la tache la plus proche comme variables explicatives (Hahn *et al.*, 2013 ; Taylor et Hoffman, 2014). Dans ce cadre, les métriques d'ARH pourraient s'avérer pertinentes pour expliquer des indices locaux de diversité et de différenciation génétiques (Foltête *et al.*, 2020). Les calculer selon plusieurs scénarios de coût permettrait également de comprendre l'influence de la matrice sur la relation entre structure des habitats et structure génétique.

Comme nous l'avons présenté précédemment, utiliser des graphes génétiques conjointement aux graphes paysagers permettrait de calculer des indices de diversité et de différenciation génétiques au même niveau que celui auquel les métriques d'ARH sont calculées. L'élagage des liens des graphes génétiques offre aussi la possibilité de tenir compte de la topologie du réseau de populations et de l'échelle à laquelle le flux génétique a une influence significative sur la structure génétique (cf. Chapitre 1). DiLeo et Wagner (2016) ont mis en évidence le fait que les analyses de type *node level* et *neighbourhood level* étaient les seules permettant d'étudier les relations entre (i) la diversité ou la différenciation génétiques et (ii) la surface ou la configuration des habitats. Nous mettons en œuvre de telles analyses dans ce chapitre de la thèse en comparant des graphes génétiques et paysagers.

Par cette approche, nous souhaitons répondre aux questions suivantes :

- 1. Les métriques d'ARH sont-elles pertinentes pour expliquer la structure génétique ?**
- 2. Est-ce que l'ARH influence autant et à la même échelle spatiale la diversité et la différenciation génétiques ?**
- 3. Est-ce que la résistance de la matrice entre les taches d'habitat influence la diversité et la différenciation génétiques de la même manière ?**

Pour cela, nous avons calculé des indices de diversité et de différenciation génétiques à partir de graphes génétiques. En parallèle, nous avons calculé des métriques d'ARH à plusieurs échelles spatiales et selon plusieurs scénarios de résistance de la matrice à partir de graphes paysagers. À des fins de comparaison, nous avons également calculé des métriques d'habitat communément utilisées, à savoir la distance d'une population à la tache d'habitat la plus proche et la quantité d'habitat dans un voisinage défini par une distance autour de chaque population. Enfin, nous avons évalué les relations entre ces indices génétiques et ces métriques d'habitat à l'aide d'analyses de corrélation et de régressions PLS (*Partial Least Squares*).

2 Matériel et méthodes

2.1 Espèce étudiée et zone d'échantillonnage

Nous avons analysé des données empiriques concernant le Criquet ensanglanté (*Stethophyma grossum*, Figure 45) acquises et décrites par Keller *et al.* (2013) et Van Strien *et al.* (2014). Cette espèce d'orthoptère forme des populations discrètes dans les prairies inondables et les zones humides ouvertes (Bönsel et Sonneck, 2011 ; Reinhardt *et al.*, 2005 ; Sonneck *et al.*, 2008). Elle peut se disperser à travers des zones ouvertes ou des prairies ne constituant pas son habitat (Marzelli, 1994) et traverser des ruisseaux mais atteint difficilement des taches d'habitat entourées de zones arborées (Reinhardt *et al.*, 2005). Elle peut exceptionnellement parcourir jusqu'à 1500 m lorsque le milieu est favorable au déplacement (Griffioen, 1996). Ces caractéristiques en font un modèle biologique idéal pour comprendre comment le patron spatial des zones d'habitat influence la structure génétique.



FIGURE 45 – Criquet ensanglanté (*Stethophyma grossum*)

Keller *et al.* (2013) ont modélisé l'habitat potentiel de l'espèce aux alentours de Langenthal dans la région de l'Oberaargau en Suisse (Figure 46). Dans cette zone de 180 km², on trouve des surfaces forestières et bâties au milieu de surfaces agricoles cultivées intensivement. En juillet et août 2010, 39 populations de criquets ont été échantillonnées en parcourant de façon exhaustive les zones d'habitat potentiel modélisées. Un segment de patte de chaque individu capturé a été prélevé pour les analyses génétiques.

2.2 Analyse des données génétiques

Les analyses génétiques se sont basées sur 7 marqueurs microsatellites dont le protocole de génotypage est décrit par Keller *et al.* (2013). Nous avons considéré 34 populations comptant chacune au moins 12 individus, pour un total de 886 individus.

2.2.1 Indices de structure génétique

À l'échelle de chaque population, nous avons estimé la richesse allélique totale et la richesse allélique privée à l'aide du logiciel ADZE (Szpiech *et al.*, 2008). Nous avons utilisé pour cela une méthode

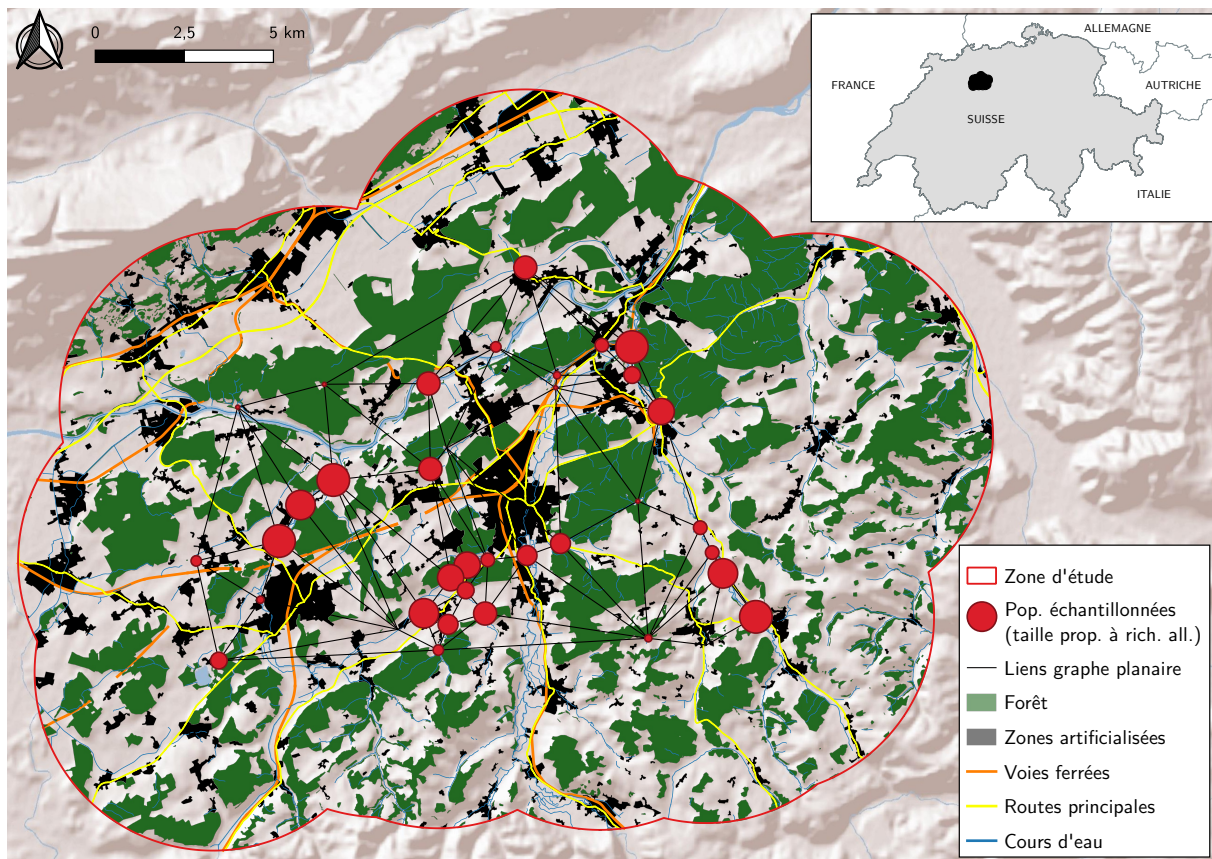


FIGURE 46 – Localisation des populations de Criquet ensanglanté (*Stethophyma grossum*) échantillonnées aux alentours de la ville de Langental dans la région de l’Oberaargau en Suisse.

basée sur un principe de raréfaction pour tenir compte des nombres différents d’individus échantillonnés dans chaque population. La richesse allélique privée indique le nombre d’allèles présents dans une population et nulle part ailleurs. Il mesure donc à la fois la diversité génétique de cette population et sa différenciation génétique par rapport aux autres.

Pour évaluer la différenciation génétique, nous avons calculé la matrice des valeurs de D_{PS} entre populations (Bowcock *et al.*, 1994) car cette distance génétique est censée répondre rapidement aux changements paysagers, la rendant utile pour estimer le flux génétique contemporain (Murphy *et al.*, 2010b). Nous avons également calculé la matrice de F_{ST} (Weir et Cockerham, 1984), qui est elle censée refléter le flux génétique historique (Latta, 2006 ; Murphy *et al.*, 2010b).

Nous avons ensuite construit des graphes génétiques dont les nœuds correspondaient aux populations de criquets. Leurs liens étaient pondérés avec les valeurs de D_{PS} ou de F_{ST} . Nous avons construit des graphes complets mais aussi des graphes élagués. Afin d’éviter d’obtenir des corrélations artefactuelles entre les métriques issues des graphes génétiques et paysagers, nous avons utilisé une méthode d’élagage ne tenant compte que des distances génétiques. Elle consistait à identifier la distance génétique associée au lien dont la suppression entraînait la déconnexion du graphe ("seuil de percolation") et à l’utiliser comme distance seuil pour l’élagage du graphe. Keller *et al.* (2013) ont montré que le flux génétique était fréquent mais limité spatialement dans cette zone. Nous avons donc considéré qu’au-delà de cette distance seuil, les estimations de différenciation génétique reflétaient peu l’effet du paysage sur le flux génétique. À partir des graphes génétiques, nous avons calculé la moyenne de l’inverse des

ponds des liens connectés à chaque nœud, notée $MIW_{g.comp. D_{PS}}$, $MIW_{g.élag. D_{PS}}$, $MIW_{g.comp. F_{ST}}$ ou $MIW_{g.élag. F_{ST}}$ selon l'élagage et la distance génétique utilisés.

2.3 Estimation de la quantité d'habitat atteignable (ARH)

Nous avons utilisé des données d'occupation du sol rasterisées à une résolution de 10 m et datant de l'année de l'échantillonnage en considérant une zone tampon de 5 km autour de chaque site (Figure 46). Dans cette zone, nous avons décrit le patron spatial des taches d'habitat avec plusieurs métriques d'ARH issues de graphes paysagers (Figure 47).

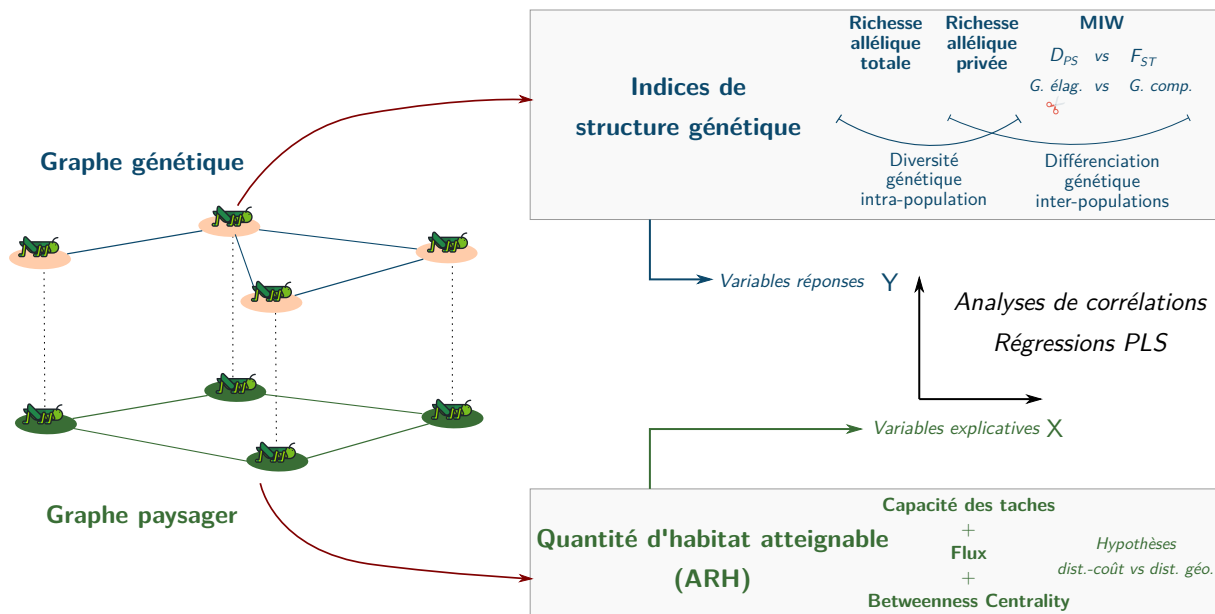


FIGURE 47 – Indices génétiques et métriques d'ARH calculés dans les graphes génétiques et paysagers pour décrire à la fois la structure génétique (diversité, différenciation) et la quantité d'habitat atteignable (ARH) et pour étudier leurs relations via des analyses de corrélation et des régressions PLS (*Partial Least Squares*). "G. élag." : graphe élagué, "G. comp." : graphe complet, "Dist.-coût" : distances-coût, "Dist. géo." : distances géodésiques euclidiennes

2.3.1 Construction des graphes paysagers

Nous avons créé des graphes paysagers à partir d'une surface de résistance incluant six types d'occupation du sol : (i) zones d'habitat potentiel, (ii) forêts, (iii) zones bâties, (iv) zones agricoles, (v) zones humides, cours et plans d'eau et (vi) routes et voies ferrées. Les zones d'habitat potentiel modélisées par Keller *et al.* (2013) incluaient toutes les surfaces agricoles ouvertes, proches des cours et plans d'eau (≤ 500 m) et au niveau desquelles l'eau pouvait s'accumuler. Comme l'échantillonnage était exhaustif, les nœuds des graphes étaient les 37 sites au niveau desquels plusieurs individus avaient été capturés.

Pour calculer les distances-coût pondérant les liens entre ces populations, nous avons distingué cinq scénarios de coût (Table 4). Dans les quatre premiers, nous avons fait varier le coût attribué aux zones humides, cours et plans d'eau (50 ou 1000, noté W50 ou W1000) et celui attribué aux routes et voies ferrées (50 ou 1000, noté R50 ou R1000) car nous voulions tester l'influence de ces potentielles barrières paysagères au flux génétique. Les autres valeurs de coût étaient fixées en considérant que l'espèce se déplace aisément dans ses zones d'habitat ou dans les zones ouvertes, tandis qu'il lui est plus difficile de traverser des zones forestières ou anthropisées (Bönsel et Sonneck, 2011 ; Griffioen, 1996

; Marzelli, 1994). Le cinquième scénario consistait à calculer des distances géodésiques euclidiennes entre populations. Les graphes paysagers avaient pour les cinq scénarios une topologie planaire car l'espèce est censée se disperser via des *stepping-stones* (Figure 46).

TABLE 4 – Scénarios de coût utilisés pour construire les graphes paysagers. Les valeurs de coût étaient attribuées aux pixels selon le type d'occupation du sol auquel ils correspondent pour calculer les distances-coût entre les taches d'habitat.

Scénario	Habitat potentiel	Forêt	Zones artif.	Zones agricoles	Eau, zones humides	Routes, voies ferrées
W50-R50	1	1000	1000	50	50	50
W1000-R50	1	1000	1000	50	1000	50
W50-R1000	1	1000	1000	50	50	1000
W1000-R1000	1	1000	1000	50	1000	1000

2.3.2 Métriques d'ARH

Dans un premier temps, nous avons calculé la capacité des taches (nœuds des graphes) comme la surface d'habitat potentiel atteignable à l'intérieur de la tache d'habitat en elle-même. Aussi, pour chaque nœud, nous avons attribué à chaque pixel d'habitat potentiel un poids qui diminuait avec la distance-coût qui le séparait du site d'échantillonnage auquel correspondait le nœud. Le poids du pixel j situé à une distance-coût d_{ij} du site i était égal à $e^{-\alpha d_{ij}}$, et la capacité $Capa_i$ de la tache i à :

$$Capa_i = \sum_{j=1}^N e^{-\alpha d_{ij}},$$

avec N le nombre total de pixels d'habitat. Nous avons fixé la valeur α de sorte que $p = e^{-\alpha d_{ij}} = 0.05$ pour une distance d_{ij} égale à 1500 m. Nous avons converti cette distance géodésique en distance-coût à l'aide d'une régression de type log-log, d'après Tournant *et al.* (2013). Cette distance donnait les meilleurs résultats par la suite et est du même ordre de grandeur que la distance de dispersion maximale de l'espèce. Étant donné que le Criquet ensanglanté occupe des taches d'habitat discrètes et de superficie réduite, cette métrique reflète la quantité d'habitat atteignable par les individus (ARH) à l'échelle de la tache d'habitat occupée par leur population. On peut donc s'attendre à ce qu'elle soit une bonne estimation de la taille efficace de la population et reflète l'intensité de la dérive génétique (DiLeo et Wagner, 2016). Nous l'avons calculée pour chacun des scénarios de coût.

Pour une tache donnée, les deux métriques F et BC reflétaient différemment la quantité d'habitat atteignable (ARH) de par la présence d'autres taches d'habitat dans le réseau :

— La métrique Flux (F) était calculée avec la formule suivante :

$$F_i = \sum_{j=1; j \neq i}^n Capa_j^\beta e^{-\alpha d_{ij}}$$

avec i l'indice de la tache considérée et j celui de chacune des n autres taches, d_{ij} la distance entre les taches i et j , $Capa_j$ la capacité de la tache j . β indique si la capacité de la tache est considérée dans le calcul ($\beta = 1$) ou non ($\beta = 0$). Lorsque $\beta = 0$, la métrique F est essentiellement topologique et reflète l'influence du nombre et de la proximité des taches atteignables. La valeur α était calculée en faisant plusieurs hypothèses sur les capacités de dispersion de l'espèce. Aussi,

nous avons utilisé des valeurs de α telles que $p = e^{-\alpha d_{ij}} = 0.05$ pour des distances d_{ij} comprises entre 1500 et 7500 m (avec un pas de 500 m), considérant ainsi l'ARH à une échelle plus large que celle à laquelle nous avons calculé la capacité des taches. Par souci de concision, nous ferons référence à ces distances comme des distances maximales de dispersion (MDD) et les exprimerons en unités métriques équivalentes lorsqu'il s'agit de distances-coût.

- La métrique de *Betweenness Centrality* (BC) représente le nombre de fois qu'une tache donnée (c'est-à-dire un nœud, une population) se trouve sur le chemin de moindre coût indirect entre deux autres taches du graphe lorsqu'on considère l'intégralité des paires de taches n'incluant pas la tache focale en elle-même. Chaque fois que cette tache se trouve sur un tel chemin, le terme de la somme constituant cette métrique est pondéré par le produit des capacités des taches impliquées (si $\beta = 1$) et par la probabilité qu'un épisode de dispersion ait lieu entre elles. La formule utilisée était donc la suivante :

$$BC_i = \sum_j \sum_k Capa_j^\beta Capa_k^\beta e^{-\alpha d_{jk}}$$

$$j, k \in \{1, \dots, n\}, k < j, i \in P_{jk}$$

avec P_{jk} l'ensemble des taches localisées sur le chemin de moindre coût entre les taches j et k . Nous avons utilisé les mêmes valeurs α et β que celles utilisées pour le calcul de F.

Comme le calcul de ces trois métriques complémentaires (capacité, F, BC) permet de couvrir une large gamme de paramètres de calcul, d'autres métriques d'habitat déjà utilisées dans la littérature (Capurucho *et al.*, 2013 ; Peterman *et al.*, 2015 ; Taylor et Hoffman, 2014) sont des cas particuliers de calcul de ces métriques. Afin d'évaluer l'intérêt des graphes paysagers en tant que cadre flexible et unifié dans lequel des métriques d'ARH complémentaires peuvent être calculées, nous avons calculé d'autres métriques d'habitat pour pouvoir établir une comparaison. Nous avons tout d'abord calculé la quantité d'habitat potentiel dans un voisinage circulaire autour de chaque population, en considérant des échelles similaires à celles utilisées pour le calcul des métriques d'ARH. Lorsque le rayon considéré était compris entre 100 et 500 m (avec un pas de 100 m), ces métriques étaient relativement similaires à la capacité des taches tandis que pour un rayon plus large (de 1000 à 5000 m avec un pas de 500 m), elles reflétaient davantage le calcul de F. Nous les noterons respectivement métriques de type "buffer local" ou "buffer large" car le terme "buffer" est souvent utilisé dans la littérature (Kindlmann et Burel, 2008 ; Moilanen et Nieminen, 2002) pour désigner de telles métriques calculées dans une zone tampon (*buffer*) autour d'un site. Nous avons également calculé cette quantité d'habitat dans un voisinage non-circulaire autour des populations en considérant des rayons en distances-coût, selon chacun des scénarios de coût testés. Enfin, nous avons calculé la distance entre chaque population et la population la plus proche, selon chaque scénario de coût.

2.4 Analyses de la relation entre les métriques d'ARH et les indices de structure génétique

2.4.1 Analyses de corrélation

Premièrement, nous avons évalué la corrélation entre les métriques d'habitat et les indices génétiques (Figure 47). Comme toutes les variables n'avaient pas une distribution normale, nous avons

calculé le coefficient de corrélation de Spearman basé sur les rangs des valeurs. Nous avons testé sa significativité en ajustant les p -valeurs à l'aide de la méthode de [Benjamini et Hochberg \(1995\)](#) pour contrôler le taux d'erreur de type I.

2.4.2 Régressions Partial Least Squares (PLS)

Les analyses de corrélation permettaient d'identifier les métriques d'habitat, les échelles spatiales et les scénarios de coût les plus fortement liés à chacun des indices génétiques. Toutefois, elles ne permettaient pas de comprendre précisément les relations complexes existant entre les indices génétiques et l'ensemble des métriques d'ARH considérées. Nous avons donc mis en œuvre des régressions PLS (*Partial Least Squares*) dans lesquelles les indices génétiques étaient les variables réponses tandis que les métriques d'ARH étaient les variables explicatives (Figure 47). Ce type de régression est particulièrement adapté aux cas dans lesquels les variables explicatives sont colinéaires et constitue alors une bonne alternative à la régression linéaire multiple et à la régression en composantes principales ([Roy et al., 2015](#) ; [Wold et al., 2001](#)). La régression PLS consiste à créer un espace factoriel tout en maximisant la part de variance des variables réponses expliquée par des composantes factorielles créées à partir des variables explicatives ([Long, 2013](#)) (cf. Annexe B2). Les variables réponses ont été transformées en considérant les rangs des valeurs car elles n'étaient pas distribuées normalement. Pour évaluer l'intérêt de la complémentarité des métriques d'ARH, nous avons testé toutes les combinaisons de variables explicatives impliquant une métrique de capacité des taches, une métrique F et une métrique BC.

Nous avons calculé l'indice Q^2 pour évaluer à quel point chaque composante améliorait la prédiction de la variable réponse lorsqu'on réalisait une validation croisée de type *Leave-One-Out Cross-Validation*. Nous ne décrivons que les modèles dans lesquels au moins une composante améliorait cette prédiction de façon significative. Dans ce cas, la valeur Q^2 associée à au moins une composante était supérieure à 0.0975 (cf. Annexe B2). Nous avons également évalué si chaque variable explicative contribuait de façon significative à la création des composantes factorielles.

3 Résultats

3.1 Corrélations entre métriques d'ARH et indices génétiques

La distance à la population la plus proche n'était jamais corrélée de façon significative à un indice génétique (Table 5). Même si la quantité d'habitat dans un voisinage local autour des populations était corrélée positivement aux indices génétiques, cette corrélation n'était jamais significative. À une échelle plus large, la quantité d'habitat autour des populations n'était corrélée significativement qu'à la richesse allélique et en considérant le scénario de coût attribuant un faible coût aux zones d'eau. Globalement, ces métriques d'habitat particulières étaient moins corrélées aux indices génétiques que ne l'étaient les métriques d'ARH issues des graphes paysagers.

La richesse allélique était corrélée positivement à la capacité des taches ($r = 0.447$). Cette corrélation n'était significative qu'en considérant le scénario W50-R1000 attribuant un coût élevé aux routes et voies ferrées et relativement faible aux zones d'eau. Ainsi, la diversité génétique d'une population était plus forte lorsque cette population occupait une tache au niveau de laquelle une surface d'ha-

TABLE 5 – Coefficients de corrélation de Spearman entre indices génétiques et métriques d’ARH selon le scénario de coût utilisé, la MDD considérée et le poids accordé aux capacités des taches dans le calcul des métriques (valeur de β). Le coefficient de corrélation le plus élevé obtenu pour chaque indice génétique, métrique d’habitat et valeur de β est affiché. La colonne "Signif." indique si la corrélation est significative après l’ajustement des p -valeurs (* : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$). Se référer à la Table 4 pour la description des scénarios de coût. 'Dist. NN' signifie 'Distance au voisin le plus proche' (NN : *Nearest Neighbour*). $MIW_{g,comp,DPS}$ et $MIW_{g,élag,DPS}$ font référence aux indices MIW-DPS calculés à partir de graphes génétiques complets et élagués respectivement (notation similaire pour $MIW-F_{ST}$).

Indice génét.	Métrique d’habitat		Corrélation			
	Métrique	Sc. coût	MDD	β	$r_{Spearman}$	Signif.
Rich. all.	Capacité	w50-R1000	1500	\emptyset	0.447	*
Rich. all.	Buffer local	Euc.	200	\emptyset	0.347	
Rich. all.	Dist. NN	w50-R1000	\emptyset	\emptyset	-0.057	
Rich. all.	F	w50-R1000	1500	0	0.538	*
Rich. all.	F	w50-R1000	1500	1	0.566	*
Rich. all.	Buffer large	w50-R1000	1000	\emptyset	0.482	*
Rich. all.	BC	w1000-R1000	7500	0	0.382	
Rich. all.	BC	w1000-R1000	7500	1	0.387	
Rich. all. priv.	Capacité	w50-R1000	1500	\emptyset	0.238	
Rich. all. priv.	Buffer local	w50-R1000	300	\emptyset	0.289	
Rich. all. priv.	Dist. NN	w50-R50	\emptyset	\emptyset	-0.181	
Rich. all. priv.	F	Euc.	5500	0	-0.609	**
Rich. all. priv.	F	Euc.	2500	1	-0.593	**
Rich. all. priv.	Buffer large	w50-R1000	2000	\emptyset	0.391	
Rich. all. priv.	BC	Euc.	3000	0	-0.437	*
Rich. all. priv.	BC	Euc.	1500	1	-0.387	
$MIW_{g,comp,DPS}$	Capacité	w50-R1000	1500	\emptyset	0.203	
$MIW_{g,comp,DPS}$	Buffer local	w50-R50	500	\emptyset	0.274	
$MIW_{g,comp,DPS}$	Dist. NN	w50-R50	\emptyset	\emptyset	0.126	
$MIW_{g,comp,DPS}$	F	Euc.	3000	0	0.500	*
$MIW_{g,comp,DPS}$	F	Euc.	3500	1	0.467	*
$MIW_{g,comp,DPS}$	Buffer large	Euc.	5000	\emptyset	-0.348	
$MIW_{g,comp,DPS}$	BC	Euc.	1500	0	0.380	
$MIW_{g,comp,DPS}$	BC	Euc.	1500	1	0.335	
$MIW_{g,élag,DPS}$	Capacité	w50-R1000	1500	\emptyset	0.176	
$MIW_{g,élag,DPS}$	Buffer local	w50-R50	500	\emptyset	0.279	
$MIW_{g,élag,DPS}$	Dist. NN	w50-R50	\emptyset	\emptyset	0.113	
$MIW_{g,élag,DPS}$	F	Euc.	3000	0	0.632	**
$MIW_{g,élag,DPS}$	F	Euc.	3500	1	0.597	**
$MIW_{g,élag,DPS}$	Buffer large	Euc.	5000	\emptyset	-0.356	
$MIW_{g,élag,DPS}$	BC	Euc.	1500	0	0.453	*
$MIW_{g,élag,DPS}$	BC	Euc.	1500	1	0.411	
$MIW_{g,comp,FST}$	Capacité	w50-R1000	1500	\emptyset	0.329	
$MIW_{g,comp,FST}$	Buffer local	w50-R50	500	\emptyset	0.303	
$MIW_{g,comp,FST}$	Dist. NN	Euc.	\emptyset	\emptyset	-0.053	
$MIW_{g,comp,FST}$	F	w50-R50	3000	0	0.663	**
$MIW_{g,comp,FST}$	F	w50-R50	1500	1	0.602	**
$MIW_{g,comp,FST}$	Buffer large	w50-R1000	1000	\emptyset	0.370	
$MIW_{g,comp,FST}$	BC	w50-R1000	7000	0	0.472	*
$MIW_{g,comp,FST}$	BC	w1000-R1000	1500	1	0.441	*
$MIW_{g,élag,FST}$	Capacité	w50-R1000	1500	\emptyset	0.327	
$MIW_{g,élag,FST}$	Buffer local	w50-R50	500	\emptyset	0.310	
$MIW_{g,élag,FST}$	Dist. NN	w1000-R1000	\emptyset	\emptyset	0.059	
$MIW_{g,élag,FST}$	F	w50-R50	2000	0	0.686	**
$MIW_{g,élag,FST}$	F	w50-R50	1500	1	0.624	**
$MIW_{g,élag,FST}$	Buffer large	w50-R1000	1000	\emptyset	0.358	
$MIW_{g,élag,FST}$	BC	w50-R1000	7000	0	0.507	*
$MIW_{g,élag,FST}$	BC	w1000-R1000	1500	1	0.454	*

bitat importante était atteignable sans avoir à traverser de routes ou voies ferrées. Au contraire, la capacité des taches n'était corrélée significativement ni à la richesse allélique privée ni aux indices MIW.

Tous les indices génétiques étaient plus fortement corrélés au F qu'au BC (Table 5). La richesse allélique et la richesse allélique privée étaient corrélées respectivement positivement et négativement à ces deux métriques. Alors que la richesse allélique était plus fortement corrélée au F calculé en considérant des distances-coût, en particulier selon le scénario W50-R1000 ($r = 0.538$ ou $r = 0.566$ pour MDD = 1500 m et $\beta = 0$ ou $\beta = 1$ respectivement, Table 5), la richesse allélique privée était plus fortement corrélée à cette métrique lorsqu'elle était calculée en considérant des distances géodésiques euclidiennes ($r = -0.609$ pour MDD = 5500 m et $\beta = 0$). Dans l'ensemble, la MDD avait peu d'influence sur les corrélations, tout comme l'inclusion des capacités dans le calcul des métriques.

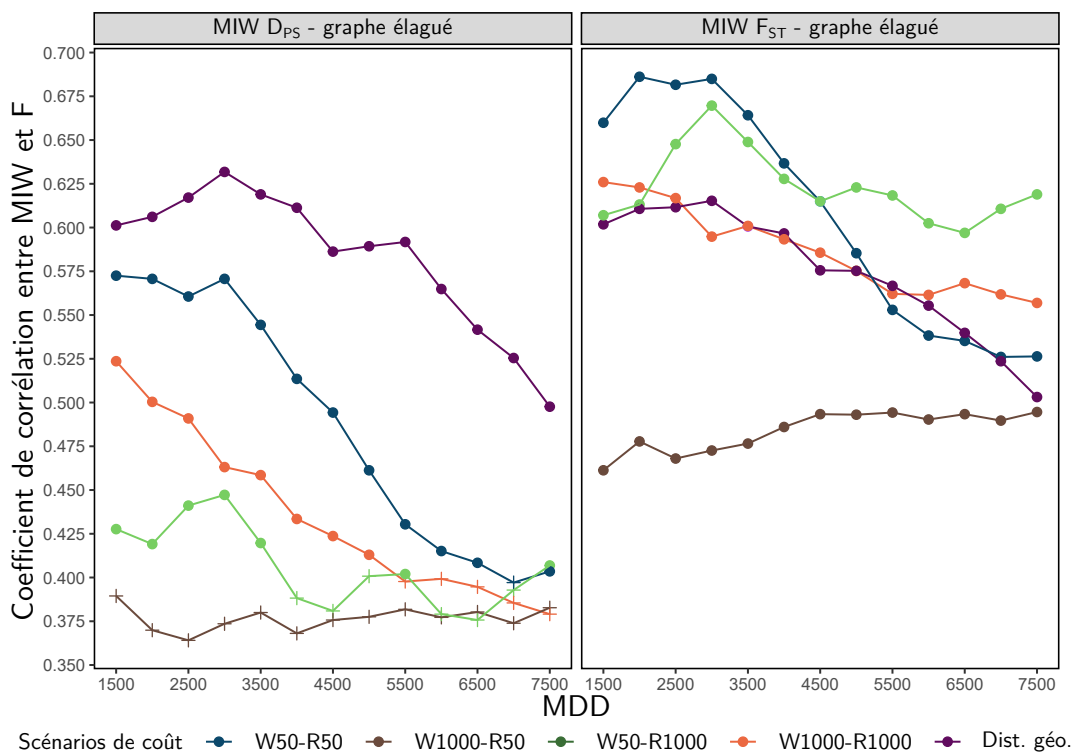


FIGURE 48 – Variations du coefficient de corrélation de Spearman entre les indices MIW calculés à partir des graphes génétiques et la métrique F, en fonction du type de distance génétique, du scénario de coût et des distances de dispersion considérées. L'axe des abscisses indique la MDD (distance maximale de dispersion) à laquelle la probabilité de dispersion est fixée à 0.05 pour paramétrer le calcul des métriques. Cette figure présente les résultats obtenus avec les métriques F calculées sans pondérer les taches par leurs capacités ($\beta = 0$). Les couleurs des points correspondent aux scénarios de coût utilisés pour calculer les distances-coût (cf. Table 4). Les panneaux de gauche et droite représentent les variations observées en calculant l'indice MIW à partir d'un graphe génétique aux liens pondérés par des valeurs de D_{PS} et de F_{ST} , respectivement. Les croix indiquent que la corrélation n'est pas significative après ajustement des p -valeurs.

Les indices MIW étaient positivement et le plus souvent significativement corrélés aux métriques F et BC (Table 5). Les populations localisées dans des taches d'habitat proches d'un nombre important de taches d'habitat de grande capacité avaient donc tendance à être plus similaires aux autres d'un point de vue génétique que des populations plus isolées (Figure 48). Ces corrélations étaient plus fortes lorsque les indices MIW étaient issus de graphes génétiques élagués (Table 5), en particulier si leurs liens étaient pondérés par le D_{PS} . En outre, alors que les indices MIW- D_{PS} étaient plus fortement corrélés aux métriques F calculées à partir de distances géodésiques euclidiennes, les indices MIW- F_{ST}

l'étaient davantage en calculant cette métrique à partir de distances-coût (selon les scénarios W50-R50 et W50-R1000 en particulier, Figure 48).

3.2 Régressions Partial Least Squares (PLS)

Dans quasiment toutes les régressions PLS intégrant une combinaison des métriques de capacité, F et BC, seule une composante factorielle expliquait significativement les indices génétiques. Les meilleurs modèles étaient obtenus lorsque les capacités des tâches n'étaient pas prises en compte dans le calcul de F et, à de rares exceptions près, prises en compte dans le calcul de BC. Néanmoins, ces différences étaient subtiles (Table 6). De plus, pour un indice génétique donné, les modèles étaient similaires donc nous ne décrivons que les meilleurs modèles obtenus pour chacun d'entre eux (Table 6).

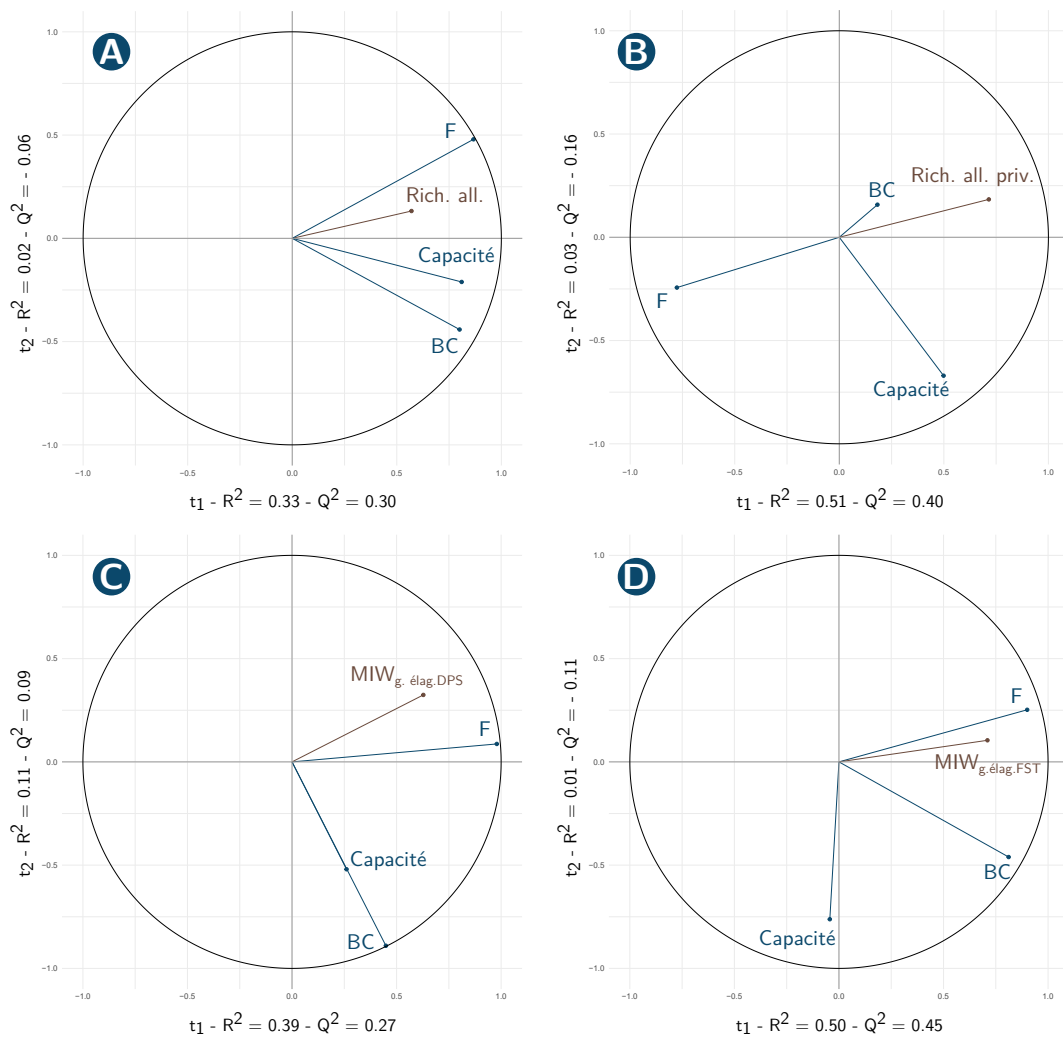


FIGURE 49 – Projection dans le premier plan factoriel des meilleurs modèles de régression PLS pour chaque indice génétique (sur la base des valeurs de Q^2) de la variable réponse (indice génétique) et des variables explicatives (métriques d'ARH : capacité, F, BC). (A) Variable réponse : richesse allélique. Variables explicatives : capacité calculée selon le scénario de coût w50-R1000, F calculé selon le scénario de coût w50-R1000 avec une MDD de 2000 m et $\beta = 0$, BC calculé selon le scénario de coût w1000-R1000 avec une MDD de 7500 m et $\beta = 1$. (B) Variable réponse : richesse allélique privée. Variables explicatives : capacité calculée selon le scénario de coût w50-R1000, F calculé selon le scénario de distance géodésique euclidienne avec une MDD de 2500 m et $\beta = 0$, BC calculé selon le scénario de coût w1000-R1000 avec une MDD de 6000 m et $\beta = 1$. (C) Variable réponse : $MIW_{g, \text{élag. DPS}}$. Variables explicatives : capacité calculée selon le scénario de coût w50-R1000, F calculé selon le scénario de distance géodésique euclidienne avec une MDD de 3000 m et $\beta = 0$, BC calculé selon le scénario de coût w1000-R1000 avec une MDD de 7500 m et $\beta = 1$. (D) Variable réponse : $MIW_{g, \text{élag. FST}}$. Variables explicatives : capacité calculée selon le scénario de coût w1000-R50, F calculé selon le scénario de coût w50-R1000 avec une MDD de 2500 m et $\beta = 0$, BC calculé selon le scénario de distance géodésique euclidienne avec une MDD de 3000 m et $\beta = 1$.

TABLE 6 – Résultats des régressions PLS des indices génétiques en fonction des trois métriques d'ARH (capacité, F, BC). Pour chaque indice génétique et paramètre de pondération des taches par la capacité lors du calcul de F et BC (valeurs β), les caractéristiques du meilleur modèle d'après la valeur de Q^2 associée à la première composante de la régression

PLS sont affichées (la ligne correspondant à la valeur maximale de Q^2 pour chaque indice génétique est notée en italique). Quand β est égal à 1, les capacités des taches sont incluses dans le calcul des métriques, et pas sinon ($\beta = 0$). La colonne "MDD" indique la distance à laquelle la probabilité de dispersion est fixée à 0.05 pour paramétrer le calcul des métriques. Se référer à la Table 4 pour le détail des scénarios de coût. La colonne "r.t1" indique le coefficient de corrélation de Pearson entre les premières composantes PLS (t1) et les métriques d'ARH. Ces valeurs sont notées en gras lorsque les métriques contribuent de façon significative à la construction des composantes PLS. Les valeurs R^2 et Q^2 indiquent respectivement la part de variance de la variable réponse expliquée par chaque composante PLS et cette même part calculée suite à une validation croisée. Les valeurs Q^2 supérieures à 0.0975 indiquent que la composante PLS a un effet significatif sur la variable réponse et sont notées en gras. $MIW_{g,comp,DPS}$ et $MIW_{g,élag,DPS}$ font référence aux indices $MIW-D_{PS}$ calculés à partir de graphes génétiques complets et élagués respectivement (notation similaire pour $MIW-F_{ST}$).

Indice génét.	Capacité			F			BC			Ajustement modèles			
	Sc. coût	r.t1	β	MDD	Sc. coût	r.t1	β	MDD	Sc. coût	r.t1	$Q^2.t1$	$R^2.t1$	$Q^2.t2$
Rich. all.	w50-R1000	0.848	0	2000	w50-R1000	0.885	0	1500	w1000-R50	0.331	0.273	0.326	-0.035
<i>Rich. all.</i>	<i>w50-R1000</i>	<i>0.829</i>	0	<i>2000</i>	<i>w50-R1000</i>	<i>0.860</i>	1	<i>7500</i>	<i>w1000-R1000</i>	<i>0.787</i>	<i>0.280</i>	<i>0.325</i>	<i>-0.098</i>
Rich. all.	w50-R1000	0.847	1	2000	w50-R1000	0.916	0	5500	w1000-R1000	0.870	0.214	0.246	-0.095
Rich. all.	w50-R1000	0.860	1	2000	w50-R1000	0.880	1	7500	w1000-R1000	0.807	0.230	0.270	-0.136
Rich. all. priv.	w50-R1000	0.201	0	1500	Euclid.	-0.947	0	5500	w1000-R1000	-0.396	0.317	0.443	-0.003
<i>Rich. all. priv.</i>	<i>w50-R1000</i>	<i>0.427</i>	0	<i>2500</i>	<i>Euclid.</i>	<i>-0.813</i>	1	<i>6000</i>	<i>w1000-R1000</i>	<i>0.160</i>	<i>0.411</i>	<i>0.511</i>	<i>-0.147</i>
Rich. all. priv.	w50-R1000	0.221	1	2000	Euclid.	-0.943	0	7000	w1000-R1000	-0.385	0.309	0.437	-0.006
Rich. all. priv.	w50-R1000	0.448	1	2500	Euclid.	-0.790	1	7000	w1000-R1000	0.182	0.410	0.508	-0.141
$MIW_{g,comp,DPS}$	W1000-R50	-0.216	0	4000	Euclid.	0.952	0	2500	Euclid.	0.880	0.148	0.217	-0.076
<i>$MIW_{g,comp,DPS}$</i>	<i>W50-R1000</i>	<i>0.390</i>	0	<i>4500</i>	<i>Euclid.</i>	<i>0.929</i>	1	<i>6500</i>	<i>w1000-R1000</i>	<i>0.492</i>	<i>0.178</i>	<i>0.291</i>	<i>0.063</i>
$MIW_{g,comp,DPS}$	W1000-R50	-0.178	1	2500	Euclid.	0.927	0	3500	Euclid.	0.905	0.110	0.182	-0.077
$MIW_{g,comp,DPS}$	W50-R1000	0.491	1	3500	Euclid.	0.904	1	5500	w1000-R1000	0.542	0.098	0.211	0.025
$MIW_{g,élag,DPS}$	W1000-R50	-0.131	0	3000	Euclid.	0.958	0	2000	Euclid.	0.901	0.312	0.366	-0.081
<i>$MIW_{g,élag,DPS}$</i>	<i>W50-R1000</i>	<i>0.284</i>	0	<i>3500</i>	<i>Euclid.</i>	<i>0.974</i>	1	<i>7500</i>	<i>w1000-R1000</i>	<i>0.477</i>	<i>0.314</i>	<i>0.394</i>	<i>0.101</i>
$MIW_{g,élag,DPS}$	W1000-R50	-0.093	1	3000	Euclid.	0.938	0	1500	Euclid.	0.899	0.270	0.325	-0.077
$MIW_{g,élag,DPS}$	W1000-R50	-0.084	1	3500	Euclid.	0.947	1	1500	Euclid.	0.891	0.241	0.303	-0.065
<i>$MIW_{com,g,FST}$</i>	<i>W1000-R50</i>	<i>0.117</i>	0	<i>2000</i>	<i>w50-R1000</i>	<i>0.969</i>	0	<i>3500</i>	<i>Euclid.</i>	<i>0.618</i>	<i>0.381</i>	<i>0.445</i>	<i>-0.115</i>
$MIW_{com,g,FST}$	W50-R1000	0.387	0	3500	Euclid.	0.942	1	2500	w1000-R50	0.091	0.355	0.448	-0.070
$MIW_{com,g,FST}$	W1000-R50	0.228	1	2500	w50-R1000	0.974	0	3000	Euclid.	0.615	0.324	0.393	-0.091
$MIW_{com,g,FST}$	W1000-R50	0.234	1	2500	w50-R1000	0.969	1	2000	Euclid.	0.701	0.295	0.365	-0.065
<i>$MIW_{g,élag,FST}$</i>	<i>W1000-R50</i>	<i>0.147</i>	0	<i>2500</i>	<i>w50-R1000</i>	<i>0.972</i>	0	<i>3500</i>	<i>Euclid.</i>	<i>0.643</i>	<i>0.443</i>	<i>0.504</i>	<i>-0.103</i>
$MIW_{g,élag,FST}$	W1000-R50	0.164	0	2500	w50-R1000	0.965	1	3000	Euclid.	0.714	0.408	0.474	-0.076
$MIW_{g,élag,FST}$	W1000-R50	0.226	1	2500	w50-R1000	0.970	0	3500	Euclid.	0.633	0.368	0.436	-0.083
$MIW_{g,élag,FST}$	W1000-R50	0.238	1	2500	w50-R1000	0.965	1	2500	Euclid.	0.717	0.327	0.402	-0.062

La richesse allélique était le mieux expliquée par un modèle de régression PLS incluant les trois métriques suivantes : la capacité calculée selon le scénario de coût W50-R1000, la métrique F calculée selon ce même scénario avec $\beta = 0$ et une MDD de 2000 m, et la métrique BC calculée selon le scénario W1000-R1000 avec une MDD de 7500 m et $\beta = 1$. Ces variables étaient fortement et positivement corrélées à la première composante factorielle dont l'indice Q^2 valait 0.280 (Table 6 et Figure 49A).

En revanche, la seule variable contribuant significativement à la première composante issue du meilleur modèle expliquant la richesse allélique privée était la métrique F calculée à partir de distances géodésiques euclidiennes, avec $\beta = 0$ et une MDD de 1500 ou 2000 m (Table 6). Cette métrique était corrélée négativement à la première composante indiquant que la richesse allélique privée est plus faible lorsque les taches d'habitat sont entourées par de nombreuses et grandes autres taches (Figure 49B). La valeur Q^2 associée à la première composante valait dans ce cas 0.415, une valeur supérieure à celle du cas précédent.

Les valeurs de MIW issues de graphes génétiques élagués étaient mieux modélisées lors des régressions PLS que celles issues de graphes complets (Table 6). L'ajustement des modèles était meilleur pour les indices MIW-F_{ST} que MIW-D_{PS}. Les meilleures valeurs de Q^2 étaient comprises dans ces cas-là entre 0.314 et 0.443. La métrique F était là encore la seule qui contribuait de façon significative à la première composante (Figures 49C et 49D). Alors qu'il valait mieux calculer cette métrique à partir de distances géodésiques euclidiennes pour expliquer le MIW-D_{PS}, il valait mieux le faire à partir de distances-coût selon le scénario W50-R1000 et $\beta = 0$ pour expliquer le MIW-F_{ST}. Dans tous les cas, la métrique F était fortement et positivement corrélée à la première composante, elle-même positivement corrélée au MIW.

4 Discussion

Nous avons comparé des graphes génétiques et paysagers pour évaluer l'influence de la quantité d'habitat atteignable (ARH) sur la structure génétique des populations. Premièrement, les métriques d'ARH issues du cadre unifié et flexible que constituent les graphes paysagers se sont révélées pertinentes pour expliquer les différentes composantes de la structure génétique des populations, c'est-à-dire la diversité et la différenciation génétiques. Elles s'avéraient pour cela meilleures que d'autres métriques plus communément utilisées dans la littérature. Deuxièmement, alors que la richesse allélique était expliquée de façon significative par les trois métriques d'ARH complémentaires dans les régressions PLS, la richesse allélique privée et les indices MIW étaient essentiellement liés aux métriques d'ARH considérant des surfaces d'habitat situées en dehors de la tache focale. Finalement, en considérant plusieurs scénarios de coût, nous avons mis en évidence le fait que la diversité génétique semblait être influencée négativement par les routes et voies ferrées et positivement par les surfaces d'eau. En revanche, ces éléments paysagers n'influençaient pas la différenciation génétique de la même façon selon qu'elle était mesurée à l'aide du D_{PS} ou du F_{ST}.

4.1 Les métriques d'ARH sont-elles adaptées à l'analyse de la structure génétique ?

Tous les indices décrivant la structure génétique des populations de criquets étaient significativement corrélés à, et expliqués de façon significative dans les régressions PLS par, au moins une des

trois métriques d'ARH. Comparativement, les autres métriques calculées étaient peu liées aux indices génétiques. Nos résultats confirment ainsi l'intérêt de ces métriques issues de graphes paysagers pour caractériser la façon dont le patron spatial des habitats influence les processus de dérive et flux génétique. Ils rappellent également les résultats de [Moilanen et Nieminen \(2002\)](#). En effet, même si ces auteurs étudiaient des réponses biologiques différentes, leurs résultats et les nôtres révèlent pareillement le pouvoir explicatif moindre de métriques telles que la quantité d'habitat dans un voisinage circulaire ou la distance à la tache la plus proche, par rapport à des métriques considérant les probabilités de dispersion entre taches voisines.

Le calcul des métriques d'ARH étant flexible, elles incluent des métriques d'habitat particulières déjà utilisées dans de précédentes études. Par exemple, [Capurucho *et al.* \(2013\)](#) identifiaient la surface des zones d'habitat situées à moins de 15 km de chaque site étudié comme la meilleure variable explicative de la richesse allélique des populations d'une espèce d'oiseau tropicale (voir [Keyghobadi *et al.* \(2005\)](#), [Millette et Keyghobadi \(2015\)](#) ou [Peterman *et al.* \(2015\)](#) pour d'autres exemples). L'utilisation de plusieurs métriques d'ARH complémentaires dans cette étude et d'autres études similaires auraient potentiellement permis d'obtenir des relations statistiques plus fortes et des éléments de compréhension supplémentaires. Aussi, même si d'autres métriques peuvent expliquer la structure génétique, les graphes paysagers offrent un cadre particulièrement adapté à la compréhension de l'influence du patron des habitats sur la structure génétique.

L'inclusion des capacités des taches dans le calcul des métriques F et BC n'influçait que marginalement nos résultats. Cela signifie que le nombre de taches atteignables au sein d'un réseau d'habitat était souvent à lui seul une bonne variable explicative de la structure génétique. Un tel résultat rappelle celui de [Peterman *et al.* \(2015\)](#) montrant que l'isolement relatif d'une tache par rapport aux autres était la meilleure variable explicative de la différenciation génétique moyenne entre la population occupant une tache et les autres. Par conséquent, l'intérêt des métriques d'ARH issues des graphes paysagers pourrait être lié à la prise en compte de la topologie des populations, paramètre important des patrons spatiaux de dispersion et de flux génétique ([Saura *et al.*, 2014](#) ; [Van Strien, 2017](#)).

4.2 Est-ce que l'ARH influence la diversité et la différenciation génétiques de façon aussi importante et à la même échelle spatiale ?

À l'instar d'autres études ([Balkenhol *et al.*, 2013](#) ; [Keyghobadi *et al.*, 2005](#) ; [Kierepka *et al.*, 2020](#) ; [Taylor et Hoffman, 2014](#)), notre étude montre que la différenciation génétique et la diversité génétique locale ne sont pas autant influencées ni à la même échelle spatiale par le patron spatial des habitats. La richesse allélique était corrélée significativement au F et à la capacité des taches, et le seul indice génétique expliqué de façon significative par la capacité des taches dans les régressions PLS. La richesse allélique privée et les indices MIW n'étaient quant à eux liés de façon significative qu'au F et au BC. Ainsi, la diversité génétique locale était influencée par l'ARH mesurée à l'échelle de la tache focale et au-delà de cette tache, alors que la différenciation génétique était influencée par l'ARH mesurée en ne considérant que les surfaces d'habitat situées au-delà de la tache focale. Bien que ces deux composantes soient influencées toutes deux par la dérive et le flux génétique, [DiLeo et Wagner \(2016\)](#) suggéraient qu'une influence plus forte de la quantité d'habitat locale sur la diversité génétique pourrait s'expliquer par le lien direct entre quantité d'habitat et taille de population. En revanche, l'influence qu'a le patron des habitats à une échelle large sur les taux de migration semble davantage affecter la dif-

férenciation génétique que l'effet qu'exerce la surface des habitats sur la dérive (Cushman *et al.*, 2012).

La différenciation génétique relative entre populations était mieux expliquée par les métriques d'ARH lorsqu'elle était estimée à partir de graphes génétiques élagués. L'intérêt de l'élagage des graphes génétiques pour les analyses basées sur des distances avait déjà été suggéré par Wagner et Fortin (2013) et mis en évidence dans le premier chapitre de cette thèse. Ce résultat suggère que l'élagage peut également être pertinent pour le calcul d'indices génétiques à l'échelle des populations (*neighbourhood level*). Il confirme également les résultats obtenus par Keller *et al.* (2013) qui montraient qu'il valait mieux ne considérer que les paires de populations séparées par une distance limitée pour identifier l'influence du paysage sur la différenciation génétique. Ils interprétaient cela comme étant lié au fait que le Criquet ensanglanté est actuellement en voie d'expansion en Suisse suite à plusieurs décennies de déclin. Ainsi, la différenciation génétique n'a probablement pas atteint son équilibre. Dans ce contexte, le patron de différenciation génétique s'apparente au patron d'IBD de type IV de la typologie d'Hutchison et Templeton (1999), ce qui peut expliquer l'intérêt de l'élagage du graphe génétique (cf. Chapitre 1).

4.3 Est-ce que le coût de déplacement à travers la matrice paysagère affecte la diversité et la différenciation génétiques de la même manière ?

La richesse allélique et la différenciation génétique relative estimée à l'aide du F_{ST} étaient davantage liées aux métriques d'ARH considérant des distances-coût. En revanche, il était préférable de considérer des distances géodésiques euclidiennes pour expliquer la richesse allélique privée et la différenciation génétique relative estimée à l'aide du D_{PS} . Ces différences pourraient s'expliquer par (i) les différentes échelles temporelles auxquelles la diversité et la différenciation génétiques répondent aux changements paysagers et (ii) la capacité des indices de différenciation génétique à refléter l'influence du paysage sur le flux génétique historique ou contemporain.

Premièrement, la différenciation génétique atteint théoriquement son état d'équilibre plus rapidement que la diversité génétique (Varvio *et al.*, 1986). Par exemple, Keyghobadi *et al.* (2005) ont détecté un lien positif entre la présence de forêts apparues récemment et la différenciation génétique chez une espèce de papillons se dispersant à travers des zones ouvertes et évitant les forêts, tandis que la diversité génétique locale était mieux expliquée par une métrique d'isolement basée sur des distances géodésiques. Ces dernières reflétaient les moindres contraintes exercées sur la dispersion avant que les forêts n'apparaissent. Ainsi, les résultats que nous avons obtenus pourraient s'interpréter à partir des hypothèses suivantes. Tout d'abord, la relation plus forte entre diversité génétique locale et métriques d'ARH calculées en considérant des distances-coût plutôt que des distances géodésiques pourrait refléter l'influence passée de la matrice sur la dispersion. Ensuite, la relation plus forte entre la richesse allélique privée ou les indices de $MIW-D_{PS}$ et les métriques d'ARH calculées à partir de distances géodésiques pourrait indiquer une influence plus faible de la matrice sur la dispersion contemporaine. Ces hypothèses sont compatibles avec l'expansion actuelle de l'espèce.

Deuxièmement, de précédentes études de génétique du paysage ont montré que le D_{PS} reflète l'influence actuelle du paysage sur la structure génétique alors que le F_{ST} reflète davantage son influence plus ancienne (Holzhauer *et al.*, 2006 ; Murphy *et al.*, 2010b ; Storfer *et al.*, 2010). Cela pourrait expliquer pourquoi les indices de différenciation génétique calculés à l'aide du F_{ST} étaient plus corrélés aux

métriques d'ARH tenant compte de la forte résistance de certains éléments paysagers à la dispersion. Ces explications sont difficilement vérifiables mais signifieraient également que la matrice paysagère est devenue plus perméable aux déplacements de cette espèce ces dernières années, expliquant potentiellement son expansion.

Enfin, [Holzhauer *et al.* \(2006\)](#) ont observé que les routes et voies ferrées étaient des barrières pour le Criquet ensanglanté alors que les zones d'eau ne l'étaient pas. Ici, le scénario selon lequel les routes et voies ferrées avaient un coût faible et les zones d'eau un coût élevé (W1000-R50) n'avait jamais un pouvoir explicatif élevé. Par contre, celui selon lequel ces infrastructures de transport limitaient fortement la dispersion et les zones d'eau étaient relativement perméables (W50-R1000) expliquait bien la richesse allélique et la différenciation génétique historique (F_{ST}). Ce résultat semble contraire à celui de [Keller *et al.* \(2013\)](#) qui montraient une influence positive des routes sur la dispersion chez cette espèce. Néanmoins, ces auteurs n'avaient considéré qu'un indice de différenciation génétique reflétant le flux génétique contemporain. Or, dans nos analyses, les indices de MIW- D_{PS} étaient mieux expliqués par les métriques d'ARH ne tenant pas compte de la résistance de la matrice.

4.4 Limites et perspectives de recherche

La relation entre le patron spatial des habitats et la structure génétique est dynamique et met du temps à atteindre un équilibre ([Slatkin, 1993](#)). Dans ce contexte, la topologie du réseau d'habitat a une forte influence sur la structure génétique ([Van Strien, 2017](#)). En effet, même sous l'hypothèse où seule la quantité d'habitat autour d'un site influence les patrons de diversité ([Fahrig, 2013](#)), la configuration des habitats les affecte également ([Saura, 2021](#)). Ainsi, la modification des propriétés de traversabilité d'un réseau d'habitat influencerait les flux génétiques à longue distance, ce qui influencerait la structure génétique. Cela explique probablement la pertinence des métriques d'ARH pour expliquer la structure génétique. Nous sommes néanmoins conscients que les effets relatifs de l'ARH ici mesurés sur les différentes composantes de la structure génétique pourraient être spécifiques à ce cas d'étude. Bien que nos résultats encouragent l'utilisation future de l'ARH dans des études empiriques, leur utilisation dans des études basées sur des simulations permettrait de comprendre à quel point la taille, la topologie, les capacités de dispersion des populations ainsi que le patron des habitats et la résistance de la matrice influencent le lien entre ARH et structure génétique.

Enfin, nos résultats sont difficilement comparables à ceux obtenus en distinguant les effets de la quantité d'habitat de ceux de la configuration des habitats ([Cushman *et al.*, 2012](#) ; [Jackson et Fahrig, 2015b](#) ; [Millette et Keyghobadi, 2015](#)). La plupart de ces études se basaient sur des analyses mettant des distances en relation, tandis que nous avons ici utilisé une approche basée sur des métriques calculées au niveau des nœuds de graphes. Par ailleurs, dans la plupart des études précédentes, des métriques de configuration des habitats telles que les distances entre taches ou l'isolement des taches étaient fortement corrélées à la quantité d'habitat. Cela aurait dû dissuader d'en tirer des conclusions sur le fait que la configuration des habitats exerce une influence plus forte sur la structure génétique que la quantité d'habitat ([Jackson et Fahrig, 2015b](#)). Nous nous sommes donc focalisés ici sur des métriques d'ARH complémentaires issues de graphes paysagers et nous montrons qu'elles permettent de comprendre comment le patron spatial des habitats influence à la fois la dérive et le flux génétique, à plusieurs échelles spatiales et temporelles et en considérant plusieurs scénarios de coût.

Transition

La comparaison de graphes génétiques et paysagers mise en œuvre dans le chapitre précédent confirme que des métriques de connectivité des habitats issues d'un graphe paysager et basées sur le concept de quantité d'habitat atteignable sont pertinentes pour expliquer les différentes composantes de la structure génétique des populations. Cela semble confirmer l'utilité des graphes paysagers pour identifier les taches d'habitat et les secteurs de la matrice paysagère qui garantissent la connectivité écologique des habitats. Néanmoins, face à la diversité des méthodes de construction des graphes paysagers, les utilisateurs y ayant recours pour guider leurs prises de décision dans un cadre opérationnel manquent d'éléments précis concernant la validité écologique de ces graphes en fonction de leurs modalités de construction. Le chapitre qui suit vise à clarifier cette question méthodologique. Nous y évaluons la validité écologique de graphes paysagers construits à dire d'expert ou à partir de données de présence/absence. Pour cela, nous les confrontons à des données génétiques censées refléter les déplacements des individus et la taille de leurs populations plus directement que les données biologiques habituellement confrontées à ce type de modèles. Les variables génétiques considérées dans les analyses sont notamment issues de graphes génétiques, ou sont des distances génétiques habituellement associées aux liens de ces graphes. Nous nous focalisons dans ce chapitre sur la connectivité écologique des habitats forestiers de la Paruline caféïette, passereau forestier de l'île de Guadeloupe.

Chapitre 4

Validation de graphes paysagers à l'aide de données génétiques

Résumé

La connectivité écologique des habitats est fréquemment modélisée à l'aide de graphes paysagers construits à partir de dires d'expert ou de modèles de distribution d'espèces (SDM). Ces outils ont rarement été validés à partir de données empiriques reflétant plus directement la connectivité fonctionnelle des habitats. Les données génétiques sont censées refléter l'influence de la connectivité fonctionnelle sur les tailles de populations et le déplacement des individus entre elles. Nous avons donc cherché à répondre à la question suivante : est-ce que des graphes paysagers construits à dires d'expert ou de données de présence/absence modélisent correctement les processus écologiques influencés par la connectivité des habitats ? Pour cela, nous avons modélisé le réseau d'habitat de la Paruline caféïette (*Setophaga plumbea*), passereau forestier de Guadeloupe, à l'aide de graphes paysagers construits de trois manières : à partir de dires d'expert, d'indices de spécialisation ou d'un SDM. En parallèle, à partir des données génétiques de 712 oiseaux issus de 27 populations, nous avons calculé des indices de diversité et de différenciation génétiques à l'échelle des populations et des distances génétiques entre populations. Nous avons ensuite étudié la relation entre (i) les distances ou les indices génétiques et (ii) les distances-coûts calculées sur le graphe ou des métriques de connectivité, à l'aide de modèles mixtes de distances de type MLPE ou de corrélations de Spearman entre métriques. Dans l'ensemble, les graphes paysagers reflétaient correctement l'influence de la connectivité des habitats sur la structure génétique des populations. Néanmoins, la relation entre la validité écologique des graphes, leurs besoins en données et la complexité des méthodes mises en œuvre était parfois contre-intuitive. En effet, les graphes construits à partir du SDM, approche la plus complexe, reflétaient moins bien certaines variables génétiques que les autres graphes. Le développement d'une nouvelle méthode de validation croisée et des analyses de sensibilité nous ont permis d'identifier les avantages et les limites de chacun des graphes. En conclusion, les graphes paysagers reflètent la réalité écologique mais nous appelons à une meilleure considération du ratio entre leurs apports et les coûts potentiellement associés à leur construction.

Cet article est en préparation pour une soumission dans la revue *Conservation Biology* en 2021 :

Savary, P.^{*}, Daniel, A.^{*}, Foltête, J. C., Khimoun, A., Faivre, B., Ollivier, A., Moal, H., Éraud, C., Vuidel, G. & Garnier, S. Validating graph-based connectivity models for a forest tropical bird species using independent presence and genetic datasets. In prep. for *Conservation Biology*. (* : contributions égales des auteurs)

*Le territoire nous apparaissait d'abord comme une configuration spatiale,
identifiable parce qu'installée de manière relativement pérenne dans l'espace.
Je me rends compte qu'il n'y a en fait rien de plus mouvementé qu'un territoire,
aussi stables pourraient être ses frontières, aussi fidèle à celui-ci soit son résident.*

*Parce que le territoire n'est pas tant espace que distances,
la territorialisation est l'acte littéral et expressif de "marquer ses distances".
La distance n'est pas une mesure, mais une intensité, un rythme.*

Vinciane Despret

1 Introduction

Grâce à l'amélioration des capacités de calcul et à la plus grande disponibilité des données spatiales et biologiques, la modélisation de la connectivité des habitats a bénéficié de méthodes d'une grande complexité, capables de représenter plus fidèlement la réalité écologique (voir par exemple [Pe'er et al. \(2011\)](#)). Néanmoins, les données et ressources nécessaires à la mise en œuvre de telles approches sont rarement disponibles, renforçant ainsi l'intérêt d'approches plus simples et "pragmatiques" ([Fagan et Calabrese, 2006](#)). Parmi celles-ci, les graphes paysagers ont été présentés comme un compromis optimal pour obtenir un modèle pertinent à partir de méthodes simples et peu exigeantes en données ([Calabrese et Fagan, 2004](#)). Néanmoins, malgré leurs avantages, les graphes paysages ont aussi des limites ([Moilanen, 2011](#)). Le besoin de fixer des valeurs de coût pour modéliser les chemins de dispersion potentiels des espèces étudiées ([Beier et al., 2008](#) ; [Zeller et al., 2012](#)) est une première limite. En effet, l'attribution des coûts est souvent basée sur des dires d'expert mais leur validité écologique est rarement testée ([Sawyer et al., 2011](#)). Par ailleurs, les nœuds des graphes sont définis à partir de données spatiales et sont censés représenter les taches d'habitat de l'espèce étudiée, mais la validité de cette délimitation pose aussi question ([Moilanen, 2011](#)). Enfin, nous avons vu qu'il était possible de calculer une multitude de métriques de connectivité à partir des graphes paysagers ([Baranyi et al., 2011](#) ; [Laita et al., 2011](#) ; [Rayfield et al., 2011](#)), mais la relation entre ces métriques et les réponses biologiques qu'elles sont censées refléter est rarement évaluée ([Moilanen, 2011](#)).

Pour dépasser ces limites, il est possible d'intégrer différents types de données biologiques dans la modélisation. Le couplage des données biologiques et des graphes paysagers a fait l'objet de la synthèse bibliographique présentée en introduction de cette thèse. Dans cette étude, nous avons montré que des données de présence étaient fréquemment utilisées dans un premier temps pour réaliser un modèle de distribution d'espèce (noté SDM par la suite), utilisé pour délimiter les taches d'habitat à partir de seuils de probabilité de présence et pour fixer les valeurs de coût à partir des probabilités de présence ([Clauzel et Godet, 2020](#) ; [DufLOT et al., 2018](#) ; [Tarabon et al., 2019](#)). Néanmoins, bien que basée sur des observations empiriques, la validité d'une telle modélisation de la connectivité est potentiellement limitée car la probabilité de présence d'une espèce dans une zone reflète souvent mal la résistance de cette zone aux mouvements de l'espèce ([Keeley et al., 2017](#)). Récemment, [Bourdouxhe et al. \(2020\)](#) et [Godet et Clauzel \(2021\)](#) ont créé des graphes paysagers à partir de dires d'expert, d'un SDM ou d'une combinaison de ces deux sources d'information afin de comparer les résultats obtenus. Ces auteurs ont décrit de façon détaillée la sensibilité des graphes à la nature des informations utilisées pour les construire mais ne disposaient pas de données biologiques indépendantes reflétant la réponse des espèces à la connectivité fonctionnelle des habitats ; les empêchant ainsi d'identifier le modèle le plus fidèle à la réalité écologique.

La synthèse bibliographique présentée en introduction de cette thèse concluait que la confrontation *a posteriori* de données biologiques n'ayant pas servi à la construction d'un graphe paysager avec des variables issues de ce graphe serait l'approche la plus à même d'évaluer la validité des graphes paysagers. Les données génétiques constituent les données idéales pour cela car elles sont influencées par la connectivité écologique des habitats ([Zeller et al., 2018](#)). Selon [Beier et al. \(2008\)](#), l'utilisation de données génétiques ou de mesures directes de déplacements animaux serait la plus adaptée pour fixer des valeurs de coût. Lorsqu'elles sont disponibles, ces données sont ainsi utilisées pour fixer les coûts, empêchant alors de les confronter *a posteriori* à un modèle indépendant de toute information

génétique. [Beier et al. \(2008\)](#) considéraient par ailleurs que les données de présence étaient moins adaptées au paramétrage des modèles de connectivité, et que les dires d’expert ou les informations issues de la littérature scientifique étaient à utiliser en dernier recours. La validité relative de ces deux dernières sources d’information reste néanmoins à vérifier. Comparer la capacité de différents graphes paysagers à expliquer ou prédire la structure génétique des populations permettrait de le faire. Cela permettrait également d’évaluer dans quelle mesure les graphes paysagers modélisent l’influence du paysage sur la diversité génétique et les flux génétiques, deux objectifs clés des mesures de conservation de la biodiversité ([Hoban et al., 2020](#)).

Dans ce chapitre, nous avons utilisé des données génétiques afin de répondre à la question suivante : est-ce que des graphes paysagers construits à dires d’expert ou à partir de données de présence/absence modélisent correctement les processus écologiques influencés par la connectivité des habitats ? Pour cela, nous avons modélisé le réseau d’habitat de la Paruline caféïette (*Setophaga plumbea*), passereau forestier de Guadeloupe sensible à la connectivité de son habitat ([Khimoun et al., 2016a, 2017](#)).

2 Matériel et méthodes

La méthodologie mise en œuvre pour cette étude est résumée par la Figure 50. Nous avons construit des graphes paysagers selon trois méthodes basées sur (i) des dires d’expert, (ii) des indices de spécialisation pour différents types d’occupation du sol ou (iii) un SDM. Nous avons ensuite calculé les distances-coût le long des liens de ces graphes et différentes métriques de connectivité. En parallèle, les données génétiques acquises dans une trentaine de populations ont permis de calculer des indices génétiques au niveau de chaque population et les distances génétiques entre elles. Nous avons enfin étudié la relation entre les réponses génétiques et les différents graphes paysagers construits afin de comparer leurs validités écologiques. Pour cela, nous nous sommes basés sur deux variables issues des graphes paysagers, à savoir (i) les distances-coût associées aux liens, mises en relation avec les distances génétiques, et (ii) les métriques de connectivité calculées au niveau des nœuds, mises en relation avec les indices génétiques calculés à ce même niveau. Nous faisons l’hypothèse que le graphe issu du SDM refléterait le mieux les réponses génétiques car sa construction nécessite une étape de modélisation supplémentaire incluant des données empiriques et plusieurs types de données spatiales.

2.1 Zone et modèle d’étude

L’île de Guadeloupe (1,713 km²) se situe dans les Petites Antilles, à l’est de la mer des Caraïbes (Figure 51). Basse-Terre (BT) est la partie ouest de l’île sur laquelle un relief marqué et un climat plus humide ont limité l’agriculture et l’urbanisation. Un isthme et un étroit cours d’eau salé sépare Basse-Terre de Grande-Terre (GT), partie est de l’île sur laquelle les surfaces de forêt ont été fortement réduites par le passé. Ainsi, les forêts recouvrent 44 % de la surface de la Guadeloupe mais leur répartition est hétérogène étant donné qu’elles représentent respectivement 60 % et 22 % de la surface de BT et GT aujourd’hui.

La Paruline caféïette (*Setophaga plumbea*, Figure 52) est une espèce d’oiseau forestière endémique des Petites Antilles. Chez cette espèce spécialiste de l’habitat forestier ([Curson, 2014](#) ; [Khimoun et al., 2016b](#) ; [Leblond, 2008](#) ; [Lovette et al., 1998](#)), il a été montré que les flux génétiques dépendent de la

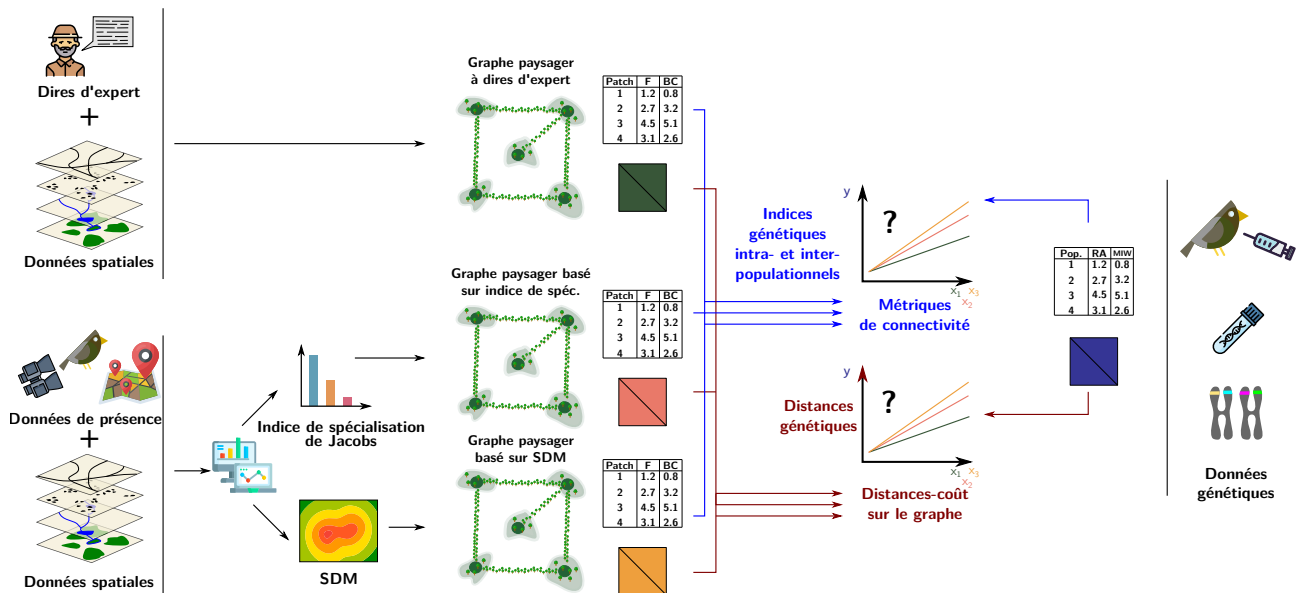


FIGURE 50 – Méthode de validation des graphes paysagers à l'aide de données génétiques

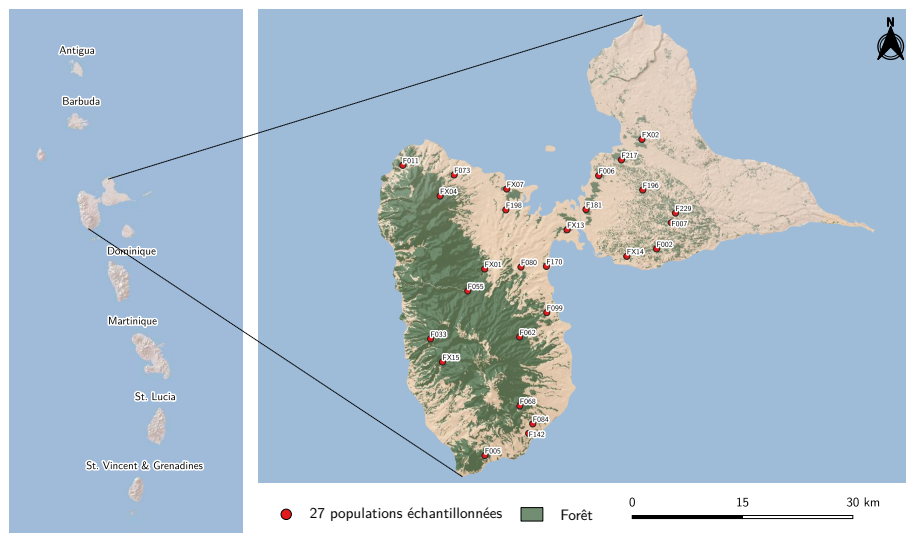


FIGURE 51 – Topographie et distribution spatiale des forêts de l'île de Guadeloupe, située dans les Petites Antilles

connectivité des forêts et de la résistance de la matrice paysagère (Khimoun *et al.*, 2016b, 2017). Cela a pour corollaire que la structure génétique de ses populations reflète la connectivité de l'habitat forestier, faisant de cette espèce un modèle biologique idéal pour valider les modèles de connectivité. Par ailleurs, bien que cette espèce soit affectée par la dégradation des forêts, elle est présente en Guadeloupe à des niveaux d'abondance permettant de détecter et d'échantillonner suffisamment d'individus pour décrire la structure génétique des populations et modéliser leur distribution (Éraud *et al.*, 2012).

2.2 Modélisation de la connectivité écologique à l'aide de graphes paysagers

2.2.1 Données spatiales et climatiques

Pour créer le SDM et les graphes paysagers, nous avons utilisé des données spatiales concernant l'occupation du sol et le micro-climat. Les données d'occupation du sol de la base Corine Land Cover de 2012 ont été complétées par les routes et les éléments bâtis de la BD TOPO de l'IGN. De plus, nous avons utilisé la cartographie de 17 formations végétales ligneuses réalisée en 2010 par le Conseil Dépar-



FIGURE 52 – La Paruline caféïette (*Setophaga plumbea*), espèce de passereau de la famille des Parulidae. Elle est forestière et endémique des Petites Antilles.

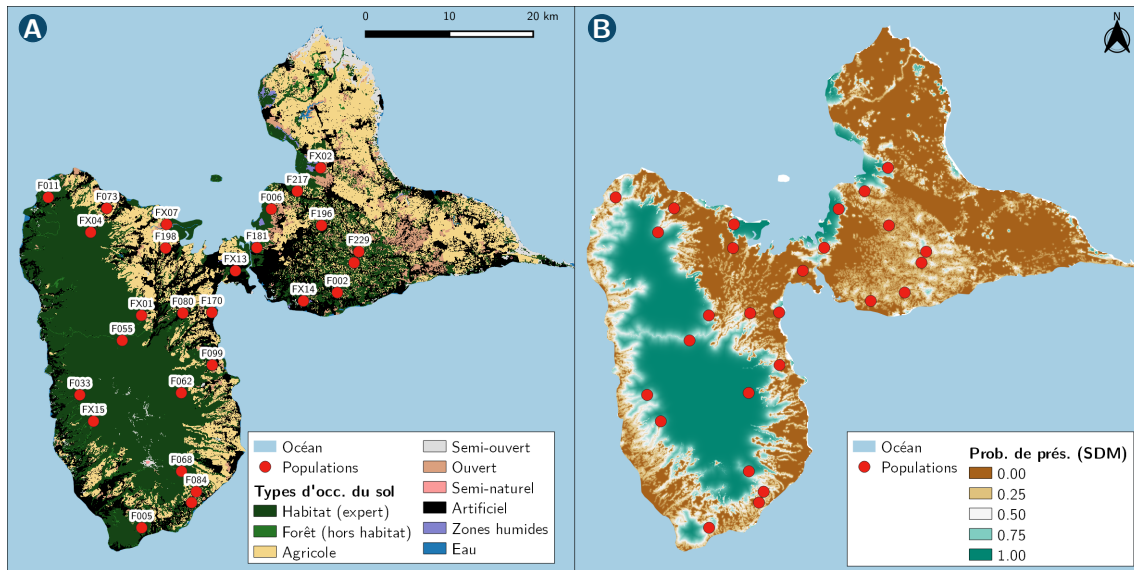


FIGURE 53 – (A) Occupation du sol de la Guadeloupe selon une classification en dix catégories à partir des données du Corine Land Cover, de la BD TOPO de l'IGN et de la carte des formations végétales de l'IGN, l'ONF et le Conseil Départemental de Guadeloupe. Cette carte a servi de base aux graphes construits à dire d'expert ou à partir des indices de spécialisation de Jacobs. (B) Probabilités de présence issues du modèle de distribution de la Paruline caféïette. Ces valeurs ont servi de base à la construction du graphe basé sur ce modèle.

temental de Guadeloupe, l'IGN et l'ONF¹. Des couches rasters indiquant les précipitations annuelles moyennes et la température moyenne à une résolution de 20 m nous ont été transmises. Elles ont été obtenues par interpolation des données de 61 stations Météo-France, suivant la méthode décrite par Castel *et al.* (2017) et Joly *et al.* (2012).

Nous avons créé un raster d'occupation du sol (résolution : 20 m) en classifiant l'occupation du sol en dix catégories² à partir de ces données (Figure 53A). Pour réaliser les modèles de distribution, nous avons également créé des couches rasters indiquant pour chaque pixel la distance au pixel d'habitat forestier (selon dire d'expert, voir section suivante), à la route, à l'élément bâti ou à la parcelle agricole la plus proche, et la proportion de chacun de ces éléments dans un rayon de 500 m autour du pixel.

1. La liste des 17 formations végétales ligneuses figure dans la Table 14 de l'annexe A6.

2. Le détail de cette classification figure dans la Table 13 de l'annexe A6.

2.2.2 Dires d'experts

Pour construire un graphe paysager à dire d'expert, nous avons demandé à un ornithologue local de :

1. Identifier les formations végétales constituant l'habitat de la Paruline caféïette. Cela nous a ensuite permis de délimiter les taches d'habitat correspondant aux nœuds des graphes construits à dire d'expert (Table 7). Ces zones constituaient une catégorie d'occupation du sol par la suite (Figure 53A).
2. Attribuer à chacun des dix types d'occupation du sol une valeur de coût pour le calcul des distances-coût associées aux liens des graphes construits à dire d'expert.

2.2.3 Indices de spécialisation

Pour calculer un indice de spécialisation pour chaque type d'occupation du sol, nous avons extrait de 991 comptages ponctuels de l'avifaune les informations relatives à la présence ou à l'absence de la Paruline caféïette. Le protocole de ces observations, réalisées entre 2009 et 2011 par l'ONCFS, le Parc National de Guadeloupe et des associations locales, est décrit par Éraud *et al.* (2012). Il est à noter que le plan d'échantillonnage stratifié permettant de couvrir les différentes formations végétales était modifié dans les zones d'accès difficile de BT de façon à suivre les sentiers et les voies d'accès aux massifs forestiers montagneux, ce qui a conduit à plusieurs ensembles de points de comptage ayant une distribution linéaire.

Pour chaque type d'occupation du sol, nous avons calculé l'indice de spécialisation de Jacobs (1974) (voir formule dans l'annexe A6), qui varie entre -1 et 1 selon que l'espèce évite ou est spécialiste du type d'occupation du sol considéré, respectivement. Une valeur nulle indique que l'espèce a une attitude "neutre" vis-à-vis de ce type d'occupation du sol. Les taches d'habitat du graphe paysager construit à partir de cet indice correspondaient au type d'occupation du sol obtenant l'indice de spécialisation maximal (Table 7). De plus, les indices étaient convertis en valeurs de coût en utilisant la formule suivante, adaptée à partir de Bourdouxhe *et al.* (2020) :

$$Cout_i = e^{\frac{-\ln(Cout_{max}) \times (Jacobs_i + 1)}{Jacobs_{habitat} + 1}} \times Cout_{max}$$

avec $Cout_i$ et $Jacobs_i$ la valeur de coût et l'indice de Jacobs respectivement associés au type d'occupation du sol i et $Cout_{max}$ la valeur de coût maximale. $Jacobs_{habitat}$ est l'indice maximal, associé au type d'occupation du sol constituant l'habitat de l'espèce. Cette formule assurait que la valeur de coût associée à l'habitat valait 1. L'utilisation d'une fonction exponentielle négative pour la conversion des indices en valeurs de coût se justifie par le fait qu'une espèce peut souvent se déplacer facilement dans une zone qui ne pourrait pas constituer son habitat (Keeley *et al.*, 2016, 2017 ; Wright *et al.*, 2021). Le coût maximal a été fixé à 1000, reflétant l'ordre de grandeur de valeurs de coût précédemment estimées à partir de données empiriques (Gurrutxaga *et al.*, 2010).

2.2.4 Modèle de distribution de l'espèce

Nous avons créé un modèle de distribution de la Paruline caféïette à l'aide d'une régression logistique car nous disposons de données ponctuelles concernant à la fois la présence et l'absence de l'espèce (cf. Annexe B3). Dans ce cas, cette méthode donne de meilleurs résultats que l'utilisation

de l’algorithme Maxent (Guillera-Arroita *et al.*, 2014), fréquemment utilisé pour construire les SDM servant de base aux graphes paysagers lorsque seules des données de présence sont disponibles (Bourdouxhe *et al.*, 2020 ; Godet et Clauzel, 2021 ; Tarabon *et al.*, 2019). De plus, les éventuels biais liés au plan d’échantillonnage sont mieux compensés lorsqu’à la fois des points de présence et d’absence sont inclus dans les modèles (Fletcher et Fortin, 2018).

Des analyses préliminaires nous ont permis d’identifier un ensemble optimal de cinq variables explicatives de la distribution de l’espèce : (i) la proportion de pixels agricoles ou (ii) artificialisés dans un rayon de 500 m autour des points, la distance (iii) au pixel d’habitat forestier ou (iv) à la route la plus proche des points et (v) les précipitations annuelles moyennes. Les pixels d’habitat forestier considérés étaient les formations végétales identifiées par l’expert comme faisant partie de l’habitat de l’espèce.

Afin de limiter les biais liés à la modification de l’échantillonnage sur BT, nous avons calculé le modèle plusieurs fois en ne sélectionnant aléatoirement à chaque itération que deux points parmi chaque ensemble de points de comptage distribués de façon linéaire sur BT. Une procédure de validation croisée nous a indiqué que cela améliorait le modèle. Les coefficients associés aux variables étant relativement stables pour les différents modèles créés, nous avons finalement considéré un de ces échantillons aléatoires pour construire le SDM final.

Nous avons ensuite évalué la qualité du modèle en calculant l’AUC (cf. Annexe B3). Puis, nous avons identifié la probabilité de présence seuil au-delà de laquelle considérer que l’espèce était présente permettait de bien prédire les observations de présence ; tout en nous assurant que les pixels ayant des probabilités de présence inférieures à cette valeur seuil incluaient principalement des observations d’absence. Par la suite, les pixels ayant des probabilités de présence supérieures à cette valeur seuil constituaient les zones d’habitat utilisées pour créer les nœuds du graphe paysager issu du SDM (Table 7). Enfin, nous avons converti les probabilités de présence issues du SDM en valeurs de coût en utilisant une formule similaire à celle utilisée pour convertir les indices de spécialisation :

$$Cost_i = e^{\frac{-\ln(Cost_{max}) \times P_i}{P_{thr.}}} \times Cost_{max}$$

avec $Cost_i$ le coût associé à chaque pixel selon sa probabilité de présence P_i , $P_{thr.}$ la probabilité de présence seuil et $Cost_{max}$ le coût maximal valant 1000 dans ce cas également.

2.2.5 Graphes paysagers

Construction des graphes

En utilisant le logiciel Graphab, nous avons créé quatre graphes dont les nœuds et les liens étaient définis de trois façons différentes (Table 7). La taille minimale des taches était d’1 ha car c’est la surface minimale nécessaire à l’établissement d’un couple reproducteur (Éraud *et al.*, 2012). La capacité des taches était équivalente à leur surface. Certaines taches avaient une surface nettement supérieure à l’échelle spatiale à laquelle la distribution de l’habitat est censée affecter la dynamique d’une population. Laroche *et al.* (2020) ont montré que la prise en compte de taches d’habitat de trop grande

taille limitait le pouvoir explicatif des métriques de connectivité issues des graphes paysagers. Nous avons donc subdivisé les taches pour que leur emprise spatiale ne dépasse pas 2500 m de côté. Les liens des graphes étaient pondérés par les distances-coût calculées entre les taches selon chaque scénario (Table 7) et avaient une topologie planaire.

Type de graphe paysager	Délimitation des nœuds	Scénarios de coût pour la pondération des liens
<i>Basé sur les dires d'expert (1 et 2)</i>	Formations végétales constituant l'habitat de l'espèce selon l'expert	Valeurs de coût données par l'expert, ré-échelonnées de deux façons (Expert 1 et 2)
<i>Basé sur les indices de spécialisation</i>	Type d'occupation du sol associé à l'indice de spécialisation maximal	Conversion des indices de spécialisation en valeurs de coût d'après Bourdouxhe et al. (2020)
<i>Basé sur le SDM</i>	Pixels dont la probabilité de présence dépasse la probabilité seuil optimale pour prédire la présence de l'espèce	Conversion des probabilités de présence en valeurs de coût d'après Bourdouxhe et al. (2020)

TABLE 7 – Définition des éléments des graphes paysagers selon les différentes méthodes de construction

Métriques de connectivité

À partir des graphes, nous avons calculé trois métriques reflétant de façon complémentaire la contribution des taches à la connectivité du réseau d'habitat dans son ensemble ([Baranyi et al., 2011](#) ; [Rayfield et al., 2011](#)). Ces trois métriques étaient les mêmes que celles considérées dans le chapitre 3 concernant le lien entre la quantité d'habitat atteignable et la structure génétique de populations de Criquet ensanglanté, à savoir la capacité et les métriques F et BC (cf. Introduction générale, Figure 17).

La métrique F reflétait le rôle d'une tache pour les flux de dispersion et était calculée avec la formule suivante :

$$F_i = \sum_{j=1, j \neq i}^n \text{Capa}_j \times e^{-\alpha \times d_{ij}}$$

avec i l'indice de la tache pour laquelle la métrique est calculée, j l'indice des n autres taches et d_{ij} la distance entre ces taches. Capa_j est la capacité de la tache j . α a été calculé en considérant que la probabilité de dispersion $p(d_{ij}) = e^{-\alpha d_{ij}}$ à une distance d_{ij} était égale à 0.05 pour des distances d_{ij} comprises entre 500 et 15000 m (avec un pas de 500 m). Ces distances géodésiques étaient converties en distances-coût en utilisant une régression de type log-log ([Tournant et al., 2013](#)), comme dans les précédents chapitres de cette thèse.

La métrique BC était calculée avec la formule suivante :

$$BC_i = \sum_j \sum_k \text{Capa}_j \times \text{Capa}_k \times e^{-\alpha \times d_{jk}}$$

$$j, k \in \{1, \dots, n\}, k < j, i \in P_{jk}$$

tel que P_{jk} représente l'ensemble des taches d'habitat traversées par les chemins les plus courts entre les taches j et k incluant la tache i , pour que ces chemins uniquement soient pris en compte dans le calcul de la métrique pour la tache i . α variait de la même façon que pour le calcul de F. La métrique BC est censée identifier les taches qui jouent le rôle de *stepping stones* dans le réseau

d’habitat. Selon [Zetterberg et al. \(2010\)](#), elle devrait donc prendre des valeurs élevées au niveau des taches d’habitat maintenant de forts niveaux de diversité génétique, bien que cette relation reposant sur l’influence des flux génétiques sur la diversité génétique ait rarement été mise en évidence. Les trois métriques ont été calculées pour chaque graphe paysager.

2.3 Analyse de données génétiques

2.3.1 Échantillonnage et génotypage

Un échantillon sanguin a été prélevé sur 712 oiseaux capturés au filet dans 27 taches de forêt (9 sur GT et 18 sur BT, [Figure 51](#))³. 20 sites avaient déjà été échantillonnés en 2015 et l’échantillonnage de sept nouveaux sites en 2020 a complété ce premier échantillonnage. Pour déterminer si ces données pouvaient être regroupées, quatre sites ont été échantillonnés aux deux dates et nous avons testé la différenciation génétique de chacune de ces populations entre ces deux dates. Elle n’était pas significative pour trois des quatre sites. Bien que significative pour un site, le niveau de différenciation génétique relative de ce site par rapport aux autres n’a pas évolué entre 2015 et 2020. Nous avons donc regroupé ces deux jeux de données génétiques. Après extraction de l’ADN, 12 *loci* microsatellites ont été amplifiés par PCR et génotypés à l’aide d’un séquenceur automatique selon le protocole décrit par [Khimoun et al. \(2017\)](#).

2.3.2 Indices de structure génétique

Après vérification de l’équilibre de Hardy-Weinberg dans chaque population et de l’absence de déséquilibre de liaison, la diversité génétique de chaque population a été estimée par sa richesse allélique raréfiée, calculée avec le logiciel ADZE ([Szpiech et al., 2008](#)). Le niveau de différenciation génétique relatif de chaque population par rapport aux autres a été estimé en créant un graphe génétique complet. Ses liens étaient pondérés par des valeurs de F_{ST} ([Weir et Cockerham, 1984](#)), indices de fixation que nous assimilons à des distances génétiques dans cette étude. Nous avons calculé l’indice MIW (*Mean Inverse Weight*, cf. Introduction générale, [Figure 33](#)) pour chacun des 27 nœuds du graphe comme la moyenne de l’inverse des valeurs de F_{ST} entre chaque population et toutes les autres. Des valeurs de MIW élevées indiquent qu’une population est globalement peu différenciée génétiquement des autres, ce qui peut signifier qu’elle échange beaucoup de migrants avec les autres, comme l’ont montré [Koen et al. \(2016\)](#).

2.4 Validation de la modélisation basée sur les graphes paysagers

Nous avons validé les graphes paysagers en nous focalisant (i) sur la relation entre les distances-coût associées à leurs liens et la distance génétique entre les populations échantillonnées séparées par ces liens, et (ii) sur la relation entre les métriques de connectivité calculées au niveau des nœuds et les indices génétiques calculés au niveau des populations occupant ces nœuds ([Figure 50](#)). Dans ces deux cas, une relation étroite était censée indiquer que les graphes paysagers représentaient correctement la réalité écologique.

3. Je n’ai pas participé personnellement à cet échantillonnage, ni aux analyses de biologie moléculaire, qui ont été assurées par Alexandrine Daniel, Stéphane Garnier, Aurélie Khimoun, Bruno Faivre, Anthony Ollivier et une équipe de techniciens de terrain dans le cadre des projets Frag&Binv (2015) et CANON (2020).

2.4.1 Validation des distances-coût associées aux liens des graphes paysagers

Nous voulions identifier le scénario de coût reflétant au mieux la façon dont chaque type d'occupation du sol affecte les déplacements de la Paruline caféïette. Les données génétiques sont censées refléter de tels déplacements lorsqu'ils sont suivis de flux génétiques. Nous avons donc considéré que les distances-coût expliquant le mieux les distances génétiques entre populations représentaient le mieux la connectivité des paires de populations. Pour chaque graphe paysager, nous avons calculé les distances-coût entre populations le long des liens de ces graphes.

Nous avons ensuite modélisé la relation entre les distances-coût issues de chaque graphe paysager et les distances génétiques en utilisant des modèles linéaires mixtes de type MLPE (Clarke *et al.*, 2002). Ces modèles expliquent la distance génétique en fonction de la distance-coût en incluant un effet aléatoire relatif à l'identité d'une des populations constituant chaque paire pour tenir compte de la non-indépendance statistique des valeurs d'une matrice de distances⁴. Il a déjà été montré dans le cadre d'analyses basées sur des simulations que leur utilisation en génétique du paysage permet d'identifier correctement le scénario de résistance à l'origine d'un patron de différenciation génétique (Shirk *et al.*, 2017b).

Par ailleurs, nous avons montré dans le chapitre 1 que l'échelle spatiale à laquelle les distances génétiques ont atteint un équilibre reflétant l'influence conjointe de la dérive et du flux génétique ainsi que la topologie des connexions considérées affectent l'analyse de la relation entre distances génétiques et paysagères. Ainsi, lorsque nous avons évalué la capacité des distances-coût associées aux liens des différents graphes à prédire les distances génétiques, nous avons également évalué l'influence sur les résultats de l'analyse de l'échelle et de la topologie des connexions prises en compte. Pour cela, nous avons réalisé les modèles en considérant différents ensembles de liens :

- Toutes les paires de populations, reflétant la pratique la plus courante en génétique du paysage.
- Les paires de populations situées sur la même partie de l'île de Guadeloupe, c'est-à-dire les paires de type "BT-BT" ou "GT-GT". Cette sélection reflète les analyses parfois faites en séparant la zone d'étude en plusieurs ensembles (Angelone et Holderegger, 2009 ; Reed *et al.*, 2017 ; Wang *et al.*, 2008).
- Les paires de populations séparées par une distance inférieure à un seuil de distance-coût, c'est-à-dire les populations connectées sur un graphe génétique élagué en utilisant ce seuil de distance-coût. Cette sélection reflète les méthodes d'élagage itératif parfois employées pour identifier l'échelle à laquelle le paysage influence la structure génétique (Angelone *et al.*, 2011 ; Emaresi *et al.*, 2011).

Afin d'évaluer la validité écologique de chaque scénario de coût tout en contrôlant le risque de sur-paramétrage (*overfitting*) des modèles, nous avons adapté la méthode de validation croisée de type *Leave One Out Cross Validation* (LOOCV) aux objectifs de notre étude et à nos données. Pour cela, nous avons itérativement supprimé du jeu de données une des 27 populations et les valeurs de distances l'impliquant compte tenu de la topologie des liens considérés avant de calibrer les modèles MLPE⁵. Chaque modèle alors créé était utilisé pour prédire *a posteriori* les distances génétiques non prises

4. Chaque population intervient dans plusieurs paires de populations. Les observations statistiques ne sont donc pas indépendantes les unes des autres. Par exemple, les paires A-B et B-C sont moins indépendantes que ne le sont les paires A-B et C-D.

5. Soit 26 valeurs en moins lorsqu'on considère une topologie complète.

en compte pour sa calibration. La distance génétique associée à chaque paire de populations était donc prédite deux fois et nous avons estimé l'erreur moyenne de prédiction associée à chaque paire de populations en calculant la racine de la moyenne des différences au carré entre les valeurs prédites et observées (RMSE). Ainsi, en plus du coefficient de détermination R^2 calculé à partir du modèle intégrant toutes les populations (R^2 de calibration), nous avons calculé le R^2 de validation en nous basant sur ces erreurs de prédictions (cf. Annexe A6 pour plus de détail sur ce calcul). Enfin, nous avons calculé pour chaque population la moyenne des erreurs de prédictions des distances génétiques entre cette population et toutes celles auxquelles elle est connectée compte tenu de la topologie des liens considérés. Nous avons ainsi pu représenter ces erreurs sur une carte pour comprendre dans quelles zones les modèles calibrés avec les distances-coût issues de chacun des graphes paysagers étaient les plus valides.

2.4.2 Validation des métriques de connectivité

En plus de valider les scénarios de coût utilisés pour calculer les liens des graphes paysagers, nous visions également dans cette étude à identifier parmi les métriques de connectivité calculées à partir des différents graphes paysagers celles qui reflétaient le mieux les indices génétiques calculés au niveau des populations. Pour cela, nous avons associé à chaque population l'ensemble des métriques calculées au niveau de la tache dans laquelle elle a été échantillonnée avant de calculer le coefficient de corrélation de Spearman entre les deux indices génétiques (Rich.all., MIW) d'une part et chacune des métriques de connectivité (Capacité, F, BC) d'autre part.

Enfin, après avoir évalué le lien entre les différentes métriques et les indices génétiques, nous avons voulu illustrer l'influence qu'aurait le choix d'un paramètre de calcul d'une métrique de connectivité différent de celui conduisant à la corrélation la plus forte entre cette métrique et les indices génétiques. L'objectif était alors de mettre en évidence l'erreur potentielle qui peut être faite en termes de priorisation des taches lorsqu'on ne dispose pas de données génétiques. Pour cela, nous avons comparé les rangs des taches échantillonnées en termes de connectivité selon que cette connectivité était estimée par (i) la métrique la plus fortement corrélée de manière générale aux indices génétiques ou (ii) par cette même métrique calculée avec un paramètre de pondération des liens entre taches différent et reflétant le choix qu'un utilisateur pourrait faire sans disposer de données empiriques ($d = 5000$ m pour l'exemple). Puis, nous avons localisé sur une carte les taches d'habitat les plus affectées par ce changement de paramètres.

3 Résultats

3.1 Graphes paysagers

3.1.1 Paramètres de construction des graphes paysagers

Dires d'expert

Selon l'expert interrogé, toutes les formations végétales ligneuses à l'exception des fourrés littoraux et des friches à ligneux bas constituent l'habitat de la Paruline caféïette. Les valeurs de coût attribuées aux types d'occupation du sol variaient entre 1 et 10, les plus faibles valeurs étant attribuées aux forêts, suivies des zones agricoles, semi-ouvertes et humides, et les plus fortes à l'océan et aux zones artificialisées. Pour rendre la comparaison avec les autres scénarios de coût possible, nous

avons ré-échelonné ces valeurs de coût pour fixer la valeur maximale à 1000. Pour cela, nous avons (i) multiplié toutes les valeurs par 100, conservant ainsi les mêmes contrastes (scénario Expert 1) ou (ii) ré-échelonné les valeurs entre 1 et 1000, conservant ainsi la même gamme de variation que dans les autres scénarios (scénario Expert 2) (Table 8 et Figures 54A et 54B) ⁶.

Type d'occupation du sol	Valeurs initiales		Valeurs de coût			
	Exp.	Ind.spéc.	Exp.1	Exp.2	Ind.spéc.	SDM
Habitat	1	0.73	100	1	1	9 ± 84
Forêt	2	0.46	200	112	3	127 ± 308
Zones semi-naturelles	4	-0.26	400	334	52	314 ± 367
Zones semi-ouvertes	3	0.29	300	223	6	577 ± 450
Zones ouvertes	4	-0.83	400	334	509	544 ± 428
Zones agricoles	3	-0.68	300	223	241	598 ± 420
Zones humides	3	-1	300	223	1000	109 ± 264
Cours et plans d'eau	4	-1	400	334	1000	459 ± 441
Océan	10	-1	1000	1000	1000	993 ± 76
Zones artificialisées	10	-0.26	1000	1000	52	510 ± 405

TABLE 8 – Indices de spécialisation de Jacobs et valeurs de coût associés à chaque type d'occupation du sol selon les différents scénarios. Les valeurs initiales "Exp." correspondent aux valeurs de coût données par l'expert. Les scénarios "Exp.1" et "Exp.2" sont les deux scénarios de coût issus de ces valeurs données par l'expert. "Ind.spéc." fait référence aux indices de spécialisation, pour lesquels à la fois les valeurs brutes initiales et les valeurs de coût calculées à partir de celles-ci sont reportées dans deux colonnes. La colonne "SDM" indique les valeurs de coût issues du modèle de distribution de la Paruline caféïette. La gamme des valeurs de coût était continue dans ce dernier cas et nous reportons ici les valeurs de coût moyennes (\pm écart-type) calculées au sein de chaque type d'occupation.

Indices de spécialisation

L'indice de spécialisation de Jacobs maximal était obtenu pour le type d'occupation du sol identifié comme étant l'habitat de l'espèce par l'expert (0.73) et les valeurs minimales pour les zones ouvertes (-0.83), agricoles (-0.68) et artificialisées (-0.26) (Table 8). Il n'y avait aucun point dans des plans ou cours d'eau, des zones humides ou dans l'océan, et nous leur avons donc attribué un indice de spécialisation égal à -1 . Il n'y en avait pas non plus dans les zones semi-naturelles, majoritairement constituées d'espaces verts urbains et auxquelles nous avons attribué le même indice qu'aux zones artificialisées. Les valeurs de coût obtenues par conversion de ces indices étaient contrastées, valant par exemple 52 et 241 pour les zones artificialisées et agricoles, respectivement (Table 8 et Figure 54C).

Modèle de distribution de l'espèce

Le SDM réalisé avait une bonne capacité prédictive (AUC : 0.918) et la probabilité de présence optimale au-delà de laquelle nous avons considéré qu'un pixel correspondait à l'habitat de l'espèce valait 0.328. Une augmentation de la distance au pixel forestier le plus proche influençait négativement la probabilité de présence. Cette variable était celle qui avait le plus d'effet. Les proportions de pixels agricoles et artificialisés influençaient négativement la probabilité de présence, tandis qu'elle augmentait avec la distance à la route la plus proche et avec la quantité de précipitations annuelles.

Les valeurs de coût issues du SDM étaient contrastées et relativement élevées (Figure 54D). Il n'y a qu'au sein des zones d'habitat délimitées par l'expert que ces coûts prenaient une valeur moyenne inférieure à 100 (Table 8). Les coûts associés aux zones agricoles, qui occupent 25.4 % de la surface de l'île, étaient nettement plus élevés lorsqu'ils étaient issus du SDM que dans les scénarios basés sur les

6. Pour cela, soit $Valeurs_{init}$ les valeurs initiales et $Valeur_{init_i}$ une de ces valeurs, sa valeur ré-échelonné $Valeur_{nouvelle_i}$ est égale à : $1 + \frac{Valeur_{init_i} - \min(Valeurs_{init})}{\max(Valeurs_{init}) - \min(Valeurs_{init})} \times (1000 - 1)$.

dières de l'expert (1 et 2) ou sur les indices de spécialisation (598 *vs* 300, 223 et 241, respectivement). Les zones artificialisées (20.8 % de la surface) étaient en moyenne associées à une valeur de coût plus faible (510) que dans les scénarios à dières d'expert (1000) mais plus forte que dans le scénario basé sur les indices de spécialisation (52, Table 8).

3.1.2 Caractéristiques des éléments des graphes

Les graphes paysagers construits à dières d'expert ou à partir des indices de spécialisation partageaient les mêmes nœuds. Ils étaient plus nombreux mais en moyenne plus petits (1653 nœuds, surface moyenne : 39 ha) que ceux du graphe basé sur le SDM (621, surface moyenne : 102 ha). L'inclusion de la distance à la forêt dans le SDM avait tendance à intégrer aux taches d'habitat des pixels non forestiers mais situés en marge de forêts, créant alors de plus grandes taches que dans les autres scénarios. En outre, en raison de l'influence négative sur la probabilité de présence d'autres variables explicatives telles que la proportion de pixels artificialisés autour d'un pixel, cette probabilité avait tendance à prendre des valeurs inférieures au seuil permettant d'intégrer un pixel aux taches d'habitat dans des zones intégrées à l'habitat selon les autres scénarios (par exemple au niveau de taches de forêt situées dans des secteurs majoritairement artificialisés ou agricoles). Ces deux effets faisaient conjointement diminuer le nombre de taches issues du SDM tout en augmentant leur surface moyenne en comparaison avec les taches délimitées à partir de la carte d'occupation du sol.

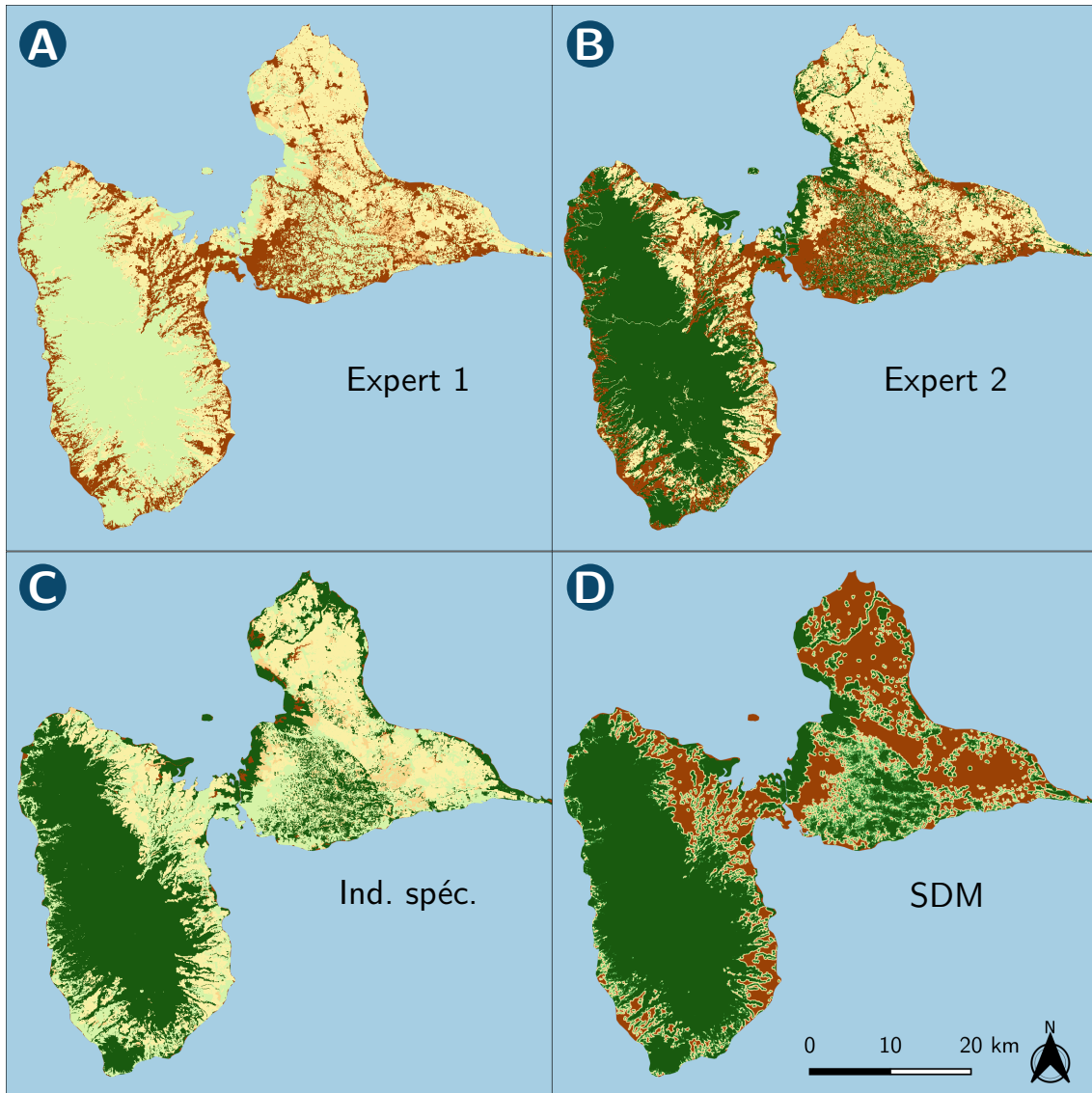
Lorsque le SDM servait de base à la construction du graphe, la distribution des distances-coût entre populations le long des liens du graphe était différente de celle obtenue avec les autres scénarios (cf. Annexe A6, Figure 41). En particulier, les distances-coût entre les populations de BT et GT séparées par l'isthme (BT-GT) étaient nettement plus élevées. Cela s'expliquait dans ce scénario par les valeurs de coût relativement élevées associées à la fois aux zones agricoles et artificialisées présentes dans cette zone.

3.2 Structure génétique des populations

Même si les populations de BT et GT avaient des niveaux de diversité génétique moyens assez similaires, la richesse allélique était nettement plus variable sur BT (4.63 ± 0.34) que sur GT (4.82 ± 0.13) (Table 9). Sur BT, les principales différences s'observaient entre les populations situées au sein des plus grandes taches forestières et celles situées en marge de celles-ci en périphérie de l'île. De façon similaire, les indices de différenciation génétique relative (MIW) avaient une distribution bimodale sur BT et prenaient des valeurs intermédiaires sur GT (>50 ou <26 sur BT, entre 26 et 50 sur GT, Table 9). Les valeurs de MIW les plus faibles concernaient les populations de BT situées en marge des plus grandes taches forestières (Table 9).

3.3 Validation des distances-coût associées aux liens

Les modèles expliquant les distances génétiques en fonction des distances-coût entre populations calculées le long des liens des graphes avaient des performances contrastées selon les scénarios de coût



Valeurs de coût



FIGURE 54 – Cartes des coûts de déplacement selon chaque scénario de coût : (A) scénario à dire d’expert 1, (B) scénario à dire d’expert 2, (C) scénario basé sur les indices de spécialisation de Jacobs, (D) scénario basé sur le SDM. La même gamme de couleurs est utilisée pour les quatre cartes.

Pop.	Partie de l'île	Nb.ind.	Rich. all.	MIW
F005	BT	8	4.14	15.86
F011	BT	6	4.33	25.22
F033	BT	8	5.26	39.38
F055	BT	28	4.89	83.58
F062	BT	27	4.89	388.55
F068	BT	30	4.88	250.20
F073	BT	52	4.67	36.01
F080	BT	52	4.52	16.16
F084	BT	15	4.83	67.19
F099	BT	29	4.57	134.50
F142	BT	23	4.28	20.04
F170	BT	26	3.99	15.49
F198	BT	12	4.23	15.88
FX01	BT	27	4.98	146.66
FX04	BT	19	4.89	88.90
FX07	BT	36	4.51	16.79
FX13	BT	21	4.88	25.55
FX15	BT	12	4.73	20.87
F002	GT	48	4.58	26.26
F006	GT	33	4.95	30.04
F007	GT	28	4.94	34.39
F181	GT	36	4.97	40.64
F196	GT	22	4.75	28.04
F217	GT	47	4.79	34.76
F229	GT	30	4.78	33.76
FX02	GT	19	4.90	31.06
FX14	GT	18	4.76	27.15

TABLE 9 – Indices génétiques calculés dans chacune des 27 populations pour décrire la diversité génétique locale (Rich. all. : richesse allélique raréfiée) et la différenciation génétique relative par rapport à l'ensemble des autres populations (MIW : *Mean Inverse Weight*, moyenne de l'inverse des valeurs de F_{ST} entre une population et toutes les autres). La colonne "Pop." indique l'identifiant des populations. "Partie de l'île" indique si une population se situe sur Basse-Terre (BT) ou Grande-Terre (GT) et "Nb.ind." indique le nombre d'individus échantillonnés dans chaque population.

et les ensembles de liens considérés (Figure 55). Indépendamment du scénario de coût, les valeurs des R^2 de calibration et de validation des modèles considérant l'intégralité des paires de populations étaient faibles. Les R^2 de calibration les plus élevés étaient obtenus avec les scénarios de coût issus du SDM (0.36) ou des indices de spécialisation (0.35), suivis du scénario Expert 2 (0.28) (Figure 55).

L'ensemble de liens à considérer pour maximiser l'ajustement du modèle n'était pas le même selon les distances-coût utilisées (Figures 55 et 56). À partir du scénario issu du SDM, les distances génétiques étaient le mieux prédites à partir d'un ensemble de distances-coût inférieures à une valeur seuil. Élaguer le graphe des populations en considérant ce seuil les séparait en deux ensembles : (i) les populations de GT et une population de BT située proche de l'isthme et des taches de forêt de GT (FX13) et (ii) toutes les autres populations de BT (Figure 56C). Cela signifie que les fortes distances-coût calculées à partir du SDM entre FX13 et les autres populations de BT, et donc associées à des chemins traversant des zones essentiellement agricoles et artificialisées ne permettaient pas d'expliquer correctement les distances génétiques observées. Ces distances génétiques avaient tendance à être surestimées. Au contraire, lorsque les distances-coût étaient basées sur les indices de spécialisation, le meilleur modèle expliquait les distances génétiques entre les paires de populations situées sur la même partie de l'île (paires BT-BT ou GT-GT, Figure 56B). Cet ensemble incluait donc des paires de populations séparées par de grandes zones défavorables à l'espèce, telles que les paires incluant la

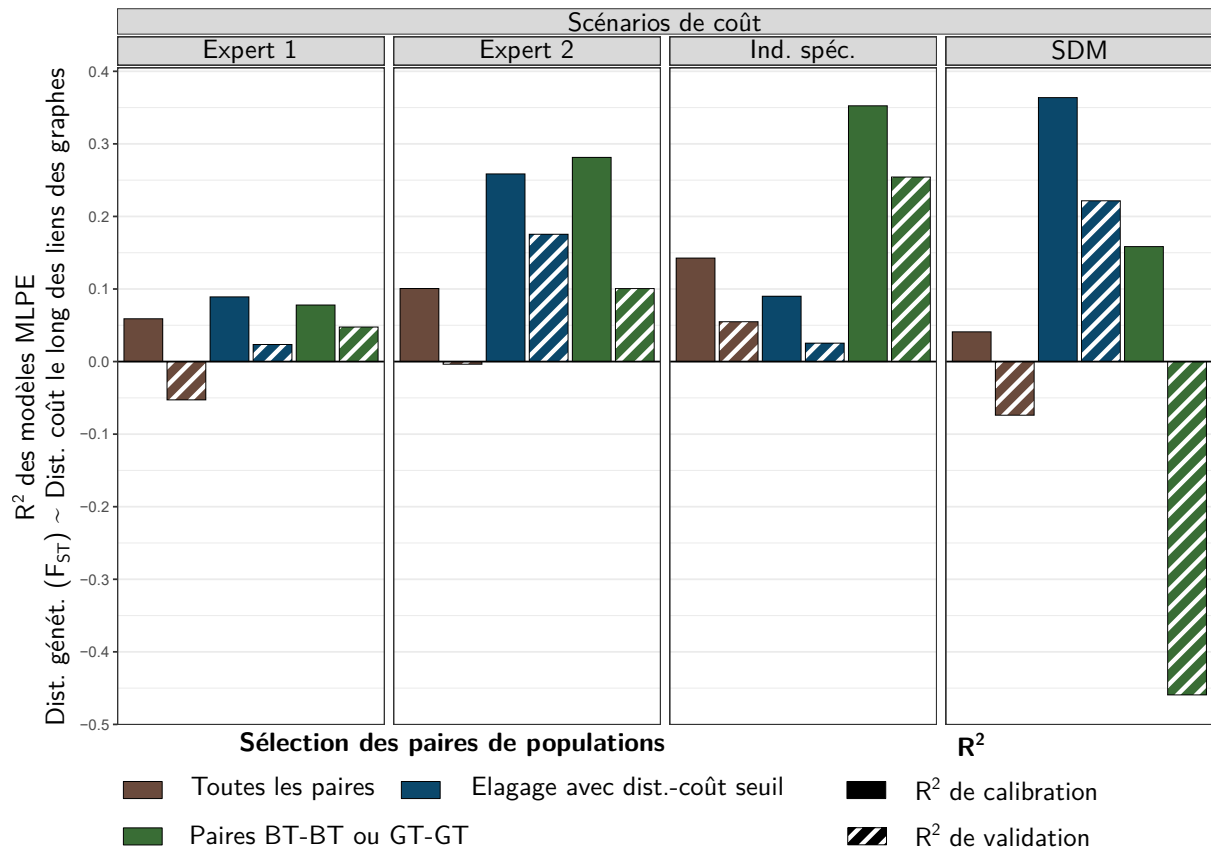


FIGURE 55 – Ajustement (R^2) des modèles mixtes de type MLPE expliquant les distances génétiques entre populations (F_{ST}) en fonction des distances-coût entre ces populations, calculées le long des liens des graphes paysagers. Les séparations verticales distinguent les différents scénarios de coût utilisés pour calculer les distances-coût lors de la création des graphes. "Expert 1" et "Expert 2" : scénarios de coût à dire d'expert, "Ind. spéc." : scénario basé sur les indices de spécialisation de Jacobs, "SDM" : scénario basé sur la conversion des probabilités de présence issues du SDM en valeurs de coût. La couleur des barres indique quel ensemble de liens (paires de populations) est inclus dans le modèle. Pour chaque scénario de coût, seuls les résultats du meilleur modèle obtenu en considérant un ensemble de liens défini par un élagage selon une distance-coût seuil sont présentés. Les barres unies et striées indiquent respectivement les R^2 de calibration et de validation.

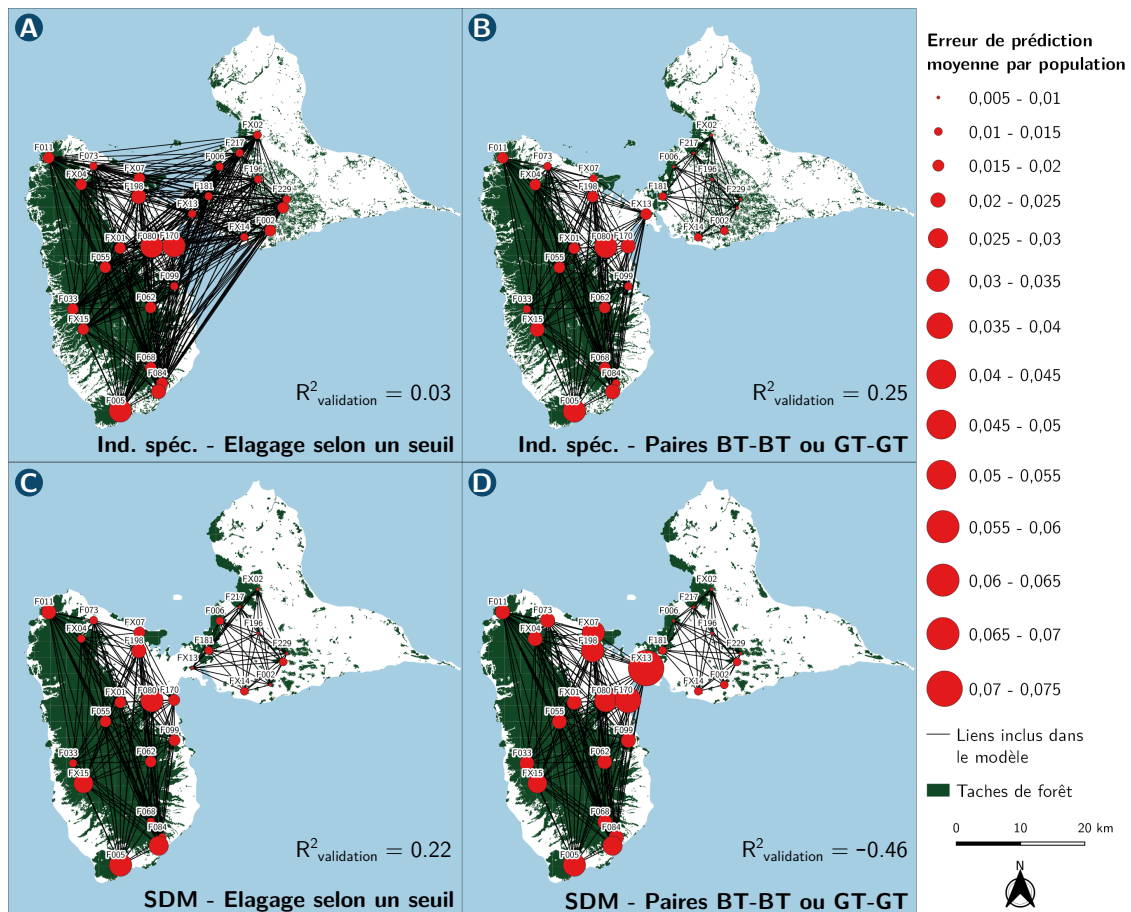


FIGURE 56 – Cartes des erreurs de prédiction des distances génétiques associées aux modèles de type MLPE. La moyenne des erreurs de prédiction des distances génétiques (RMSE) impliquant chaque population et les populations auxquelles elle est connectée dans la topologie considérée a été calculée pour chaque population. L'erreur de prédiction considérée est la racine de la moyenne du carré des différences entre les distances génétiques prédites et observées (deux valeurs par paire de populations). Chaque prédiction a été faite à partir du modèle calibré en excluant une des deux populations de la paire concernée. Le R^2 de validation du modèle MLPE correspondant est reporté sur chaque carte. Les résultats affichés concernent les modèles incluant les distances-coût issues des scénarios basés sur les indices de spécialisation de Jacobs (A, B) ou sur le SDM (C, D), et considérant des paires de populations sélectionnées selon une distance seuil (A, C) ou situées sur la même partie de l'île de Guadeloupe (paires BT-BT ou GT-GT). Les liens correspondent aux paires de populations prises en compte pour la réalisation des modèles dans chaque cas et ont donc des topologies différentes.

population FX13 et les autres populations de BT. Cela illustre la validité relative des scénarios de coût selon les zones dans lesquelles on les utilise pour expliquer la distance génétique.

Les valeurs du R^2 de validation confirmaient le fait que les distances-coût calculées à partir des différents graphes paysagers étaient parfois de bons prédicteurs des distances génétiques (Figure 55). En effet, les valeurs obtenues étaient relativement élevées si l'on considère qu'une distance génétique dépend pour partie de l'effet de la dérive qui est aléatoire et donc difficilement prédictible. Les différences de performance des modèles selon les liens considérés dépendaient en partie du fait que les distances génétiques impliquant la population FX13 et les autres populations de BT étaient mal prédites par les distances-coût issues du SDM, comme le confirmait le R^2 de validation négatif obtenu en considérant uniquement les paires BT-BT ou GT-GT dans ce cas. Au contraire, les distances génétiques entre cette population et les populations de GT étaient bien prédites en utilisant des distances-coût calculées à partir du même scénario.

Dans l'ensemble, les erreurs de prédiction faites par les meilleurs modèles étaient supérieures lorsque les distances génétiques impliquaient des populations de BT plutôt que de GT. Enfin, le scénario Expert 2 permettait d'obtenir la troisième plus forte valeur de R^2 de validation (0.18), dans le cas où un ensemble de paires de populations était sélectionné avec un élagage basé sur une distance-coût seuil. Cela signifie que par rapport aux scénarios basés sur l'utilisation de données empiriques (Ind.spéc., SDM), un scénario basé sur des dires d'expert apportait des prédictions de distances génétiques satisfaisantes.

3.4 Validation des métriques de connectivité

Les deux indices génétiques (richesse allélique, MIW) étaient chacun corrélés significativement à toutes les métriques de connectivité (Capacité, F, BC). Pour chaque méthode de construction des graphes paysagers (Exp.1, Exp.2, Ind.spéc., SDM), nous ne reportons que (i) la plus forte corrélation obtenue en testant plusieurs distances d pour lesquelles nous considérons une probabilité de dispersion égale à 0.05 pour fixer la valeur α et (ii) la corrélation correspondante obtenue avec $d = 5000$ m (Table 10). Ces corrélations étaient fortes dans l'ensemble ($r_{Sp.}$ allant jusqu'à 0.72). Les métriques étaient globalement plus fortement corrélées à l'indice MIW qu'à la richesse allélique. De plus, la distance d optimale pour le calcul des métriques était toujours inférieure ou égale à 3500 m.

Les corrélations obtenues avec les métriques issues du graphe basé sur le SDM étaient globalement différentes et le plus souvent plus faibles que celles obtenues en considérant les métriques issues des trois autres graphes (Table 10). À l'exception de la corrélation entre le MIW et la capacité des taches, qui était plus forte en considérant la capacité des taches issues du SDM, les métriques issues du graphe créé selon le scénario Expert 1 étaient les plus corrélées aux indices génétiques. Néanmoins, des corrélations très similaires étaient obtenues avec les métriques issues des graphes Expert 2 ou basé sur les indices de spécialisation. Dans ce cadre, il est difficile d'établir si ces différences de corrélation proviennent de différences entre les graphes paysagers concernant la capacité de leurs taches ou les distances-coût associées à leur liens.

Par construction, la métrique F calculée en considérant que $d = 5000$ m était moins fortement corrélée aux indices génétiques que lorsqu'elle était calculée avec la valeur d optimale, même si ces corrélations restaient significatives dans la plupart des cas (Table 10). Cela signifie que les rangs des taches d'habitat en termes de connectivité reflétaient moins bien dans ce cas les réponses génétiques mesurées dans les populations occupant ces taches. De façon intéressante, les plus grandes différences de rang observées en calculant la métrique F avec la valeur d optimale ou avec $d = 5000$ concernaient soit des taches situées dans la partie la plus forestière du nord de GT, soit des taches situées en marge des grandes taches forestières de BT (Figure 57). Dans ce premier cas, utiliser une valeur d élevée sous-estimait le niveau de connectivité des taches de GT. En effet, les métriques ne reflétaient le fait qu'elles sont entourées d'importantes surfaces forestières qu'à condition de les calculer en considérant une échelle spatiale restreinte. Au contraire, dans le cas des taches de BT situées en marge des grandes taches forestières, considérer une échelle spatiale large avait tendance à sur-estimer leur connectivité en limitant l'influence de leur déconnexion vis-à-vis de ces grandes taches forestières.

Ind.génét.	Mét.connect.	Type de graphe	$r_{Sp. - Optim.}$	$d_{Optim.}$	$r_{Sp. - 5000 m}$	
MIW	Capacité	Exp. 1/ Exp.2 / Ind.spéc.	0.50	*		
MIW	Capacité	SDM	0.57	**		
Rich.all.	Capacité	Exp. 1/ Exp.2 / Ind.spéc.	0.55	*		
Rich.all.	Capacité	SDM	0.45			
MIW	F	Expert 1	0.72	***	1000	0.55 **
MIW	F	Expert 2	0.68	***	1000	0.55 **
MIW	F	Ind.spéc.	0.71	***	500	0.53 **
MIW	F	SDM	0.58	**	3500	0.54 **
Rich.all.	F	Expert 1	0.53	**	500	0.22
Rich.all.	F	Expert 2	0.42	*	1000	0.20
Rich.all.	F	Ind.spéc.	0.48	*	500	0.26
Rich.all.	F	SDM	0.37		3500	0.31
MIW	BC	Expert 1	0.68	***	2500	0.58 **
MIW	BC	Expert 2	0.63	***	1000	0.55 **
MIW	BC	Ind.spéc.	0.68	***	500	0.49 *
MIW	BC	SDM	0.63	***	3500	0.62 ***
Rich.all.	BC	Expert 1	0.50	**	2500	0.44 *
Rich.all.	BC	Expert 2	0.51	**	500	0.39 *
Rich.all.	BC	Ind.spéc.	0.50	**	500	0.41 *
Rich.all.	BC	SDM	0.42	*	1500	0.35

TABLE 10 – Coefficients de corrélation de Spearman entre les indices génétiques ("Ind.génét.") et les métriques de connectivité ("Mét.connect.") issues des graphes paysagers. Les valeurs sont reportées pour chaque indice génétique ("Rich.all." : richesse allélique raréfiée, "MIW" : *Mean Inverse Weight*, moyenne de l'inverse des valeurs de F_{ST} entre une population et toutes les autres), chaque métrique de connectivité (Capacité, F, BC), chaque scénario utilisé pour délimiter les taches et calculer les distances-coût associées aux liens des graphes ("Expert 1" et "Expert 2" : selon dire d'expert, "Ind. spéc." : selon les indices de spécialisation de Jacobs, "SDM" : selon les probabilités de présence issues du SDM) et en fonction de la distance d à laquelle on considère une probabilité de dispersion égale à 0.05 lors du calcul des métriques. Pour chaque combinaison d'un indice génétique et d'une métrique de connectivité, nous reportons les corrélations obtenues en utilisant la distance d optimale, ainsi qu'une distance d égale à 5000 m. Le coefficient de corrélation maximum obtenu pour chaque combinaison est noté en gras. Les étoiles indiquent le degré de significativité : * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$.

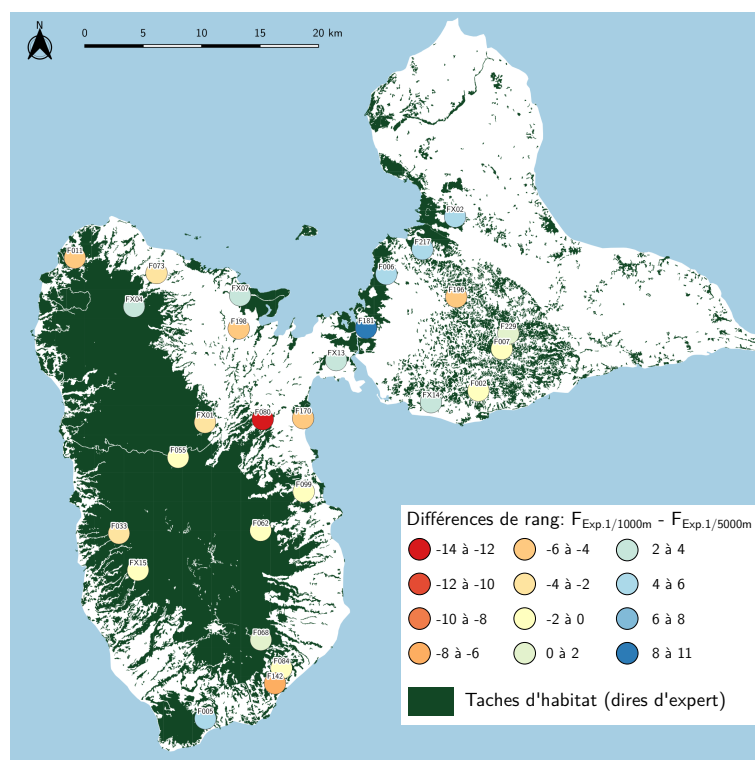


FIGURE 57 – Différences entre les rangs des taches d'habitat en termes de connectivité selon les paramètres utilisés pour calculer la métrique F. Pour chaque tache, nous avons calculé la différence entre ses rangs d'après les valeurs des métriques F calculées selon le scénario à dire d'expert 1 et en considérant que la probabilité de dispersion est égale à 0.05 à des distances d égales à 1000 m ou 5000 m. $d = 1000$ m était le paramètre optimal d'après la Table 10. Une différence de rang négative indique que la connectivité de la tache est sur-estimée avec $d = 5000$ m.

4 Discussion

Les modèles de connectivité basés sur les graphes paysagers ont rarement été validés *a posteriori* à partir de données empiriques (Foltête *et al.*, 2020 ; Godet et Clauzel, 2021 ; Kadoya, 2009). En utilisant des données génétiques, nous avons démontré dans ce chapitre que des graphes paysagers construits à partir de dires d’expert ou de données de présence/absence pouvaient refléter correctement l’influence de la connectivité des habitats sur la structure génétique des populations. Néanmoins, la relation entre la validité des graphes, leur besoin en données et la complexité des méthodes à mettre en œuvre pour les construire était parfois contre-intuitive. En effet, le graphe basé sur le SDM, construit en utilisant le plus de données et la méthode la plus complexe, reflétait parfois moins bien la structure génétique des populations que les autres. Dans les sections suivantes, nous discutons ces résultats et leurs implications pour la conservation de la biodiversité.

4.1 Les graphes paysagers sont validés empiriquement par les données génétiques

Les distances-coût associées aux liens des graphes peuvent expliquer et prédire des distances génétiques convenablement

Nous avons tout d’abord montré qu’indépendamment des méthodes de construction des graphes, les distances-coût associées à leurs liens pouvaient expliquer une part relativement importante (> 25 %) de la variance des distances génétiques entre les populations échantillonnées au niveau de leurs nœuds ; compte tenu de l’influence aléatoire de la dérive sur les distances génétiques. Elles avaient également un pouvoir prédictif non négligeable dans la plupart des cas. La relation entre distances génétiques et distances-coût avait principalement été mise en évidence avec des analyses de corrélation simples jusqu’à présent (Balbi *et al.*, 2018 ; Creech *et al.*, 2014 ; Wang *et al.*, 2008). Nous confirmons ces résultats à l’aide d’une nouvelle méthode de validation croisée appliquée à des modèles linéaires mixtes. Cette approche illustre le fait que les graphes paysagers pourraient être utiles dans le cadre d’analyses prédictives en génétique du paysage. De telles analyses renforceraient l’intérêt de cette discipline pour la conservation de la biodiversité (Keller *et al.*, 2015 ; Richardson *et al.*, 2016), mais ont rarement été réalisées (mais voir Van Strien *et al.* (2014)). Elles permettraient par exemple de prédire la distance génétique entre l’ensemble des taches d’habitat d’un réseau d’habitat sans avoir à acquérir des données génétiques dans l’intégralité de ces taches.

Les métriques de connectivité issues des graphes reflètent correctement la structure génétique des populations

Moilanen (2011) soulignait la signification écologique incertaine associée aux nombreuses métriques de connectivité développées à partir des graphes paysagers. Des auteurs ont néanmoins montré qu’un nombre réduit de métriques pourrait être suffisant pour refléter la façon dont la connectivité des habitats influence des processus écologiques tels que le recrutement local et les flux de dispersion à des échelles spatiales et temporelles plus ou moins grandes (Baranyi *et al.*, 2011 ; Rayfield *et al.*, 2011 ; Urban et Keitt, 2001). Nous avons par conséquent considéré trois métriques de connectivité complémentaires (Capacité, F, BC) et montré qu’elles étaient chacune corrélées de façon significative à la diversité génétique locale des populations et à leur degré de différenciation génétique vis-à-vis des autres populations.

Contrairement à des analyses ayant précédemment mis en évidence de telles relations (Castillo *et al.*, 2016 ; Creech *et al.*, 2014), les graphes paysagers que nous avons construits étaient indépendants de toute information issue des données génétiques. Cela renforce ainsi la significativité de ces relations entre réponses génétiques et métriques de connectivité. De plus, ces corrélations significatives impliquaient à la fois des indices de diversité génétique et de différenciation génétique relative issus de graphes génétiques. Elles confirment donc de précédents résultats obtenus séparément concernant les relations entre des métriques de connectivité similaires d'une part et la diversité génétique (Bertin *et al.*, 2017 ; Capurcho *et al.*, 2013) ou des indices de différenciation génétique calculés localement d'autre part (Peterman *et al.*, 2015). Ainsi, en accord avec la théorie (DiLeo et Wagner, 2016), les métriques associées aux nœuds et aux liens des graphes paysagers reflètent correctement les facteurs déterminant l'intensité de la dérive et des flux génétiques.

4.2 Quelles relations entre la validité écologique des graphes, leur besoin en données et la complexité des méthodes de construction et d'analyse ?

En plus de valider la modélisation basée sur les graphes paysagers, nous souhaitions comparer la validité écologique de graphes construits à partir de dires d'expert ou de données empiriques. Nous faisons l'hypothèse que le graphe issu du SDM serait le plus réaliste car cette approche permet théoriquement de s'affranchir de la subjectivité associée aux dires d'expert (Dufflot *et al.*, 2018). Néanmoins, cette hypothèse n'est que partiellement vérifiée.

Utiliser des données empiriques et des méthodes de construction complexes et coûteuses ne garantit pas un bénéfice en termes de validité écologique des graphes

Pour construire le SDM et le graphe correspondant, nous avons mobilisé 991 observations ponctuelles, des données d'occupation du sol et des données climatiques et mis en œuvre des analyses relativement complexes. Pour calculer les indices de spécialisation, nous avons utilisé ces mêmes observations, des données d'occupation du sol et une simple formule. Enfin, pour construire les graphes à dires d'expert, nous n'avons eu besoin que des données d'occupation du sol. Les méthodes de construction des graphes différaient donc à la fois par leur besoin en données et par leur complexité, le tout déterminant le coût global associé à la modélisation. Malgré ces différences, les métriques de connectivité issues des graphes construits à dires d'expert étaient souvent les plus fortement corrélées aux indices génétiques, même si les métriques issues des autres graphes étaient également corrélées significativement à ces dernières. Par ailleurs, même si le meilleur modèle expliquant les distances génétiques incluait des distances-coût basées sur les coûts issus du SDM, les distances-coût calculées à partir des indices de spécialisation permettaient d'obtenir un modèle aux performances similaires. De plus, les prédictions faites avec des distances-coût calculées à dires d'expert étaient également convenables. Tout cela semble indiquer qu'en comparaison avec l'approche basée sur le SDM, des approches plus simples basées ou non sur des données empiriques pourraient s'avérer pertinentes dans de nombreux cas. Ainsi, pour un coût de modélisation moindre, leur bénéfice en termes de réalisme écologique était à peu près équivalent.

De façon similaire, Poor *et al.* (2012) avaient comparé la coïncidence entre différents modèles de corridors de migration et des données de télémétrie obtenues en période de migration et avaient montré que des modèles de corridors de migration réalisés à dires d'expert constituaient l'approche la plus pragmatique compte tenu de l'amélioration négligeable apportée dans ce cas par l'utilisation d'un

SDM. Ainsi, même si nous confirmons que l'intégration de données empiriques à la modélisation permet d'obtenir des modèles réalistes (Kadoya, 2009), nous montrons que le ratio le plus favorable entre le coût associé à la modélisation et le bénéfice que représente sa validité écologique n'est pas associé aux modèles basés sur un SDM⁷.

Les méthodes de construction ont autant d'importance que la nature des données mobilisées

La qualité du SDM ne permet pas d'expliquer pourquoi le graphe issu de celui-ci n'était pas systématiquement le plus réaliste. En effet, le SDM avait une excellente capacité prédictive ($AUC > 0.9$), reflétant la qualité de tels modèles lorsqu'ils sont utilisés pour modéliser l'habitat d'espèces spécialistes telles que la Paruline caféïette (Hernandez *et al.*, 2006 ; McPherson et Jetz, 2007). De plus, la disponibilité de données concernant à la fois la présence et l'absence de l'espèce, et notre prise en compte d'un potentiel biais d'échantillonnage, sont d'autres facteurs expliquant potentiellement la qualité du SDM obtenu (Fletcher et Fortin, 2018). Il est à noter que l'utilisation ou non des données de présence/absence ne semble pas non plus expliquer les principales différences entre les graphes obtenus. En effet, nous avons utilisé ces mêmes données pour calculer les indices de spécialisation. Or, les métriques issues du graphe basé sur ces indices étaient plus similaires à celles issues des graphes construits à dire d'expert qu'à celles issues du graphe basé sur le SDM. Ainsi, les choix faits pour délimiter les taches et pour fixer les valeurs de coûts semblent aussi importants, sinon plus, que les données utilisées. Un tel résultat semble en contradiction avec l'idée selon laquelle l'utilisation de données empiriques permettrait de s'affranchir de choix arbitraires.

Délimiter les taches d'habitat en se basant sur les probabilités de présence produisaient moins de taches mais elles avaient une plus grande taille en raison notamment des contrastes progressifs entre les probabilités de présence dans des zones dominées par des types d'occupation du sol relativement favorables à l'espèce. Un tel résultat a déjà été observé (Bourdouxhe *et al.*, 2020 ; Godet et Clauzel, 2021), mais le résultat opposé est également possible selon la nature des effets des variables incluses dans le SDM (Stevenson-Holt *et al.*, 2014). Il peut être évité en ajoutant au SDM final des éléments paysagers tels que des infrastructures de transport pour fragmenter les taches (Ziółkowska *et al.*, 2012). Les différences de corrélations entre les indices génétiques et les métriques de connectivité issues des graphes s'expliquent potentiellement par ces différences de délimitation des nœuds.

En outre, les méthodes utilisées pour fixer les valeurs de coût avaient également une influence considérable sur les résultats. Premièrement, les scénarios Expert 1 et Expert 2 différaient uniquement par l'amplitude des valeurs de coût et par leurs contrastes. Ces deux scénarios ont néanmoins conduit à des degrés de corrélations entre métriques et à des ajustements des modèles de distances différents. Ces différences étaient plus fortes dans le cas des modèles de distances et le scénario ayant l'amplitude la plus large (1-1000) permettait d'expliquer le mieux les distances génétiques.

Par ailleurs, l'utilisation de la fonction exponentielle négative introduite par Keeley *et al.* (2016) pour convertir les probabilités de présence en valeurs de coût conduisait à des coûts relativement élevés dans les zones agricoles et artificialisées. Or, ces zones représentent des surfaces importantes en Guadeloupe. Ainsi, les distances-coût entre les populations séparées par de vastes zones agricoles

7. Cf. section 2.1.2 de la discussion générale pour une réflexion plus large concernant le rapport coût-bénéfice associé à la modélisation de la connectivité des habitats à l'aide de graphes paysagers.

et artificialisées étaient très élevées et ne permettaient pas de prédire convenablement les distances génétiques entre ces populations. Pourtant, les distances-coût basées sur les indices de spécialisation obtenues avec la même fonction exponentielle négative permettaient de mieux prédire ces mêmes distances génétiques. Cela illustre premièrement le fait que cette différence n'est pas due à la fonction exponentielle négative en elle-même mais plutôt au gradient de valeurs auquel on l'applique. De plus, cela illustre le fait que deux scénarios de coût peuvent être utiles pour prédire la différenciation génétique entre des populations de la même espèce, mais dans des zones différentes, compte tenu des erreurs de prédictions relativement faibles obtenues entre certaines paires de populations avec le scénario de coût issu du SDM. De façon similaire, [Reed et al. \(2017\)](#) concluaient que des scénarios de coût différents, basés dans leur étude sur des données de télémétrie ou des dires d'expert, pouvaient être complémentaires pour analyser la connectivité car ils étaient chacun validés par l'utilisation de données génétiques, mais pas dans les mêmes zones d'étude.

L'échelle spatiale des analyses influence fortement les résultats

Enfin, l'échelle à laquelle les liens entre populations ou taches sont considérés avait une forte influence sur les résultats. Calculer les métriques de connectivité en attribuant un poids significatif aux connexions entre taches à une échelle spatiale plus large que l'échelle optimale faisait diminuer la corrélation entre les métriques de connectivité et les indices génétiques. Selon la localisation des taches, cela tendait à sous- ou sur-estimer la connectivité des taches. De même, l'ensemble de liens considéré pour prédire les distances génétiques à partir des distances-coût influençait la qualité des prédictions, et de façon différente selon le scénario de coût. Ces résultats complexes sont en partie liés à la topographie particulière de l'île de Guadeloupe et à la répartition des forêts sur cette île. Néanmoins, ils rappellent l'importance de l'échelle à laquelle le paysage influence les réponses biologiques ("*scale of effect*") en écologie du paysage ([Jackson et Fahrig, 2012](#) ; [Stuber et Gruber, 2020](#)) et de la façon dont la topologie des populations est prise en compte pour analyser la relation entre distances génétiques et distances paysagères (cf. Chapitre 1).

4.3 Implications pour la conservation de la biodiversité

Nos résultats renforcent la pertinence des nombreuses méthodes basées sur les graphes paysagers et destinées à la conservation de la biodiversité ([Foltête et al., 2014](#) ; [Tarabon et al., 2019](#) ; [Zetterberg et al., 2010](#)). De plus, ils présentent de nouvelles façons d'utiliser et d'analyser ces graphes pour répondre aux enjeux de conservation, que des données génétiques soient disponibles ou non.

Les données génétiques nous ont ici permis d'évaluer la validité écologique de différents éléments issus des graphes. Ainsi, nous avons par exemple mis en évidence le fait que la diversité génétique des populations de Paruline caféïette est affectée positivement par la présence de surfaces forestières dans un rayon limité autour des populations. Par ailleurs, les faibles diversités génétiques mesurées dans des taches de la périphérie de BT indiquait qu'elle était affectée négativement par la subdivision des taches forestières, quand bien même de très grandes surfaces forestières sont présentes à une distance relativement faible. L'ensemble de ces résultats basés sur les graphes paysagers indiquent également les mesures complémentaires qui pourraient être prises pour conserver la variabilité génétique des populations de Paruline caféïette en Guadeloupe. Elles consisteraient par exemple à (i) protéger les taches d'habitat situées au nord-ouest de GT, (ii) à restaurer la connectivité entre les fragments forestiers situés en périphérie de BT et le massif forestier de cette partie de l'île et (iii) à restaurer des

taches d'habitat à l'ouest de l'isthme séparant BT et GT. Ces trois types de mesures reflètent la diversité des usages des graphes paysagers dans un contexte opérationnel présentée par [Foltête *et al.* \(2014\)](#).

Nous avons aussi pu distinguer des populations dont les réponses génétiques s'expliquent bien par la connectivité des habitats et la résistance de la matrice paysagère, et d'autres populations pour lesquelles les erreurs de prédiction des distances génétiques étaient systématiquement élevées. Dans ce dernier cas, des analyses ou des inventaires de terrain complémentaires pourraient se focaliser sur ces populations pour identifier d'autres facteurs potentiellement à l'origine d'une réduction des tailles de populations (des interactions interspécifiques antagonistes par exemple) ou d'une diminution des flux génétiques (par exemple une barrière paysagère non modélisée) pouvant expliquer ces réponses génétiques. Ce cas particulier illustre le fait que de façon générale, si des données génétiques sont disponibles, les limites des modèles basés sur les graphes paysagers peuvent être spatialisées pour approfondir les analyses et améliorer les mesures et inventaires destinés à la conservation des espèces.

Lorsque les données génétiques ne sont pas disponibles, les paramètres d'analyse et de construction ne peuvent pas être optimisés mais nous avons ici montré que la plupart des paramètres permettaient de prédire convenablement les réponses génétiques. Par conséquent, même s'il serait nécessaire de reproduire le même type d'analyses concernant des espèces moins spécialistes ou occupant d'autres types d'habitat ou en utilisant des données de télémétrie pour valider les graphes, ces outils semblent pertinents pour répondre aux enjeux de conservation de la connectivité des habitats.

Néanmoins, nous insistons sur le fait que la disponibilité et la qualité respectives des informations fournies par des experts ou par la littérature et des données empiriques doivent déterminer la méthode de construction des graphes. Au vu de nos résultats, il semble par exemple que l'utilisation d'un SDM pour construire un graphe paysager est à réserver aux situations dans lesquelles des dires d'expert et de simples données d'occupation du sol ne permettent pas de délimiter l'habitat de l'espèce ni même de paramétrer les coûts de déplacement relatifs associés aux différents types d'occupation du sol. En effet, les approches empiriques ne sont pas nécessairement les plus pertinentes au vu du coût associé à leur mise en œuvre et du bénéfice parfois réduit qui en résulte.

Transition

Dans la première partie de cette thèse, nous avons identifié les méthodes de construction et d'analyse des graphes génétiques et paysagers les plus adaptées aux objectifs poursuivis, tout en développant un outil pour les mettre en œuvre. Dans le cadre de deux études empiriques, nous avons comparé ces deux types de graphes et confirmé la validité écologique des graphes paysagers en mettant en évidence leurs relations avec des graphes génétiques construits de façon indépendante (comparaison *a posteriori*). Nous avons également montré que les graphes paysagers pouvaient améliorer notre compréhension des processus déterminant la structure génétique des populations. Ces deux types de graphes sont donc véritablement complémentaires.

Cette première partie a toutefois révélé la sensibilité de la modélisation de la connectivité écologique aux valeurs de coût associées à la matrice paysagère et utilisées pour construire les graphes paysagers. Elle a également mis en évidence l'influence parfois importante de la dérive génétique aléatoire sur la structure génétique. Cette influence peut se révéler problématique lorsqu'on analyse la structure génétique pour évaluer l'effet du paysage sur les flux génétiques.

Pour aller au-delà d'une simple approche comparative, la seconde partie de cette thèse visera à intégrer les graphes génétiques et paysagers en tirant profit de leur complémentarité. Nous réaliserons cette intégration en mobilisant conjointement les données génétiques et paysagères (intégration *a priori*) afin d'identifier les valeurs de coût reflétant au mieux la structure génétique des populations. L'apport des graphes dans ce contexte viendra de l'utilisation de variables caractérisant à la fois leurs nœuds et leurs liens dans des modèles gravitaires. Cette approche permettra de tester l'influence potentielle de l'hétérogénéité spatiale de la dérive génétique sur l'inférence. Pour la mise en œuvre de l'ensemble de la démarche, nous nous baserons essentiellement sur des données issues de simulations.

Il convient d'évaluer l'intérêt de cette approche dans des contextes où l'inférence est réalisable. Sa réalisation dépend notamment de la sensibilité des chemins de moindre coût et des distances-coût aux variations des scénarios de coût que l'on vise à distinguer lors de l'inférence. Par conséquent, nous analysons cette sensibilité dans le chapitre suivant qui constitue une étape préalable à l'intégration des graphes génétiques et paysagers pour l'inférence de la résistance du paysage.

Partie II

Intégration de graphes génétiques et paysages

Chapitre 5

Analyse de la sensibilité des chemins de moindre coût et des distances-coût aux scénarios de coût

Résumé

Le succès des mesures de conservation de la biodiversité basées sur le maintien de la connectivité écologique dépend de la qualité de la modélisation spatiale des mouvements des espèces. Le calcul de chemins de moindre coût permet d'identifier les chemins de dispersion les plus probables et apporte deux types d'information : (i) la localisation spatiale des chemins de moindre coût et (ii) le coût accumulé le long de ces chemins, aussi appelé distance-coût. Cette modélisation spatiale nécessite qu'un coût soit attribué à chaque type d'occupation du sol. Nous avons évalué la sensibilité des chemins et des distances-coût aux scénarios de coût en considérant un ensemble de paysages réels et une gamme de scénarios variée. Les chemins de moindre coût étaient plus sensibles aux scénarios de coût que ne l'étaient les distances-coût. De plus, des matrices de distances-coût fortement corrélées entre elles étaient parfois obtenues avec des scénarios de coût très différents. Même si les valeurs de coût et les propriétés de chaque scénario influençaient les résultats de la modélisation de façon très significative, les variables décrivant la composition et la configuration des paysages expliquaient également leurs variations. À partir de ces résultats, nous formulons des conseils méthodologiques pour l'application de la modélisation basée sur les chemins de moindre coût aux études d'écologie théorique d'une part, et à la conservation de la biodiversité d'autre part.

Cet article a été re-soumis après modifications mineures à l'*International Journal of Geographical Information Science* en septembre 2021 :

Savary, P., Foltête, J. C. & Garnier, S. Cost distances and least cost paths respond differently to cost scenario variations - A sensitivity analysis of ecological connectivity modeling. Submitted to *International Journal of Geographical Information Science*

*Je sillonne tant de terres, passe au pied des collines où fument le camp des hommes.
Je nage dans une rivière immense, le courant m'entraîne vers l'autre rive. Je ressors couvert de boue.
Des maisons partout, des chiens aboient à tue-tête, des engins crachent leurs bruits.
Je passe sans me faire voir, à grandes enjambées. Je dors contre les fourrés.*

Caroline Audibert

1 Introduction

Pour modéliser la connectivité écologique des habitats, il est fréquent de chercher au sein d'un paysage les chemins censés minimiser le coût total de déplacement des individus entre leurs taches d'habitat (Adriaensen *et al.*, 2003). Ces chemins de moindre coût (*Least Cost Paths*, noté LCPs par la suite) apportent deux types d'information : (i) leur localisation spatiale et (ii) le coût accumulé le long de ces chemins, aussi appelé distance-coût (*Cost Distance*, noté CD par la suite). Même si des approches basées sur des principes similaires ont été développées (Marrotte et Bowman, 2017 ; McRae, 2006 ; Panzacchi *et al.*, 2016), l'utilisation encore fréquente de la modélisation basée sur les chemins de moindre coût reflète leur pertinence pour la localisation de passages à faune (Beier *et al.*, 2008 ; Carroll *et al.*, 2012 ; Sawyer *et al.*, 2011) ou pour les analyses statistiques visant à tester des hypothèses en écologie (Balbi *et al.*, 2019 ; Mony *et al.*, 2018).

Dans ce cadre, les valeurs de coût associées aux différents types d'occupation du sol sont une donnée d'entrée clé (Zeller *et al.*, 2012). Lorsque leur choix se base sur la littérature scientifique ou sur des dires d'expert (voir par exemple Clevenger *et al.* (2002) ou Pullinger et Johnson (2010)), il est en partie arbitraire, ce qui peut influencer les résultats obtenus. D'une part, des scénarios différents de par les valeurs des coûts attribués aux types d'occupation du sol et par l'ampleur de leurs contrastes peuvent produire des différences notables dans les résultats (Gonzales et Gergel, 2007 ; Murekatete et Shirabe, 2018 ; Rayfield *et al.*, 2010). D'autre part, l'ordre de ces valeurs influence significativement les CDs et LCPs (Beier *et al.*, 2009). Cet ordre reflète la préférence des individus pour un type d'occupation du sol plutôt qu'un autre et est donc un paramètre ayant une signification écologique importante. Néanmoins, il faudrait tester l'éventualité que des matrices de CDs fortement corrélées ou des ensembles de LCPs proches spatialement proviennent de scénarios classant les coûts des types d'occupation du sol dans des ordres différents. En effet, bien que Murekatete et Shirabe (2018) aient évalué l'influence de la distribution des valeurs de coût sur les LCPs, leur étude était basée sur des paysages simulés et des gradients continus de variation des coûts. En somme, une évaluation de la sensibilité de la modélisation aux changements de scénarios mériterait d'être effectuée à partir de paysages réels et de valeurs de coût discrètes, reflétant les pratiques les plus couramment mises en œuvre.

Outre les valeurs de coût, la structure du paysage en elle-même influence la sensibilité des distances-coût aux scénarios de coût, d'après les quelques études menées à partir de paysages simulés (Bowman *et al.*, 2020 ; Cushman *et al.*, 2013 ; Marrotte et Bowman, 2017 ; Murekatete et Shirabe, 2018 ; Rayfield *et al.*, 2010 ; Simpkins *et al.*, 2017). Par exemple, les distances-coût sont moins affectées par les variations de scénarios de coût lorsque les zones d'habitat sont fortement connectées (Cushman *et al.*, 2013). Cependant, réaliser de pareilles évaluations à partir d'une gamme diversifiée de paysages réalistes permettrait d'identifier plus précisément les contextes dans lesquels la modélisation est la plus sensible aux scénarios de coût.

De la même façon, la sensibilité de la localisation spatiale des LCPs aux variations de scénarios a fait l'objet de peu d'études. Beier *et al.* (2009) ont modélisé le corridor assurant la connectivité entre deux aires protégées californiennes en utilisant plusieurs scénarios de coût alternatifs issus de dires d'experts. Ils concluaient que ces corridors coïncidaient le plus souvent. Cependant, les valeurs de coût utilisées étaient assez similaires et systématiquement comprises entre 1 et 10, une gamme de variation réduite au regard des coûts basés sur des observations de terrain (Khimoun *et al.*, 2017 ;

Ruiz-González *et al.*, 2014). Pullinger et Johnson (2010) sont eux arrivés à des conclusions plus pessimistes en cherchant à reproduire des traces GPS suivies par des caribous à l'aide de LCPs calculés selon plusieurs scénarios de coût, suggérant que les LCPs y étaient relativement sensibles. Toutefois, peu de scénarios différents étaient utilisés dans ces travaux et il faudrait les reproduire en utilisant des gammes de scénarios plus variées.

Afin de renforcer le réalisme écologique des valeurs de coût, il est souvent conseillé de les déterminer à partir de la relation entre des données biologiques et des variables environnementales (Kadoya, 2009 ; Pressey, 2004). Alors que l'utilisation de données de présence pour ce faire est sujette à précautions (cf. Chapitre 4), l'utilisation de données génétiques ou de télémétrie fournit des résultats satisfaisants d'après l'analyse de Zeller *et al.* (2018). Des approches statistiques ont donc été développées pour inférer des valeurs de coût à partir de la relation entre des distances-coût et des distances génétiques (Peterman *et al.*, 2019 ; Peterman et Pope, 2020 ; Zeller *et al.*, 2016). On suppose dans ce cas que le scénario de coût maximisant l'ajustement de cette relation statistique reflète les distances-coût telles qu'elles sont perçues par les individus lors des mouvements de dispersion à l'origine des flux génétiques. Les valeurs de coût associées à chaque type d'occupation du sol dans ce scénario optimisé deviennent alors un résultat supplémentaire de la modélisation. Leurs valeurs absolues, leur ordre et leurs contrastes sont en eux-mêmes des informations interprétables écologiquement. Par exemple, ils indiquent qu'un type d'occupation du sol est plus résistant qu'un autre, et combien de fois plus résistant il est, entre autres interprétations possibles (voir Khimoun *et al.* (2017) ou Ruiz-González *et al.* (2014)).

Cependant, de telles approches statistiques devraient être incapables d'identifier le scénario de coût reflétant le plus fidèlement la réalité écologique lorsque plusieurs scénarios de coût alternatifs mènent à des matrices de CDs fortement corrélées. En effet, on pourrait dans ce cas s'attendre à ce qu'elles soient corrélées de façon similaire aux distances génétiques utilisées pour l'inférence (Zeller *et al.*, 2016), rendant celle-ci impossible ou du moins incertaine. Par ailleurs, nous ne savons pas à quel point un scénario de coût issu de l'optimisation de la relation statistique entre distances-coût et distances génétiques est à même de localiser de façon correcte les chemins empruntés par les individus. Deux matrices de CDs fortement corrélées mais obtenues avec des scénarios différents pourraient tout à fait correspondre à des LCPs spatialement distincts. Utiliser l'une ou l'autre de ces matrices dans une analyse statistique intégrant des données génétiques n'affecterait que marginalement les résultats et on leur affecterait une vraisemblance similaire. Au contraire, le choix d'un des deux scénarios de coût affecterait significativement les LCPs, et conduirait potentiellement à des mesures de conservation inadéquates. Évaluer l'éventualité d'une telle différence de sensibilité des LCPs et des CDs aux scénarios de coût et en identifier les causes est donc nécessaire.

Dans ce chapitre, nous avons évalué la sensibilité aux variations des scénarios de coût de la localisation spatiale des chemins de moindre coût (LCPs) d'une part, et des distances-coût (CDs) qui leur sont associées d'autre part. Pour ce faire, nous avons échantillonné aléatoirement des paysages réels, localisé des points au sein d'eux et calculé les LCPs et leurs CDs selon une gamme contrastée de scénarios de coût. Nous avons considéré arbitrairement qu'un de ces scénarios reflétait la "réalité" écologique afin d'évaluer (i) la coïncidence spatiale entre les "vrais" LCPs et les LCPs alternatifs et (ii) la corrélation entre la "vraie" matrice de CDs et les matrices de

CDs alternatives. Nous avons enfin mis en œuvre des analyses statistiques pour identifier les facteurs de la sensibilité des LCPs et CDs aux scénarios de coût. Elles ont permis d'identifier les contextes paysagers et les caractéristiques des scénarios de coût influençant cette sensibilité.

2 Méthodes

2.1 Sélection des paysages

Pour que nos résultats se traduisent en préconisations pour la modélisation de LCPs dans des contextes réels, nous avons considéré des paysages de France métropolitaine échantillonnés aléatoirement à partir d'une base de données d'occupation du sol homogène à l'échelle de ce territoire (Inglada *et al.*, 2017). L'application d'un ensemble de critères de sélection (cf. Annexe A7) a restreint notre échantillon initial à 77 paysages de 30 km × 30 km avec une résolution spatiale de 10 m. Ils comprenaient cinq types d'occupation du sol : (1) forêts, (2) prairies, (3) cultures agricoles, (4) zones artificialisées et (5) autres (cours et plans d'eau et autres types d'occupation du sol).

2.2 Création des scénarios de coût

La sensibilité de la modélisation aux scénarios de coût est plus directement évaluable en considérant un scénario de coût de référence. Assimiler arbitrairement cette référence à une réalité écologique permet de déterminer à quel point les résultats de la modélisation s'éloignent de cette réalité lorsque les scénarios de coût varient. Nous avons donc considéré un scénario de référence (noté "vrai" scénario) dans lequel les coûts associés aux forêts, prairies, cultures, zones artificielles et autres types d'occupation du sol étaient respectivement égaux à 1, 10, 100, 1000 et 100. Aussi bien leurs valeurs absolues que leur gamme de variation reflétaient les capacités de dispersion dans le paysage d'une espèce forestière spécialiste, telles qu'estimées lors d'études empiriques (Gurrutxaga *et al.*, 2010 ; Khimoun *et al.*, 2017 ; Pérez-Espona *et al.*, 2008 ; Ruiz-González *et al.*, 2014 ; Schadt *et al.*, 2002 ; Wang *et al.*, 2008).

En parallèle, nous avons constitué 100 scénarios de coût alternatifs en tâchant de faire varier à la fois l'ordre des types d'occupation du sol et les contrastes entre les valeurs de coût. Pour cela, nous avons considéré cinq ensembles de quatre valeurs : [1, 1, 11, 1000], [1, 4, 11, 1000], [4, 63, 317, 1000], [63, 250, 563, 1000], [250, 500, 750, 1000] (cf. Annexe A7). En utilisant chacun d'eux, nous avons attribué leurs valeurs aléatoirement aux forêts, prairies, cultures et zones artificialisées et choisi aléatoirement 100 scénarios parmi ces combinaisons. La catégorie "autres types d'occupation du sol" conservait un coût égal à 100 dans tous les scénarios pour limiter le nombre de combinaisons d'une part, et car son absence de certains paysages aurait limité l'interprétation des résultats d'autre part.

2.3 Modélisation des chemins de moindre coût

Nous avons sélectionné aléatoirement 50 points séparés d'au moins 500 m dans les taches de forêt de chaque paysage. Nous avons ensuite calculé les LCPs et les matrices de CDs entre chacune des paires de points pour chaque scénario de coût dans chacun des paysages (Figure 58A). Nous avons créé des zones tampons de 200 m de part et d'autre de chaque LCP avant de convertir ces zones en polygones de largeurs égales (400 m).

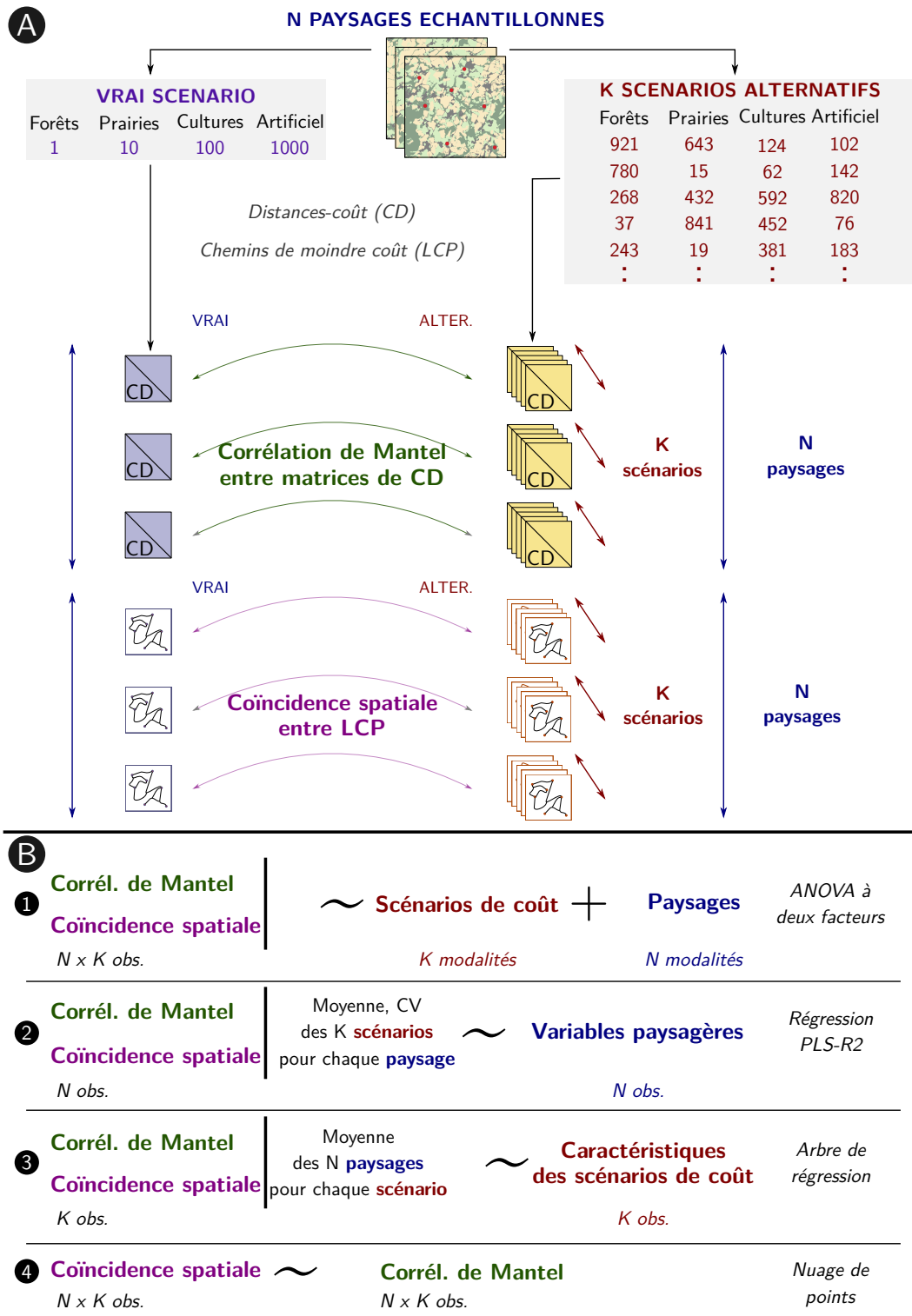


FIGURE 58 – Protocole d’analyse de la sensibilité des chemins de moindre coût (LCP, *least-cost path*) et distances-coût (CD, *cost distance*) aux scénarios de coût. (A) Comparaisons spatiales ou basées sur les corrélations de Mantel entre les LCPs et les matrices de CDs, respectivement, obtenus avec le vrai scénario (VRAI) et les $K = 100$ scénarios alternatifs (ALTER) dans $N = 77$ paysages. (B) Analyses statistiques réalisées pour évaluer et identifier les facteurs de la sensibilité de la modélisation aux scénarios de coût. (1) ANOVA à deux facteurs évaluant l’influence respective du paysage et du scénario sur (i) la coïncidence spatiale des LCPs et (ii) les corrélations de Mantel entre matrices de CDs. (2) Régression PLS-R2 évaluant l’influence de la composition et de la configuration du paysage sur la sensibilité de la modélisation aux scénarios dans chaque paysage (moyenne et coefficient de variation pour les différents scénarios pour chaque paysage). (3) Arbre de régression identifiant les caractéristiques des scénarios de coût expliquant les coïncidences spatiales et les corrélations de Mantel (valeurs moyennes pour les différents paysages pour chaque scénario). (4) Nuage de points représentant la relation entre coïncidences spatiales et corrélations de Mantel pour chaque combinaison d’un paysage et d’un scénario.

2.4 Coïncidences spatiales des LCPs et corrélations entre matrices de CDs

Tout d'abord, nous avons mesuré la proportion de la surface de chaque zone tampon autour des vrais LCPs intersectée spatialement par les zones tampons créées autour de chaque LCP alternatif. Nous avons calculé ce degré de coïncidence spatiale moyen pour chaque scénario alternatif dans chaque paysage (100 valeurs par paysage). En parallèle, nous avons évalué le coefficient de corrélation de Mantel r entre chaque matrice de CDs alternative et la vraie matrice de CDs (100 valeurs r par paysage) (Figure 58A).

2.5 Description des contextes paysagers et des scénarios de coût

Pour évaluer la sensibilité de la modélisation aux scénarios de coût en fonction du contexte paysager, nous avons calculé les variables suivantes pour décrire la composition et la configuration des 77 paysages :

- Proportions de forêts, prairies, cultures et zones artificialisées dans le paysage
- Indice de Shannon mesurant la diversité de l'occupation du sol (divisé par $\log(n)$ avec n le nombre de types d'occupation du sol)
- Indice de contigüité mesurant le degré d'agrégation des pixels d'un même type d'occupation du sol (une valeur pour le paysage dans son ensemble)
- Degré de fragmentation ("clumpy"), complexité de forme et nombre des taches de forêt et de zones artificialisées
- Coût total de l'ensemble des pixels du paysage

En parallèle, pour comprendre quelles propriétés des différents scénarios de coût étaient responsables des variations de résultats de modélisation lorsqu'elles étaient modifiées d'un scénario à un autre, nous avons décrit chaque scénario à l'aide des critères suivants :

- Le scénario classe-t-il les coûts des types d'occupation du sol dans le même ordre que le vrai scénario ? (variable binaire : Oui/Non)
- Les contrastes entre valeurs sont-ils les mêmes que dans le vrai scénario ? (variable binaire : Oui/Non) Nous avons considéré que les valeurs [1, 4, 101, 1000] reflétaient le mieux les vrais contrastes.
- Les différences relatives de coûts entre les paires de types d'occupation du sol sont-elles les mêmes que dans le vrai scénario ? (par exemple : Forêt < Prairies : Oui ou Non)

2.6 Analyses statistiques des facteurs de la sensibilité de la modélisation aux scénarios de coût

Les valeurs du (i) coefficient de corrélation de Mantel entre les matrices de CDs vraies et alternatives et (ii) les proportions de coïncidence spatiale entre les LCPs alternatifs et les vrais LCPs étaient supposées refléter la sensibilité de la modélisation aux scénarios de coût. Plus la variation de ces métriques était grande dans un paysage donné pour les différents scénarios, plus la sensibilité de la modélisation était forte dans ce paysage. De même, pour un scénario de coût donné, ces métriques prenaient des valeurs au sein des différents paysages reflétant la similarité globale de ce scénario vis-à-vis du vrai scénario de coût. En calculant ces deux métriques dans chaque paysage pour chaque scénario

($2 \times 77 \times 100$ valeurs), nous pouvions comparer leurs sensibilités, évaluer dans quelle mesure le contexte paysager les expliquait et identifier les caractéristiques des scénarios de coût qui les rendaient similaires au vrai scénario. Par conséquent, nous avons tout d'abord réalisé deux ANOVA à deux facteurs, une pour chaque métrique, en considérant le scénario et le paysage comme des variables qualitatives expliquant leurs variations (Figure 58 B1). Cette analyse nous a permis d'évaluer les contributions respectives de la structure du paysage et des scénarios de coût aux variations des deux métriques.

Nous avons ensuite cherché quelles variables paysagères pouvaient expliquer les différences de sensibilité aux scénarios. Pour ce faire, nous avons calculé la moyenne et le coefficient de variation des corrélations de Mantel et des coïncidences spatiales obtenues pour tous les scénarios dans chacun des paysages (2×77 valeurs). Une valeur moyenne forte indiquait qu'indépendamment du scénario de coût, les LCPs et les CDs alternatifs étaient relativement proches des vrais LCPs et CDs dans un paysage donné. Un fort coefficient de variation indiquait que dans un paysage donné, les variations entre scénarios étaient importantes. Ces deux indicateurs numériques ont chacun été modélisés en fonction des variables paysagères à l'aide d'une régression PLS-R2 (Figure 58B, Annexe B2). Une valeur de Q^2 associée à une composante de l'espace factoriel issu de cette analyse supérieure à 0.0975 indiquait que cette composante avait un effet significatif sur les variables réponses.

Finalement, nous avons cherché à identifier les caractéristiques des scénarios de coût alternatifs responsables des variations par rapport aux résultats obtenus avec le vrai scénario. Nous avons calculé les valeurs moyennes des corrélations de Mantel et des coïncidences spatiales pour un scénario de coût donné dans les différents paysages (2×100 valeurs). De fortes valeurs indiquaient qu'indépendamment du paysage, un scénario alternatif donné engendrait des LCPs et CDs similaires à ceux obtenus avec le vrai scénario. Nous avons modélisé l'influence des valeurs absolues des coûts, de leur ordre, de leurs contrastes, du gradient des valeurs et des différences entre chaque scénario et le vrai (cf. section 2.5) sur les valeurs moyennes des métriques. Nous avons pour cela réalisé des arbres de régression (un par métrique). Cette méthode consiste à délimiter l'espace des variables explicatives en un nombre limité de branches et feuilles terminales. La variable réponse prend pour chaque branche ou feuille la valeur moyenne des observations concernées par cette région de l'espace des variables explicatives (cf. Annexe B4).

3 Résultats

3.1 Structure des paysages sélectionnés

La composition et la configuration de l'occupation du sol des 77 paysages sélectionnés étaient très contrastées (Figure 59). Certains paysages étaient très diversifiés avec des taches d'occupation du sol relativement petites, tandis que d'autres comportaient de vastes surfaces forestières ou agricoles, entre autres contrastes.

3.2 Influence relative des valeurs de coût et de la structure du paysage sur la sensibilité de la modélisation

Les corrélations de Mantel étaient comprises entre -0.40 et 0.99 et les coïncidences spatiales entre 7 et 94 % (Figure 60). La variance des coïncidences spatiales pour les différentes combinaisons d'un

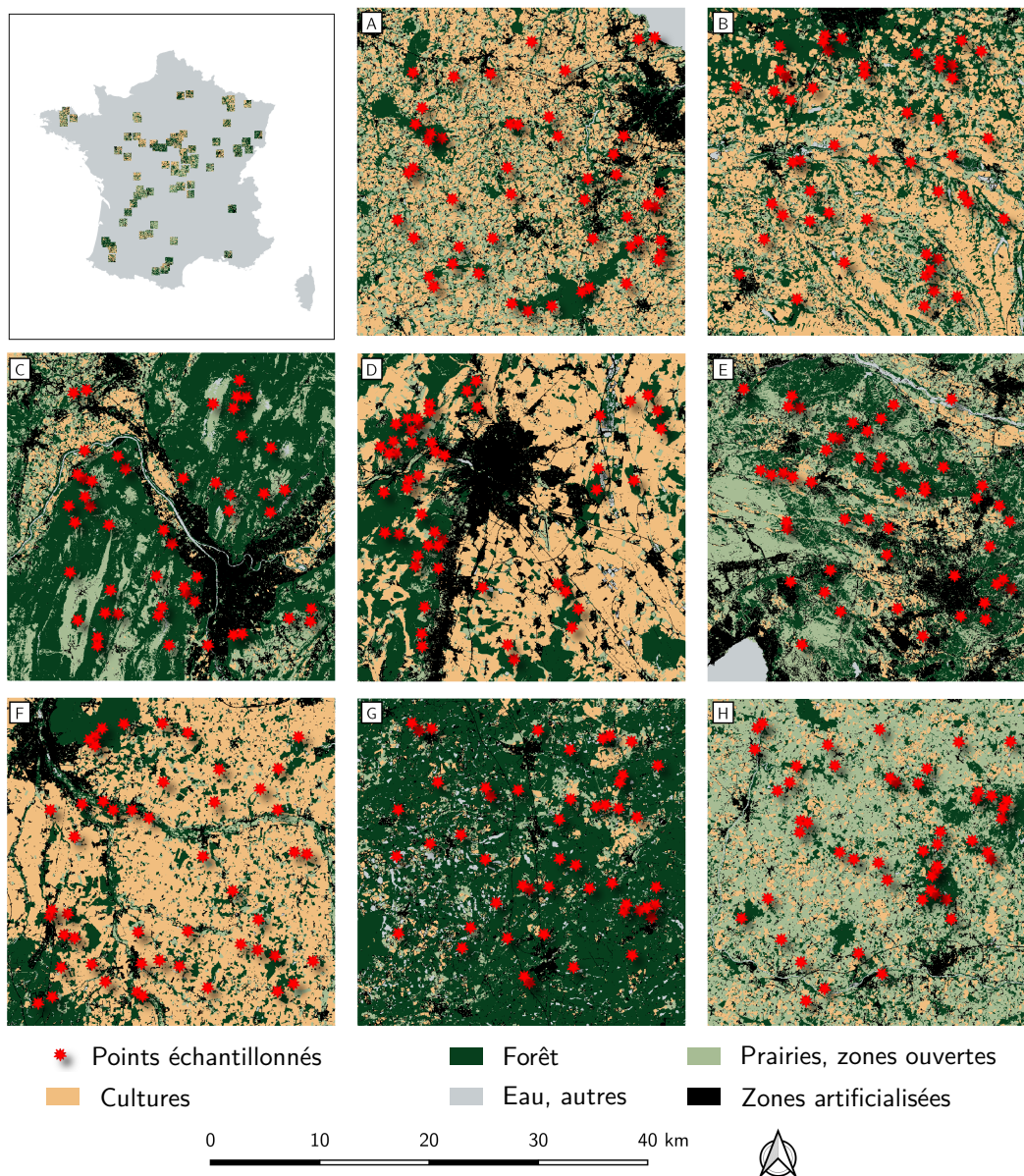


FIGURE 59 – Résultats de la sélection de paysages : 77 paysages de 30 km × 30 km avec une résolution spatiale de 10 m ont été échantillonnés aléatoirement en France métropolitaine, en appliquant les critères de sélection suivants à l'échantillonnage initial : > 15 % de forêts, > 5 % de prairies, > 5 % de cultures agricoles, > 2 % de zones artificialisées. 50 points ont été échantillonnés aléatoirement dans les zones de forêt, séparés d'au moins 500 m. Les cartes de cette figure présentent huit exemples contrastés (A à H).

paysage et d'un scénario de coût était davantage liée au scénario de coût qu'au paysage dans lequel il était utilisé (valeurs F issues de l'ANOVA : 498.17 *vs* 33.71, respectivement). Néanmoins, ces deux facteurs avaient chacun une influence très significative et expliquaient à eux deux 87 % de la variance totale. Cependant, l'utilisation d'un scénario de coût dans différents paysages produisait des variations de coïncidence spatiale des LCPs moins importantes que celles de différents scénarios dans un même paysage (Figures 60A et 60B). Ces deux facteurs expliquaient une part moindre de la variabilité des corrélations de Mantel (79 %). Le scénario de coût avait là également une influence supérieure à celle du paysage (valeurs F : 235.39 *vs* 67.34, respectivement, toutes deux très significatives, Figures 60C et 60D).

La plus forte sensibilité des coïncidences spatiales par rapport aux corrélations de Mantel s'illustre par une diminution rapide des coïncidences spatiales médianes calculées pour chaque scénario (Figure 60A) par rapport à celle plus lente observée pour les corrélations de Mantel médianes (Figure 60B). En revanche, en considérant la distribution de ces deux métriques dans un paysage donné pour les différents scénarios, on observait moins de variations en termes de coïncidences spatiales qu'en termes de corrélations de Mantel (Figures 60B et 60D). En effet, les coïncidences spatiales calculées dans un paysage donné pour les différents scénarios étaient globalement faibles (Figure 60B). Les corrélations de Mantel étaient plus variables. Elles prenaient des valeurs systématiquement élevées dans certains paysages tandis que leur gamme de variation était bien plus large dans d'autres.

3.3 Influence de la structure du paysage sur la sensibilité de la modélisation

Les régressions PLS-R2 ont identifié les variables paysagères responsables du fait que la modélisation était plus sensible aux scénarios de coût dans certains paysages que dans d'autres. Une seule composante expliquait la moyenne et le coefficient de variation des métriques (coïncidence spatiale, corrélations de Mantel) dans les deux régressions faites. Le degré moyen de coïncidence spatiale des LCPs calculés pour différents scénarios de coût dans un paysage donné tendait à être plus élevé dans les paysages comportant de vastes zones forestières et des taches d'occupation du sol relativement grandes. À l'inverse, il était plus faible lorsque le coût total du paysage, la diversité de l'occupation du sol, le nombre de taches forestières et artificialisées et la proportion de zones artificialisées étaient importants (Figure 61A).

De façon similaire, dans un paysage donné, la moyenne des corrélations de Mantel mesurées entre la vraie matrice de CDs et les matrices de CDs alternatives tendait à augmenter avec le degré d'agrégation des types d'occupation du sol. Leur coefficient de variation était corrélé négativement à leur moyenne. Cela signifie que c'est dans les paysages où les matrices de CDs alternatives étaient en moyenne les moins corrélées à la vraie matrice qu'elles variaient le plus pour différents scénarios. Ces variations tendaient à être élevées, et les corrélations moyennes faibles, lorsque le coût total du paysage était élevé et que l'occupation du sol y était diversifiée.

3.4 Influence des caractéristiques des scénarios de coût sur la sensibilité de la modélisation

Les arbres de régression ont identifié les caractéristiques des scénarios de coût dont les variations affectaient le plus les LCPs et les CDs dans l'ensemble des paysages (Figure 62). La première sépa-

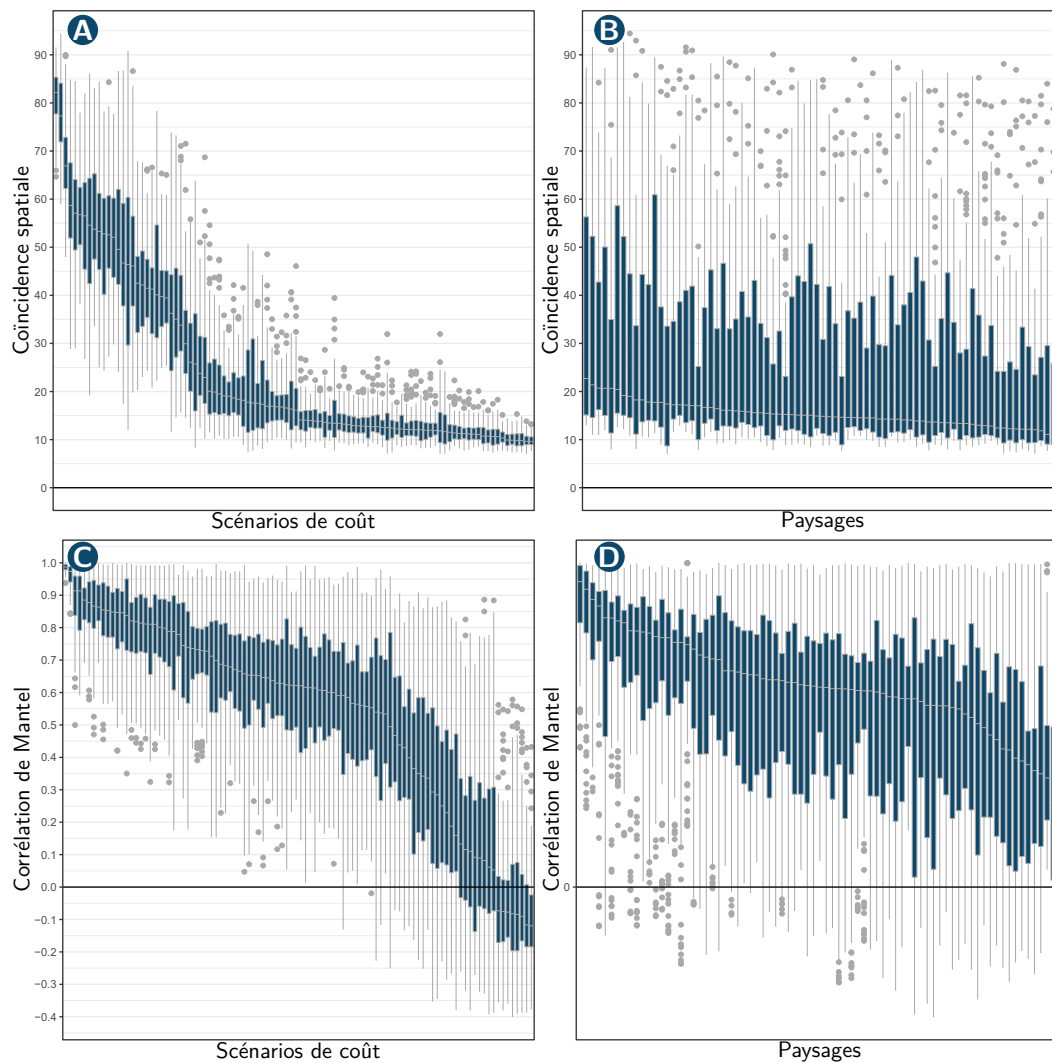


FIGURE 60 – Distribution statistique des coïncidences spatiales entre LCPs (A, B) et des corrélations de Mantel entre CDs (C, D) selon le scénario ou le paysage. Les scénarios de coût (A, C) et les paysages (B, D) sont placés dans l'ordre décroissant des valeurs médianes de coïncidence spatiale (A, B) ou de corrélations de Mantel (C, D) le long de l'axe des abscisses. Quand la distribution est affichée pour un scénario de coût donné (A, C), les 100 boîtes à moustaches incluent 77 valeurs correspondant chacune à un paysage, tandis que lorsqu'elle est affichée pour un paysage donné (B, D), les 77 boîtes à moustaches incluent 100 valeurs correspondant à un scénario de coût chacune.

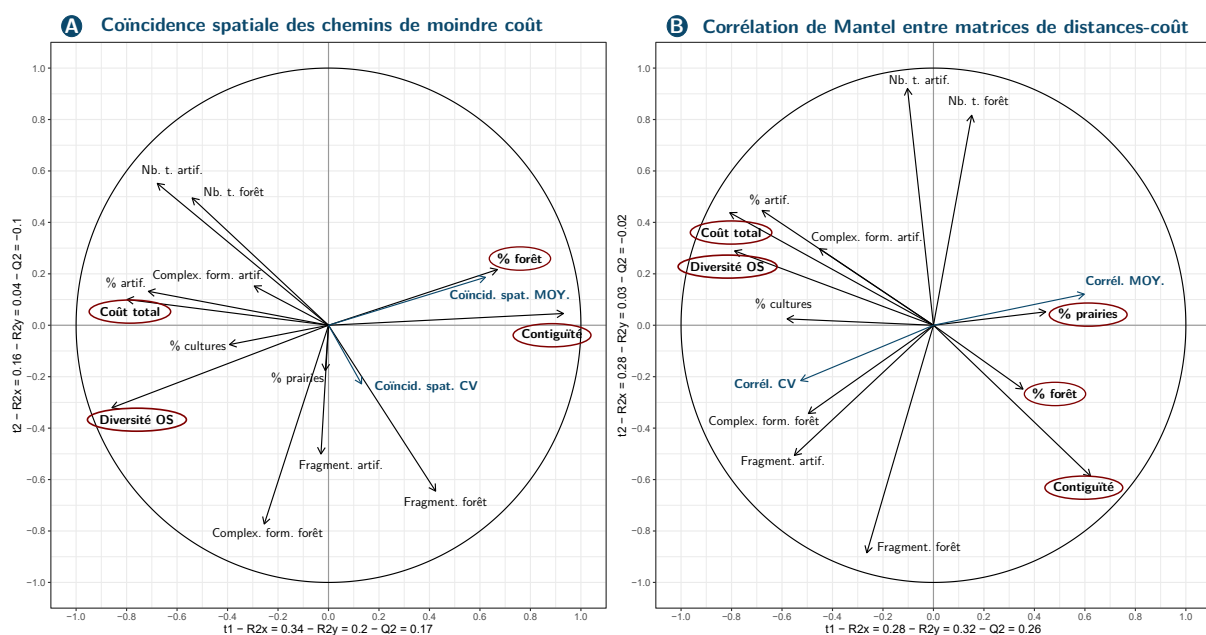


FIGURE 61 – Projection des variables réponses (notées en bleu) et explicatives (notées en noir) des régressions PLS-R2 dans l'espace factoriel défini par les deux premières composantes (t1, t2). Résultats obtenus en modélisant la moyenne et le coefficient de variation de la coïncidence spatiale des LCPs (A) ou du coefficient de corrélation de Mantel entre matrices de CDs (B) dans chaque paysage pour les différents scénarios de coût en fonction des variables paysagères.

ration dans les deux arbres de régression créés concernait les valeurs de coût attribuées aux forêts. Lorsqu'un scénario attribuait une valeur de coût inférieure à 82 aux forêts, la coïncidence spatiale était de l'ordre de 40 % (Figure 62A). Au contraire, lorsqu'il attribuait un coût inférieur à 875 (donc différent de 1000) aux forêts, la corrélation de Mantel moyenne entre les différents paysages valait en moyenne 0.63 (Figure 62B). Dans le cas où ce coût était égal à 1000, si les coûts attribués aux prairies, cultures et zones artificialisées étaient issus des gradients [63, 250, 563] ou [250, 500, 750], les corrélations de Mantel prenaient tout de même une valeur moyenne de 0.55. Ces corrélations chutaient à 0.03 en moyenne lorsque ces autres coûts étaient issus des gradients [1, 1, 11], [1, 4, 101] ou [4, 63, 317], faisant du gradient des valeurs de coût le second critère le plus important déterminant les corrélations de Mantel entre CDs.

Au contraire, le second critère déterminant la coïncidence spatiale des LCPs était le coût relatif attribué aux deux types d'occupation du sol les plus favorables, à savoir les forêts et les prairies (Figure 62A). Lorsque les forêts étaient plus résistantes que les prairies, la coïncidence spatiale entre les vrais LCPs et les LCPs alternatifs était faible. Dans ces deux arbres de régression, les variables binaires caractérisant les différences entre chaque scénario alternatif et le vrai scénario n'étaient pas conservées comme des critères clés.

3.5 Relation entre la coïncidence spatiale entre LCPs et la corrélation entre matrices de CDs

Le degré de coïncidence spatiale des LCPs et la corrélation entre matrices de CDs mesurés pour chaque combinaison d'un paysage et d'un scénario étaient liés statistiquement par une relation non linéaire (Figure 63). La coïncidence spatiale entre les vrais LCPs et les LCPs alternatifs n'était supérieure à 65 % que lorsque la matrice de CDs alternative correspondante était fortement corrélée à la vraie matrice de CDs ($r > 0.5$). Toutefois, le degré de corrélation entre ces matrices n'était pas

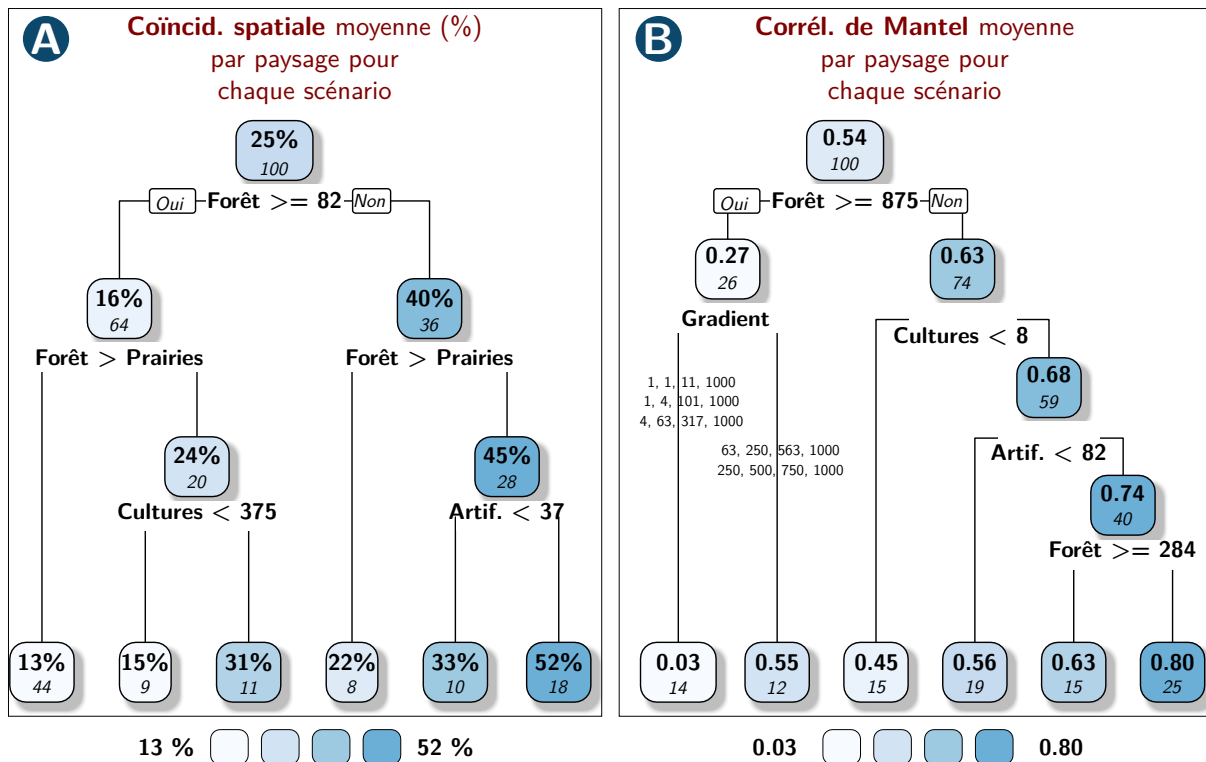


FIGURE 62 – Arbres de régression expliquant la coïncidence spatiale moyenne (A) ou le coefficient de corrélation de Mantel moyen (B), calculés pour les différents paysages pour chaque scénario de coût, en fonction des caractéristiques des scénarios de coût. À chaque nœud (séparation de branches), le critère affiché est vérifié dans toutes les feuilles issues de la branche située à gauche de la séparation, tandis que son contraire est vérifié à droite de la séparation. Les boîtes colorées au niveau des nœuds indiquent la valeur moyenne (en gras) de la variable réponse dans toutes les feuilles issues de ces nœuds, et le nombre d'observations, c'est-à-dire de scénarios de coût, correspondantes (en italique).

une bonne indication du degré de coïncidence spatiale des chemins. En effet, des coïncidences spatiales inférieures à 20 % étaient fréquemment obtenues alors que les matrices de CDs correspondantes étaient très fortement corrélées ($r > 0.9$, Figure 63 et Table 11). En outre, des coïncidences spatiales supérieures à 80 % étaient principalement observées lorsque les types d'occupation du sol étaient classés selon le même ordre de résistance que dans le vrai scénario de coût. On note néanmoins que des scénarios ordonnant les coûts de ces types d'occupation du sol de façon incorrecte permettaient parfois d'atteindre des coïncidences spatiales supérieures à 90 % (Figure 63).

Le scénario de coût alternatif produisant les plus fortes corrélations de Mantel avec la vraie matrice de CDs attribuait des coûts égaux à 1, 4, 101 et 1000 aux forêts, prairies, cultures et zones artificialisées, respectivement (Table 11). Il était en apparence le plus proche du scénario de référence ([1, 10, 100, 1000]) en termes d'ordre et de contraste des valeurs. Pourtant, ce n'est pas ce scénario mais le scénario [4, 63, 317, 1000] qui produisait les coïncidences spatiales les plus fortes. Alors que les scénarios produisant les coïncidences spatiales les plus fortes attribuaient toujours un coût plus faible aux forêts qu'aux prairies, parmi les dix scénarios produisant les plus fortes corrélations de Mantel, deux scénarios attribuaient un coût plus faible aux prairies qu'aux forêts.

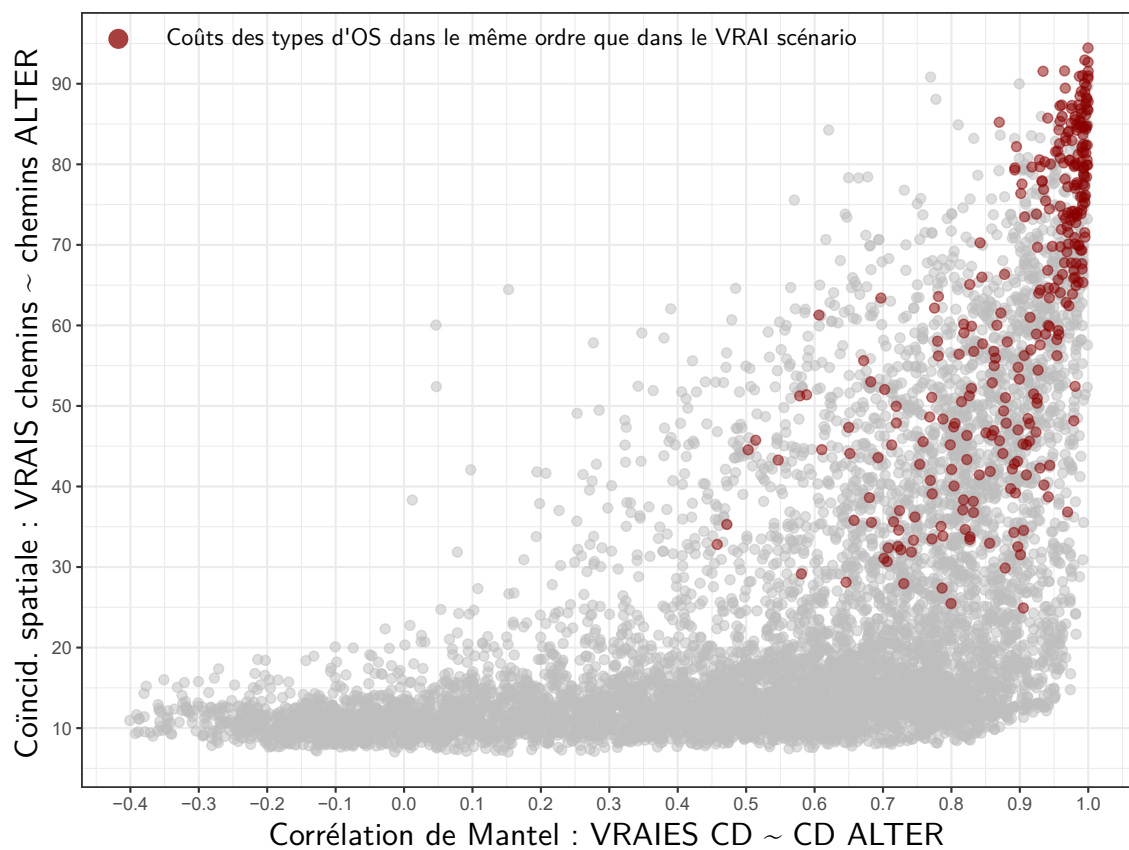


FIGURE 63 – Relation entre la corrélation de Mantel entre matrices de CDs (axe des abscisses) et la coïncidence spatiale entre LCPs (axe des ordonnées). La coïncidence spatiale entre LCPs est exprimée comme la proportion de la zone tampon de 200 m de large de chaque côté du LCP de référence qui coïncide avec la même zone tampon créée à partir de chaque LCP alternatif. $n = 7700$, correspondant au nombre de combinaisons de 100 scénarios de coût dans 77 paysages. Les points rouges correspondent aux valeurs obtenues avec les scénarios de coût classant les types d'occupation du sol dans le même ordre que le vrai scénario.

Forêts	Prairies	Cultures	Artif.	For.<Prair.	Ordre	% coïnc. spat.	<i>r</i> Mantel
1	4	101	1000	Oui	Même	77.81	0.98
4	63	317	1000	Oui	Même	81.00	0.96
1	4	1000	101	Oui	Diff.	68.04	0.88
4	1	101	1000	Non	Diff.	25.78	0.88
4	317	63	1000	Oui	Diff.	58.83	0.87
1	1	11	1000	Non	Diff.	36.86	0.87
63	250	563	1000	Oui	Même	56.08	0.86
63	563	1000	250	Oui	Diff.	55.52	0.84
63	563	250	1000	Oui	Diff.	54.59	0.84
1	101	4	1000	Oui	Diff.	54.44	0.83
63	1000	563	250	Oui	Diff.	52.25	0.83
63	1000	250	563	Oui	Diff.	51.60	0.83
1	11	1	1000	Oui	Diff.	45.86	0.81
4	317	1000	63	Oui	Diff.	57.19	0.81
250	500	1000	750	Oui	Diff.	40.09	0.79
250	750	1000	500	Oui	Diff.	41.67	0.79
4	1000	63	317	Oui	Diff.	53.70	0.79
250	500	750	1000	Oui	Même	39.80	0.79
250	1000	750	500	Oui	Diff.	41.89	0.79
250	750	500	1000	Oui	Diff.	40.13	0.78
4	1000	317	63	Oui	Diff.	52.70	0.76
1	11	1000	1	Oui	Diff.	43.99	0.74
1	1000	4	101	Oui	Diff.	48.63	0.67
1	1000	101	4	Oui	Diff.	47.03	0.62

TABLE 11 – Différences entre les modélisations obtenues avec le vrai scénario et les scénarios alternatifs. Nous avons inclus les scénarios de coût présentant les 20 coïncidences spatiales moyennes ou coefficients de corrélation de Mantel (*r* Mantel) moyens les plus forts (24 scénarios). Ils sont classés dans l'ordre décroissant des coefficients de corrélation de Mantel (*r* Mantel). Leur coïncidence spatiale avec les LCPs obtenus avec le vrai scénario de coût est affichée (% coïnc. spat.). Les 10 valeurs les plus fortes des colonnes "*r* Mantel" et "% coïnc. spat." sont affichées en gras. Les valeurs moyennes obtenues pour tous les LCPs, les matrices de CDs et les paysages ont été calculées pour chaque scénario. Les valeurs de coût associées à chacun des quatre types d'occupation du sol sont incluses dans les colonnes "Forêts", "Prairies", "Cultures" et "Artif.". La colonne "For. < Prair." indique si la valeur de coût associée aux forêts est plus faible que celle associée aux prairies. La colonne "Ordre" indique si les valeurs de coût associées aux types d'occupation du sol dans chaque scénario alternatif suivent le même ordre que celles du vrai scénario.

4 Discussion

En utilisant une large gamme de scénarios de coût dans des paysages réels et contrastés, nous voulions analyser de façon précise la sensibilité de la localisation spatiale des chemins de moindre coût (LCPs) d'une part et des distances-coût (CDs) d'autre part au choix des valeurs de coût. Ces deux résultats de la modélisation étaient sensibles aux scénarios de coût mais leurs sensibilités ne dépendaient pas des mêmes caractéristiques de ces scénarios et ne s'observaient pas de la même manière dans tous les paysages. À partir de ces résultats, nous formulons quelques recommandations pour l'inférence des valeurs de coût et la modélisation basée sur les chemins de moindre coût plus généralement.

4.1 Sensibilité des LCPs et CDs aux scénarios de coût

La coïncidence spatiale des LCPs est très sensible aux scénarios de coût

Nous avons calculé le degré de coïncidence spatiale entre des LCPs de référence (vrais LCPs) et des LCPs obtenus avec plusieurs scénarios alternatifs pour mesurer la sensibilité de ces chemins au scénario utilisé. Cette sensibilité était forte car seuls quelques scénarios permettaient de reproduire fidèlement le tracé des chemins de référence. Alors que Pullinger et Johnson (2010) ont montré une

faible adéquation entre différents LCPs et des traces GPS de caribous, les différents LCPs calculés par [Beier et al. \(2009\)](#) coïncidaient fortement avec un corridor reliant deux aires protégées. Nos résultats apportent des éléments d'explication à ces résultats en apparence contradictoires. En effet, même si le choix du scénario de coût est le facteur qui influence le plus la localisation des LCPs, dans les paysages les plus favorables aux déplacements, peu diversifiés et comportant de grandes taches d'occupation du sol, la coïncidence spatiale entre les LCPs était globalement plus élevée. Dans ce cas, l'utilisation de scénarios similaires de par l'ordre des valeurs de coût et leur gamme de variation conduisait à des degrés de coïncidence spatiale très élevés. C'est peut-être la raison du résultat obtenu par [Beier et al. \(2009\)](#) en considérant un ensemble de scénarios de coût plausibles et peu contrastés. En revanche, dans des paysages plus diversifiés, comportant de petites taches et des proportions importantes des types d'occupation du sol défavorables aux déplacements, les tracés des LCPs divergeaient davantage. Cette situation reflète quant à elle le contexte de l'étude de [Pullinger et Johnson \(2010\)](#).

Des matrices de CDs fortement corrélées peuvent être obtenues avec des scénarios très différents

Nous avons étudié la sensibilité de la distribution statistique des distances-coût aux scénarios en calculant les corrélations entre la vraie matrice de CDs et chacune des matrices issues des scénarios alternatifs. Nos résultats montrent que des matrices de CDs fortement corrélées à la vraie matrice peuvent être obtenues avec plusieurs scénarios en apparence très différents du scénario de référence. Même si les caractéristiques des scénarios de coût alternatifs influençaient notablement le degré de corrélation des matrices alternatives avec celle de référence, le contexte paysager dans lequel la modélisation était réalisée avait aussi une influence significative. Dans les paysages les plus favorables aux déplacements et comportant de grandes taches, quel que soit le scénario considéré, les matrices de CDs étaient fortement corrélées à la matrice de référence. Au contraire, les matrices se distinguaient davantage les unes des autres lorsque les taches d'occupation du sol étaient diversifiées et de forme complexe. Ce résultat rappelle ceux de [Cushman et al. \(2013\)](#) montrant que des distances euclidiennes et des distances-coût expliquaient de façon équivalente la différenciation génétique lorsque la proportion d'habitat dans le paysage était importante et le contraste entre les valeurs de coût faible.

Des matrices de CDs fortement corrélées peuvent correspondre à des LCPs spatialement distincts

Notre travail apporte pour la première fois des informations sur la relation existant entre la coïncidence spatiale des LCPs et le degré de corrélation des matrices de CDs correspondantes, ainsi que sur les facteurs de cette relation. Nous montrons tout d'abord qu'il est possible que deux matrices de CDs soient fortement corrélées alors que les LCPs dont elles sont issues sont spatialement distincts. Cependant, ce n'est pas réciproque dans la mesure où une forte coïncidence spatiale entre LCPs ne s'observe qu'à condition que les matrices de CDs correspondantes soient très corrélées.

Ce résultat s'explique en partie par le fait que les chemins et les distances-coût ne sont pas influencés par les mêmes caractéristiques des scénarios de coût. Tandis que le coût relatif attribué aux types d'occupation du sol les plus favorables aux déplacements détermine en grande partie la localisation des chemins, les distances-coût dépendent davantage du gradient des valeurs de coût. Cela paraît logique vu que les coûts relatifs déterminent le passage d'un chemin dans un type d'occupation du sol plutôt que dans un autre, alors que le gradient des valeurs détermine la distribution statistique des distances-coût. Ainsi, lorsque le coût maximal était attribué au type d'occupation du sol le moins résistant dans le scénario de référence (forêts : 1000), limiter le contraste avec les autres valeurs de coût permettait

malgré tout d'obtenir des matrices de CDs fortement corrélées à la matrice de référence.

Par ailleurs, le scénario paraissant le plus proche du scénario de référence ([1, 4, 101, 1000]) produisait les matrices de CDs les plus fortement corrélées aux matrices de référence mais ne reproduisait pas aussi bien le tracé des LCPs de référence que le scénario [4, 63, 317, 1000]. Cela pourrait s'expliquer par la sensibilité des chemins aux contrastes entre les valeurs de coût les plus faibles. En effet, les ratios 4/63 et 63/317 sont tous deux plus proches de 1/10 et 10/1000 que ne le sont 1/4 et 4/100.

4.2 Implications pour l'inférence des valeurs de coût et la modélisation basée sur les LCPs

L'interprétation écologique et l'utilisation de valeurs de coût inférées sont sujettes à précaution

Nos résultats ont des implications pour l'inférence et l'interprétation des valeurs de coût, ainsi que pour l'utilisation de modélisations basées sur les chemins de moindre coût à des fins appliquées ou de compréhension théorique. Nous avons vu que deux matrices de CDs fortement corrélées peuvent correspondre à des chemins spatialement distincts obtenus avec des scénarios dont les valeurs de coût s'interprètent très différemment d'un point de vue écologique. Étant donné que l'inférence des valeurs de coût basée sur des données empiriques dépend essentiellement de la distribution statistique des distances-coût, il faut donc être prudent lors de l'interprétation des valeurs de coût inférées et lors de l'utilisation des trajectoires spatiales qu'elles permettent de modéliser.

Préconisations pour l'utilisation de valeurs de coût issues d'une inférence

La modélisation de chemins de moindre coût répond parfois à un besoin opérationnel, tel que la localisation de passages à faune (Clevenger *et al.*, 2002 ; Mimet *et al.*, 2016). Les chemins étant fortement sensibles aux scénarios de coût, nos résultats suggèrent que ce type d'approche peut être plus incertain et moins robuste qu'il n'y paraît, quand bien même les valeurs de coût sont inférées à partir de données de terrain. Nous tâchons néanmoins de fournir des solutions potentielles au problème que nous identifions ici.

Tout d'abord, le scénario de coût permettant de reproduire au mieux la localisation des chemins de référence faisait toujours partie des scénarios produisant les matrices de CDs les plus corrélées à la matrice de référence. On peut donc s'attendre à ce que l'ensemble des scénarios reflétant le plus les véritables contraintes aux déplacements partagent des propriétés statistiques et soient retenus comme étant les plus vraisemblables lors de l'inférence statistique. Nos résultats remettent donc en cause une pratique courante consistant à identifier un seul "meilleur" scénario de coût lors d'une optimisation. Plutôt que de n'en retenir qu'un, il pourrait être intéressant de considérer l'ensemble des scénarios ayant une certaine vraisemblance statistique pour maximiser les chances que celui qui reflète le plus la réalité écologique en fasse partie. Alors que l'utilisation du logiciel Circuitscape (McRae et Beier, 2007) permet de considérer plusieurs chemins alternatifs entre deux points à partir d'un scénario unique, une solution alternative consisterait à localiser un chemin unique entre deux points en considérant tour à tour plusieurs scénarios vraisemblables. Cette solution fait écho à l'usage croissant de méthodes d'inférence basées sur plusieurs modèles lorsque le meilleur modèle est à lui seul une mauvaise approximation de la réalité écologique stochastique (Burnham et Anderson, 2004). De façon similaire, un ensemble de scénarios de coût au degré de vraisemblance équivalent pourrait être identifié sur la base d'un critère d'AIC par exemple. On peut toutefois s'attendre à ce que cette méthode conduise dans

certaines situations à des chemins divergeant fortement. Dans un tel cas, ce sont leurs intersections qui fourniraient les informations les plus utiles à la conservation de la connectivité.

Une autre solution consisterait à limiter le nombre de scénarios considérés pour maximiser les contrastes entre les matrices de distances-coût. Idéalement, il faudrait faire en sorte que leurs différences reflètent plusieurs hypothèses relatives au comportement de l'espèce étudiée. Cette recommandation rappelle celle de [Zeller *et al.* \(2016\)](#), qui incitaient à ne pas inclure de matrices de CDs fortement corrélées entre elles dans une optimisation de scénarios de coût. Néanmoins, même si les valeurs de coût issues d'une telle inférence doivent être utilisées avec précaution pour localiser des LCPs, les matrices de CDs correspondantes sont censées refléter la distribution statistique de variables empiriques, des distances génétiques par exemple, et peuvent donc servir de base au calcul de métriques de connectivité au niveau des taches d'habitat (issues de graphes paysagers par exemple). Cette utilisation des résultats de l'inférence serait moins incertaine tout en permettant de spatialiser les résultats.

Perspectives méthodologiques pour la modélisation basée sur les LCPs

Alors que les scénarios de coût considérés peuvent être choisis jusqu'aux derniers stades d'une étude, le paysage dans lequel l'étude est menée est quant à lui choisi lors des stades initiaux. Compte tenu de (i) l'influence du contexte paysager sur la sensibilité de la modélisation aux scénarios et (ii) de l'importance de cette sensibilité pour l'inférence d'un scénario de coût, il pourrait être utile de l'évaluer au préalable pour choisir au mieux la zone d'étude, comme l'avaient déjà suggéré [Cushman *et al.* \(2013\)](#). Nous avons pour cela inclus la fonction `link_compar()` dans le package `graph4lg` (cf. Chapitre 2). Elle calcule la coïncidence spatiale entre différents LCPs et permet de reproduire nos analyses tout en ajustant la largeur des zones tampons considérées autour des LCPs. Nous avons ici utilisé une largeur constante et égale à 400 m, en estimant qu'elle reflétait l'échelle à laquelle les mesures de conservation sont mises en œuvre suite à ce type de modélisation ([Ford *et al.*, 2020](#) ; [Spackman *et Hughes*, 1995](#)). Cela limitait également le risque de surestimation de la coïncidence spatiale pour les chemins les plus courts.

La plus faible sensibilité de la modélisation observée dans les paysages favorables aux déplacements et incluant les plus grandes taches peut être due à l'échantillonnage des points entre lesquels les chemins étaient calculés dans les forêts. Ce choix vient du fait que les analyses de connectivité visent le plus souvent à identifier des chemins favorables entre des zones d'habitat. Il influençait les résultats dans la mesure où les chemins débutaient et aboutissaient dans les zones de forêt, indépendamment du coût attribué à ces forêts. L'utilisation de la théorie des circuits ([McRae, 2006](#)) aurait augmenté les différences entre les distances de résistance obtenues en attribuant un coût faible ou fort aux zones de forêt. Au contraire, elle aurait probablement augmenté le degré de coïncidence spatiale entre les scénarios de coût les plus similaires car des chemins alternatifs potentiellement partagés par plusieurs scénarios auraient été considérés.

Transition

Dans le chapitre précédent, nous avons montré que des matrices de distances-coût pouvaient être fortement corrélées bien que calculées à partir de scénarios de coût différents. Cela constitue une difficulté potentielle pour l'identification du scénario de coût reflétant le mieux la réalité écologique à partir de données génétiques. De plus, l'utilisation d'un scénario de coût différent du scénario le plus réaliste pourrait orienter vers des mesures de conservation inadéquates. Dans ce contexte, il convient d'évaluer la qualité des inférences de valeurs de coût basées sur des données génétiques.

Par ailleurs, la dérive génétique n'affecte pas la structure génétique des populations de façon homogène car elle dépend de la taille des populations. Ne pas tenir compte de l'effet hétérogène de la dérive lors de l'inférence de valeurs de coût à partir de données génétiques pourrait également limiter la qualité de cette inférence. Pour cette raison, nous testons dans le chapitre suivant l'intérêt de l'inclusion de variables calculées à la fois à l'échelle des nœuds et des liens de graphes génétiques et paysagers dans l'inférence. Nous réalisons l'intégration de ces deux types de graphes à l'aide de modèles gravitaires.

Chapitre 6

Inférer la résistance du paysage au flux génétique en intégrant les nœuds et les liens de graphes génétiques et paysagers dans des modèles gravitaires

Résumé

Des méthodes de génétique du paysage permettent d'inférer les coûts de la matrice paysagère à partir de la relation entre différenciation génétique et distances-coût. L'hétérogénéité spatiale des tailles de populations est habituellement négligée dans ces inférences alors qu'il est avéré qu'elle influence la différenciation génétique. De même, les taux de migration et la distribution spatiale des populations influencent potentiellement cette inférence. Notre objectif était d'évaluer la qualité de l'inférence des valeurs de coût pour plusieurs taux de migration, distributions spatiales des populations et degrés d'hétérogénéité des tailles de populations. Nous voulions aussi évaluer à quel point l'inclusion dans les modèles de variables calculées au niveau des nœuds de graphes génétiques ou paysagers pouvait améliorer l'inférence basée sur les distances calculées au niveau de leurs liens. Pour cela, nous avons simulé plusieurs intensités de flux génétique entre des ensembles de populations de tailles différentes et ayant des distributions spatiales variées. Nous avons ensuite réalisé des modèles gravitaires expliquant les distances génétiques (*link level*) en fonction des "vraies" distances-coût déterminant la simulation ou d'autres distances-coût (*link level*), et de variables calculées au niveau des populations et de leurs taches d'habitat (*node level*) telles que les tailles des populations et les capacités des taches. Nous souhaitions identifier les conditions dans lesquelles l'identification du "vrai" scénario de coût était possible, et évaluer la contribution des variables calculées au niveau des nœuds à l'atteinte de cet objectif. La qualité de l'inférence des valeurs de coût était la plupart du temps satisfaisante mais était réduite lorsque la migration était restreinte, les tailles de populations fortement hétérogènes et que certaines populations formaient des agrégats spatialement. Dans ces situations, l'inclusion de variables mesurées au niveau des nœuds dans des modèles gravitaires améliorait l'inférence des valeurs de coût à partir de données génétiques.

Cet article est en préparation pour une soumission dans une revue scientifique :

Savary, P., Foltête, J. C., Moal, H., Vuidel, G. & Garnier, S. 2021. Inferring landscape resistance to gene flow using gravity models. In prep.

*We now must ask about the soundness of this appropriation of Newtonian force.
Should selection and drift be treated in this way, or not ?*

Charles H. Pence

1 Introduction

La connectivité écologique des habitats se modélise en représentant la matrice paysagère sous la forme d'une surface de résistance sur laquelle on calcule les chemins de dispersion potentiels entre les taches d'habitat (Zeller *et al.*, 2012). Une telle modélisation n'est valide qu'à condition que les valeurs de coût attribuées aux différents types d'occupation du sol reflètent correctement la façon dont ils facilitent ou limitent les déplacements de l'espèce étudiée. Bien que ces valeurs puissent être fixées à partir de dires d'experts, des données biologiques sont fréquemment utilisées pour renforcer leur réalisme écologique (Zeller *et al.*, 2018). Une comparaison de plusieurs méthodes visant à calibrer les coûts à partir de données biologiques (cf. Chapitre 4) nous a montré que l'utilisation de données de présence dans ce cadre ne garantissait pas la qualité des modèles de connectivité. En revanche, l'utilisation de données génétiques a souvent été conseillée pour inférer les valeurs de coût (Beier *et al.*, 2008) et son intérêt a été mis en évidence (Zeller *et al.*, 2018).

L'inférence de valeurs de coût à partir de données génétiques repose sur le modèle d'Isolement par la Résistance du Paysage (IBLR), extension du modèle d'Isolement par la Distance (IBD) aux situations dans lesquelles la matrice exerce un effet hétérogène sur la dispersion (Cushman *et al.*, 2006 ; Guillot *et al.*, 2009 ; McRae, 2006). Dans une situation d'IBLR, il existe une relation statistique linéaire entre la différenciation génétique entre populations et les distances-coût ou les distances de résistance (théorie des circuits) calculées entre ces populations en appliquant un certain scénario de coût. Utiliser plusieurs scénarios pour identifier celui qui maximise l'ajustement de cette relation est alors un moyen d'obtenir des valeurs de coût censées refléter l'effet de la matrice sur la dispersion (Cushman *et al.*, 2006 ; Graves *et al.*, 2013 ; Peterman, 2018).

Cette approche soulève néanmoins plusieurs interrogations. D'une part, le chapitre précédent nous a montré que plusieurs scénarios de coût en apparence différents pouvaient engendrer des matrices de distances-coût fortement corrélées. Cela pourrait limiter la qualité de l'inférence basée sur la relation entre distances-coût et distances génétiques. Il est donc nécessaire d'évaluer la qualité de cette inférence dans un contexte où l'on connaît le scénario de coût responsable de la différenciation génétique, et où l'on cherche à le distinguer de scénarios de coût alternatifs à l'aide de données génétiques.

D'autre part, la capacité des modèles à identifier le scénario de coût reflétant le mieux la résistance du paysage pourrait être altérée par des facteurs déterminant la différenciation génétique entre populations indépendamment de la distance paysagère entre ces populations. Tout d'abord, les méthodes d'inférence basées sur l'IBLR partent souvent du principe que la matrice paysagère et les flux génétiques qui la traversent ont une influence prépondérante sur la différenciation génétique (Richardson *et al.*, 2016). Or, cette dernière dépend également de la dérive génétique, en particulier lorsqu'une espèce forme des populations subdivisées et de petite taille et que les taux de migration entre elles sont très réduits (Frankham, 1996b ; Frankham *et al.*, 2004 ; Hartl *et al.*, 1997). Dans ce contexte, l'hétérogénéité spatiale des tailles de populations détermine de façon significative le patron de différenciation génétique (Serrouya *et al.*, 2012 ; Weckworth *et al.*, 2013). Pour cette raison, Prunier *et al.* (2017) ont développé des indices mesurant l'hétérogénéité spatiale des tailles de populations (SHNe, pour *Spatial Heterogeneity in Population Sizes*) et ont montré qu'ils pouvaient expliquer la différenciation génétique entre populations tout autant que la distance entre ces populations. La part de variance expliquée par ces indices était d'autant plus importante que la taille des populations était hétérogène

et que les taux de migration étaient faibles.

Par ailleurs, selon [Graves et al. \(2013\)](#), lorsque des individus forment des agrégats spatialement et que les flux génétiques entre de tels agrégats sont très limités, la dérive exerce au sein de chacun d'eux un effet indépendant sur la variabilité génétique. L'effet de la distance paysagère sur la différenciation génétique est alors difficile à estimer lorsque l'intégralité des individus de la zone d'étude est prise en compte dans l'analyse. De même, [Ueno et al. \(2000\)](#) ont montré chez une espèce de plante (*Camellia japonica*) formant des *clusters* d'individus que cette agrégation spatiale était à l'origine d'un patron d'IBD ne s'observant qu'à une échelle spatiale réduite. Il reste toutefois à vérifier qu'un phénomène similaire se produit lorsque les individus forment plusieurs populations discrètes.

En somme, en influençant la relation entre distances paysagères et différenciation génétique, la distribution spatiale des populations, le taux de migration et l'hétérogénéité spatiale des tailles de populations (SHNe) pourraient limiter la qualité de l'inférence des valeurs de coût. Néanmoins, si la SHNe limite la qualité de l'inférence des coûts, on peut se demander si sa prise en compte explicite dans l'analyse ne pourrait pas atténuer cette influence négative sur l'inférence. L'estimation des tailles de populations est un pré-requis à cette prise en compte de la SHNe. Elle est difficile à faire ([Wang, 2005](#)) mais des variables environnementales mesurées au niveau des taches d'habitat peuvent constituer des approximations relativement fiables de la capacité de charge de ces taches ([Prunier et al., 2017](#)), et éviter d'avoir à réaliser des inventaires de terrain coûteux. Ces variables se calculent au niveau des nœuds des graphes génétiques et paysagers étudiés dans le cadre de cette thèse. Nous avons ainsi fait l'hypothèse que leur inclusion dans l'analyse de la relation entre les liens de ces graphes pourrait améliorer la qualité de l'inférence des coûts. Une telle approche vise à inférer les valeurs de coût nécessaires à la construction des graphes paysagers à l'aide des données empiriques utilisées pour construire les graphes génétiques. L'utilisation des données génétiques lors de la construction du graphe paysager (*a priori*) la distingue donc des comparaisons *a posteriori* mises en œuvre dans la première partie de cette thèse, et l'apparente davantage à une intégration de ces outils.

Ce chapitre avait donc pour objectifs (1) d'évaluer la qualité de l'inférence des valeurs de coût réalisée à partir de données génétiques et (2) d'identifier les situations dans lesquelles l'inclusion dans les modèles utilisés de variables mesurées au niveau des populations améliore l'inférence. Nous nous attendions à ce que la qualité de cette inférence soit limitée pour des taux de migration faibles, des tailles de populations hétérogènes et lorsque des populations formaient des ensembles agrégés spatialement. Dans ces cas-là, nous faisons l'hypothèse que l'inclusion dans les modèles de variables mesurées au niveau des populations, telles que la taille de ces populations ou la surface des taches qu'elles occupent, améliorerait la qualité de l'inférence

Pour atteindre ces objectifs, nous avons simulé plusieurs intensités de flux génétique entre plusieurs ensembles de populations ayant des distributions spatiales variées et différents degrés d'hétérogénéité de leurs nombres d'individus. Nous avons ensuite réalisé des modèles gravitaires afin d'expliquer les distances génétiques simulées (liens) en fonction (i) des distances-coût (liens) déterminant les simulations ou de distances-coût alternatives, et (ii) de variables mesurées au niveau des populations (nœuds) telles que la surface des taches ou la taille des populations. Nous souhaitons déterminer les conditions dans

lesquelles l'inférence des coûts est fiable et celles nécessitant l'inclusion de variables calculées au niveau des nœuds pour identifier correctement le "vrai" scénario de coût déterminant les flux génétiques.

2 Méthodes

La méthodologie générale mise en œuvre dans ce chapitre pour atteindre les objectifs 1 et 2 est illustrée par les figures 64 et 65, respectivement. Nous avons simulé le patron de différenciation génétique résultant de plusieurs générations de flux génétique chez une espèce aux capacités de dispersion limitées. Nous connaissions les "vraies" valeurs de coût associées à chaque type d'occupation du sol, fixées arbitrairement en amont des simulations, et les distances-coût correspondantes. Ces "vraies" distances-coût déterminaient l'intensité des flux génétiques entre populations dans les paysages simulés. Notre premier objectif était d'évaluer la capacité des modèles de génétique du paysage à identifier le "vrai" scénario de coût au sein d'un ensemble de scénarios de coût "alternatifs" divergeant plus ou moins du "vrai" scénario (Figure 64). En utilisant des arbres de régression, nous avons également identifié précisément les conditions dans lesquelles des modèles gravitaires incluant à la fois (i) des distances-coût calculées au niveau des liens entre populations et (ii) des variables calculées à l'échelle des nœuds correspondant aux populations amélioreraient la qualité de l'inférence (Figure 65).

2.1 Simulations

2.1.1 Simulations des paysages et des populations

Nous avons simulé 200 paysages de 60×60 km ayant une résolution spatiale de 100 m et comptant chacun quatre types d'occupation du sol. Nous avons pour cela utilisé la même méthode qu'au chapitre 1 (Schlather *et al.*, 2015) pour obtenir des paysages réalistes tout en nous assurant qu'il y avait une certaine variabilité entre les matrices de distances-coût ensuite calculées dans ces paysages selon plusieurs scénarios de coût alternatifs.

Nous souhaitions simuler les déplacements d'une espèce forestière spécialiste car ce type d'espèce fait souvent l'objet d'analyses de connectivité des habitats. Ainsi, les forêts occupaient 20 % des paysages simulés et constituaient le type d'occupation du sol le plus favorable à la dispersion (coût : 1). Les parts de l'occupation du sol et les coûts associés aux autres types d'occupation du sol ont été choisis de façon à reproduire les contraintes à la dispersion d'une espèce forestière dans un paysage hétérogène : 27 % de prairies (coût : 10), 27 % de cultures agricoles (coût : 100) et 26 % de zones artificialisées (coût : 1000)¹.

Un de nos objectifs était d'évaluer si l'hétérogénéité spatiale des tailles de populations influençait l'inférence des valeurs de coût. Cette hétérogénéité dépend en partie de la distribution spatiale des taches d'habitat car la surface d'une tache est souvent corrélée positivement à sa capacité de charge et par conséquent à la taille de la population qu'elle héberge. Nous nous sommes donc assurés de prendre en compte une large gamme de distributions de tailles de populations et de capacités des taches d'habitat. Nous avons échantillonné aléatoirement 60 pixels forestiers dans chaque paysage, séparés d'au moins 1000 m les uns des autres. Chacun de ces pixels a été utilisé pour établir une population. Les taches d'habitat occupées par chaque population correspondaient aux pixels de forêt inclus dans la

1. Notons que le scénario de coût de référence utilisé dans ce chapitre est identique à celui utilisé dans le chapitre 5

Objectif 1 : L'inférence des valeurs de coût à partir de données génétiques est-elle correcte ?

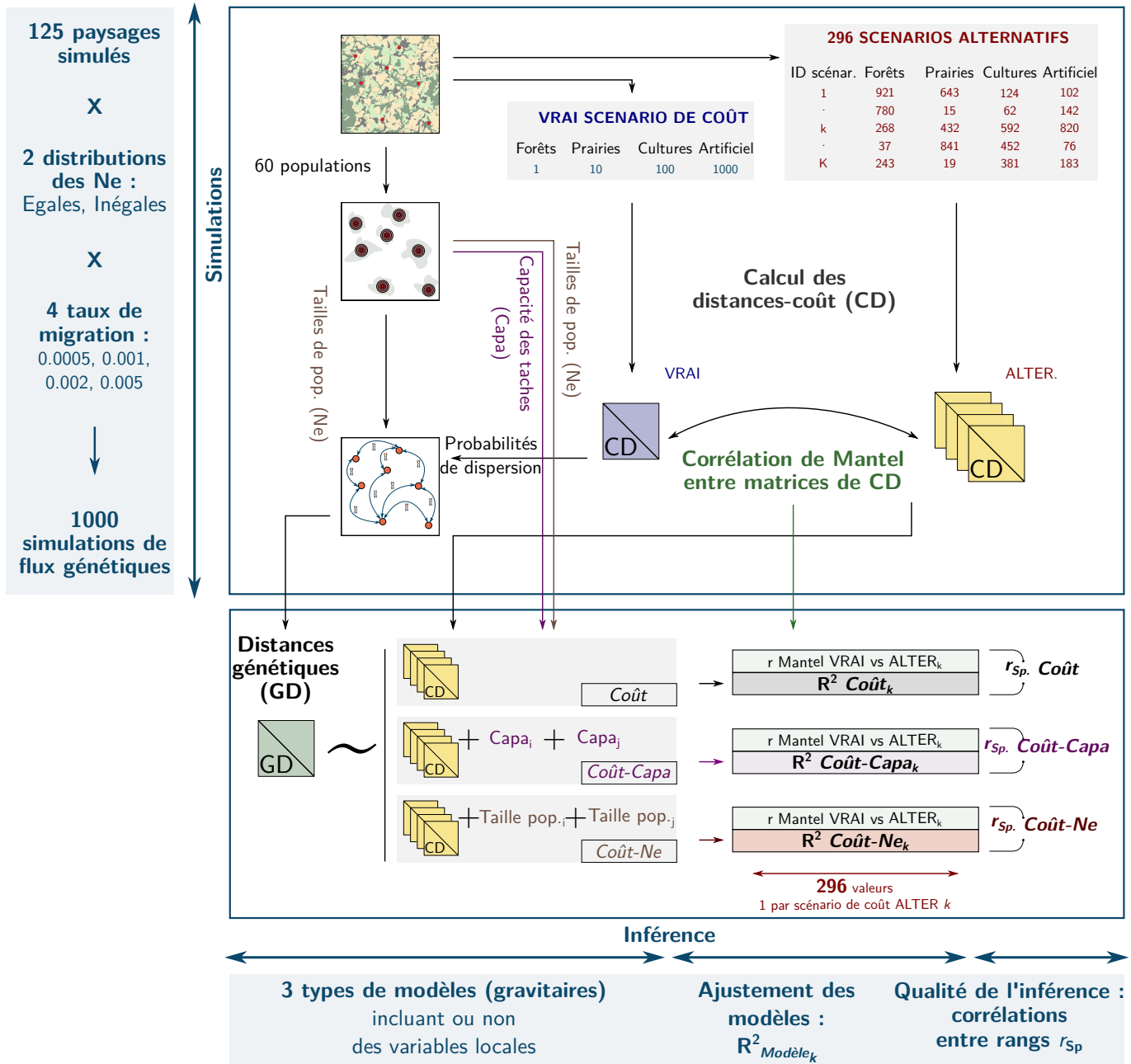


FIGURE 64 – Méthodologie générale des simulations de flux génétiques réalisées pour évaluer la capacité de différents modèles à identifier le "vrai" scénario de coût déterminant les flux au sein d'un ensemble de scénarios de coût "alternatifs" divergeant plus ou moins de ce "vrai" scénario (objectif 1).

Objectif 2 : Dans quelles situations l'inclusion de variables locales dans les modèles améliore l'inférence ?

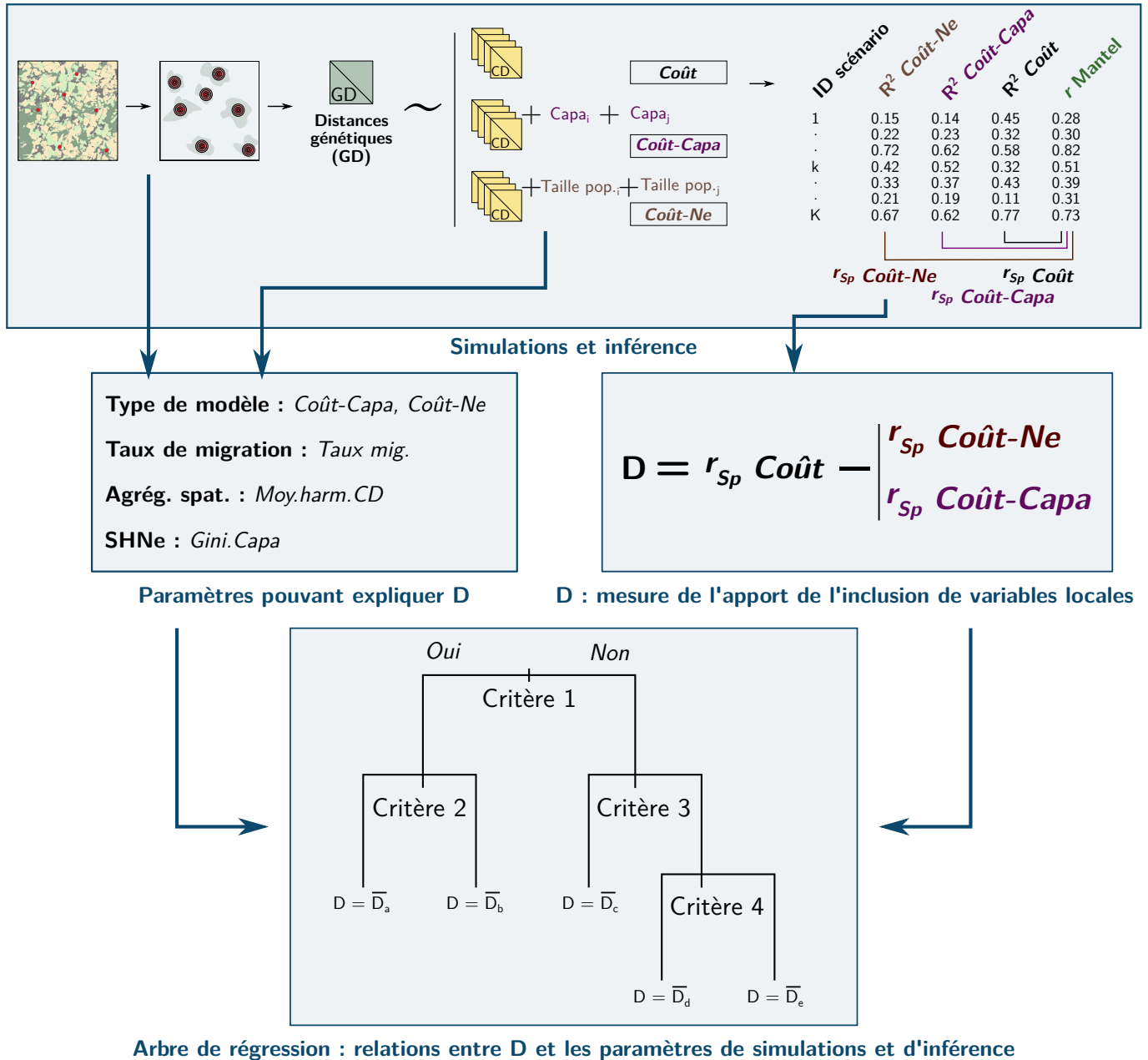


FIGURE 65 – Identification des situations dans lesquelles l'inclusion de variables locales dans les modèles améliore l'inférence (objectif 2). SHNe : hétérogénéité spatiale des tailles de populations, Agrég. spat. : degré d'agrégation spatiale des populations, Moy.harm.CD : moyenne harmonique des distances-coût entre populations, Gini.Capa : indice de Gini calculé à partir des capacités des taches d'habitat. Dans chaque feuille de l'arbre de régression délimitée à partir de critères binaires appliqués aux variables explicatives, la valeur de D modélisée est égale à la valeur moyenne de D dans la feuille (\bar{D}_a pour la feuille a par exemple).

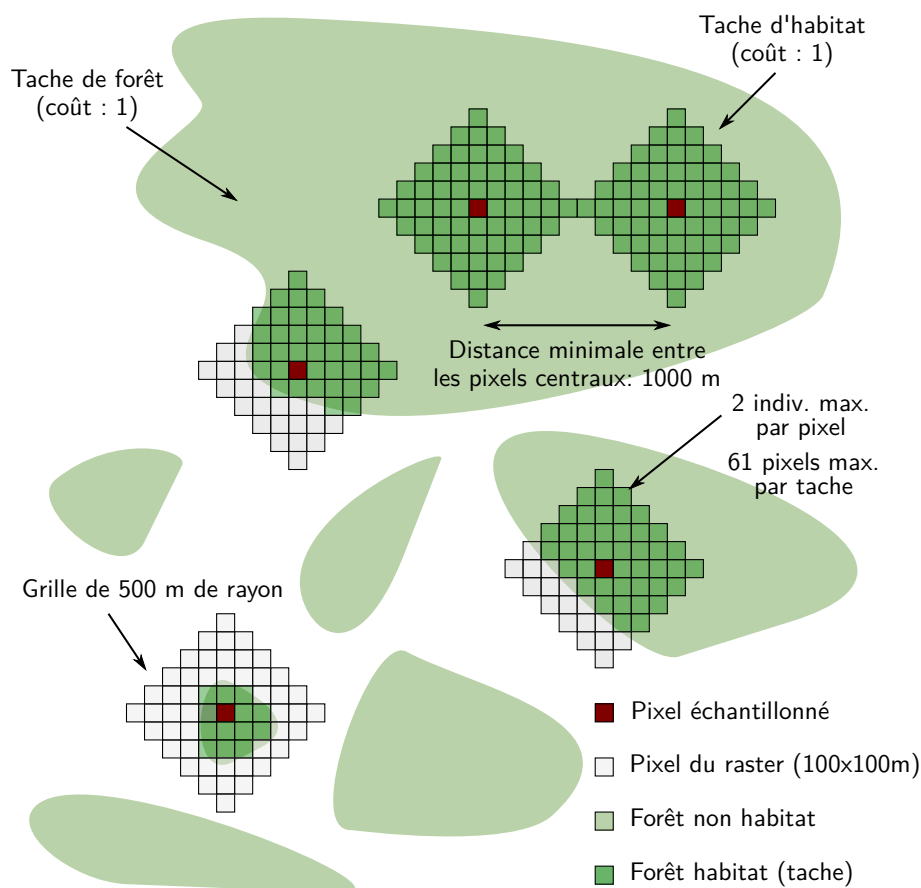


FIGURE 66 – Calcul de la capacité des taches d'habitat simulées à partir des pixels forestiers échantillonnés. Les taches d'habitat et les taches de forêt ont la même résistance car nous faisons l'hypothèse que l'espèce forestière simulée se déplace aussi facilement dans son habitat que dans la forêt, même quand cette dernière ne réunit pas les conditions lui permettant d'en faire son habitat.

zone de 500 m de rayon autour de chaque pixel échantillonné (Figure 66), et la capacité de ces taches était égale au nombre de ces pixels (équivalente à une surface). Cela nous a permis de faire varier la capacité des taches d'habitat en fonction de la structure du paysage simulé.

Nous avons ensuite fixé le nombre d'individus dans chaque population entre 10 et 100 pour un total de 3300 individus en distinguant deux distributions spatiales différentes des tailles de populations :

1. Tailles de populations égales et contenant chacune 55 individus (distribution nommée "Égales").
2. Tailles de populations hétérogènes et dépendantes de la capacité des taches qu'elles occupent (distribution nommée "Inégales").

La distribution "Égales" pour laquelle les tailles de populations sont égales représentait une référence rendant possible l'évaluation de l'effet de la SHNe sur l'inférence par comparaison à la distribution "Inégales". Pour cette seconde distribution, nous avons fait en sorte de couvrir un large gradient de SHNe tout en reproduisant des conditions réalistes dans lesquelles les tailles de populations dépendent de la capacité des taches d'habitat. Pour cela, nous avons généré des séries de 60 valeurs comprises entre 10 et 100 et dont la somme valait 3300 ± 50 . Nous avons classé ces séries selon leur degré d'hétérogénéité à l'aide de l'indice d'inégalité de Gini (Gini, 1912) (cf. Annexe A8, Figure 55) pour une illustration). En parallèle, nous avons calculé l'indice de Gini décrivant l'hétérogénéité des capacités des taches d'habitat échantillonnées dans chaque paysage. Nous avons ensuite associé à chaque paysage la distribution des tailles de populations correspondant à un degré d'hétérogénéité équivalent

à celui des capacités des taches échantillonnées dans ce paysage. Les plus grands nombres d'individus étaient ensuite attribués aux populations échantillonnées dans les plus grandes taches. Dans chaque cas, la densité d'individus maximale était égale à deux individus par hectare de forêt. Nous avons enfin calculé les distances-coût (CDs) entre chaque paire de populations dans chaque paysage.

En plus de simuler des situations contrastées en termes d'hétérogénéité des tailles de populations et des capacités des taches, nous nous sommes également assurés que les distributions spatiales des populations échantillonnées étaient variées car nous souhaitions tester l'influence de ce paramètre sur l'inférence. En effet, la présence d'ensembles relativement agrégés de populations entre lesquelles les échanges d'individus sont plus fréquents qu'avec d'autres populations plus éloignées pourrait limiter notre capacité à inférer la résistance du paysage au flux génétique (Graves *et al.*, 2013). Pour tester l'existence d'un tel phénomène, nous avons mesuré le degré d'agrégation spatiale des populations comme la moyenne harmonique des valeurs de CDs entre l'ensemble des paires de populations. Cette valeur est d'autant plus faible que les faibles valeurs de CD sont fréquentes², ce que l'on observe lorsque des populations sont proches les unes des autres et forment des ensembles au sein desquels le flux génétique s'effectue préférentiellement à courte distance. Enfin, à l'issue de ce paramétrage, pour pouvoir tester au mieux l'influence sur l'inférence de la SHNe et de l'agrégation spatiale des populations selon les modèles utilisés (cf. section 2.3), nous avons sélectionné 125 paysages maximisant les contrastes en termes d'indices de Gini des tailles de populations et de moyenne harmonique des CDs.

2.1.2 Simulations de flux génétiques

Nous avons utilisé le logiciel CDPOP (Landguth et Cushman, 2010) pour simuler des flux génétiques en utilisant une approche similaire à celle mise en œuvre dans le chapitre 1. Les tailles de populations et le sex-ratio (égal à 1) restaient constants durant les simulations. Elles duraient 500 générations pour que l'état d'équilibre du patron de différenciation génétique soit atteint. À chaque génération, les individus se reproduisaient dans leur propre population et les juvéniles pouvaient se disperser pour s'établir dans d'autres populations. Lors de la reproduction, chaque femelle pouvait avoir trois descendants en moyenne mais les individus surnuméraires étaient éliminés pour que chaque population conserve la taille qui lui était attribuée au début des simulations. Chaque individu était caractérisé par son génotype au niveau de 20 *loci* comptant initialement 30 allèles différents.

Les probabilités de dispersion p_{ij} entre les populations i et j dépendaient de la distance-coût CD_{ij} entre ces populations de sorte que : $p_{ij} = e^{-\alpha CD_{ij}}$. La valeur α était fixée afin que la probabilité de dispersion à une distance équivalente à 1000 unités de coût soit égale à 0.01.

Prunier *et al.* (2017) ont montré que la contribution de la SHNe à la différenciation génétique dépendait du taux de migration. Des analyses préliminaires nous avaient montré que pour des taux de migration supérieurs à 0.005, le flux génétique était trop important pour que la SHNe influence l'inférence, tandis que pour des taux de migration inférieurs à 0.0005, l'intensité de la dérive était trop forte pour que l'inférence soit possible. Nous avons donc réalisé les simulations pour des taux de migration égaux à 0.0005, 0.001, 0.002 et 0.005, pour identifier l'influence de ce taux sur l'inférence

2. Cela vient des propriétés de la moyenne harmonique qui accorde un poids plus important aux valeurs les plus faibles.

des coûts. Au total, 1000 simulations ont été réalisées (c'est-à-dire 125 paysages \times 2 distributions des tailles de populations \times 4 taux de migration, Figure 64).

Suite aux simulations, nous avons calculé les matrices de D_{PS} entre populations à partir des génotypes obtenus en génération 500 (Bowcock *et al.*, 1994) car nous avons vu que cette distance reflétait bien l'influence du paysage sur le patron de différenciation génétique dans le chapitre 1.

2.1.3 Scénarios de coût alternatifs

Pour identifier les situations dans lesquelles l'inférence permet de distinguer le "vrai" scénario de coût de scénarios alternatifs, il est nécessaire que les matrices de CDs obtenues à partir de ces scénarios alternatifs aient une certaine variabilité (Cushman *et al.*, 2013). Pour cela, nous avons utilisé une approche similaire à celle mise en œuvre dans le chapitre 5 pour générer aléatoirement des scénarios alternatifs. Nous avons utilisé la formule de Shirk *et al.* (2010) pour obtenir quatre valeurs de coût selon différents gradients (déterminés par la valeur x) et différentes valeurs de coût maximales (C_{max}) :

$$C_i = \left(\frac{Rang_i}{Rang_{max}} \right)^x \times C_{max}$$

avec C_i la valeur de coût comprise entre 1 et C_{max} associée au type d'occupation du sol i . $Rang_i$ est le rang du type d'occupation du sol i compris entre 1 et $Rang_{max} = 4$. Comme la valeur de coût maximale apporte une information sur le contraste entre les types d'occupation du sol les plus et moins favorables à la dispersion, nous avons utilisé des valeurs de C_{max} égales à 100, 1000 ("vraie" valeur maximale) ou 10000. Les valeurs de x déterminant le gradient des valeurs valaient 1, 2, 4, 8 ou 16 (cf. Chapitre 5). Pour chaque combinaison d'une valeur de x et d'une valeur de C_{max} (15 au total), les quatre valeurs obtenues étaient attribuées aléatoirement et sans remise aux quatre types d'occupation du sol. Chacune de ces combinaisons aléatoires constituait un scénario de coût. Nous avons généré l'ensemble des scénarios alternatifs possibles avant d'en sélectionner 296 aléatoirement³. Nous avons ensuite utilisé chacun des 296 scénarios pour calculer des matrices de CDs alternatives dans chaque paysage entre les populations échantillonnées. Nous avons enfin calculé le coefficient de corrélation de Mantel r entre chacune des matrices de CDs alternatives et la "vraie" matrice de CDs correspondante afin d'évaluer à quel point elles différaient de cette matrice de référence (Figure 64).

2.2 Modèles gravitaires

Les modèles gravitaires ont initialement été utilisés en géographie et en économie (Anderson, 1979 ; Fotheringham et O'Kelly, 1989 ; Schneider *et al.*, 1998) pour modéliser plusieurs types d'interactions spatiales. Leur application en écologie (Bossenbroek *et al.*, 2001, 2007 ; Ferrari *et al.*, 2006 ; Kong *et al.*, 2010 ; Xia *et al.*, 2004) et en génétique du paysage (DiLeo *et al.*, 2014 ; Moran-Lopez *et al.*, 2016 ; Murphy *et al.*, 2010a ; Robertson *et al.*, 2018b ; Watts *et al.*, 2015 ; Zero *et al.*, 2017) est plus récente. Ils modélisent des interactions spatiales ou des flux en fonction des variables caractérisant les objets impliqués dans l'interaction et des distances entre ces objets (masses et distances dans la théorie de la gravitation de Newton) (cf. Annexe B5). Ici, nous les avons utilisés pour modéliser la distance génétique G_{ij} entre les populations i et j (variable réponse, *link level*) en fonction de variables explicatives calculées à différents niveaux :

3. Nous avons obtenu 296 scénarios et non pas 300 car il y avait moins de 100 combinaisons possibles lorsque C_{max} valait 10000.

- Au niveau des populations (*node level*) :
 - Capacités des taches : $Capa_i, Capa_j$
 - Tailles de populations : N_i, N_j
- Entre les populations (*link level*) :
 - Distances-coût CD_{ijk} calculées entre les populations i et j selon le scénario k

Nous avons ainsi réalisé trois types de modèles pour tester nos hypothèses :

Modèle "Coût" : $G_{ij} \sim c \times CD_{ijk}^m$

Modèle "Coût-Capa" : $G_{ij} \sim c \times CD_{ijk}^m \times Capa_i^n \times Capa_j^o$

Modèle "Coût-Ne" : $G_{ij} \sim c \times CD_{ijk}^m \times N_i^p \times N_j^q$

c était une constante. Nous avons calculé ces trois modèles en utilisant les CDs obtenues avec chacun des scénarios, "vrai" ou "alternatifs". L'application d'une fonction logarithme à l'expression de ces modèles permettait de les transformer en des modèles de régression multiple dont les paramètres c, m, n, o, p et q pouvaient être estimés. Pour tenir compte de l'absence d'indépendance des valeurs des matrices de distances, nous avons inclus dans les modèles un effet aléatoire correspondant à l'identité de chaque population i (modèles de type MLPE (Clarke *et al.*, 2002) également utilisés dans le chapitre 4).

2.3 Évaluation de la performance des modèles

Nous avons évalué la qualité de l'inférence des coûts et identifié les situations dans lesquelles l'ajout de variables locales (modèles "Coût-Capa" ou "Coût-Ne") l'améliorait (Figure 64). À partir de ces résultats, nous avons formulé des conseils méthodologiques pour l'inférence des valeurs de coût.

Pour cela, nous avons tout d'abord évalué l'ajustement des modèles en calculant le R_β^2 d'Edwards *et al.* (2008) car d'autres critères de sélection de modèles n'étaient pas adaptés aux modèles mixtes utilisés (Van Strien *et al.*, 2012). À l'issue de chaque simulation, chacun des trois types de modèle était réalisé en incluant chacune des 296 matrices de CDs alternatives tour à tour. Compte tenu de notre protocole, pour qu'un type de modèle ("Coût", "Coût-Capa", "Coût-Ne") puisse être considéré comme satisfaisant pour l'inférence des coûts, il fallait que les 296 valeurs de R_β^2 obtenues reflètent le degré de similarité entre ces 296 scénarios alternatifs et le "vrai" scénario de coût. Nous avons mesuré cette similarité à l'aide du r de Mantel calculé entre les "vraies" matrices de CDs et les matrices de CDs "alternatives". Par conséquent, pour mesurer la performance des modèles dans chaque situation, nous avons calculé le coefficient de corrélation de Spearman r_{Sp} entre les séries de 296 valeurs de R_β^2 et de r de Mantel. De plus, nous nous attendions à ce que la différence D entre la valeur r_{Sp} associée aux modèles "Coût-Capa" ou "Coût-Ne" et celle associée au modèle "Coût" soit positive lorsque l'inclusion des variables mesurées au niveau des populations améliorait l'inférence des coût (Figure 65).

Enfin, nous souhaitons identifier les conditions d'hétérogénéité spatiale des tailles de populations, de distribution spatiale des populations et les taux de migration pour lesquels l'inclusion de variables locales dans les modèles gravitaires améliorait l'inférence (Figure 65). Nous avons donc modélisé l'amélioration des performances des modèles due à l'inclusion des variables locales, c'est-à-dire la variable

D , en fonction du taux de migration, du type de modèle, de l'indice de Gini des capacités des taches et de la moyenne harmonique des vraies CDs. Pour ce faire, nous avons construit des arbres de régression (Breiman *et al.*, 1984) (cf. Annexe B4). Cette méthode d'analyse était plus adaptée à nos analyses que des modèles linéaires (des ANOVA par exemple). De plus, elle présentait l'avantage de nous fournir un arbre de décision délimitant les situations dans lesquelles l'utilisation de modèles gravitaires améliorait l'inférence. Nous avons également testé si les valeurs de D incluses dans chaque feuille des arbres de régression étaient significativement différentes de 0 en faisant un test de Student.

3 Résultats

3.1 Résultats des simulations

Globalement, les paramètres de simulation des paysages nous ont permis de faire varier considérablement les degrés d'hétérogénéité des capacités des taches et des tailles de populations et les degrés d'agrégation spatiale des populations. Par ailleurs, les coefficients de corrélation de Mantel entre les matrices de CDs obtenues en appliquant les scénarios de coût alternatifs et les "vraies" matrices de CDs étaient compris entre -0.350 et 0.999 (médiane : 0.628). Ainsi, la structure des paysages et la diversité des scénarios alternatifs permettaient d'obtenir la variabilité des matrices de CDs nécessaire à nos analyses.

Lors des simulations de flux génétique, il y avait en moyenne 2.0, 3.7, 7.1 et 17.3 migrants entre les 60 populations à chaque génération pour des taux de migration valant respectivement 0.0005, 0.001, 0.002 et 0.005. Les épisodes de dispersion à longue distance étaient plus fréquents pour les taux de migration les plus élevés. En effet, après 50 générations, le nombre de chemins de dispersion suivis par les individus parmi les 1770 possibles valait en moyenne 84 (± 15), 145 (± 16), 240 (± 26) et 423 (± 68) pour des taux de migration respectivement égaux à 0.0005, 0.001, 0.002 et 0.005. Les influences respectives de la dérive et du flux génétique sur le patron de différenciation génétique étaient donc contrastées, tout comme l'influence sur ce patron des épisodes de dispersion à longue distance.

3.2 Modèles gravitaires

Pour une simulation et un type de modèle donnés, les valeurs de R^2_β variaient fortement lorsque les CDs incluses dans le modèle étaient issues de différents scénarios (coefficients de variation compris entre 0.49 et 0.78). Cela signifie que les modèles permettaient de différencier les scénarios. Tous modèles, paramètres de taille des populations (Égales, Inégales) et scénarios de coût confondus, l'ajustement des modèles était minimal pour le taux de migration le plus faible et maximal pour le plus élevé (Table 12). Tandis que les valeurs de R^2_β médianes obtenues dans les 125 paysages étaient systématiquement plus fortes lorsque les "vraies" valeurs de CDs étaient incluses dans le modèle plutôt que les valeurs alternatives, une tendance opposée s'observait pour les valeurs de R^2_β maximales (Table 12). En effet, dans la plupart des cas, le modèle ayant le meilleur ajustement n'était pas celui qui incluait les "vraies" CDs déterminant le flux génétique simulé. Les quelques scénarios alternatifs responsables de ce résultat

différait du "vrai" scénario par les valeurs absolues des coûts, par l'ordre des types d'occupation du sol en termes de coût, ou par ces deux critères (par exemple [1, 4, 1000, 101], [1, 40, 1002, 10000], [1, 40, 10000, 1002], [40, 625, 10000, 3165] au lieu de [1, 10, 100, 1000]). Ils attribuaient généralement des coûts faibles aux forêts et aux prairies et des coûts élevés aux zones agricoles et artificialisées.

Tailles de pop.	Taux de mig.	Modèle	R_β^2 médian		R_β^2 max.	
			VRAI	ALTER	VRAI	ALTER
Égales	0.0005	Coût	0.053	0.036	0.117	0.525
Égales	0.0005	Coût-Capa	0.055	0.039	0.117	0.547
Égales	0.001	Coût	0.147	0.107	0.294	0.463
Égales	0.001	Coût-Capa	0.151	0.111	0.294	0.475
Égales	0.002	Coût	0.337	0.227	0.519	0.584
Égales	0.002	Coût-Capa	0.344	0.233	0.521	0.590
Égales	0.005	Coût	0.558	0.380	0.774	0.776
Égales	0.005	Coût-Capa	0.562	0.388	0.774	0.776
Inégales	0.0005	Coût	0.087	0.063	0.210	0.432
Inégales	0.0005	Coût-Capa	0.100	0.074	0.239	0.439
Inégales	0.0005	Coût-Ne	0.099	0.073	0.236	0.442
Inégales	0.001	Coût	0.182	0.133	0.344	0.446
Inégales	0.001	Coût-Capa	0.194	0.142	0.367	0.458
Inégales	0.001	Coût-Ne	0.193	0.142	0.372	0.471
Inégales	0.002	Coût	0.334	0.234	0.546	0.543
Inégales	0.002	Coût-Capa	0.345	0.242	0.548	0.545
Inégales	0.002	Coût-Ne	0.345	0.242	0.547	0.545
Inégales	0.005	Coût	0.539	0.366	0.732	0.735
Inégales	0.005	Coût-Capa	0.554	0.372	0.734	0.736
Inégales	0.005	Coût-Ne	0.554	0.373	0.733	0.737

TABLE 12 – Ajustement des modèles gravitaires mesuré par le R_β^2 selon les paramètres d'hétérogénéité des tailles de populations (Égales, Inégales), le taux de migration (0.0005, 0.001, 0.002, 0.005), les variables incluses dans les modèles et les scénarios de coût utilisés pour obtenir les distances-coût incluses dans les modèles. VRAI signifie que les modèles incluent les "vraies" valeurs de distances-coût déterminant les simulations tandis que ALTER signifie que les modèles incluent les valeurs de distances-coût alternatives. Pour les colonnes correspondant au cas ALTER, les valeurs médianes et maximales sont calculées pour chaque ligne à partir de 37000 valeurs (125 paysages \times 296 scénarios de coût), tandis que pour les colonnes correspondant au cas VRAI, elles sont calculées pour chaque ligne à partir de 125 valeurs.

Les coefficients de corrélation de Spearman r_{Sp} entre les valeurs de R_β^2 obtenues en utilisant les CDs alternatives dans les modèles et les corrélations de Mantel entre les "vraies" CDs et ces CDs alternatives prenaient des valeurs élevées, comprises en moyenne entre 0.782 et 0.944 (Table 13). Cela signifie que les modèles avec les meilleurs ajustements incluaient le plus souvent des CDs calculées selon des scénarios similaires au "vrai" scénario de coût. Les modèles permettaient donc d'inférer correctement les valeurs de coût, y compris dans les cas où les valeurs de R_β^2 étaient relativement faibles, comme on peut le constater en comparant les valeurs de R_β^2 de la Table 12 aux valeurs de r_{Sp} correspondantes dans la Table 13.

Quand les tailles des populations étaient hétérogènes et dépendaient des capacités des taches, en particulier pour des taux de migration faibles (0.0005 ou 0.001), les valeurs de r_{Sp} variaient davantage (Table 13). Pour des taux de migration et des modèles équivalents ("Coût", "Coût-Capa"), ces valeurs étaient légèrement plus faibles que lorsque les tailles de populations étaient égales. Cela signifie que la SHNe avait globalement un effet négatif sur la qualité de l'inférence des coûts. En outre, les différences entre les valeurs de r_{Sp} obtenues avec d'une part les modèles "Coût" et d'autre part les modèles "Coût-

Capa" ou "Coût-Ne" pour la distribution "Inégales" (valeurs D) étaient plus importantes pour les taux de migration les plus faibles (Table 13). Plus particulièrement, même si les valeurs maximales de r_{Sp} obtenues avec les trois types de modèles ("Coût", "Coût-Capa" et "Coût-Ne") étaient relativement similaires, les valeurs médianes et moyennes de r_{Sp} issues des modèles "Coût-Capa" et "Coût-Ne" étaient supérieures à celles issues du modèle "Coût" (Table 13). Il y avait par conséquent une certaine variabilité dans les performances relatives de ces modèles selon les paysages. En d'autres termes, si dans certains paysages l'inclusion des variables locales dans les modèles n'améliorait que marginalement l'inférence, cette amélioration était plus significative dans d'autres paysages. Nous nous focalisons donc désormais sur les résultats obtenus avec les taux de migration les plus faibles pour identifier les caractéristiques des paysages à l'origine de ces améliorations de l'inférence dues à la prise en compte de variables calculées au niveau des nœuds.

Tailles de pop.	Taux de mig.	Modèle	r_{Sp} min.	r_{Sp} médian	r_{Sp} moy.	r_{Sp} max.
Égales	0.0005	Coût	-0.526	0.838	0.782	0.980
Égales	0.0005	Coût-Capa	-0.606	0.848	0.790	0.982
Égales	0.001	Coût	-0.265	0.910	0.874	0.983
Égales	0.001	Coût-Capa	-0.392	0.919	0.878	0.984
Égales	0.002	Coût	0.259	0.938	0.915	0.988
Égales	0.002	Coût-Capa	0.186	0.943	0.917	0.984
Égales	0.005	Coût	0.628	0.957	0.944	0.989
Égales	0.005	Coût-Capa	0.646	0.961	0.944	0.988
Inégales	0.0005	Coût	-0.452	0.816	0.727	0.971
Inégales	0.0005	Coût-Capa	-0.397	0.872	0.798	0.974
Inégales	0.0005	Coût-Ne	-0.431	0.871	0.797	0.974
Inégales	0.001	Coût	-0.046	0.889	0.837	0.981
Inégales	0.001	Coût-Capa	-0.149	0.907	0.866	0.982
Inégales	0.001	Coût-Ne	-0.227	0.910	0.865	0.982
Inégales	0.002	Coût	0.118	0.932	0.904	0.986
Inégales	0.002	Coût-Capa	0.021	0.943	0.914	0.986
Inégales	0.002	Coût-Ne	-0.027	0.940	0.912	0.986
Inégales	0.005	Coût	0.656	0.960	0.943	0.989
Inégales	0.005	Coût-Capa	0.661	0.960	0.944	0.988
Inégales	0.005	Coût-Ne	0.666	0.960	0.944	0.988

TABLE 13 – Coefficients de corrélation de Spearman (r_{Sp}) entre les rangs des valeurs de R_{β}^2 des modèles gravitaires et les coefficients de corrélation de Mantel entre les "vraies" valeurs de CDs et les valeurs de CDs alternatives, selon les paramètres d'hétérogénéité des tailles de populations (Égales, Inégales), le taux de migration (0.0005, 0.001, 0.002, 0.005) et les variables incluses dans les modèles. Plus les valeurs de r_{Sp} sont élevées, plus les modèles sont capables d'identifier le scénario de coût le plus proche du "vrai".

3.3 Arbres de régression

Quand les tailles de populations dépendaient de la capacité des taches ("Inégales"), la différence de performance D entre les modèles n'incluant que les valeurs de CDs ("Coût") et les modèles gravitaires incluant à la fois des valeurs de CDs et des variables calculées au niveau des nœuds telles que les capacités des taches ou les tailles de populations ("Coût-Capa", "Coût-Ne") valaient en moyenne 0.050 pour des taux de migration égaux à 0.0005 ou 0.001 (Figure 67). Ces valeurs variaient entre -0.330 et 0.590. Le meilleur arbre de régression expliquant les valeurs de D incluait le taux de migration, la moyenne harmonique des valeurs de CDs et l'indice de Gini des capacités des taches comme variables explicatives, mais n'incluait pas le type de modèle. En effet, la différence entre les valeurs de D obtenues pour les modèles "Coût-Capa" et "Coût-Ne" était négligeable (Table 13). Cet arbre de

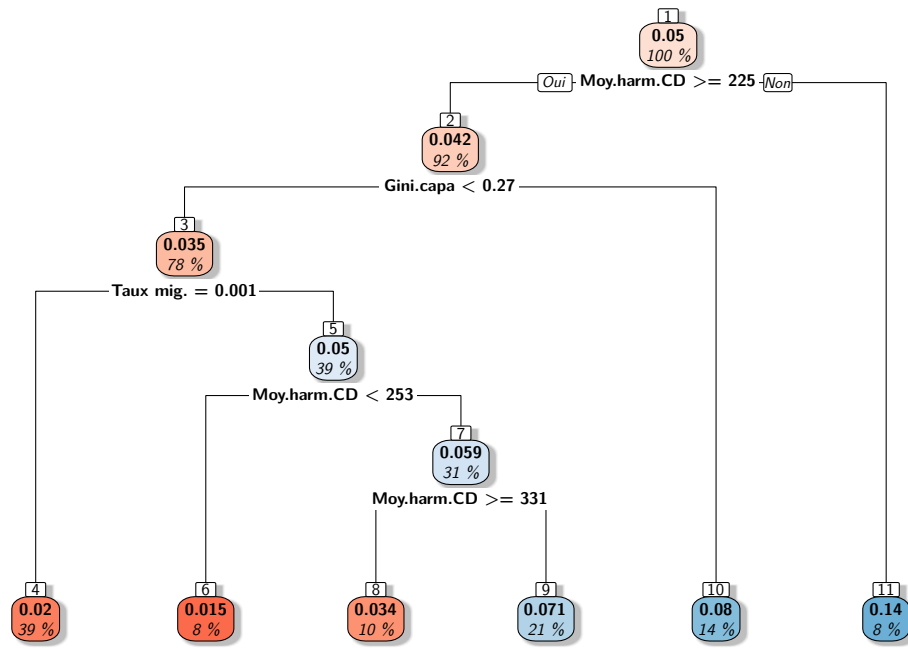


FIGURE 67 – Arbre de régression obtenu en considérant cinq variables explicatives (modèle, taux de migration (Taux mig.), moyenne harmonique des distances-coût (Moy.harm.CD), indice de Gini des capacités des tâches (Gini.capa)) pour expliquer la variable réponse D . Seuls les cas correspondant au scénario dans lequel les rangs des tailles des populations et des capacités des tâches d’habitat sont corrélés, et les taux de migration égaux à 0.0005 ou 0.001 sont considérés. L’arbre a été élagué de façon à avoir au moins 40 observations par feuille. Il y a 500 observations au total. Les nombres notés en gras dans les boîtes présentes à chaque nœud sont les valeurs moyennes de D pour les observations appartenant aux branches et feuilles situées en aval de chaque nœud. Les pourcentages indiquent la proportion des 500 observations appartenant à chaque branche ou feuille terminale.

régression expliquait 86 % de la variabilité de D et scindait l’espace des variables explicatives en six feuilles (Figures 67 et 68). Les valeurs de D appartenant à cinq de ces feuilles étaient positives et significativement différentes de 0 (Figure 68).

D’après les critères de séparation des branches de l’arbre de régression (Figure 67), quand les faibles valeurs de CDs étaient fréquentes ($\text{Moy.harm.CD} < 225$), l’ajout de variables locales dans les modèles gravitaires améliorait l’inférence des coûts. Dans ce dernier cas, D valait en moyenne 0.140. Selon le second critère de séparation des branches, l’avantage apporté par l’inclusion des variables locales était supérieur lorsque les capacités des tâches étaient les plus hétérogènes. En effet, lorsque l’indice de Gini des capacités des tâches était supérieur à 0.27, D valait en moyenne 0.080 contre 0.035 pour des degrés d’hétérogénéité inférieurs. Dans ce dernier cas, les valeurs moyennes de D étaient de 0.020 et 0.050 pour des taux de migration valant respectivement 0.001 et 0.0005 (Figure 67). Les modèles gravitaires amélioreraient donc davantage l’inférence lorsque le taux de migration était minimal. Pour un taux de migration égal à 0.0005, même si D prenait de faibles valeurs (0.015) lorsque la moyenne harmonique des CDs était inférieure à 253, sa valeur atteignait 0.071 pour des moyennes harmoniques des CDs comprises entre 253 et 331 (Figure 67). En outre, lorsque cette moyenne harmonique dépassait 331, D valait seulement 0.034 en moyenne. Ainsi, l’inclusion des variables locales semblait améliorer l’inférence dans plusieurs situations où la moyenne harmonique des distances-coût indiquait que les populations étaient relativement agrégées spatialement.

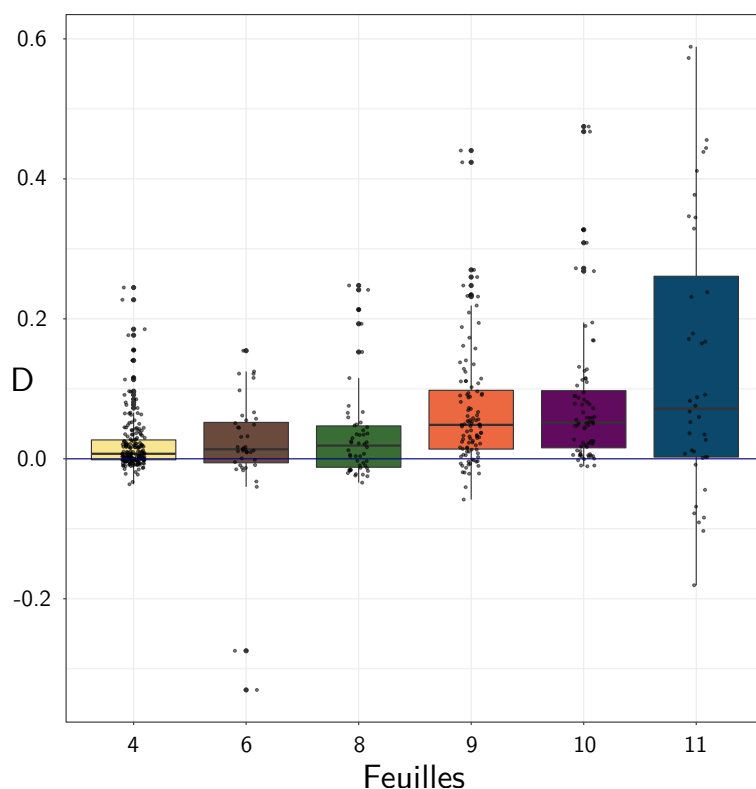


FIGURE 68 – Distribution statistique des valeurs de D dans les feuilles de l'arbre de régression présenté sur la figure 67 (les numéros des feuilles correspondent à ceux indiqués sur cette figure). Seuls les cas correspondant au scénario dans lequel les rangs des tailles des populations et des capacités des taches sont corrélés, et les taux de migration égaux à 0.0005 ou 0.001 sont considérés.

4 Discussion

4.1 L'inférence de valeurs de coût à partir de données génétiques est-elle correcte ?

Dans l'ensemble, les modèles permettaient d'inférer correctement les valeurs de coût. Cela confirme que les données génétiques sont pertinentes pour ce type d'analyse, comme l'avaient suggéré [Beier *et al.* \(2008\)](#) et démontré empiriquement [Zeller *et al.* \(2018\)](#). C'est lorsque les taux de migration étaient les plus faibles que nous avons obtenu les modèles ayant les plus faibles ajustements (faibles valeurs de R^2_β). Dans ces cas-là, les modèles n'identifiaient pas aussi bien les scénarios de coût les plus réalistes que pour des taux de migration plus élevés (plus faibles valeurs de r_{Sp} en comparaison). L'influence prépondérante de la dérive comparativement au flux génétique dans de tels cas pourrait expliquer ce résultat. Toutefois, même pour des taux de migration très faibles, les modèles parvenaient souvent à classer les scénarios de coût d'une façon qui reflétait relativement bien leurs similitudes avec le "vrai" scénario de coût (r_{Sp} moyens et médians élevés). Aussi, l'inférence de valeurs de coût à partir de données génétiques semble possible quand bien même le ratio entre le signal dû à l'effet du flux génétique et le bruit lié à l'effet aléatoire de la dérive est faible.

Néanmoins, les scénarios de coût "alternatifs" produisant les distances-coût les plus fortement corrélées aux "vraies" distances-coût étaient souvent identifiés comme étant les "meilleurs" selon un critère purement statistique d'ajustement des modèles. Ces scénarios différaient du "vrai" scénario de par leurs valeurs absolues et la résistance relative des types d'occupation du sol à la dispersion. L'utilisation d'un scénario optimisé sur la base d'un critère statistique pour modéliser spatialement des chemins

de dispersion pourrait conduire à des stratégies de conservation inadéquates. En effet, nous avons vu dans le chapitre 5 que des matrices de distances-coût fortement corrélées pouvaient correspondre à des chemins de moindre coût aux tracés très différents, en particulier lorsque les scénarios de coût dont elles sont issues classent les types d'occupation du sol dans un ordre différent. De plus, nous avons obtenu un tel résultat alors que nous avons fait en sorte qu'il y ait une certaine variabilité entre les matrices de distances-coût, condition nécessaire à la distinction des scénarios sur un critère statistique (Cushman *et al.*, 2013 ; Graves *et al.*, 2013). On peut donc s'attendre à ce que plusieurs scénarios de coût "alternatifs" fournissent des résultats équivalents dans des situations où la structure du paysage ou les caractéristiques des scénarios comparés ne permettent pas d'observer une telle variabilité (cf. Chapitre 5). Il semble que cette limite soit difficilement évitable, quelle que soit la méthode d'inférence utilisée dès lors qu'elle mobilise des matrices de CDs, car elle a pour origine la forte corrélation entre les matrices correspondant aux hypothèses mises en concurrence lors de l'analyse. Une solution potentielle consisterait à décrire le patron spatial des populations et l'hétérogénéité de la matrice paysagère qui les sépare autrement qu'avec des distances paysagères, par exemple avec des vecteurs propres issus de *Moran's Eigenvector Maps* à l'instar de Galpern *et al.* (2014) et plus récemment d'Hein *et al.* (2021).

Nous avons néanmoins observé que les modèles incluant les "vraies" distances-coût faisaient systématiquement partie des meilleurs modèles en termes d'ajustement. Considérer l'ensemble des scénarios de coût les plus vraisemblables d'après les résultats de l'inférence assurerait donc de considérer le scénario le plus réaliste, renforçant ainsi les préconisations formulées à la fin du chapitre 5. Ces différents scénarios pourraient toutefois fournir des chemins de moindre coût aux trajectoires spatiales assez distinctes. Dans le cas où ces trajectoires spatiales sont utilisées comme support d'une prise de décision en matière de conservation, il faudrait alors se focaliser sur les endroits où elles coïncident spatialement car ils indiquent les zones les plus probablement parcourues par l'espèce étudiée.

4.2 La SHNe, la distribution spatiale des populations et le taux de migration influencent-ils l'inférence ?

D'une part, conformément à notre hypothèse, l'hétérogénéité des tailles de populations avait tendance à réduire la qualité de l'inférence des coûts quand les taux de migration étaient les plus faibles (5×10^{-4} , 10×10^{-4}). Cette inférence nécessite que la différenciation génétique reflète l'influence du paysage sur le flux génétique (cf. Chapitre 1). Elle est donc contrainte par le bruit ajouté par la dérive au signal reliant différenciation génétique et flux génétique, en particulier lorsque les taux de migration sont faibles et que les populations ont des effectifs limités (Frankham, 1996b). Nous mettons ici en évidence une difficulté supplémentaire liée au fait que ce "bruit" n'est pas distribué de façon homogène dans l'espace. Cela rend plus difficile encore l'inférence de la résistance du paysage au flux génétique à partir du patron de différenciation génétique.

D'autre part, en accord avec notre hypothèse, l'agrégation spatiale des populations avait tendance à diminuer la qualité de l'inférence. Nous avons utilisé la moyenne harmonique des distances-coût pour distinguer les paysages dans lesquels les flux génétiques s'effectuaient préférentiellement à une échelle spatiale restreinte car les populations formaient plusieurs agrégats spatiaux. Lorsque les tailles des populations sont hétérogènes et dépendent de la capacité des taches, l'agrégation spatiale de populations dans les zones les plus favorables du paysage augmente la fréquence des flux génétiques entre des populations voisines et de grande taille. Cela fait augmenter la différenciation génétique de ces

populations vis-à-vis (i) d'autres populations petites et isolées et (ii) de grandes populations situées dans d'autres "clusters" de populations. Il devient alors difficile de relier le patron global de différenciation génétique à la résistance de la matrice paysagère. Ce point avait déjà été suggéré par [Graves et al. \(2012\)](#) et [Graves et al. \(2013\)](#), mais n'avait pas été étudié dans le contexte de l'inférence des coûts.

L'hétérogénéité des tailles de populations et leur degré d'agrégation spatiale dépendent tous deux directement de la surface et de la configuration des taches d'habitat et de la distribution spatiale des populations dans ces taches. Ces deux paramètres influençaient de façon significative l'inférence des coûts. Cela signifie qu'indépendamment de l'espèce étudiée et de son taux de migration, la structure du paysage et en particulier le patron spatial des habitats sont des paramètres à prendre en compte lorsqu'on envisage d'inférer la résistance d'un paysage, comme l'avaient suggéré [Cushman et al. \(2013\)](#).

La prise en compte d'un tel résultat lors de la réalisation d'une étude empirique de génétique du paysage est limitée par le fait que les valeurs absolues des variables incluses dans l'arbre de régression (indice de Gini, moyenne harmonique des CDs) permettent difficilement de se représenter les situations réelles dans lesquelles on pourrait les obtenir. Toutefois, si une étude visant à inférer la résistance du paysage peut être réalisée dans différentes zones, nos résultats indiquent qu'il est préférable de la mener dans celle où on s'attend à ce que (i) les tailles de populations soient les plus homogènes (en se basant sur la variabilité des tailles des taches d'habitat par exemple) et que (ii) les populations échantillonnées aient la distribution spatiale la plus régulière. Enfin, notons que même si nous avons simulé les flux génétiques entre les populations d'une espèce forestière virtuelle dans un paysage hétérogène, nos résultats sont applicables à l'inférence de la résistance du paysage à la dispersion pour tout type d'espèce dès lors qu'elle forme des populations discrètes dans un type d'habitat délimitable et se déplace au sein d'un paysage hétérogène.

4.3 Quand devrait-on inclure des variables intra-populationnelles dans des modèles gravitaires pour inférer des valeurs de coût ?

Notre second objectif était d'évaluer l'intérêt qu'il y a à prendre en compte des variables locales représentant l'hétérogénéité des tailles de populations dans l'inférence des valeurs de coût. Nous avons pour cela réalisé des modèles gravitaires dans lesquels des distances-coût calculées au niveau des liens entre populations et des variables locales calculées au niveau des nœuds, telles que les capacités des taches ("Coût-Capa") et les tailles des population ("Coût-Ne"), étaient des variables explicatives des distances génétiques entre populations (liens). Lorsque l'hétérogénéité des tailles de populations influençait l'inférence, l'inclusion de variables locales dans les modèles gravitaires en améliorait la qualité (valeurs de D positives), en accord avec notre hypothèse. Nos résultats élargissent ainsi la portée de la recommandation de [Prunier et al. \(2017\)](#) de tenir compte de l'hétérogénéité des tailles de populations au contexte spécifique de l'inférence des valeurs de coût.

L'apport de l'inclusion de variables locales dans les modèles gravitaires pour inférer des valeurs de coût ne dépendait pas uniquement du taux de migration et de l'hétérogénéité des tailles de populations, il dépendait également du degré d'hétérogénéité des tailles de populations et de leur degré d'agrégation spatiale. D'une part, les valeurs de D étaient plus élevées lorsque les capacités des taches, et les tailles de populations dépendant de celles-ci, étaient les plus hétérogènes. Ce résultat rappelle

celui de [Prunier et al. \(2017\)](#), qui avaient montré que la SHNe expliquait significativement le patron de différenciation génétique à condition que l'hétérogénéité spatiale des populations soit importante et que le taux de migration soit faible. Cependant, ces auteurs évaluaient l'hétérogénéité des tailles de populations à l'aide de leur coefficient de variation et non pas de l'indice d'inégalité de Gini ici utilisé.

Par ailleurs, nous avons montré que l'inclusion dans les modèles gravitaires de variables paysagères calculées au niveau des populations à partir du patron spatial des habitats pourrait améliorer l'inférence des coûts. En effet, inclure soit la capacité des taches, soit les tailles de populations dans ces modèles apportait un résultat équivalent lorsqu'une telle inclusion améliorait l'inférence. Compte tenu de nos paramètres de simulation, les valeurs de ces deux variables n'étaient pas directement proportionnelles. Seuls les rangs des valeurs étaient corrélés. Une telle situation se rencontre très probablement dans la réalité lorsque la capacité des taches détermine la taille des populations qui les occupent. De façon similaire, [Prunier et al. \(2017\)](#) avaient estimé la taille de populations de goujons occitans (*Gobio occitaniae*) à partir de la largeur des portions de rivières qu'elles occupaient. En utilisant ces estimations, ils avaient mesuré le degré d'hétérogénéité des tailles de populations et montré qu'il expliquait de façon significative le patron de différenciation génétique. Ces résultats signifient que de coûteux inventaires de terrain peuvent être évités lorsque des variables environnementales déterminant la taille des populations sont connues et estimables.

4.4 Limites et perspectives

Les taux de migration pour lesquels nous avons observé une influence significative de l'hétérogénéité des tailles de populations sur l'inférence étaient faibles. Néanmoins, ils reflètent des situations réalistes étant donné que les taux de migration inférés à partir de données génétiques sont souvent bien plus faibles que les taux de mouvement observés entre des taches d'habitat (0.5 % *versus* 7-32 % respectivement, selon [Riley et al. \(2006\)](#) ; voir aussi [Meirmans \(2014\)](#)). De plus, ce résultat est similaire à celui de [Prunier et al. \(2017\)](#), même si nous avons utilisé une gamme de taux de migration n'incluant que les taux les plus faibles utilisés par ces auteurs. Toutefois, nos résultats montrent que les variables locales améliorent l'inférence des coûts uniquement lorsque le flux génétique est très limité mais ils ne signifient pas que l'influence de l'hétérogénéité des tailles de populations n'est pas significative lorsque les flux génétiques sont plus fréquents.

En outre, nous avons considéré que nous avons échantillonné l'ensemble exhaustif des populations. En pratique, une telle intensité d'échantillonnage est difficilement atteignable, bien que fortement recommandée ([Van Strien, 2017](#)). Nous pouvons donc nous demander à quel point nos résultats seraient affectés par un échantillonnage partiel. Dans ce cas, les modèles gravitaires pourraient s'avérer utiles pour prédire les distances génétiques entre des populations non échantillonnées. Une telle utilisation consisterait à intégrer des graphes génétiques et paysagers comportant des nombres différents de nœuds pour prédire les distances génétiques entre l'ensemble des taches d'habitat (nœuds) du graphe paysager. Ces prédictions se baseraient sur les données génétiques d'un nombre limité de populations correspondant aux nœuds du graphe génétique. Il serait alors nécessaire que le modèle gravitaire basé sur les nœuds et liens de ces deux graphes ait un bon ajustement statistique, garant de la qualité des prédictions faites à partir de ce modèle. Dans ce chapitre, nous avons réalisé de tels modèles sans objectif prédictif et à partir de graphes génétiques et paysagers de même dimension. Nos résultats nous

ont montré que l'ajout des variables locales améliorerait l'ajustement des modèles dans des situations où les effets de la dérive étaient prédominants. Bien qu'amélioré, l'ajustement des modèles était toutefois limité, ce qui rendraient les prédictions basées sur ces modèles fortement incertaines.

L'utilisation de modèles gravitaires pour intégrer des graphes génétiques et paysagers à des fins prédictives nécessite donc des recherches supplémentaires. Elle serait probablement plus adaptée aux situations dans lesquelles les variables locales incluses dans les modèles influencent la fréquence des flux génétiques et non pas uniquement l'intensité de la dérive aléatoire (Pflüger et Balkenhol, 2014 ; Watts *et al.*, 2015). C'est par exemple le cas lorsque la surface d'une tache d'habitat ou la densité des individus qui l'occupent détermine son caractère attracteur pour les individus qui se dispersent (Clobert *et al.*, 2012)⁴. D'autres métriques locales issues de graphes paysagers pourraient potentiellement être incluses dans les modèles gravitaires pour cela. Il serait également envisageable de calculer des variables mesurant l'hétérogénéité des capacités des taches d'un graphe ou leur degré d'agrégation spatiale. De telles variables permettraient d'identifier les contextes dans lesquels l'hétérogénéité des tailles de populations pourrait avoir une influence significative sur le patron de différenciation génétique.

5 Conclusion

Les études de génétique du paysage ont très tôt pris en compte l'hétérogénéité de la matrice paysagère pour inférer la résistance du paysage au flux génétique. En revanche, elles ont rarement considéré l'influence des taux de migration, de l'hétérogénéité des tailles de populations et de leur degré d'agrégation spatiale sur cette inférence. Ici, nous montrons que l'inférence des valeurs de coûts à partir de données génétiques est satisfaisante dans une large gamme de situations, bien que dépendante des taux de migration, de l'hétérogénéité des tailles de populations et de la distribution spatiale des populations dans le paysage. Ce chapitre illustre l'intérêt de l'intégration de graphes génétiques et paysagers dans le cas particulier de l'inférence des coûts. Les variables associées aux nœuds et aux liens de ces deux outils peuvent être intégrées à des modèles gravitaires pour améliorer l'inférence dans un certain nombre de situations. L'utilisation de ces modèles à des fins prédictives justifierait néanmoins de futures recherches.

4. Le lien entre la densité d'une population et le caractère attracteur d'une tache dépend fortement de l'espèce étudiée et du contexte. L'ouvrage de Clobert *et al.* (2012) décrit les résultats des travaux menés sur les mécanismes complexes impliqués dans la dispersion des individus.

Discussion générale

*La science cherche le mouvement perpétuel. Elle l'a trouvé; c'est elle-même.
Tout remue en elle, tout change, tout fait peau neuve.
Ce qu'on acceptait hier est remis à la meule aujourd'hui.
Elle n'est jamais satisfaite; elle est insatiable du mieux.
La science va sans cesse se raturant elle-même. Ratures fécondes.
La science est l'asymptote de la vérité. Elle approche sans cesse, et ne touche jamais.
Cherchons encore. Cette agitation est superbe.*

Victor Hugo

Dans le cadre de cette thèse, nous avons étudié les modalités d'utilisation conjointe des graphes génétiques et paysagers pour l'analyse de la connectivité écologique des habitats. Notre travail s'est structuré en deux parties, reflétant le fait que :

- ces deux graphes peuvent être comparés après avoir été construits indépendamment (comparaison *a posteriori*, première partie),
- ou que les données et méthodes mobilisées pour la construction et l'analyse de ces graphes peuvent être utilisées conjointement (intégration *a priori*, seconde partie).

En introduction, nous avons fait l'hypothèse que ces deux approches permettraient chacune de répondre à des questions théoriques en génétique du paysage. D'autre part, leur utilisation conjointe nous a permis d'évaluer la pertinence de l'utilisation des graphes dans le cadre de la conservation de la connectivité. Dans les sections suivantes, nous discutons de l'apport des comparaisons et intégrations de graphes génétiques et paysagers mises en œuvre pour atteindre les objectifs formulés au départ. Nous mentionnons à cette occasion les limites de notre travail, leurs solutions potentielles et les perspectives de recherche auxquelles cette thèse ouvre la voie.

1 Les graphes génétiques et paysagers permettent-ils d'améliorer notre compréhension des relations entre connectivité des habitats et structure génétique ?

Les graphes génétiques et paysagers représentent respectivement la structure génétique d'un réseau de populations et le patron de connectivité d'un réseau de taches d'habitat. En mettant en relation de plusieurs manières les propriétés de leurs nœuds et/ou de leurs liens, **nous avons répondu à des questions théoriques propres à la génétique du paysage ou héritées de l'écologie du paysage et de la génétique des populations. Nous revenons sur les apports de ces analyses tout en évoquant leurs limites.**

1.1 Enseignements tirés de la mise en relation de métriques de connectivité et d'indices génétiques issus de graphes

1.1.1 Mise en évidence empirique de l'influence de la quantité d'habitat atteignable

La définition même du concept de connectivité écologique des habitats et sa mesure font débat (Pascual-Hortal et Saura, 2006 ; Saura et Rubio, 2010 ; Taylor *et al.*, 2006 ; Tischendorf et Fahrig, 2000). Ces questions fondamentales s'inscrivent dans un débat plus large au sein de la communauté de l'écologie scientifique, relatif à l'influence de la quantité d'habitat et de la configuration des taches d'habitat sur la biodiversité (Didham *et al.*, 2012 ; Fahrig, 2013 ; Haddad *et al.*, 2017a ; Saura, 2021). Dans ce cadre, nous sommes partis du principe que la connectivité des habitats pouvait être assimilée à la quantité d'habitat atteignable du point de vue de l'espèce étudiée et à plusieurs échelles spatiales et temporelles. Nous avons également souligné l'importance de considérer un ensemble de métriques complémentaires pour caractériser ses différentes composantes. Or, les métriques issues des graphes paysagers telles que la capacité des taches et les métriques de flux (F) et de *betweenness centrality* (BC) sont justement censées refléter ses différentes composantes, à savoir la quantité d'habitat atteignable au sein de la tache focale elle-même, à partir de cette tache ou atteignable du fait de sa position spécifique dans le réseau. Le caractère intégrateur de ces métriques revêt un intérêt particulier lorsqu'il s'agit d'expliquer la structure génétique des populations dans la mesure où la dérive est théoriquement influencée par la capacité des taches d'habitat tandis que les flux génétiques sont supposés dépendre davantage de la quantité d'habitat atteignable en dehors des taches.

Afin de valider la pertinence de ces métriques, nous les avons confrontées à des réponses génétiques dans le cadre de deux études empiriques concernant une espèce d'orthoptère en Suisse et une espèce de passereau forestier en Guadeloupe (Chapitres 3 et 4). Dans ces deux cas, des indices étaient calculés au niveau des nœuds de graphes génétiques pour refléter la diversité génétique locale des populations (richesse allélique) et leur niveau de différenciation génétique relativement aux autres populations (indice MIW). Ces différentes réponses génétiques étaient chacune corrélées significativement aux métriques issues des graphes dans ces deux études. En plus de confirmer l'intérêt de ces métriques, ces analyses confirment qu'assimiler la connectivité à une quantité d'habitat atteignable, comme nous l'avons fait en nous basant entre autres sur les travaux de Santiago Saura (Saura et de la Fuente, 2017 ; Saura,

2018), est pertinent pour comprendre les processus écologiques à l'origine de la structure génétique, à savoir la dérive et les flux génétiques.

L'intérêt de la complémentarité de ces métriques s'illustre également par le fait que les degrés de corrélation obtenus étaient différents selon les cas d'étude et les métriques considérés. Cela permet dans chaque cas de formuler des hypothèses sur les déterminants de la structure génétique des populations et sur sa dynamique. Par exemple, la capacité des taches n'était corrélée significativement qu'à la richesse allélique des populations de Criquet ensanglanté, tandis qu'une telle corrélation s'observait à la fois pour la richesse allélique et le MIW dans le cas de la Paruline caféïette. De plus, les corrélations étaient légèrement plus faibles et moins souvent significatives dans le premier cas que dans le second. Les traits écologiques contrastés des deux espèces étudiées (un orthoptère de milieux ouverts et un passereau forestier) et les histoires évolutives de leurs populations dans ces différents contextes paysagers en sont peut-être la cause. En effet, les milieux essentiellement agricoles constituant l'habitat du Criquet ensanglanté sont soumis à des perturbations fréquentes liées aux pratiques agricoles et aux choix d'assolement. À l'inverse, la distribution spatiale des forêts occupées par la Paruline caféïette a peu évolué ces dernières décennies, en particulier sur Basse-Terre où la majeure partie des forêts est protégée par la réglementation du Parc National de Guadeloupe. En outre, la Paruline caféïette est une espèce forestière spécialiste qui semble peu affectée par la structure de l'habitat dès lors qu'il est de nature forestière (Khimoun *et al.* (2016b) ; voir aussi Chapitre 3). Cela facilite la modélisation et la délimitation de son habitat à partir de variables d'occupation du sol. De plus, dans le contexte insulaire de la Guadeloupe, les métriques de connectivité calculées à partir des taches d'habitat ne sont pas soumises aux effets de bord qui peuvent parfois biaiser les analyses (Koen *et al.*, 2010). Dans le cas du Criquet ensanglanté en revanche, la structure de la végétation et le caractère inondable des milieux ouverts déterminent de façon importante leur utilisation en tant qu'habitat. Néanmoins, nous ne disposons pas de données spatiales concernant ces caractéristiques. De plus, l'aire de distribution de l'espèce s'étend au-delà des limites de la zone d'étude, ce qui peut influencer le calcul des métriques de connectivité. Enfin, alors que le Criquet ensanglanté effectue son cycle de vie en une année (Bönsel *et Sonneck*, 2011), la Paruline caféïette peut vivre plusieurs années⁵. En somme, la connectivité des habitats du Criquet ensanglanté est plus difficile à cartographier précisément que celle des habitats de la Paruline caféïette, ce qui a potentiellement limité notre capacité à analyser son influence sur la structure génétique des populations. Ajoutons que la structure génétique des populations de Criquet ensanglanté est probablement éloignée de son état équilibre dans ce contexte paysager dynamique. L'ensemble de ces raisons pourrait expliquer le lien plus faible dans le cas du Criquet ensanglanté entre la structure génétique des populations et des métriques reflétant l'influence de la distribution des taches d'habitat modélisées.

Les coûts de déplacement attribués aux différents éléments de la matrice paysagère avaient une influence notable sur les corrélations mises en évidence dans ces deux études empiriques. Mony *et al.* (2018) et Foltête *et Giraudoux* (2012), entre autres, avaient déjà mis en évidence la relation entre des métriques calculées à partir de graphes paysagers en tenant compte de l'hétérogénéité de la matrice paysagère et des réponses biologiques telles que la composition de communautés d'espèces végétales ou l'abondance de populations de micro-mammifères, respectivement. Nos résultats confirment ainsi

5. Des individus bagués au début du programme de recherche Frag& Binv dans la continuité duquel s'inscrit cette étude ont été recapturés plusieurs années plus tard (Garnier, S., communication personnelle).

l'influence de l'hétérogénéité de la matrice paysagère sur la dispersion à partir de données génétiques censées refléter ce processus plus directement, à l'instar de précédents travaux (Coulon *et al.*, 2004 ; McRae, 2006). Toutefois, peu de travaux avaient auparavant montré comment prendre en compte le comportement de déplacement des espèces à travers la matrice pour mesurer la quantité d'habitat atteignable et expliquer son influence sur les flux génétiques. Les approches basées sur les graphes permettent une telle prise en compte car elles intègrent des distances-coût dans le calcul des métriques. En outre, les analyses concernant le Criquet ensanglanté (Chapitre 3) ont révélé que des métriques plus classiquement calculées pour décrire la distribution spatiale des habitats, telles que la distance à la tache d'habitat la plus proche ou la quantité d'habitat à un rayon défini autour d'un site, n'expliquaient pas aussi bien les réponses génétiques que les métriques issues des graphes paysagers.

1.1.2 La quantité d'habitat atteignable affecte différemment la diversité et la différenciation génétiques

La façon dont évoluent d'une part la diversité génétique locale et d'autre part la différenciation génétique entre populations est au cœur des réflexions menées en génétique des populations (Keyghobadi *et al.*, 2005 ; Varvio *et al.*, 1986). Ici, nous avons comparé l'influence relative de la quantité d'habitat atteignable sur ces deux composantes de la structure génétique grâce à l'utilisation de graphes génétiques dans le cadre de deux études empiriques (Chapitres 3 et 4). Les indices calculés au niveau de leurs nœuds représentaient (i) la diversité génétique au sein des populations ou (ii) la différenciation génétique entre une population et toutes celles auxquelles elle est connectée sur ce graphe. Les réponses de ces deux composantes étaient différentes dans le cas du Criquet ensanglanté et nous avons fait l'hypothèse que cela pouvait s'expliquer par des différences temporelles dans la réponse aux évolutions passées du paysage. Inversement, ce n'était pas le cas pour la Paruline caféïette pour laquelle ces réponses étaient relativement similaires, probablement en raison de la plus grande stabilité de l'habitat forestier de cette espèce.

Au vu des résultats différents obtenus dans ces deux cas, on peut également s'attendre à obtenir des résultats différents pour d'autres espèces et dans d'autres contextes paysagers. Les déterminants de la relation entre les composantes de la structure génétique et de la quantité d'habitat atteignable restent donc à étudier dans une plus large gamme de situations. Cela permettrait de tester les attendus théoriques de cette relation et le caractère généralisable des résultats obtenus. Les approches basées sur des graphes génétiques et paysagers mises en œuvre ici permettraient de traiter cette question soit (i) en échantillonnant un grand nombre d'espèces aux traits contrastés (taxons, espérance de vie, types d'habitat, distances de dispersion, niveaux d'abondance) dans différents paysages, soit (ii) en simulant les réponses génétiques au patron spatial des habitats dans une gamme variée de situations virtuelles. Ces deux approches seraient certainement complémentaires. La première demanderait des investissements conséquents pour l'obtention des données. Elle serait en outre soumise aux limites habituelles des études empiriques en écologie, à savoir un contrôle limité des conditions environnementales et des résultats basés sur des analyses essentiellement corrélatives. Toutefois, ses résultats refléteraient la réalité écologique des processus en jeu et auraient à ce titre un grand intérêt. La seconde serait plus simple à mettre en œuvre et permettrait de contrôler les conditions environnementales simulées, mais serait soumise à l'incertitude des hypothèses sous-jacentes aux simulations réalisées (Zurell *et al.*, 2010).

En conclusion, la mise en relation d'indices génétiques et de métriques de connectivité issues de graphes confirme que :

- les attendus théoriques concernant la relation entre la quantité d'habitat atteignable et la structure génétique des populations s'observent empiriquement,
- mesurer la quantité d'habitat atteignable à l'échelle de la tache focale, au-delà de cette tache et en considérant le réseau d'habitat dans son ensemble à l'aide de métriques complémentaires permet d'analyser l'influence de cette variable sur les différentes composantes de la structure génétique des populations,
- calculer ces métriques à l'aide du cadre flexible que constituent les graphes paysagers permet de s'adapter à la diversité des espèces étudiées et des contextes paysagers en intégrant dans l'analyse leurs capacités de dispersion et leur rapport à l'occupation du sol.

1.2 Intérêt des graphes pour l'analyse de la relation entre distances génétiques et paysagères

1.2.1 Élaguer les graphes pour ajuster l'échelle de l'analyse

Miele *et al.* (2019) conseillent de commencer par préciser la question posée avant d'utiliser des graphes pour y répondre⁶. Un tel conseil peut paraître trivial. Pourtant, nous avons remarqué que les objectifs poursuivis lors de l'utilisation d'un graphe génétique n'étaient pas toujours formulés explicitement, ni les méthodes employées pour y parvenir choisies en adéquation. Cela nous a conduits à comparer les méthodes de construction et d'analyse des graphes génétiques en distinguant les utilisations visant à localiser les chemins de dispersion directs de celles visant à quantifier la résistance du paysage à partir de la relation entre distances génétiques et paysagères (cf. Chapitre 1). Une des conclusions majeures de ce chapitre était que l'élagage d'un graphe génétique peut servir à sélectionner les paires de populations à considérer dans ce second cas pour améliorer la puissance des analyses, comme l'avaient suggéré Wagner et Fortin (2013). En effet, nos simulations ont montré que lorsque la différenciation génétique n'a pas atteint un état d'équilibre reflétant l'influence conjointe de la dérive et du flux génétique (patron d'IBD de type IV selon Hutchison et Templeton (1999)), il est préférable d'élaguer le graphe pour estimer au mieux la relation entre la différenciation génétique et les distances-coût déterminant le flux génétique. Ce résultat basé sur des simulations a des équivalents dans des études empiriques précédemment publiées (voir Castric et Bernatchez (2003) ou Hänfling et Weetman (2006)), bien que les conclusions de ces dernières aient rarement été interprétées en faisant référence à l'élagage d'un graphe ou à l'équilibre entre dérive et flux génétique.

Par la suite, nous avons montré l'intérêt de l'élagage des graphes dans le cadre d'analyses empiriques. En effet, dans le chapitre 4, la considération d'un sous-ensemble de paires de populations permettait d'améliorer les prédictions de la différenciation génétique entre populations de Paruline caféïette basées sur les distances paysagères entre les taches d'habitat occupées par ces populations. Un tel résultat a pu être obtenu par l'utilisation d'un modèle mixte de type MLPE, réputé plus robuste que les corrélations de Mantel utilisées dans le chapitre 1 (Shirk *et al.*, 2017b), et par le développement d'une méthode de validation croisée.

6. *Tip 1 : Formulate questions first, use networks later* (Miele *et al.*, 2019)

En outre, en élaguant les liens du graphe génétique représentant le réseau de populations de Criquet ensanglanté, nous avons obtenu des corrélations plus fortes entre l'indice de différenciation génétique relative MIW et les métriques de connectivité que dans le cas d'un graphe non élagué. Même si cette analyse n'était qu'indirectement basée sur la relation entre distances génétiques et paysagères, l'ensemble de ces résultats basés à la fois sur des simulations et sur des données empiriques illustrent l'intérêt de l'élagage des graphes pour les analyses de génétique du paysage. Notons néanmoins qu'ils suggèrent que l'élagage est principalement utile lorsque l'état d'équilibre du patron de différenciation génétique n'a pas été atteint. Aussi, nous n'avons pas élagué les matrices de distances entre populations dans le chapitre dédié à l'inférence de la résistance du paysage à l'aide de données génétiques et de modèles gravitaires (Chapitre 6) car nous avons réalisé les simulations génétiques sur un pas de temps assez long pour qu'il reflète l'état d'équilibre de la différenciation génétique.

En somme, l'intérêt de l'élagage des graphes génétiques est lié au fait qu'il permet d'identifier l'échelle spatiale à laquelle la structure génétique est influencée par le paysage. Il permet ainsi d'ajuster la topologie des paires de populations à prendre en compte pour caractériser cette influence. Dans l'introduction, nous avons mentionné le lien entre l'élagage des graphes et les analyses d'autocorrélation spatiale et nos résultats rappellent ce parallèle. En effet, à l'instar de [Van Strien *et al.* \(2015\)](#), suite aux simulations de flux génétiques réalisées pour comparer des graphes génétiques (Chapitre 1), nous avons identifié la distance maximale de corrélation (DMC) entre distances génétiques et paysagères via des élagages successifs basés sur des distances seuil. De façon similaire, [Wagner *et al.* \(2005\)](#) et [Brooks \(2006\)](#) avaient mis en évidence l'intérêt des variogrammes et des corrélogrammes de Mantel pour estimer l'échelle spatiale à laquelle un patron d'IBD s'observait, ainsi que pour analyser à cette échelle la structure génétique des populations d'une espèce de lichen (*Lobaria pulmonaria*), d'une salamandre (*Plethodon glutinosus*) ou du pathogène d'une plante (*Microbotryum violaceum*). Bien que l'élagage des graphes génétiques basé sur des distances seuil s'apparente aux approches mises en œuvre par ces auteurs, la diversité des méthodes d'élagage testées dans notre travail permet d'aller au-delà et d'exploiter davantage le potentiel de la théorie des graphes. Par exemple, les élagages basés sur un graphe de Gabriel, un seuil de percolation, ou sur la topographie de l'île de Guadeloupe sont différents des méthodes basées sur l'autocorrélation spatiale ou sur le choix d'une distance seuil mais étaient mieux adaptés aux questions posées dans les chapitres dans lesquels nous les avons utilisés.

Ainsi, les méthodes d'élagage des graphes ont un intérêt pour l'analyse de la relation entre distances génétiques et paysagères. Rappelons que c'est sur cette relation que se base l'optimisation des surfaces de résistance, de plus en plus souvent réalisée à l'aide d'algorithmes génétiques via l'utilisation du package R `ResistanceGA` ([Peterman, 2018](#) ; [Peterman *et al.*, 2019](#) ; [Winiarski *et al.*, 2019](#)). Ces analyses sont particulièrement utiles en tant qu'elles permettent de calibrer des surfaces de résistance à partir de données génétiques avant de modéliser la connectivité des habitats. Pourtant, comme la plupart des analyses de génétique du paysage, elles considèrent l'intégralité des paires de populations. Si la différenciation génétique entre certaines paires de populations dépend davantage de la dérive génétique aléatoire que du flux génétique, déterminé quant à lui par la résistance de la matrice et le patron spatial des habitats, la prise en compte de ces paires de populations pourrait bruite le signal sur lequel se base cette optimisation et en réduire ainsi la qualité. Par ailleurs, cette optimisation est le plus souvent réalisée à l'aide de modèles mixtes de type MLPE. Or, nous avons montré dans le chapitre 4 que ces modèles étaient influencés par la topologie des paires de populations de Paruline caféïette

considérées. Ainsi, pour compléter les résultats déjà obtenus, il faudrait évaluer à quel point l'élagage des graphes permet de renforcer la qualité de l'optimisation d'une surface de résistance. Cette question pourrait être traitée à l'aide de simulations similaires à celles mises en œuvre dans les chapitres 1 et 6, ou à l'aide de données empiriques en utilisant la méthode de validation croisée développée dans le chapitre 4. Elle pourrait également être traitée avec des méthodes de statistique spatiale récemment introduites en génétique du paysage. Galpern *et al.* (2014) ont développé une méthode de modélisation de la relation entre des distances génétiques et la distribution spatiale des populations basée sur les *Moran Eigenvector Maps* (Borcard et Legendre, 2002 ; Dray *et al.*, 2006 ; Dray, 2011)⁷. Cette méthode consiste à résumer des matrices de distances "élaguées" ou non⁸ sous la forme d'ensembles de vecteurs propres orthogonaux avant de les mettre en relation. Elle tient compte de l' "élagage" des matrices de distances pour résumer la structure spatiale des populations. De plus, elle permet de s'affranchir des limites associées aux méthodes d'analyses de distances plus classiques (voir Legendre et Fortin (2010)) dans la mesure où les vecteurs propres comportent autant de valeurs que de populations. Néanmoins, elle n'a jamais été mise en œuvre pour analyser des patrons d'isolement par la résistance du paysage alors qu'elle aurait un intérêt pour intégrer l'élagage des graphes dans ce cadre.

1.2.2 Considérer les nœuds et les liens des graphes pour intégrer l'hétérogénéité spatiale de la dérive

Dans le chapitre dédié à la comparaison des méthodes de construction des graphes génétiques (Chapitre 1), nous avons fait le constat que nous ne pouvons pas estimer convenablement l'influence du paysage si une partie des valeurs de différenciation génétique prises en compte dans cette analyse dépend essentiellement de la dérive génétique. L'influence de la dérive était également au cœur de la réflexion menée dans le chapitre 6, dans lequel nous nous sommes demandé si le fait de négliger son hétérogénéité spatiale pouvait constituer un obstacle à l'inférence de l'effet de la matrice paysagère sur la différenciation génétique. En effet, bien que l'issue de la dérive soit aléatoire, son intensité est inversement proportionnelle à la taille des populations qu'elle affecte. Ainsi, la dérive n'a pas le même effet sur la différenciation génétique entre deux paires de populations selon les tailles des populations qui constituent chacune de ces paires. Or, la taille des populations dépend le plus souvent de conditions environnementales spatialement hétérogènes. Récemment, Prunier *et al.* (2017) ont montré que l'hétérogénéité spatiale des tailles de populations pouvait influencer la différenciation génétique d'une façon aussi significative que la distance entre ces populations, en particulier lorsque les taux de migration sont faibles et les tailles de populations fortement hétérogènes. Néanmoins, ces auteurs n'ont pas évalué à quel point un tel effet pouvait limiter notre capacité à évaluer l'influence de la matrice paysagère sur la différenciation génétique.

L'intensité de la dérive est censée dépendre de caractéristiques des nœuds de graphes génétiques ou paysagers telles que la taille des populations ou la surface des taches d'habitat lorsque cette dernière en est une bonne approximation (Prunier *et al.*, 2017). Au contraire, l'intensité des flux génétiques est censée dépendre des distances-coût qui pondèrent les liens des graphes paysagers, tandis que la différenciation génétique entre populations est le poids associé aux liens des graphes génétiques. Partant de ce constat, nous nous sommes demandé si la prise en compte dans un même modèle de l'ensemble des variables relatives aux nœuds et aux liens des graphes permettrait d'intégrer l'hétérogénéité spatiale

7. Galpern *et al.* (2014) ont développé le package R MEMGENE pour mettre en œuvre cette méthode.

8. Borcard et Legendre (2002) parlent de *truncated distance matrix*.

de la dérive dans l'analyse de la relation entre distances-coût et différenciation génétique. Nous avons vérifié cette hypothèse en utilisant des modèles gravitaires et montré qu'une telle approche améliorerait l'inférence de la résistance de la matrice paysagère aux flux génétiques. Ce résultat met donc à nouveau en évidence l'intérêt des graphes génétiques et paysagers en génétique du paysage, en particulier dans la perspective d'un usage intégré de ces deux objets. Cette approche permet d'identifier un scénario de coût utilisable ensuite pour construire un graphe paysager reflétant *a priori* la réalité écologique du fait de l'intégration des données génétiques au choix de ses paramètres de construction.

Il convient toutefois de rappeler les conditions dans lesquelles nous avons obtenu de tels résultats. En effet, nous avons montré que les modèles gravitaires basés sur des graphes n'ont un intérêt significatif pour la calibration d'un scénario de coût que lorsque les taux de migration sont très faibles, les tailles de populations fortement hétérogènes et lorsque la distribution spatiale des populations est irrégulière du fait de la présence d'agrégats de populations. Lors de notre étude, nous connaissions à l'avance le "vrai" scénario de coût utilisé pour les simulations et l'analyse devait nous y conduire. Nous connaissions également les autres paramètres de ces simulations, tels que les taux de migration et la localisation et la taille de l'ensemble exhaustif des populations. Cependant, dans la pratique, identifier ce scénario est l'objectif final des analyses et ces paramètres sont le plus souvent inconnus. Comment savoir dans ce cas s'il est nécessaire d'utiliser des modèles gravitaires ? Bien que nous ayons réalisé un arbre de régression pour caractériser précisément les contextes dans lesquels leur utilisation est pertinente, certaines des variables clés identifiées via cette approche sont difficilement estimables en pratique. Par conséquent, il y aurait un intérêt à compléter notre analyse fondée sur des simulations par des analyses empiriques.

Des travaux empiriques (Serrouya *et al.*, 2012 ; Weckworth *et al.*, 2013) et basés sur des simulations (Prunier *et al.*, 2017 ; Pflüger et Balkenhol, 2014) ont déjà souligné l'importance d'inclure des variables intra-populationnelles dans les analyses de différenciation génétique, notamment pour tenir compte de l'influence des caractéristiques des taches sur le flux génétique ou de l'intensité locale de la dérive. Ils suggèrent que notre approche basée sur les modèles gravitaires pourrait être utile dans un grand nombre de situations car l'influence de l'hétérogénéité spatiale de la dérive que nous avons simulée s'observe également dans la réalité (Prunier *et al.*, 2017 ; Serrouya *et al.*, 2012 ; Weckworth *et al.*, 2013). Nous n'avons toutefois pas appliqué cette approche aux données empiriques dont nous disposons concernant le Criquet ensanglanté et la Paruline caféïette. D'une part, nous n'avons pas eu suffisamment de temps pour le faire et, d'autre part, cela n'entraîne pas dans les objectifs initiaux des analyses réalisées à partir de ces jeux de données. De plus, on peut s'attendre à ce que ces modèles améliorent peu notre compréhension de l'effet de la matrice paysagère sur la différenciation génétique dans ces deux cas particuliers. En effet, les relations entre distances-coût et différenciation génétique suggéraient que les taux de migration y étaient relativement élevés en comparaison avec les faibles taux de migration des simulations pour lesquelles l'utilisation de ces modèles affectait l'inférence. Cette hypothèse mériterait d'être vérifiée et de manière générale, il serait utile d'évaluer l'intérêt de modèles gravitaires intégrant des variables calculées au niveau des nœuds de graphes génétiques et paysagers pour améliorer la compréhension du lien entre distances génétiques et paysagères. Cette démarche impliquerait de mesurer à quel point des modèles intégrant ou non ces différentes variables prédisent correctement des mesures de différenciation génétique non utilisées lors de leur calibration. Une telle validation croisée reflète celle qui a été mise en œuvre dans le chapitre 4 pour évaluer la per-

tinence écologique de différents graphes paysagers. La généralisation de cette démarche pour évaluer les modèles gravitaires bénéficierait de l'intégration des fonctions la rendant possible dans le package `graph4lg`, d'autant plus que le package R `GeNetIt`, spécifiquement dédié à ce type de modèles (Evans et Murphy, 2015), n'inclut pas de telles fonctions.

1.3 Obstacles à la compréhension des relations entre connectivité des habitats et structure génétique

Si l'utilisation conjointe de graphes génétiques et paysagers permet d'approfondir notre compréhension des relations entre connectivité des habitats et structure génétique, elle se heurte toutefois à des obstacles méthodologiques. Nous abordons ces difficultés à présent en nous focalisant tout d'abord sur la façon dont les liens entre taches d'habitat sont modélisés, puis sur la modélisation de la structure génétique des populations sous la forme d'un graphe.

1.3.1 Influence de la méthode de modélisation des liens entre taches d'habitat

La variabilité des scénarios de coût affecte l'inférence de l'effet de la matrice paysagère sur la différenciation génétique

Nous avons montré qu'il est souvent possible d'identifier un scénario de coût reflétant la réalité écologique à partir de données génétiques, que l'utilisation de modèles gravitaires soit nécessaire ou non (cf. Chapitre 6). Néanmoins, le scénario permettant d'obtenir le modèle s'ajustant le mieux aux données était souvent différent du "vrai" scénario de coût utilisé pour simuler les flux génétiques, en dépit du contraste important entre les scénarios de coût considérés. Un tel décalage s'explique par le fait que deux scénarios de coût qualitativement différents et aboutissant à des chemins de moindre coût spatialement distincts peuvent engendrer des matrices de distances-coût fortement corrélées. En effet, les résultats du chapitre 5 indiquent que la localisation spatiale des chemins de moindre coût est nettement plus sensible au scénario utilisé que la distribution statistique des distances-coût correspondantes. Ce constat est problématique dans la mesure où l'inférence précise des coûts de déplacement associés à chacun des types d'occupation du sol constituant la matrice paysagère dépend essentiellement des contrastes existant entre les matrices de distances-coût mises en relation avec les distances génétiques. Nous pouvons donc en conclure que la précision de l'inférence est souvent limitée, comme l'ont déjà suggéré Cushman *et al.* (2013) et Zeller *et al.* (2016).

Nous avons également montré dans le chapitre 5 que la distribution des valeurs de distances-coût était affectée par les valeurs absolues des coûts, tandis que la localisation spatiale des chemins de moindre coût dépendait davantage de l'ordre des types d'occupation du sol en termes de coût. Ce résultat pourrait expliquer en partie le fait que les modèles expliquant la différenciation génétique entre les populations de Paruline caféïette (cf. Chapitre 4) avaient des qualités relativement différentes en fonction du scénario de coût. En effet, les deux scénarios basés sur des dires d'expert différaient uniquement par les valeurs absolues des coûts, et non pas par leurs ordres. Or, les modèles obtenus à partir de ces scénarios étaient différents. Par ailleurs, les valeurs de coût associées aux zones agricoles et artificialisées étaient toutes deux élevées lorsqu'elles étaient issues du SDM. Les distances-coût entre les populations situées de part et d'autre de l'isthme séparant la Guadeloupe en deux parties (Basse-Terre et Grande-Terre) étaient alors nettement plus fortes dans ce cas, rendant les prédictions de différenciation génétique entre certaines de ces populations très différentes de la réalité. Il est donc

nécessaire de considérer la façon dont les scénarios de coût affectent la distribution des valeurs de distances-coût, à la fois pour l'interprétation des résultats de l'inférence de ces scénarios mais également lorsque les distances-coût sont utilisées dans le cadre d'analyses prédictives. Nous espérons que les résultats des chapitres 4, 5 et 6 et les fonctions intégrées dans le package `graph4lg` pour calculer et comparer des distances-coût obtenues selon différents scénarios y contribueront.

La modélisation basée sur les chemins de moindre coût est-elle une limite ?

Les remarques précédentes découlent peut-être du fait que l'ensemble de nos travaux s'est fondé sur des distances-coût ([Adriaensen et al., 2003](#)) plutôt que sur des distances de résistance calculées à partir de la théorie des circuits ([McRae, 2006](#)). Rappelons tout d'abord que nous avons choisi d'utiliser cette méthode pour la simplicité de sa mise en œuvre, par rapport à la théorie des circuits dont l'application conduit à des calculs beaucoup plus longs. À cet égard, les calculs de chemins de moindre coût et de distances-coût réalisés aux chapitre 5 et 6 auraient été difficilement réalisables à l'aide de la théorie des circuits⁹.

Par ailleurs, lorsque nous avons simulé des flux génétiques pour comparer les méthodes de construction des graphes génétiques (Chapitre 1), nous aurions pu calculer les probabilités de dispersion à partir de distances de résistance. Toutefois, cela n'aurait probablement pas influencé les résultats de cette comparaison. En effet, la structure génétique simulée à l'aide du logiciel CDPOP ([Landguth et Cushman, 2010](#)) dépend uniquement de la distribution statistique des distances utilisées pour paramétrer la dispersion. Nous aurions d'ailleurs pu faire ces analyses en générant aléatoirement et de façon plus systématique des distributions statistiques de distances. Le choix de créer des paysages au préalable servait, d'une part, à restreindre les paramètres de simulations à des situations observables dans la réalité et, d'autre part, à illustrer le cadre concret dans lequel les analyses mises en œuvre trouvent des applications. Notons qu'il est probable que des distributions statistiques identiques à celles des distances-coût utilisées puissent être obtenues en calculant un autre type de distance à partir d'une configuration des populations légèrement différente. Ainsi, calculer les distances paysagères en considérant que l'espèce simulée se déplace en ligne droite, en suivant un chemin de moindre coût ou bien en empruntant plusieurs chemins alternatifs n'affecterait pas les principaux résultats de nos analyses. En somme, nos simulations se sont basées sur des distances calculées en faisant l'hypothèse que l'espèce simulée était forestière et se déplaçait en suivant des chemins de moindre coût. Néanmoins, la portée de nos résultats dépasse ce contexte particulier car les distributions statistiques des distances-coût déterminant les flux génétiques dans nos simulations sont peut-être tout à fait semblables aux distances paysagères déterminant ces flux entre les populations d'espèces aux capacité de dispersion et types d'habitat différents.

Lors des analyses de données empiriques (Chapitres 3 et 4), on peut néanmoins se demander si nous aurions obtenu les mêmes résultats en utilisant des distances de résistance. Notons à ce titre que [Khimoun et al. \(2017\)](#) avaient observé que les distances génétiques séparant 20 des 27 populations de Paruline caféïette étudiées dans le chapitre 4 étaient plus fortement corrélées à des distances de

9. Les calculs de distances-coût et de distances de résistance impliquent des complexités algorithmiques différentes. Dans le chapitre 5, les distances-coût étaient calculées entre 50 points dans 77 paysages comptant chacun 9 millions de pixels et selon 100 scénarios différents. Le calcul a duré un peu moins d'une journée et on peut estimer approximativement que ce temps de calcul aurait été multiplié par au moins 50 (nombre de points) si nous avions calculé des distances de résistance.

résistance qu'à des distances-coût. Ces auteurs avaient toutefois considéré l'intégralité des paires de population, une carte d'occupation du sol différente et utilisé l'algorithme génétique d'optimisation de **ResistanceGA**. Leurs analyses et résultats ne sont donc pas exactement comparables à ceux du chapitre 4, ce qui justifierait de les reproduire avec une méthodologie comparable à la nôtre. Enfin, reproduire les analyses de sensibilité des modèles de connectivité aux scénarios de coût à partir de cartes de courant et de distances de résistance basées sur la théorie des circuits apporterait sûrement des résultats complémentaires à ceux du chapitre 5, applicables aux nombreuses analyses de génétique du paysage basées sur ce type de modèle. Le développement récent d'une version du logiciel Circuitscape dans un nouveau langage informatique plus performant (Julia, [Bezanson *et al.* \(2017\)](#)) permettrait de le faire tout en conservant des temps de calcul raisonnables ([Banville *et al.*, 2021](#)).

1.3.2 Questions relatives à la modélisation de la structure génétique sous la forme d'un graphe

Dans l'ensemble de nos travaux, nous n'avons considéré que des situations dans lesquelles le phénomène d'isolement par la résistance du paysage expliquait la structure génétique des populations, que cette relation s'observe à l'échelle de l'ensemble de la zone d'étude ou non (cf. Chapitres 1, 3 et 4) et que la dérive ait une intensité homogène ou non (cf. Chapitre 6). Or, le patron d'isolement par la résistance n'est pas systématiquement celui qui reflète au mieux la réalité et il est susceptible d'évoluer au cours du temps. De plus, certaines des méthodes utilisées pour caractériser la structure génétique des populations et sa dynamique spatio-temporelle soulèvent des questions d'ordre méthodologique. Toutefois, nous pensons que l'utilisation des graphes n'est pas restreinte aux situations étudiées dans cette thèse et que les questions méthodologiques qu'elle pose peuvent obtenir des réponses. Nous discutons donc à présent des points suivants :

- l'apport potentiel de la théorie des graphes pour comprendre les déterminants des patrons de différenciation génétique entre populations,
- les méthodes permettant d'analyser les phénomènes d'isolement par des barrières paysagères à l'aide de graphes,
- les limites liées à l'utilisation du principe d'indépendance conditionnelle,
- la signification des distances et indices génétiques utilisés.

Quelles propriétés des réseaux d'habitat et de populations influencent le patron de différenciation génétique ?

Dans le chapitre 1, nous avons simulé plusieurs paysages et échantillonné aléatoirement des populations dans les taches de forêt de ces paysages. Nous avons ensuite simulé la structure génétique résultant de flux génétiques dont l'intensité dépendait uniquement des distances-coût entre ces populations. Nous avons alors obtenu des patrons de différenciation génétique fortement contrastés. Ce contraste reflétait la typologie des patrons d'IBD élaborée par [Hutchison et Templeton \(1999\)](#), ce qui nous a permis d'identifier les méthodes de construction et d'analyse des graphes génétiques à préconiser dans différents contextes selon les objectifs poursuivis. Compte tenu des résultats fournis par nos analyses, une question majeure apparaît néanmoins : pourquoi avons-nous obtenu des patrons de différenciation génétique si contrastés avec des paramètres de simulation en tous points identiques ?

Les résultats du chapitre 6, dans lequel nous avons réalisé des simulations similaires à celles du premier chapitre, apportent quelques éléments de réponse à cette question. En effet, nous avons mis en évidence le fait que le degré d'agrégation spatiale des populations influençait le patron de différenciation génétique, indépendamment du taux de migration et de la résistance de la matrice paysagère. Aussi, identifier de façon plus précise les caractéristiques du réseau de populations responsables des différents patrons de différenciation génétique permettrait de mieux comprendre la manière dont les propriétés spatiales et topologiques de ce réseau façonnent la structure génétique en influençant l'interaction entre la structure du paysage et les processus écologiques (Van Strien, 2017). Une telle analyse compléterait de précédents travaux menés sur les principaux patrons de structure génétique spatiale (Bradburd *et al.*, 2016, 2018 ; Greenbaum et Fefferman, 2017 ; Milligan *et al.*, 2018). Il serait par exemple possible de réaliser des simulations similaires à celles des chapitres 1 et 6 et d'identifier une typologie des patrons de différenciation génétique obtenus, pour comprendre, à partir de métriques décrivant le réseau des populations, les conditions à l'origine de ces patrons de différenciation génétique. Ces métriques pourraient s'inspirer de celles calculées à l'échelle d'un graphe paysager, telles que la Probabilité de Connectivité (PC), le nombre de composantes d'un graphe élagué ou encore le seuil de percolation. Elles pourraient également s'inspirer (i) des variables utilisées pour interpréter les différences de performance des modèles gravitaires pour inférer la résistance du paysage selon les paramètres de simulations (cf. Chapitre 6), telles que la moyenne harmonique des distances-coût séparant les populations, ou (ii) de métriques utilisées en statistique spatiale telle que l'indice de Ripley (1977), entre autres (Dale et Fortin, 2014).

Comment comparer des graphes génétiques et paysagers en présence d'un patron d'isolement par des barrières (IBB) ?

Il est fréquent que des populations forment des *clusters* relativement distincts les uns des autres d'un point de vue génétique, soit en raison de l'histoire phylogéographique de ces populations (Garnier *et al.*, 2004), soit parce que des barrières paysagères empêchent les flux génétiques entre ces *clusters* (Gauffre *et al.*, 2015). Dans ce second cas, on parle d'un phénomène d'Isolement par les Barrières (IBB) (Balkenhol *et al.*, 2016). Le plus souvent, ces *clusters* sont détectés à l'aide de l'algorithme basé sur une approche bayésienne intégré au logiciel STRUCTURE (Pritchard *et al.*, 2000). L'algorithme de Monmonier (1973) (Manni *et al.*, 2004) ou l'algorithme bilinéaire de Womble (1951) ont aussi été utilisés pour localiser les barrières paysagères aux flux génétiques responsables du patron d'IBB, entre autres exemples (Manel *et al.*, 2007 ; Safner *et al.*, 2011). En génétique du paysage, la dernière étape consiste alors à établir un lien entre la localisation de ces barrières et les éléments paysagers.

Nous avons mentionné en introduction la possibilité d'utiliser des graphes génétiques et paysagers pour identifier des barrières paysagères. En effet, la comparaison *a posteriori* de modules identifiés au niveau de graphes génétiques et paysagers permettrait d'étudier à quel point la configuration spatiale des taches d'habitat et la façon dont les contraintes à la dispersion sont modélisées expliquent l'existence de *clusters* génétiques. Nous n'avons pas réalisé cette analyse mais nous avons intégré dans le package `graph4lg` des fonctions qui la rendent possible (cf. Chapitre 2).

À l'issue de cette thèse, nous pouvons imaginer une seconde analyse des barrières paysagères basée sur les modules de graphes génétiques et paysagers (Figure 69). En effet, des graphes paysagers pourraient être construits à partir de différents scénarios de coût de façon itérative jusqu'à identifier

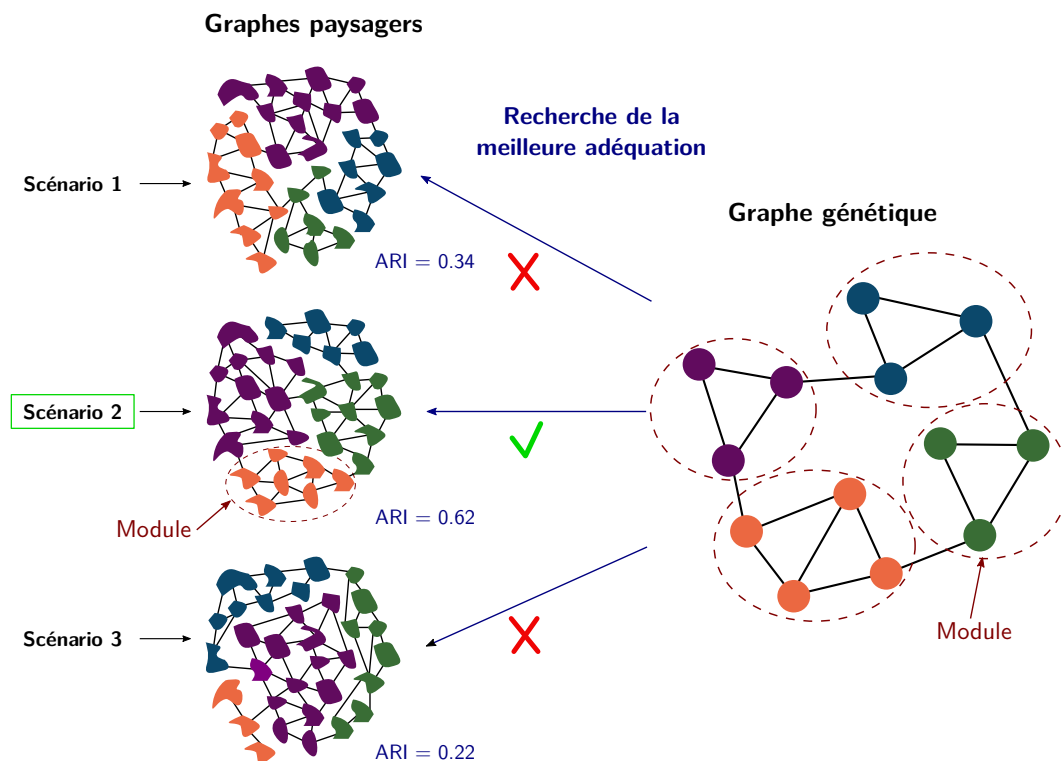


FIGURE 69 – Optimisation d’un scénario de coût basée sur la comparaison des modules de graphes génétiques et paysagers. Des graphes paysagers sont construits à partir de plusieurs scénarios de coût. Des modules de tâches d’habitat sont identifiés à l’aide d’un algorithme de modularité dans chaque cas. En parallèle, des modules de populations sont identifiés dans le graphe génétique basé sur des populations échantillonnées au sein des tâches du graphe paysager. Les deux partitions sont comparées à l’aide d’un indice tel que l’*Adjusted Rand Index* (ARI). La valeur maximale de l’ARI indique la meilleure adéquation entre les modules, et par conséquent le scénario de coût le plus réaliste.

le scénario conduisant au graphe paysager dont les modules reproduisent au mieux la partition en modules du graphe génétique. Le choix du scénario optimal se ferait en considérant un indice d’adéquation entre les partitions, tel que l’*Adjusted Rand Index* (Hubert et Arabie, 1985) intégré à `graph4lg` ou l’indice de *Normalized Mutual Information* (Danon et al., 2005). Même si l’utilisation d’un algorithme génétique permettrait de parcourir efficacement le champ des possibles pour identifier le scénario optimal, il serait dans ce cas prudent de considérer l’ensemble des scénarios de coût permettant d’obtenir une bonne adéquation des partitions compte tenu de la forte corrélation souvent mesurée entre plusieurs matrices de distances coût. L’interprétation des résultats se baserait alors essentiellement sur les points communs des résultats obtenus avec ces différents scénarios. Une telle approche se prêterait mieux aux situations caractérisées par une structuration génétique en *clusters* bien distincts que les méthodes basées sur le principe d’isolement par la résistance du paysage (cf. Chapitres 1 et 6). Elle permettrait en outre de construire un graphe paysager valide *a priori* à partir du scénario de coût optimal. Il conviendrait de s’assurer au préalable que ces *clusters* de populations n’ont pas pour origine un contact secondaire, lié par exemple à la colonisation post-glaciaire d’une même région par des populations ayant évolué génétiquement de façon indépendante avant cette colonisation (voir par exemple Garnier et al. (2004)). À ce sujet, Taberlet et al. (1998) ont identifié les régions européennes dans lesquelles de tels phénomènes ont été les plus fréquents.

Comment utiliser le principe d’indépendance conditionnelle ?

Le chapitre 1 visait entre autres à évaluer l’intérêt du principe d’indépendance conditionnelle pour l’identification des chemins de dispersion directs entre populations. C’est sur ce principe que reposent

les *population graphs* développés par [Dyer et Nason \(2004\)](#). Appliquée à des données génétiques, cette méthode est censée tester la corrélation entre les fréquences alléliques de deux populations A et B après avoir pris en compte la relation entre les fréquences alléliques de chacune et les fréquences alléliques des autres populations. Cette corrélation partielle reflète la dépendance conditionnelle entre les fréquences alléliques des populations A et B (cf. Annexe B1). En considérant qu'elle traduit l'échange direct de migrants entre les populations A et B, cette méthode permet ainsi de construire des graphes génétiques reflétant directement le réseau de dispersion. Elle présente à ce titre un intérêt évident en génétique du paysage et plus largement en biologie de la conservation.

Cependant, nous avons constaté que la formule utilisée dans le package `popgraph` pour appliquer le principe d'indépendance conditionnelle était différente de celle utilisée ou démontrée mathématiquement dans des travaux précédents ([Everitt et Hothorn, 2011](#) ; [Gower et Legendre, 1986](#) ; [Magwene, 2001](#) ; [Smouse et Peakall, 1999](#)) (Annexe B1). Nous avons alors comparé les résultats obtenus avec la formule issue de ces travaux à ceux obtenus avec le package `popgraph` à l'issue des simulations réalisées dans le premier chapitre. Les résultats obtenus ne permettaient pas d'identifier de façon nette la meilleure méthode. Même si l'utilisation de la formule démontrable mathématiquement permettait de maximiser la corrélation entre les distances génétiques associées aux liens des graphes et les distances paysagères, l'utilisation du package `popgraph` permettait généralement de reproduire plus fidèlement la topologie des liens suivis par les individus simulés. Néanmoins, le nombre de liens des graphes obtenus à l'aide de `popgraph` était étonnamment stable (entre 134 et 154 liens) quand le nombre de liens réellement suivis par les individus lors des simulations passait de 155 à 393 liens. Il serait donc utile d'identifier les conditions dans lesquelles cette méthode est pertinente, et d'explicitier les raisons de sa plus-value. Des réponses à ces questions sont peut-être à chercher dans les travaux ayant mobilisé le principe d'indépendance conditionnelle en génétique du paysage ([Peterson *et al.*, 2019](#) ; [Schaeffer *et al.*, in prep](#)) ou dans des approches similaires ([Kuismin et Sillanpää, 2017](#) ; [Kuismin *et al.*, 2020](#) ; [Whittaker, 2009](#)). La recherche de ces éléments de réponse est un enjeu d'autant plus important que ces méthodes sont fréquemment utilisées en génétique du paysage ([Dyer, 2015b](#)).

Les distances génétiques calculées reflètent-elles vraiment l'effet du paysage sur les flux génétiques ?

À travers les chapitres de cette thèse, nous avons caractérisé la structure génétique des populations à l'aide d'indices calculés au niveau des nœuds de graphes génétiques et avec différentes distances génétiques. Nous avons interprété les résultats obtenus en considérant que les réponses génétiques décrites par ces mesures sont influencées par le paysage à différentes échelles spatiales et temporelles, comme l'ont montré de précédents travaux.

Tout d'abord, nous avons remarqué dans le chapitre 1 que les valeurs de D_{PS} mesurant la différenciation génétique entre populations étaient plus fortement liées aux distances-coût déterminant les flux génétiques que les valeurs de F_{ST} . Cette différence avait tendance à s'estomper à mesure que le nombre de générations simulées augmentait. Un tel résultat peut s'expliquer par le fait que l'utilisation du F_{ST} est en principe réservée aux situations dans lesquelles un équilibre a été atteint entre la dérive et le flux génétique ([Neigel, 2002](#) ; [Whitlock et McCauley, 1999](#)). Des auteurs ont également suggéré qu'en l'absence d'équilibre, l'utilisation du D_{PS} était préférable pour les analyses de génétique du paysage ([Murphy *et al.*, 2016](#) ; [Storfer *et al.*, 2010](#)). Enfin, certaines études suggèrent que le D_{PS} refléterait davantage l'effet du paysage sur les flux génétiques contemporains, tandis que le F_{ST} inté-

grerait l'influence du paysage sur une période plus longue (Latta, 2006 ; Murphy *et al.*, 2010b). À ce titre, les résultats du chapitre 1 sont en accord avec ces travaux. Par ailleurs, dans l'étude dédiée à l'inférence de la résistance du paysage aux flux génétiques (Chapitre 6), nous avons également utilisé le D_{PS} car les conditions optimales d'utilisation du F_{ST} n'étaient probablement pas vérifiées à l'issue de toutes les simulation.

Dans le chapitre concernant le Criquet ensanglanté (Chapitre 3), nous avons utilisé à la fois le D_{PS} et le F_{ST} pour estimer la différenciation génétique entre les populations de Criquet ensanglanté. Ces deux distances génétiques ont été utilisées pour calculer l'indice de différenciation génétique relative MIW à partir d'un graphe génétique. Les indices dérivés du D_{PS} et du F_{ST} étaient corrélés de façon différente aux métriques de connectivité selon le scénario de coût utilisé pour calculer ces dernières. Nous avons interprété cela en considérant que le scénario de coût expliquant au mieux le F_{ST} reflétait la résistance du paysage ancien aux flux génétiques. À l'inverse, le lien fort entre les métriques de connectivité basées sur des distances géodésiques euclidiennes et les indices calculés à partir du D_{PS} reflétait la moindre résistance du paysage contemporain aux flux génétiques. Cette interprétation, basée sur le fait que le D_{PS} est censé refléter davantage les flux génétiques contemporains que le F_{ST} , pourrait également expliquer l'expansion actuelle du Criquet ensanglanté (Trautner et Hermann, 2008).

Par ailleurs, nous avons montré que l'élagage des graphes génétiques permettait de tenir compte de l'échelle à laquelle l'équilibre entre dérive et flux génétique était atteint pour inférer l'effet du paysage sur la différenciation génétique (cf. Chapitre 1). Par conséquent, lors de l'analyse de la structure génétique des populations de Criquet ensanglanté (Chapitre 3), la plus forte corrélation entre les métriques de connectivité et les indices MIW issus de graphes génétiques élagués suggérait une absence d'équilibre entre dérive et flux génétique à l'échelle de l'intégralité de la zone d'étude (Keller *et al.*, 2013).

Enfin, pour décrire la structure génétique des populations de Paruline caféïette (Chapitre 4), nous avons calculé le D_{PS} et le F_{ST} ainsi que des indices MIW à partir de graphes complets et élagués. Néanmoins, nous nous sommes ensuite focalisés sur le F_{ST} et sur le graphe génétique complet car c'est ainsi que les relations mesurées entre les variables génétiques et les métriques de connectivité issues des graphes paysagers étaient les plus fortes. Est-ce que ce résultat signifie que l'équilibre entre dérive et flux génétique a été atteint pour cet ensemble de populations de Paruline caféïette ? Il est difficile de l'affirmer sans investigations supplémentaires¹⁰, d'autant plus que l'élagage du graphe génétique était nécessaire pour prédire convenablement la différenciation génétique à partir de distances-coût. Dans ce cas d'étude, l'intérêt de l'élagage pourrait s'expliquer par le maintien de génotypes assez similaires dans des populations qui appartenaient historiquement à une seule grande population mais qui se trouvent aujourd'hui séparées par de grandes distances. Si ces populations ont conservé des effectifs importants au cours du temps, on peut supposer que la dérive les a peu affectées et que leur différenciation génétique est donc restée faible. Dans ce cas, considérer de telles paires de populations ne permet pas de caractériser l'influence de la distance paysagère sur la différenciation génétique et il

10. L'algorithme bayésien 2MOD développé par Ciofi *et al.* (1999) permettrait de tester si le patron de différenciation génétique s'explique par un effet prédominant de la dérive uniquement ou par un équilibre entre dérive et flux génétique. Néanmoins, cette méthode n'a pas permis à Hänfling et Weetman (2006) de mettre en évidence le patron d'IBD de type IV pourtant apparent sur le nuage de points représentant la relation d'IBD. D'autres types de méthodes basées notamment sur le principe de coalescence pourraient éventuellement apporter des réponses à cette question, en testant la vraisemblance de différents scénarios évolutifs pour ces populations (Cornuet *et al.*, 2008).

vaut mieux élaguer le graphe génétique, comme nous l'avons constaté dans nos résultats. Néanmoins, nous ne disposons pas d'éléments permettant de valider cette hypothèse.

Plus généralement, au regard de l'ensemble de ces résultats, il serait nécessaire d'approfondir les points suivants :

- Bien qu'un faisceau de résultats indique que le D_{PS} reflète davantage les flux génétiques contemporains que le F_{ST} , une telle affirmation mériterait d'être démontrée de façon théorique et testée à l'aide d'un protocole dédié. En effet, on peut se demander pourquoi deux distances calculées à l'aide des mêmes données génétiques mais avec deux formules différentes refléteraient une temporalité différente dans l'influence du flux génétique sur la différenciation génétique.
- Au-delà de l'absence d'équilibre entre dérive et flux génétique, l'élagage des graphes génétiques peut être nécessaire pour d'autres raisons pour améliorer l'acquisition de connaissances en génétique du paysage. Ces causes alternatives pourraient être liées à des phénomènes précédemment observés en génétique des populations. Par exemple, le maintien d'un polymorphisme ancestral dans des populations éloignées (Marko et Hart, 2011) pourrait limiter notre capacité à évaluer la résistance du paysage au flux génétique à partir de mesures de différenciation génétique entre ces populations. Ainsi, élaguer un graphe génétique afin d'exclure des paires de populations ayant divergé à une époque où le paysage était très différent du paysage contemporain pourrait améliorer la qualité des analyses de connectivité menées à partir de données génétiques.
- L'origine du patron d'IBD de type IV caractérisé par Hutchison et Templeton (1999), et précédemment décrit par Slatkin (1993), est l'expansion des populations d'une espèce dans une nouvelle zone géographique provoquant une absence temporaire d'équilibre entre dérive et flux génétique. Cette situation pourrait correspondre au patron observé au niveau des populations de Criquet ensanglanté et lors des simulations de flux génétiques. Néanmoins, on peut se demander si des modifications de l'occupation du sol peuvent être à l'origine d'un patron d'IBD ou d'IBR similaire en modifiant temporairement les conditions de l'équilibre entre la dérive et les flux génétiques. Investiguer cette question aiderait à déterminer les contextes dans lesquels les méthodes d'élagage des graphes génétiques testées dans notre travail seraient utiles.

Après avoir montré en quoi l'utilisation conjointe de graphes génétiques et paysagers contribue à notre compréhension de la relation entre connectivité et structure génétique, nous allons présenter en quoi elle valide et complète les méthodes utilisées à des fins de conservation de la connectivité écologique des habitats.

2 Intérêt des graphes génétiques et paysagers pour la conservation de la connectivité des habitats

Construire un graphe consiste à modéliser un réseau sous la forme d'un ensemble de nœuds et de liens. Dans le cas des graphes génétiques et paysagers, il s'agit de réseaux de populations et d'habitats qui sont représentés de façon spatialement explicite et visualisables sur une carte. Comme toute tentative de modélisation, ces représentations simplifiées sont nécessairement imparfaites du fait de la complexité et du caractère parfois stochastique des processus écologiques étudiés ainsi que de l'imbrication des échelles spatiales et temporelles auxquelles ils se produisent. Lorsque des graphes sont utilisés dans le cadre de travaux d'écologie scientifique, cette simplification est nécessaire pour tester des hypothèses précises (Vimal *et al.*, 2012), concernant l'influence de la structure spatiale des habitats sur la structure génétique dans nos travaux par exemple. En revanche, lorsqu'ils sont utilisés dans un cadre opérationnel, ils sont la représentation simplifiée de la réalité sur laquelle les acteurs se basent pour orienter l'action (Vimal *et al.*, 2012). On peut alors se demander si la modélisation de graphes paysagers convient aux deux usages.

Nous avons vu dans la section précédente qu'utiliser conjointement des graphes génétiques et paysagers permettait d'améliorer notre compréhension de la relation entre la connectivité des habitats et la structure génétique des populations. À ce titre, nous avons montré que la modélisation de réseaux sous la forme de ces deux types de graphe permet de tester des hypothèses sur des processus écologiques et contribue à l'acquisition de connaissances en écologie scientifique. Parallèlement, notre travail avait pour objectif d'évaluer l'intérêt de ces graphes pour orienter la conservation de la connectivité, contribuant ainsi aux réflexions menées en biologie de la conservation pour accompagner la mise en œuvre des politiques publiques (cf. Introduction, Figure 7). Nous revenons donc à présent sur ce second objectif.

Les graphes paysagers ont reçu une attention particulière suite à la naissance du concept de réseau écologique (Mougenot et Meliin, 2000 ; Opdam *et al.*, 2006), concomitante de l'intégration des enjeux liés à la nature ordinaire aux politiques publiques (Godet, 2010 ; Vimal *et al.*, 2012). Ils ont notamment été utilisés dans le cadre de la politique de Trame Verte et Bleue en France (Foltête *et al.*, 2012a). Néanmoins, le concept de réseau écologique et l'importance accordée à sa modélisation dans les programmes de conservation de la biodiversité ont été critiqués (Boitani *et al.*, 2007 ; Gippoliti et Battisti, 2017) (mais voir Foltête (2019)). Parmi ces critiques, on retrouve le fait que les informations nécessaires à la modélisation sont rarement disponibles et que ces modèles manquent donc de validité (Boitani *et al.*, 2007). Nous avons d'ailleurs fait le constat dans la section 2.3 de l'introduction que les validations de ces modèles avec des données indépendantes faisaient défaut dans la littérature scientifique (Foltête *et al.*, 2020). De plus, pour certains auteurs, le focus sur la connectivité des réseaux écologiques détourne l'attention d'autres paramètres d'importance majeure pour la conservation tels que la surface des habitats ou la qualité écologique de ces zones (Gippoliti et Battisti, 2017 ; Moilanen, 2011). **Dans ce contexte et au vu des résultats obtenus en utilisant conjointement des graphes génétiques et paysagers, nous tâcherons de répondre aux questions suivantes : (i) les graphes paysagers peuvent-ils orienter correctement la conservation de la connectivité ? et (ii) comment orienter cette conservation à l'aide de graphes gé-**

nétiques et paysagers ? Nous aborderons ensuite les obstacles à l'utilisation des graphes dans ce cadre.

2.1 Les graphes paysagers peuvent-ils orienter correctement la conservation de la connectivité ?

2.1.1 Validation écologique des graphes paysagers à l'aide de graphes génétiques

Les graphes paysagers sont censés représenter la connectivité fonctionnelle d'un réseau d'habitat mais sont habituellement construits à partir de dires d'expert, ou de données biologiques ne reflétant pas directement l'influence de la connectivité des habitats sur le mouvement des espèces et les flux génétiques. Ainsi, le chapitre 4 était spécialement dédié à la validation de méthodes de construction des graphes paysagers couramment utilisées dans une optique de conservation de la biodiversité. Cette étude a montré que les graphes construits à dires d'expert ou à partir d'un SDM basé sur des données de présence/absence reflétaient convenablement la structure génétique de populations de Paruline caféïette échantillonnées au niveau de leurs taches d'habitat. Parmi les rares études basées sur des graphes paysagers et sur deux jeux de données biologiques indépendants (Foltête *et al.*, 2020), il s'agit à notre connaissance de la première visant spécifiquement à évaluer la pertinence de ces outils. Les relations similaires entre les indices génétiques et les métriques de connectivité observées lors de l'analyse de la structure génétique de populations de Criquet ensanglanté (Chapitre 3) complètent cette validation.

L'ensemble de ces résultats suggère que la prise en compte de plusieurs métriques complémentaires et basées sur le concept de quantité d'habitat atteignable, même issues de graphes paysagers fondés sur des méthodes simples, peut permettre de prioriser correctement les taches d'habitat en vue de la conservation de la diversité génétique des populations d'une espèce. Ils confirment donc l'intérêt des graphes paysagers pour des analyses opérationnelles ayant pour objectifs l'identification de taches d'habitat à conserver en priorité ou à restaurer, ou l'évaluation de l'impact d'un projet affectant l'occupation du sol (Foltête *et al.*, 2014 ; Tarabon *et al.*, 2019 ; Zetterberg *et al.*, 2010).

2.1.2 Rapport coûts-bénéfices associé à la modélisation par les graphes paysagers

Lors de la comparaison de la validité de graphes paysagers construits à partir d'informations de sources différentes et selon plusieurs méthodes (Chapitre 4), nous avons obtenu un résultat contre-intuitif. En effet, le graphe paysager construit avec l'approche la plus exigeante en données biologiques et nécessitant la réalisation d'un modèle de distribution de l'espèce étudiée n'était pas toujours celui qui reflétait le mieux la structure génétique des populations. De plus, lorsque l'utilisation de ce graphe a fourni les meilleurs résultats, d'autres options plus simples ont également apporté des résultats satisfaisants. Dans la mesure où l'utilisation de graphes paysagers est principalement adaptée aux espèces spécialistes occupant des taches d'habitat subdivisées (Urban *et Keitt*, 2001), on peut s'attendre à ce que les taches d'habitat de telles espèces soient délimitables convenablement dans un grand nombre de situations. Étant donné que l'utilisation du modèle de distribution d'une espèce pour construire un graphe se justifie davantage pour identifier ses taches d'habitat que pour estimer ses coûts de déplacement (DufLOT *et al.*, 2018 ; Godet *et Clauzel*, 2021), cela limite également l'intérêt d'une telle approche pour la construction d'un graphe paysager. Ces éléments suggèrent donc que le bénéfice lié à l'utilisation d'une méthode complexe de construction d'un graphe paysager est parfois négligeable

au regard des investissements nécessaires à sa mise en œuvre. Une telle approche devrait donc être réservée aux situations dans lesquelles une des propositions suivantes se vérifie :

- l'espèce étudiée est spécialiste et forme des populations discrètes sans qu'on puisse délimiter *a priori* ses taches d'habitat,
- il n'existe aucun expert capable d'estimer les coûts de déplacement de l'espèce étudiée à travers différents types d'occupation du sol, et aucun résultat dans la littérature relatif à ces coûts.

La première situation se rencontre notamment lorsqu'il est nécessaire de combiner plusieurs types de données spatiales (topographiques, climatiques et d'occupation du sol par exemple) pour identifier l'habitat d'une espèce tout en se basant sur des résultats de modélisation pour effectuer cette combinaison. Néanmoins, lorsque les données spatiales disponibles ne renseignent pas sur les conditions environnementales définissant l'habitat de l'espèce étudiée, cette méthode ne permet pas de combler l'absence de connaissances sur la distribution de l'habitat. En outre, dans la situation où il n'est pas possible de se baser sur des dires d'expert pour calibrer les coûts de déplacement de l'espèce, nous avons montré que si des données de présence/absence sont disponibles, le calcul d'un indice de spécialisation tel que l'indice de [Jacobs \(1974\)](#) peut être suffisant pour construire un graphe paysager relativement valide (cf. Chapitre 4). Si un modèle de distribution d'espèce est toutefois réalisé, il est préférable qu'à la fois des données de présence et d'absence soient disponibles, et que ces données aient été acquises dans le cadre d'un protocole minimisant les biais d'échantillonnage. Dans tous les cas, une attention particulière doit être accordée à la distribution des points de présence et/ou d'absence car elle peut influencer de façon importante le modèle réalisé ([Fletcher et Fortin, 2018](#)) (Annexe B3).

Au vu de ces résultats, on peut s'interroger plus largement sur la place accordée à la modélisation dans les études à visée opérationnelle en fonction des données disponibles et/ou du temps et de l'argent investis pour les acquérir. Dans le cadre d'une réflexion menée sur la qualité et la pertinence des diagnostics écologiques et études d'impact préalables aux projets d'aménagement, [Moulherat \(2019\)](#) a suggéré de repenser l'allocation des ressources engagées pour réaliser ces études. Actuellement, une part importante des ressources est dédiée à des inventaires faunistiques visant à détecter la présence d'espèces protégées et une part moindre aux inventaires floristiques et d'habitat. Réduire la part des inventaires faunistiques au profit de prélèvements génétiques et de relevés visant à mieux cartographier l'occupation du sol permettrait d'intégrer une étape de modélisation plus importante à ces études, tout en fournissant les paramètres nécessaires à la réalisation de modèles pertinents. La validité des modèles que constituent les graphes paysagers mise en évidence ici, l'apport limité des données de présence/absence et enfin l'intérêt des données génétiques pour l'optimisation de surfaces de résistance (cf. Chapitre 6) confortent ces préconisations visant à optimiser le rapport coûts-bénéfices associé à la modélisation d'un réseau écologique.

2.2 Comment orienter la conservation à l'aide de graphes génétiques et paysagers ?

Nous avons vu que les graphes paysagers peuvent refléter correctement la façon dont la connectivité des habitats affecte la structure génétique des populations et, à ce titre, être utilisés pour orienter la conservation. Néanmoins, on peut se demander quelle est l'importance à accorder aux résultats de la modélisation de la connectivité dans la mise en œuvre de mesures de conservation. Nous verrons notamment que l'utilisation du concept de quantité d'habitat atteignable peut faciliter cette mise en

œuvre en évitant de se focaliser exclusivement sur les mouvements entre taches aux dépends d'autres paramètres importants. Par ailleurs, nous verrons comment l'utilisation de graphes génétiques pourrait compléter celle des graphes paysagers pour conserver la variabilité génétique.

2.2.1 Utilisation des graphes paysagers pour conserver la quantité d'habitat atteignable

Intérêt opérationnel des métriques de quantité d'habitat atteignable

Pour certains auteurs, concentrer les efforts de conservation sur le maintien de la connectivité est le plus souvent inefficace si cela se fait aux dépends de la conservation de surfaces d'habitat importantes et ayant une qualité écologique suffisante pour les espèces ou de la restauration de la matrice paysagère (Gippoliti et Battisti, 2017 ; Mortelliti *et al.*, 2010 ; Taylor *et al.*, 2006). Le maintien de la connectivité entre des taches d'habitat suite à un projet d'aménagement réduisant la surface de celles-ci ou modifiant fortement la matrice paysagère pourrait même laisser penser que l'impact de cet aménagement est négligeable. Lorsque la notion de connectivité des habitats se restreint aux mouvements *inter*-taches d'habitat sans intégrer la surface de ces taches et les mouvements *intra*-taches, toutes ces critiques sont parfaitement justifiées. En revanche, si on assimile la connectivité à une quantité d'habitat atteignable au sein et en dehors des taches, tout en tenant compte de la résistance de la matrice, de telles critiques sont moins justifiées, comme l'avait d'ailleurs fait remarquer Foltête (2019) en réponse aux arguments formulés par Gippoliti et Battisti (2017). De plus, selon Saura et Rubio (2010), cette conception de la connectivité dispense les gestionnaires d'attribuer, souvent arbitrairement, un poids relatif à la quantité d'habitat et aux mouvements entre les taches dans leurs critères de décision¹¹. Ainsi, l'intérêt de cette conception intégrée de la connectivité, mis en évidence dans la section précédente pour expliquer la structure génétique des populations, se retrouve également lorsque les métriques mesurant la quantité d'habitat atteignable sont utilisées pour orienter la conservation.

Notons que les métriques de quantité d'habitat atteignable issues d'un graphe paysager peuvent prendre en compte la qualité des taches d'habitat. Il suffit pour cela d'attribuer aux taches une capacité qui ne dépend pas uniquement de leur surface mais aussi de variables telles que la disponibilité ou la densité de ressources alimentaires et de micro-habitats, ou plus généralement un score reflétant la qualité de la tache. Dans ce cas, on peut considérer qu'on évalue la quantité effective d'habitat atteignable, en pénalisant les surfaces qui permettent difficilement à une population de s'établir dès lors qu'on ne considère pas uniquement l'occupation du sol pour les caractériser. Le fait qu'une telle information soit rarement intégrée à la modélisation ne provient pas d'un choix délibéré des modélisateurs de négliger cette information, mais plutôt du fait que les données permettant de l'intégrer sont rarement disponibles. Dans les quelques exemples existant, la capacité d'une tache est calculée comme la probabilité de présence moyenne de l'espèce étudiée dans chaque tache d'habitat d'après un modèle de distribution de cette espèce (Dufлот *et al.*, 2018 ; Hofman *et al.*, 2018 ; Pereira *et al.*, 2011).

Par ailleurs, nous avons vu dans les chapitres 3 et 4 comment intégrer la résistance de la matrice paysagère dans le calcul des métriques issues des graphes paysagers. Ainsi, l'impact d'un projet d'aménagement affectant la résistance de la matrice (une infrastructure de transport par exemple) peut être évalué en construisant un graphe paysager représentant le réseau d'habitat suite à ces modifications

11. Le titre de leur article est d'ailleurs tout à fait explicite : "A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape".

et en le comparant à la situation initiale. De même, les graphes paysagers peuvent être utilisés pour évaluer l'intérêt d'une mesure de restauration visant à créer une tache d'habitat (Foltête *et al.*, 2014) ou à diminuer l'impact négatif d'un élément de la matrice (Girardet *et al.*, 2013).

Les exemples précédents illustrent le fait que les mesures de quantité d'habitat atteignable que nous avons validées empiriquement peuvent intégrer la qualité des surfaces d'habitat et l'hétérogénéité de la matrice paysagère. Ils rappellent également le fait que les graphes paysagers sont utiles pour identifier les taches à conserver à un instant t , mais aussi pour évaluer l'intérêt ou l'impact de différents scénarios d'aménagement ou de restauration (Foltête *et al.*, 2014). Les craintes de Boitani *et al.* (2007) et Gippoliti et Battisti (2017) selon lesquelles modéliser un réseau écologique serait antagoniste avec la prise en compte de la dynamique de ce réseau et avec l'évaluation de différents scénarios d'aménagement semblent donc infondées dès lors que la modélisation se fonde sur des graphes paysagers (Foltête, 2019). Dans ce cadre, l'utilisation de métriques issues de graphes permet de comparer les scénarios en fonction de leur impact sur la quantité totale d'habitat atteignable. Cette mesure est parfaitement adaptée à l'application de la séquence Éviter-Réduire-Compenser en France par exemple, comme l'ont montré Bergès *et al.* (2020) de façon schématique. En simulant le développement de la métropole de Toulouse, Tarabon *et al.* (2020a) ont ensuite montré l'intérêt d'une telle approche pour orienter le développement territorial en réduisant l'impact sur la quantité d'habitat atteignable.

Distinguer les objectifs de compréhension et de conservation pour une utilisation adaptée des graphes paysagers

Comme indiqué précédemment, la modélisation spatiale a un statut différent selon qu'elle est utilisée pour (i) tester des hypothèses dans un objectif de compréhension ou (ii) pour représenter la réalité de façon simplifiée pour orienter l'action (Vimal *et al.*, 2012). Dans cette thèse, nous avons construit des graphes pour comprendre l'influence de la connectivité sur la structure génétique des populations (Chapitres 3 et 4) ou pour évaluer la validité des méthodes sur lesquelles repose la construction de graphes génétiques et paysagers (Chapitres 1, 5 et 6), que ceux-ci soient utilisés dans un cadre scientifique ou opérationnel. En revanche, nous n'avons pas construit de graphes pour répondre à des enjeux de conservation concrets. Nous précisons donc à présent en quoi l'utilisation de graphes paysagers dans ce cadre pourrait différer de celle que nous en avons faite lorsqu'un objectif de compréhension était poursuivi.

Rappelons tout d'abord qu'une modélisation spatiale peut apporter deux types de résultats, à savoir (i) des variables quantitatives relatives aux entités spatiales modélisées et calculées lors de la modélisation (par exemple des métriques de connectivité) et (ii) des entités spatiales nouvellement créées lors de la modélisation (par exemple des chemins de moindre coût)¹². Ces différents résultats peuvent être utilisés à des fins explicatives, prédictives ou en tant que support d'une prise de décision. Ces objectifs sont liés dans la mesure où une modélisation ayant un faible pouvoir explicatif a généralement un faible pouvoir prédictif. De plus, lorsque l'utilisateur se base sur un graphe paysager pour prendre une mesure de conservation, il part implicitement du principe que ce modèle a un bon pouvoir prédictif des réponses biologiques qu'il vise à favoriser dans ce cadre (richesse spécifique ou variabilité génétique par exemple). Pour cette raison, en mettant en évidence le pouvoir explicatif des

12. Des variables quantitatives caractérisant des entités spatiales nouvellement créées sont souvent calculées (les distances-coût associées aux chemins de moindre coût par exemple).

métriques de connectivité issues des graphes paysagers, nous validons indirectement la pertinence de leur utilisation pour orienter la conservation.

Néanmoins, les métriques à utiliser pour orienter la conservation ne sont pas nécessairement les mêmes que celles à intégrer dans des modèles explicatifs. Nous avons fait le choix d'utiliser les métriques de capacité, flux (F) et *betweenness centrality* (BC) dans les analyses explicatives réalisées aux chapitres 3 et 4. Toutefois, dans le cadre d'une utilisation opérationnelle, les delta-métriques introduites par Saura et Rubio (2010) auraient certainement un intérêt plus important, comme le reflètent leurs nombreuses utilisations dans ce cadre précis (Foltête *et al.*, 2014 ; Sahraoui *et al.*, 2017 ; Tannier *et al.*, 2016 ; Tarabon *et al.*, 2019). Bien qu'ils s'agissent de valeurs relatives au réseau considéré et qu'elles soient influencées par sa délimitation spatiale, elles répondent plus directement au besoin d'évaluation de la contribution d'un élément du réseau à sa connectivité globale. Étant donné que les métriques de capacité, F et BC sont conceptuellement équivalentes aux métriques dPC_{intra} , dPC_{flux} et $dPC_{connector}$ introduites par Saura et Rubio (2010), nos résultats confirment l'intérêt de la prise en compte d'un ensemble complémentaire de delta-métriques pour orienter la conservation.

Par ailleurs, nous avons montré que les métriques complémentaires utilisées pour mesurer la quantité d'habitat atteignable n'étaient pas liées de la même façon aux différentes réponses génétiques considérées. Par exemple, alors que la capacité n'était liée qu'à la richesse allélique des populations de Criquet ensanglanté, la différenciation génétique était davantage liée à la métrique F. Ce résultat rappelle que l'influence de la quantité d'habitat atteignable s'exerce à différentes échelles spatiales et temporelles. En effet, la prise en compte d'une seule métrique pourrait occulter le fait que certaines taches contribuent à la quantité d'habitat atteignable à l'échelle de l'ensemble du réseau en mettant en connexion des ensembles de taches éloignés les uns des autres (souligné par les métriques BC ou $dPC_{connector}$ notamment). Les composantes complémentaires de la quantité d'habitat atteignable doivent donc être considérées chacune comme un critère pour orienter la conservation. La connectivité est parfois prise en compte dans des méthodes d'optimisation des mesures de conservation basées sur plusieurs critères telles que celles intégrées aux programmes Marxan (Ball *et al.*, 2009), et sa récente extension Marxan Connect (Daigle *et al.*, 2020), ou Zonation (Moilanen *et al.*, 2009). Dans ce cadre, trois critères d'optimisation devraient donc concerner la connectivité et non un seul.

Enfin, lorsque la modélisation spatiale est utilisée à des fins explicatives, ce sont les variables quantitatives produites qui sont intégrées à des modèles. À l'inverse, lorsque la modélisation oriente la conservation, la localisation spatiale des entités modélisées sert souvent de base à la décision. Dans ce contexte, rappelons que la localisation des chemins de moindre coût constituant les liens d'un graphe paysager est très sensible au scénario de coût utilisé pour construire le graphe (cf. Chapitre 5). Par conséquent, il est préférable de ne pas exploiter directement ces trajectoires spatiales pour localiser des mesures de conservation si (i) le scénario de coût utilisé n'a pas été validé à partir de données empiriques reflétant les mouvements des espèces (Zeller *et al.*, 2018) ou si (ii) d'autres scénarios de coût conduisant à des chemins très différents sont tout aussi vraisemblables dans le cadre d'une validation statistique. Cependant, la sensibilité des chemins n'invalide pas pour autant la capacité des métriques issues d'un graphe dérivé d'un tel scénario à refléter correctement la réalité écologique. En effet, les matrices de distances-coût sont nettement moins sensibles aux scénarios de coût et les métriques sont basées à la fois sur ces distances et sur les capacités des taches.

2.2.2 Intégrer graphes génétiques et paysagers pour conserver la variabilité génétique

L'utilisation de données génétiques pour modéliser la structure génétique des populations sous la forme d'un graphe a permis de confirmer la pertinence de l'utilisation des graphes paysagers dans une optique de conservation. Quand ces données sont disponibles, elles pourraient permettre de répondre plus spécifiquement à des enjeux de conservation de la variabilité génétique en intégrant des graphes génétiques et paysagers.

Comment orienter des mesures de conservation en intégrant des graphes génétiques et paysagers ?

Alors que les graphes paysagers incluent l'intégralité des taches d'habitat potentielles de l'espèce étudiée, les graphes génétiques ne concernent le plus souvent que des populations échantillonnées dans un sous-ensemble de ces taches. L'intégration de ces graphes visait notamment à modéliser le réseau formé par toutes ces taches d'habitat à l'aide d'informations reflétant la réalité écologique, tirant ainsi parti de la complémentarité de ces outils. Dans le dernier chapitre, nous avons réalisé cette intégration en utilisant des modèles gravitaires pour identifier le scénario de coût le plus réaliste. Dans ces analyses basées sur des simulations, nous disposions de données génétiques issues de l'ensemble des populations au niveau desquelles nous avons simulé des flux génétiques. Ce type de modèles aurait un intérêt supplémentaire pour la conservation de la connectivité si leur extrapolation permettait de prédire la différenciation génétique entre des paires de populations non échantillonnées. Ils permettraient alors d'identifier dans l'ensemble du réseau d'habitat les zones au niveau desquelles la différenciation génétique devrait être la plus forte.

Les modèles gravitaires avaient la particularité d'intégrer des variables calculées au niveau des taches, telles que la surface de ces taches. Néanmoins, dans les situations où l'inclusion de ces variables améliorerait les modèles, leur pouvoir explicatif (coefficients de détermination, R^2) était très faible. L'extrapolation de tels modèles à des fins prédictives semble donc avoir un intérêt limité. S'ils n'incluent pas de variables locales, les modèles à partir desquels nous pourrions réaliser des prédictions seraient identiques à ceux utilisés pour modéliser la différenciation génétique entre les populations de Paruline caféiette (modèles mixtes de type MLPE, cf. Chapitre 4). Nous avons évalué leur qualité prédictive à l'aide d'une méthode de validation croisée. Le R^2 de validation maximal obtenu par cette approche valait 0.25. Cette valeur est relativement élevée si l'on considère que la différenciation génétique dépend pour partie de la dérive génétique aléatoire et que le modèle ne tient compte que de l'influence des flux génétiques. Néanmoins, il n'est pas certain qu'elle soit suffisante pour en conclure que le modèle a une valeur prédictive. Une telle méthode pourrait être reproduite dans d'autres contextes où l'extrapolation du modèle se justifierait peut-être davantage, notamment parce que la relation entre différenciation génétique et distance paysagère y serait forte et qu'un patron d'isolement par la résistance du paysage s'observerait à l'échelle de l'ensemble de la zone d'étude. Dans un tel cas, il serait possible de prédire convenablement la différenciation génétique entre chaque paire de populations dès lors qu'on peut calculer la distance paysagère qui les sépare et qu'on dispose de suffisamment de données génétiques pour calibrer le modèle à extrapoler.

Nous avons en outre montré que la validation croisée avait l'avantage d'identifier les paires de populations entre lesquelles la différenciation génétique est mal prédite. Cette information peut orienter de futures prospections de terrain vers ces populations pour comprendre pourquoi la connectivité des habitats, telle qu'elle est modélisée, n'est pas suffisante pour expliquer la différenciation génétique.

Cette démarche contribuerait ainsi à l'identification des facteurs locaux influençant la structure génétique des populations.

Développement d'indicateurs opérationnels à partir de graphes génétiques et paysagers

Les indices génétiques calculés à partir des graphes génétiques permettent d'estimer la diversité génétique locale d'une population ou son degré de différenciation génétique vis-à-vis des autres populations. Ces informations sont utiles à la conservation de la biodiversité dans la mesure où elles identifient les populations dont la diversité génétique a été fortement érodée. Dans ce cas, des mesures peuvent être mises en œuvre pour restaurer les conditions nécessaires aux flux génétiques, ou pour provoquer directement ces flux génétiques en introduisant des individus (*assisted gene flow*, [Frankham et al. \(2017\)](#)). Dans le premier cas, les flux génétiques ne permettront de restaurer la diversité génétique dans la population ciblée que si les migrants viennent de populations dans lesquelles des allèles absents de cette population sont présents. Il est donc nécessaire d'évaluer également la dissimilarité des pools d'allèles des populations du réseau pour estimer le niveau de diversité génétique qui pourrait être potentiellement rétabli à court terme par restauration des flux génétiques.

Des indices tels que la richesse allélique privée ([Szpiech et al., 2008](#)), utilisée pour étudier la structure génétique des populations de Criquet ensanglanté, sont basés sur la dissimilarité des pools d'allèles des différentes populations. Néanmoins, ils restent limités par une vision essentiellement statistique des données génétiques car ils comparent une population donnée à toutes les autres populations du réseau, sans intégrer leur distribution spatiale. Par ailleurs, des approches permettent d'estimer la diversité génétique d'un ensemble de populations voisines spatialement ([Shirk et Cushman, 2011](#)). Cependant, elles permettent d'évaluer leur degré de diversité et non pas la complémentarité de leurs pools d'allèles. Il serait alors intéressant de mesurer cette complémentarité en tenant compte de la topologie du réseau de populations et de la probabilité que des flux génétiques se produisent d'une population à une autre. Dans cette perspective, une telle démarche consisterait à calculer un indice de diversité génétique potentielle qui pourrait s'appuyer sur les propriétés respectives d'un graphe génétique et d'un graphe paysager. La figure 70 illustre l'application d'un tel indice, qui faciliterait le ciblage des mesures de restauration de la connectivité au niveau de populations dont on souhaite restaurer la diversité génétique.

2.3 Obstacles à l'utilisation des graphes génétiques et paysagers pour conserver la connectivité

Si l'utilisation de la modélisation spatiale et plus spécifiquement de graphes génétiques et paysagers présente un certain nombre d'avantages pour la conservation de la connectivité, cette approche se heurte néanmoins à des obstacles. Ils sont d'une part liés au coût des données nécessaires à la modélisation ; et en particulier lorsqu'il s'agit de données génétiques. D'autre part, le véritable apport de l'utilisation des graphes dépend de la place accordée à la modélisation dans le processus de décision et à son influence sur le succès des mesures de conservation à l'issue de ce processus.

2.3.1 Le coût associé à l'obtention des données génétiques

Les méthodes évoquées dans la section précédente nécessitent l'acquisition de données génétiques, ce qui est rarement à la portée des gestionnaires de la biodiversité en raison du coût élevé que cela

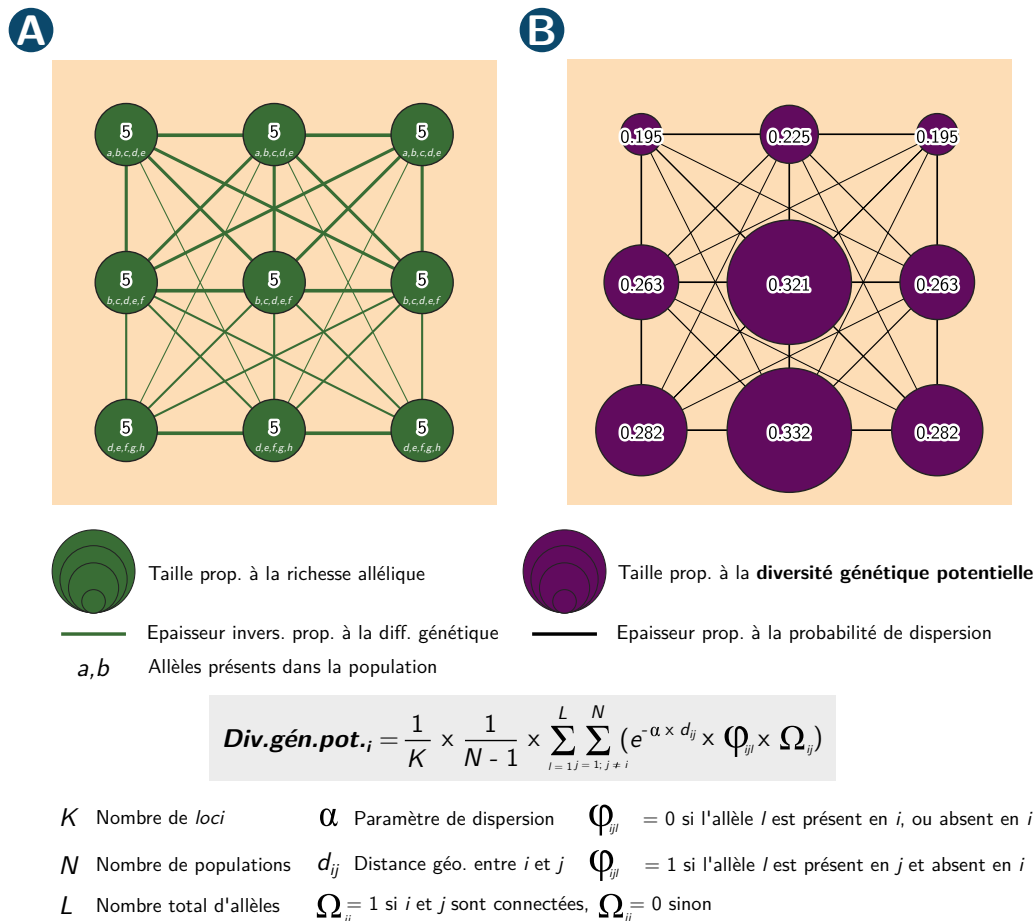


FIGURE 70 – Exemple d'utilisation d'un indice de diversité génétique potentielle basé sur des graphes génétiques et paysagers. La formule permettant de calculer l'indice tient compte des différences de composition des pools alléliques et des probabilités de dispersion calculées à partir de la distance géodésique euclidienne entre populations. Elle estime ainsi la diversité génétique que pourrait potentiellement gagner une population grâce aux flux de dispersion si ceux-ci n'étaient pas contraints par la résistance de la matrice paysagère. L'hypothèse sous-jacente est que l'absence de certains allèles dans des populations en l'état actuel est liée à la résistance qu'exerce la matrice paysagère sur la dispersion. Elle indique ainsi à quels endroits il faudrait réduire cette résistance pour que la variabilité génétique soit plus largement distribuée. Dans cette illustration, nous présentons le patron de diversité et de différenciation génétique initial (A) et les valeurs de l'indice de diversité génétique potentielle obtenues à partir des probabilités de dispersion (B). Un seul locus à huit allèles est considéré. La richesse allélique est la même partout et la dispersion dépend uniquement de la distance géodésique euclidienne entre populations, mais les populations centrales sont moins différentes de celles du haut que de celles du bas. La diversité génétique potentielle est donc plus forte dans les populations du bas, plus différentes initialement. Elle est également forte dans les populations centrales, ainsi qu'au milieu de chacun des rangs, du fait de la position de ces populations dans le réseau. Cet indice est donc sensible à la topologie du graphe et aux différences entre pools alléliques. Il indique les populations au niveau desquelles les efforts de restauration des flux génétiques auraient l'effet le plus bénéfique sur la diversité génétique locale, celles du bas et la population centrale dans ce cas.

représente (Baguette *et al.*, 2013). Néanmoins, de nouvelles technologies de séquençage tendent à faire diminuer ce coût et permettent d’obtenir des données génétiques concernant une plus large gamme de taxons (Segelbacher *et al.*, 2010). Lorsque l’acquisition de données génétiques est possible, pour limiter son coût tout en maximisant son intérêt, certains auteurs suggèrent de se focaliser sur les espèces parapluie ou reconnues comme sensibles à la connectivité identifiées dans de précédents travaux (Albert et Chaurand, 2018 ; Baguette *et al.*, 2013). Il est alors plus probable que les analyses permettent de comprendre l’influence de la connectivité et leurs résultats pourraient orienter des mesures de conservation dédiées à ces espèces sensibles.

Les situations dans lesquelles l’obtention de données génétiques est impossible devraient malgré tout rester fréquentes. Dans la mesure où le maintien de la variabilité génétique des populations est un objectif clé des programmes de conservation (Díaz *et al.*, 2019), on peut toutefois se demander si la structure génétique des populations pourrait être prise en compte de façon plus explicite pour orienter la conservation y compris lorsque les données génétiques ne sont pas disponibles. En effet, la prise en compte de certaines métriques de connectivité se justifie parfois par le fait qu’elles sont censées refléter la diversité génétique locale de la population qui occupe une tache d’habitat (Zetterberg *et al.*, 2010). Nous avons néanmoins montré dans les chapitres 3 et 4 que la structure génétique des populations dépend de plusieurs métriques complémentaires car elle est déterminée par des processus écologiques ayant lieu à différentes échelles spatiales et temporelles. Même si nous préconisons pour cette raison de prendre en compte un ensemble de métriques de connectivité complémentaires pour analyser le patron de connectivité des habitats, simuler le degré de diversité génétique des populations au niveau des taches serait un moyen direct de guider les mesures de conservation en l’absence de données empiriques. Un graphe paysager permettrait de déterminer les paramètres de simulation (probabilités de dispersion, effectifs des populations occupant les taches), à l’instar des simulations réalisées dans les chapitres 1 et 6. Une telle approche refléterait les analyses de viabilité des populations mises en œuvre en biologie de la conservation (Boyce, 1992) et renforcerait la dimension appliquée des travaux de génétique du paysage, très faible encore aujourd’hui (Bowman *et al.*, 2016). La figure 71 en présente un exemple. La réalisation de telles simulations avec CDPOP (Landguth et Cushman, 2010) impliquerait des calculs relativement longs, mais d’autres programmes permettent de faire les simulations de façon plus rapide (Adamack et Gruber, 2014 ; Andrello *et al.*, 2021 ; Hoban *et al.*, 2012)¹³.

Les graphes que nous avons validés à l’aide de données génétiques concernant la Paruline caféïette (Chapitre 4) reflétaient l’influence de la connectivité des habitats sur la dérive et les flux génétiques ayant eu lieu sur plusieurs générations. À ce titre, ils ont un intérêt pour la mise en œuvre de mesures visant à conserver la variabilité génétique des populations¹⁴. Cependant, d’autres données telles que des traces GPS pourraient refléter les mouvements quotidiens ou intra-annuels des espèces et être utilisées de façon similaire pour valider les graphes paysagers. Les premières analyses menées en ce sens pour valider des chemins de moindre coût (Driezen *et al.*, 2007 ; Laliberté et St-Laurent, 2020), et plus rarement des métriques issues de graphes (Poli *et al.*, 2020) mériteraient d’être complétées. Enfin, la méthode de validation mise en œuvre dans le chapitre 4 pourrait également être répliquée en

13. Notons que les résultats du chapitre 1 ont pu être reproduits en faisant les simulations avec le package PopGen-Report.

14. L’ouvrage de Frankham *et al.* (2017) présente de telles mesures et identifie les situations dans lesquelles elles auraient le plus d’intérêt.

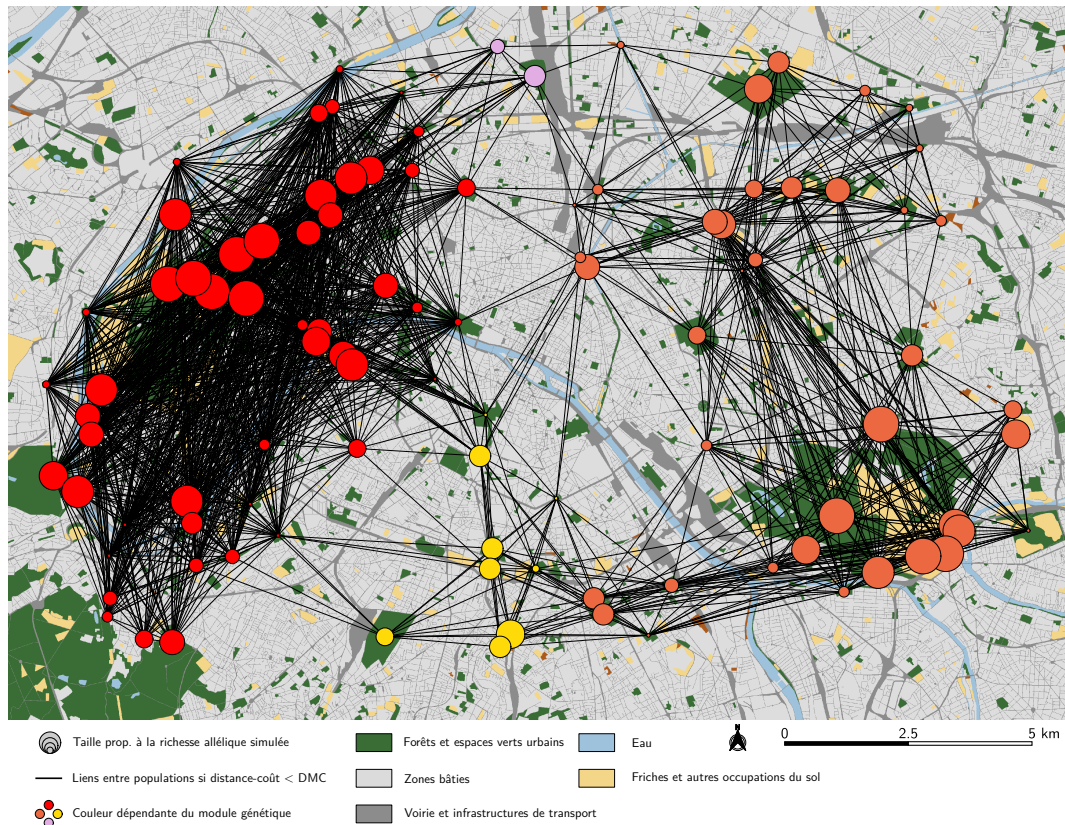


FIGURE 71 – Illustration de l'intérêt des simulations génétiques dans une optique de conservation. Pour cet exemple caricatural, nous avons localisé aléatoirement des populations dans les espaces verts urbains et les forêts de la ville de Paris, avant de calculer des distances-coût et de simuler des flux génétiques entre ces populations. Nous représentons sur cette carte la richesse allélique simulée de ces populations, la localisation des liens contribuant le plus aux flux génétiques (leur distances-coût est inférieure à la distance maximale de corrélation (DMC) calculée à partir des valeurs de D_{PS}) et l'appartenance de chaque population à un module du graphe génétique basé sur ces liens. On distingue ainsi deux ensembles de populations situées à l'ouest et à l'est de la ville. Les populations situées entre ces deux ensembles sont davantage isolées ce qui se traduit par une diversité génétique plus faible. Une telle représentation est basée sur un échantillon de populations uniquement, entre autres limites, mais pourrait apporter des informations complémentaires à l'analyse de la connectivité menée à partir d'un graphe paysager.

étudiant des espèces aux traits écologiques et degrés de spécialisation variés afin d'évaluer l'intérêt des graphes paysagers dans la large gamme de situations dans laquelle ils sont communément employés.

2.3.2 Le passage difficile de la modélisation à l'action

La validité des graphes paysagers et des métriques de connectivité qu'ils fournissent ne signifie pas nécessairement qu'ils contribuent concrètement à la conservation de la biodiversité. Pour pouvoir affirmer qu'ils ont un intérêt pour répondre à un enjeu de conservation précis, il faudrait :

1. que des graphes soient utilisés,
2. que cette utilisation réponde convenablement aux enjeux de conservation identifiés,
3. qu'elle se traduise par des mesures concrètes,
4. que ces dernières soient mises en œuvre,
5. que leur mise en œuvre améliore l'état de conservation des populations ciblées.

Ainsi, ces outils pourraient davantage contribuer à la conservation s'ils étaient utilisés plus fréquemment dans un cadre opérationnel. Bien que les graphes paysagers aident les gestionnaires à présenter les enjeux de conservation de la connectivité de façon explicite aux acteurs politiques, le manque de validation de ces outils a parfois été mentionné comme un frein à leur utilisation (Bergsten et Zetterberg, 2013 ; Galpern *et al.*, 2011). Nous espérons donc que nos travaux dissiperont les doutes concernant la pertinence de ces outils. En évitant aux gestionnaires de choisir les critères de conservation à prioriser, l'adoption plus large du concept de quantité d'habitat atteignable pourrait également favoriser l'utilisation de ces outils. Elle permettrait de rassembler les acteurs autour d'une vision commune de la connectivité n'opposant plus les partisans de la conservation de grandes taches d'habitat à ceux qui prônent la conservation de taches connectées. Or, l'absence de vision commune des enjeux est un obstacle majeur au succès des mesures de conservation de la connectivité selon Keeley *et al.* (2018).

En outre, pour que des outils tels que les graphes paysagers soient plus largement utilisés par les acteurs, il semble nécessaire de rappeler que la modélisation spatiale vise notamment à simplifier la réalité pour orienter l'action. Cela rendrait plus explicites les avantages mais aussi les limites de cette simplification. Alors que le lien mis en évidence entre les réponses génétiques mesurées et les métriques de connectivité illustre ces avantages, certains de nos résultats en illustrent également les limites. L'incertitude associée à la localisation spatiale des chemins de moindre coût en est un exemple (cf. Chapitre 5). L'écart entre les réponses génétiques mesurées dans certaines taches forestières de Guadeloupe et les valeurs des métriques de connectivité correspondantes, ou encore les fortes erreurs de prédiction de la différenciation génétique entre populations de Paruline caféïette parfois obtenues sur cette île en constituent d'autres exemples (cf. Chapitre 4). Dans ce contexte, la modélisation par les graphes paysagers serait probablement davantage acceptée par les acteurs, et notamment par les experts naturalistes, si elle était présentée comme un outil complémentaire et non concurrent de leur expertise. Cela éviterait notamment que la mise en œuvre des modèles de réseau écologique ne soit freinée par des critiques relatives à leur manque de réalisme, telles que celles rapportées par Vimal et Mathevet (2011) à partir de dialogues entre acteurs. Ces critiques sont justifiées mais l'intégration de certains éléments ou connaissances est parfois difficile et dépasse l'objet même de la modélisation qui est délibérément simplificatrice. C'est par exemple le cas de la prise en compte d'éléments difficiles à cartographier tels que des falaises ou d'éléments fins du chevelu hydrographique (Vimal et Mathevet, 2011). De tels éléments mériteraient d'être pris en compte en complément des résultats

issus des graphes paysagers, notamment pour orienter l'action dans les zones où ces modèles ont une valeur limitée. Pour cette raison, en même temps que nous avons validé les graphes paysagers dans le chapitre 4, nous avons tâché d'identifier les conditions dans lesquelles les prédictions issues de ces modèles s'écartaient des réponses mesurées.

Il est ensuite nécessaire que l'utilisation des graphes réponde aux enjeux précis de conservation. Nous avons illustré dans la section précédente la façon dont ils pourraient le faire lorsque l'enjeu principal est le maintien de la variabilité génétique des populations d'une espèce. Néanmoins, les enjeux de conservation concernent souvent la biodiversité au sens large, nécessitant de considérer les besoins de plusieurs espèces dans la modélisation. Les travaux basés sur des graphes paysagers d'[Albert *et al.* \(2017\)](#), [Bourgeois \(2015\)](#) et [Sahraoui *et al.* \(2017\)](#) apportent des solutions en ce sens.

Pour que les résultats de ces modélisations se traduisent en actions concrètes et qu'elles soient réellement mises en œuvre, il est également nécessaire d'impliquer un large panel d'acteurs. L'inclusion d'un modèle objectif tel qu'un graphe paysager dans le processus décisionnel peut être un atout, comme l'ont montré [Sahraoui *et al.* \(2021\)](#) dans le cadre d'une approche participative visant à évaluer l'impact de différents scénarios d'urbanisation de la métropole de Bordeaux avec des graphes paysagers, entre autres. Il est également utile dans ce contexte de considérer le coût financier associé aux mesures de conservation de la connectivité pour optimiser leur apport potentiel au regard du coût engagé ([Carroll *et al.*, 2012, 2021](#) ; [Naidoo *et al.*, 2006](#)), ou encore le statut foncier des terrains concernés pour se focaliser sur les mesures qu'il est possible de mettre en œuvre ([Tarabon *et al.*, 2020b](#)).

Pour finir, il convient de préciser que dans le rapport coûts-bénéfices estimé pour évaluer la validité écologique des graphes paysagers, les coûts correspondent au temps et à l'argent nécessaires à la modélisation spatiale (acquisition de données, modélisation informatique) tandis que les bénéfices correspondent au degré de réalisme écologique de cette modélisation. Néanmoins, un second type de rapport coûts-bénéfices devrait être optimisé pour améliorer la contribution des graphes paysagers à la conservation de la biodiversité. Celui-ci mettrait en rapport des coûts correspondant à l'ensemble de la démarche de mise en œuvre des mesures de conservation, de la modélisation aux actions concrètes sur le terrain, et des bénéfices qui seraient mesurés en évaluant l'état de conservation des populations ciblées. Cette démarche est rarement évaluée dans sa globalité. Néanmoins, l'acquisition diachronique de données génétiques et spatiales et la construction de graphes génétiques et paysagers avant et après la mise en œuvre des mesures de conservation permettrait de le faire. L'évolution de la structure génétique des populations caractérisée à l'aide des graphes génétiques pourrait être mise en relation avec l'évolution de la connectivité des graphes paysagers. L'utilisation conjointe de ces graphes contribuerait à évaluer le succès des mesures, mais aussi à comprendre les raisons de ce succès. Comme l'ont précisé [Baguette *et al.* \(2013\)](#), le coût d'un tel suivi génétique est néanmoins élevé et sa temporalité doit être adaptée à la durée de vie des espèces suivies ainsi qu'au rythme de modification des paysages étudiés.

3 Perspectives de recherche

Si les approches développées nous ont permis de répondre aux questions de recherche énoncées en introduction, elles laissent aussi percevoir des manières de poursuivre la démarche de recherche engagée dans cette thèse. **Les perspectives de recherche que nous présentons désormais permettraient (i) de traiter des problématiques similaires en s’inspirant de méthodes utilisées dans d’autres disciplines ou au contraire (ii) d’utiliser des méthodes similaires à celles développées dans cette thèse pour répondre à des questions posées dans d’autres disciplines.**

3.1 Vers une diversification de l’utilisation des graphes en génétique du paysage ?

3.1.1 Peut-on construire des graphes génétiques avec d’autres marqueurs génétiques ?

Les graphes génétiques que nous avons construits à partir de données empiriques (Chapitres 3 et 4) étaient basés sur des variations génétiques neutres caractérisées à partir de marqueurs de type microsatellite. En outre, lors des simulations réalisées (Chapitres 1 et 6), la sélection naturelle n’intervenait pas et les données génétiques étaient similaires à celles qui sont issues de ce type de marqueur. Ce choix vient du fait que les microsatellites reflètent principalement les processus démographiques influencés par la connectivité des habitats (Putman et Carbone, 2014). Néanmoins, des travaux de génétique du paysage concernent également la façon dont la structure du paysage influence la sélection naturelle (Balkenhol *et al.*, 2017 ; Cushman *et al.*, 2018 ; Schoville *et al.*, 2012 ; Storfer *et al.*, 2018). Par exemple, Perrin *et al.* (2021) ont récemment montré que la fragmentation des habitats n’affectait pas de la même manière des *loci* non sélectionnés et des *loci* liés aux réponses immunitaires chez le Grimpar bec-en-coin (*Glyphorynchus spirurus*), oiseau forestier de Guyane. Dans un tel contexte, il serait utile de comprendre comment les différentes composantes de la quantité d’habitat atteignable affectent aussi bien la dérive et le flux génétique que l’adaptation locale des populations. Une meilleure compréhension de ces phénomènes est également nécessaire à l’heure où le lien existant entre la diversité génétique neutre et la survie des populations est remis en cause (Teixeira et Huber, 2021). Étudier des marqueurs sélectionnés serait donc un moyen d’évaluer l’influence du paysage sur la capacité des populations à s’adapter à leur environnement, et non pas uniquement sur les déplacements des individus, qui limitent parfois cette adaptation (Lenormand, 2002). Cela contribuerait aussi bien à la compréhension des phénomènes évolutifs qu’à la conservation de conditions garantissant que les populations soient capables de s’adapter pour survivre dans un contexte changeant.

Les analyses de génétique du paysage visant à comprendre le lien entre le paysage et les processus adaptatifs se basent le plus souvent sur des marqueurs tels que les SNPs (*Single Nucleotide Polymorphisms*, Sork *et al.* (2016) par exemple), ou les AFLP (*Amplified Fragment-Length Polymorphisms*, Bertin *et al.* (2017) par exemple). De nouvelles technologies telles que la technique RAD-seq permettent d’analyser ce type de marqueur au niveau de milliers de *loci*, sans que le développement d’amorces spécifiques ne soit nécessaire au préalable (Andrews *et al.*, 2016 ; Storfer *et al.*, 2018). Les méthodes de *genome scan* permettent ensuite d’identifier parmi ces *loci* ceux qui sont soumis à la sélection (Foll et Gaggiotti, 2008). L’utilisation de ces marqueurs tend donc à se démocratiser, même si les analyses de bio-informatique à réaliser suite aux analyses de biologie moléculaire sont dans ce cas

plus complexes qu’avec des microsatellites. Nous n’avons pas utilisé de marqueurs de type SNPs pour construire des graphes génétiques mais avons rendu les fonctions du package `graph4lg` compatibles avec leur utilisation. Plusieurs méthodes permettent de créer des graphes génétiques à partir de tels marqueurs (Greenbaum *et al.*, 2016 ; Neuditschko *et al.*, 2012 ; Steinig *et al.*, 2016). À condition que les *loci* considérés ne soient pas soumis à la sélection, ces graphes pourraient être utilisés de la même façon que ceux que nous avons construits. À l’inverse, si des *loci* soumis à la sélection sont utilisés, ces graphes pourraient servir à étudier le lien entre la structure du paysage et les processus adaptatifs. Par exemple, imaginons qu’un marqueur soit sélectionné par les conditions climatiques ou la pression parasitaire locale (voir par exemple Sork *et al.* (2016)). Une forte quantité d’habitat atteignable pourrait soit (i) favoriser l’adaptation locale en limitant la disparition aléatoire des allèles conférant un avantage adaptatif sous l’effet de la dérive ou (ii) freiner cette adaptation en favorisant les flux génétiques avec des populations non soumises à la même pression sélective. Mettre en relation des indices de diversité et de différenciation génétique relative calculés à partir de marqueurs sélectionnés au niveau des nœuds d’un graphe (ex. : richesse allélique et MIW, respectivement) et des mesures complémentaires de la quantité d’habitat atteignable permettrait d’identifier l’hypothèse la plus vraisemblable.

3.1.2 Comment dépasser le graphe de populations statique et non-dirigé ?

Les graphes construits dans cette thèse étaient des graphes spatiaux, pondérés, non-dirigés, parfois élagués et dont les nœuds étaient soit des taches d’habitat soit des populations d’individus. La flexibilité des méthodes issues de la théorie des graphes a permis de créer des graphes génétiques et paysagers aux propriétés différentes pour traiter des questions de génétique du paysage. Tout d’abord, bien que la structure génétique soit le plus souvent analysée à l’échelle des populations, Prunier *et al.* (2013) ont montré que les analyses individus-centrés pouvaient être avantageuses en génétique du paysage (voir aussi Balkenhol et Fortin (2016), Luximon *et al.* (2014) ou Schwartz et McKelvey (2009) sur ce point). Elles permettraient d’échantillonner un plus grand nombre de sites pour un nombre total d’individus constant. Notre utilisation de graphes ayant des populations comme nœuds se justifiait par leur mise en relation avec des graphes paysagers représentant le réseau d’habitat d’espèces formant des populations discrètes. Néanmoins, dans d’autres contextes, les graphes ayant des individus comme nœuds, utilisés notamment par Cross *et al.* (2018) ou Greenbaum *et al.* (2016), auraient un intérêt. Ils seraient par exemple plus adaptés aux analyses de modularité visant à identifier des barrières paysagères au flux génétique.

La structure génétique des populations et les paysages étudiés sont dynamiques. Ainsi, la structure génétique évolue au cours du temps, tout comme la distribution des habitats et les possibilités de dispersion entre les taches d’habitat ne sont pas identiques d’une année sur l’autre. Pourtant, les graphes que nous avons construits représentent les réseaux d’habitat et de populations à un instant t et sont donc statiques. Ils ne permettent donc pas d’étudier l’influence de la dynamique spatio-temporelle de la connectivité sur la structure génétique. Cette influence est difficile à évaluer dans la mesure où la structure génétique des populations répond elle-même aux contraintes paysagères avec un certain délai temporel (Landguth *et al.*, 2010), mais aussi car les données spatiales anciennes ne sont pas toujours disponibles. Néanmoins, la prise en compte de la dynamique dans les modèles de connectivité a donné lieu à plusieurs formes d’analyses, qui diffèrent par leur façon d’intégrer les données multi-temporelles (Uroy *et al.*, 2021). Parmi ces approches, Martensen *et al.* (2017) ont développé une méthode basée sur des graphes paysagers permettant de calculer des métriques de connectivité spatio-temporelle en

tenant compte du fait que le déplacement des individus à l'échelle de plusieurs années dépend de la dynamique du réseau d'habitat. De telles métriques pourraient être mises en relation avec des indices génétiques calculés au niveau de populations occupant des taches d'habitat plus ou moins anciennes pour comprendre l'effet de la dynamique du réseau d'habitat sur la structure génétique. Ce type d'approche pourrait également s'inspirer de travaux d'écologie récents basés sur des graphes dynamiques (Miele et Matias, 2017).

Par ailleurs, les graphes seraient plus réalistes si leurs liens étaient dirigés car la dispersion est parfois asymétrique (Holderegger et Gugerli, 2012), certaines populations ou taches d'habitat jouant le rôle de puits quand d'autres constituent des sources (Pulliam, 1988). Il serait ainsi utile d'identifier les propriétés faisant d'une tache une source ou un puits. Par la suite, les probabilités de dispersion intégrées au calcul des métriques de connectivité pourraient tenir compte différemment des propriétés des taches d'habitat selon qu'elles sont l'origine ou la destination des flux de dispersion, augmentant ainsi leur réalisme. De premières études ont été menées en ce sens. Par exemple, Draheim *et al.* (2016) ont mis en évidence des processus de type source-puits chez des populations d'Ours noir (*Ursus americanus*) en utilisant des graphes génétiques, tout comme Andreasen *et al.* (2012) chez des populations de Puma (*Puma concolor*).

Pour intégrer ces processus aux graphes génétiques, il serait possible de pondérer leurs liens avec les nombres de migrants estimés à partir des données génétiques entre chaque paire de populations et dans chaque sens, comme le permettent les méthodes de Faubet et Gaggiotti (2008) et Sundqvist *et al.* (2016), entre autres. Ces effectifs de migrants pourraient être modélisés à l'aide de modèles gravitaires similaires à ceux réalisés dans le dernier chapitre. Cependant, contrairement à ces derniers, ces modèles gravitaires devraient intégrer deux observations par paire de populations, pour modéliser le flux de migrants allant d'une population A à une population B, et inversement. Les variables explicatives seraient alors les propriétés des taches d'habitat "d'origine" et de "destination" et les distances-coût entre elles.

3.2 Applications des méthodes développées à d'autres disciplines

Les méthodes de comparaison et d'intégration des graphes génétiques et paysagers que nous avons utilisées trouveraient des applications dans d'autres disciplines. Elles permettraient par exemple d'étudier les réseaux d'interaction entre des espèces hôtes et leurs parasites, ou d'identifier les déterminants de la composition des communautés d'espèces.

3.2.1 Comprendre l'enchevêtrement des réseaux d'hôtes, de parasites et de leurs habitats à l'aide de graphes

Les espèces que l'on peut étudier à l'aide des graphes interagissent avec d'autres espèces. Si nous disposons aujourd'hui de beaucoup de connaissances sur la façon dont les espèces sont influencées par la connectivité de leurs habitats, il en est autrement concernant l'influence du patron spatial des habitats sur les interactions entre espèces (Fardila *et al.*, 2017). Parmi elles, les interactions antagonistes de type hôte-parasite sont omniprésentes (Lafferty *et al.*, 2006 ; Sorci et Garnier, 2008). Elles ont une influence sur la survie des populations des espèces hôtes, leur structure génétique et sur l'évolution

de leur système immunitaire, entre autres (Pedersen et Fenton, 2007). Elles sont également étudiées pour comprendre les mécanismes à l'origine des zoonoses.

Des travaux ont récemment mis en évidence le lien entre le patron spatial des habitats et les interactions hôtes-parasites (Bordes *et al.*, 2015 ; Pérez-Rodríguez *et al.*, 2018 ; Perrin, 2021 ; Tylianakis *et al.*, 2007). En effet, la connectivité des habitats semble influencer les taxons impliqués et l'intensité des interactions hôtes-parasites. De plus, la diversité génétique des populations des espèces hôtes pourrait également déterminer leur propension à être parasitées. Toutefois, les réponses à ces questions sont encore partielles. Elles se posent dans un contexte où l'utilisation conjointe de différents types de graphes spatiaux ou aspatiaux trouve de plus en plus d'applications en écologie (Fontaine *et al.*, 2011 ; Hutchinson *et al.*, 2019). Ainsi, des études portant sur les interactions hôtes-parasites mobilisent parfois des graphes bipartites et aspatiaux pour décrire le réseau d'interaction à l'aide de métriques issues de la théorie des graphes, telles que des indices de modularité ou de connectance du graphe (voir Thébault et Fontaine (2010)). Si de tels graphes sont construits à partir de données acquises en plusieurs sites, les métriques peuvent être reliées statistiquement à des variables environnementales mesurées au niveau de ces sites pour identifier les déterminants environnementaux des interactions hôtes-parasites. Des métriques de connectivité des habitats pourraient ainsi être incluses à une telle analyse pour évaluer le rôle de la connectivité sur la structure de graphes d'interaction construits au niveau de plusieurs nœuds d'un graphe paysager. De façon similaire, des indices issus de graphes génétiques et décrivant la structure génétique des populations hôtes permettraient de comprendre le lien entre la diversité génétique et la présence des parasites dans ces populations. Ces approches laissent donc percevoir l'intérêt qu'auraient les graphes génétiques et paysagers dans ce contexte.

3.2.2 Contribuer à la compréhension de la structure des communautés d'espèces et à la conservation de la diversité spécifique à l'aide de graphes

Si la diversité génétique est au cœur de programmes de conservation de certaines espèces, la diversité spécifique reste sans doute le niveau de biodiversité central dans ce type de programme, en particulier dans un contexte d'érosion massive. Nous terminerons donc en évoquant les parallèles existant entre les questions auxquelles cherchent à répondre les analyses de génétique du paysage et d'écologie des communautés, et la façon dont les méthodes développées dans cette thèse permettraient de répondre à ces dernières.

La structure des communautés d'espèces et la structure génétique des populations sont soumises à l'influence de processus écologiques comparables. En effet, on peut considérer que la colonisation est l'analogie au niveau communautaire du flux génétique pour une espèce donnée (Vellend et Geber, 2005). De même, l'extinction peut correspondre à la dérive, la spéciation à la mutation et les filtrages environnementaux et interactions inter-spécifiques favorisant ou excluant certaines espèces des communautés à la sélection naturelle. Espèces et allèles ne seraient alors que différentes entités identifiées pour décrire et comprendre la façon dont les patrons de biodiversité sont soumis à des processus analogues, mais à différents niveaux d'organisation du vivant.

Nous avons décrit la diversité génétique des populations au niveau des nœuds des graphes génétiques et pondéré les liens de ces graphes par des mesures de différenciation génétique. En écologie des communautés, on distingue de façon similaire la diversité α et la diversité β . La première cor-

respond au nombre d'espèces présentes dans une communauté, pondéré ou non par leurs fréquences respectives, tandis que la seconde correspond aux différences de composition des communautés d'espèces. La richesse spécifique mesurée à l'échelle d'un ensemble de communautés, c'est-à-dire d'une méta-communauté, dépend de ces deux variables et constitue la diversité γ . L'influence des conditions environnementales et de la structure du paysage sur ces différentes composantes de la structure des communautés fait encore l'objet de nombreuses questions à l'heure actuelle, par exemple :

- À quel point l'hétérogénéité des habitats contribue à la diversité γ d'une méta-communauté ? (Allouche *et al.*, 2012 ; Ben-Hur et Kadmon, 2020 ; Stein *et al.*, 2014)
- Y-a-t-il davantage d'espèces dans un ensemble de petites taches que dans une grande tache d'habitat à surfaces totales équivalentes ? (Fahrig *et al.*, 2020)
- Quelle est l'influence de la connectivité des taches d'habitat sur les diversités spécifiques α , β et γ ? (Chisholm *et al.*, 2011 ; Economo et Keitt, 2008, 2010)

Des graphes similaires à ceux utilisés dans cette thèse pourraient aider à répondre à ces questions. En effet, il est possible de créer des graphes de communautés dont les nœuds correspondraient à des communautés d'espèces au niveau desquelles on pourrait mesurer la diversité α et dont les liens seraient pondérés par des indices de diversité β . Economo et Keitt (2008) et Economo et Keitt (2010) ont mis en œuvre une approche semblable dans le cadre de simulations, mais elle mériterait d'être appliquée avec intérêt à des données empiriques, en considérant des communautés d'espèces occupant un type d'habitat similaire (des amphibiens par exemple). Les résultats de cette thèse indiquent que ce type de graphe pourrait être mis en relation avec un graphe paysager pour comprendre comment la connectivité des habitats influence la structure des communautés en déterminant les taux de colonisation et d'extinction des espèces. Des analyses ont déjà mis en relation des indices de diversité α (Laroche *et al.*, 2020 ; Mony *et al.*, 2018 ; Ribeiro *et al.*, 2011) ou β (Martín-Queller et Saura, 2013) avec des métriques issues de graphes paysagers, sans que la structure des communautés ne soit toutefois décrite sous la forme d'un graphe. Ces études pourraient donc être complétées en calculant l'équivalent de l'indice MIW à partir d'indices de diversité β . Par ailleurs, les méthodes d'élagage pourraient trouver des applications dans ce contexte car le lien entre la similarité de composition des communautés et les distances paysagères ne devrait également s'observer que jusqu'à une certaine distance seuil (Nekola et White, 1999). Les méthodes utilisées en écologie des communautés pour modéliser cette relation telles que le *Generalized Dissimilarity Modelling* (Ferrier *et al.*, 2007) intègrent d'ailleurs cela et on peut se demander si elles ne pourraient pas être utilisées pour modéliser les patrons d'isolement par la résistance en génétique du paysage.

Enfin, nous avons établi un parallèle entre la structure génétique des populations et la composition des communautés d'espèces. Néanmoins, la diversité génétique et la diversité spécifique peuvent aussi bien être corrélées positivement que négativement (Vellend et Geber, 2005). En effet, plusieurs hypothèses permettent d'expliquer la relation entre ces deux composantes de la biodiversité. Selon l'une d'elles (Wright, 1983), un grand nombre d'espèces peut subsister si chacune de leurs populations compte un nombre important d'individus, limitant ainsi leur risque d'extinction. Dans un tel cas, chaque population est peu affectée par la dérive et devrait conserver une diversité génétique importante. On peut alors s'attendre à une corrélation positive entre diversités génétique et spécifique sous cette hypothèse. Selon une seconde hypothèse (Allouche *et al.*, 2012 ; Stein *et al.*, 2014), l'hétérogénéité des conditions environnementales permet à un nombre important d'espèces de survivre, mais cela se fait aux dépens de l'abondance de chacune d'elles (mais voir Ben-Hur et Kadmon (2020)). Dans un

tel cas, on s'attend à observer une corrélation négative entre diversités génétique et spécifique. Ces différentes hypothèses ont récemment été testées. [Schmidt *et al.* \(2021\)](#) ont observé une corrélation négative entre la diversité génétique et la diversité spécifique en considérant un grand nombre d'espèces de mammifères à l'échelle de l'Amérique du Nord. À l'inverse, [Decocq *et al.* \(2021\)](#) ont mesuré une corrélation positive entre ces deux mesures de biodiversité en considérant des espèces de plantes forestières. Si la vérification d'une hypothèse plutôt qu'une autre dépend probablement du milieu et des taxons considérés et mérite davantage de recherches, nous pouvons néanmoins supposer que la comparaison de graphes génétiques et de graphes de communautés tels qu'imaginés dans le paragraphe précédent contribuerait à l'avancée de telles recherches. De façon complémentaire, elle permettrait d'évaluer à quel point la différenciation génétique et la diversité β sont corrélées et d'identifier les déterminants et l'échelle spatiale d'une telle relation.

Conclusion

Dans le contexte actuel d'érosion de la biodiversité, identifier l'influence de la connectivité écologique des habitats sur la structure génétique des populations est un enjeu majeur. En effet, si les activités humaines ont bouleversé la biodiversité en modifiant l'occupation des sols et la distribution des habitats, l'aménagement du territoire constitue un des principaux leviers sur lesquels reposent les mesures de conservation mises en œuvre actuellement pour limiter les impacts passés et à venir. Le succès d'une telle démarche dépend néanmoins des connaissances disponibles pour identifier les mesures de conservation les plus efficaces et de l'existence d'outils valides et accessibles permettant aux acteurs en charge de la conservation de planifier ces mesures.

Dans ce contexte, nous avons étudié l'intérêt des approches basées sur des graphes génétiques et paysagers. Ces deux outils sont basés sur le même cadre conceptuel, à savoir la théorie des graphes. Elle permet de représenter les réseaux de taches d'habitat et de populations sous la forme d'un ensemble de nœuds et de liens. La simplicité de cette modélisation permet de se focaliser sur la structure spatiale des réseaux et sur les propriétés des nœuds et liens qui les composent pour tester des hypothèses précises et acquérir des connaissances en écologie scientifique. Sa flexibilité rend quant à elle possible l'analyse des relations entre leurs entités à plusieurs échelles spatiales et en intégrant l'hétérogénéité de l'espace qui les entoure. Contrairement à des approches basées sur le calcul de distances entre l'ensemble des taches ou populations étudiées ou de variables locales décrivant le voisinage de ces entités indépendamment du réseau qu'elles forment, la théorie des graphes fait de la topologie de ce réseau une variable à part entière. De plus, elle intègre cette topologie pour caractériser la distribution spatiale des variables associées aux entités du réseau. Cet apport singulier des approches basées sur les graphes dans le cadre d'une démarche scientifique s'illustre notamment par l'intérêt de l'élagage des graphes génétiques pour comprendre l'influence du paysage sur les flux génétiques ou par la complémentarité des métriques de connectivité mesurant la quantité d'habitat atteignable pour analyser la structure génétique des populations.

Modéliser la réalité écologique sous la forme d'un graphe est également une façon de simplifier cette réalité pour orienter l'action. Toute tentative de modélisation est nécessairement imparfaite. Cependant, cette simplification est pertinente et plus largement utilisée dès lors qu'elle s'appuie sur des théories scientifiques validées empiriquement. Une telle validation faisait défaut en ce qui concerne l'utilisation des graphes paysagers pour modéliser la connectivité écologique des habitats et son influence sur la structure génétique des populations. Pour pallier à ce manque, nous avons validé la pertinence écologique de ces outils en les confrontant à des graphes génétiques construits à l'aide de données empiriques. Nous avons également amélioré les méthodes permettant de calibrer leurs paramètres de construction lorsque des données génétiques sont disponibles. Ainsi, bien qu'un graphe soit essentiellement un ensemble de points reliés par des traits, la complexité que la théorie des graphes permet d'intégrer dans cette représentation d'un réseau rapproche ces outils de la réalité écologique. Néanmoins, cette complexité n'est pas toujours gage de qualité, des approches simples reflétant parfois correctement la réalité écologique. Ce constat est un signal positif car le succès de la conservation de la biodiversité dépend de la qualité d'analyses mises en œuvre à partir de ressources le plus souvent limitées.

Glossaire

Biologie de la conservation

Science appliquée dédiée au maintien de la biodiversité (Hunter Jr et Gibbs, 2006). Discipline de crise selon l'un de ses fondateurs Michael E. Soulé (Soulé, 1986), elle a également un caractère fortement pluri-disciplinaire dans la mesure où elle se base sur les connaissances acquises en génétique, physiologie animale et végétale, biologie des populations mais aussi sur les sciences sociales et la philosophie pour identifier les conditions permettant de conserver la biodiversité et les mesures à mettre en œuvre pour les garantir (Soulé, 1985). Elle constitue une interface entre l'écologie scientifique et les politiques de conservation de la biodiversité.

Chemin de moindre coût

Chemin dont le tracé minimise le coût total lié au déplacement d'un individu à travers la matrice paysagère en fonction de la résistance attribuée aux différents éléments paysagers (Adriaensen *et al.*, 2003). Cette trajectoire spatiale, ou ligne dans un SIG, est le résultat de la modélisation du déplacement d'une espèce, le plus souvent animale.

Connectivité écologique des habitats

Dans cette thèse, nous définissons la connectivité écologique des habitats comme la surface d'habitat atteignable lors du déplacement des individus au sein de leurs taches d'habitat et vers les autres taches. Elle se mesure à partir de chaque tache d'habitat de la zone d'étude à plusieurs échelles spatiales définies par les capacités de dispersion de l'espèce à l'échelle temporelle d'une ou plusieurs générations. Aux échelles spatiales les plus larges, la connectivité d'une tache d'habitat tient compte de son rôle dans la connectivité du réseau d'habitat dans son ensemble.

Dérive génétique

Évolution aléatoire des fréquences alléliques d'une génération à une autre causée par la finitude du nombre d'individus, de gamètes et d'allèles qui contribuent à la génération suivante. L'intensité de la dérive génétique augmente lorsque la taille de la population qu'elle affecte diminue (Hamilton, 2011).

Différenciation génétique

Différence de fréquences alléliques entre deux populations d'individus. Cette différence peut se mesurer à l'aide de nombreux indicateurs, notamment des indices de fixation tels que le F_{ST} , que l'on nomme parfois distances génétiques. La significativité de la différenciation génétique entre deux populations peut être testée à l'aide d'un test basé sur une table de contingence génotypique.

Dispersion

Déplacement d'un individu de son lieu de naissance vers le lieu où il se reproduira si celui-ci est différent. Ce déplacement implique trois étapes : la décision par un individu de quitter un habitat favorable, le déplacement à travers une matrice paysagère plus ou moins favorable et l'établissement dans un nouvel habitat déjà occupé ou non par une population (Clobert *et al.*, 2012).

Distance-coût

Coût accumulé des pixels traversés le long du chemin de moindre coût entre deux points, pondéré ou non par la distance parcourue.

Diversité génétique

Au sens large, variabilité du matériel génétique des êtres vivants et niveau d'organisation le plus fin de la biodiversité. Au sens restreint, variabilité génétique observée entre les individus d'une même espèce formant une population, c'est-à-dire la diversité génétique intra-spécifique et intra-population.

Élagage

Suppression d'un ou plusieurs liens connectant les nœuds d'un graphe.

Écologie du paysage

Discipline qui étudie les interactions entre (i) l'hétérogénéité spatiale de paysages plus ou moins anthropisés et (ii) les patrons et processus écologiques qu'on y observe à différentes échelles spatiales.

Flux génétiques

Échanges d'allèles entre populations causés par le déplacement d'individus ou de leurs gamètes et la participation de ces derniers à la reproduction dans leur population de destination (Hamilton, 2011). Nous utilisons le terme "flux génétique" et non les termes "flux de gènes" ou "flux génique", également utilisés, pour éviter toute ambiguïté dans la mesure où les échanges génétiques que nous étudions impliquent des changements de fréquences alléliques et non pas des modifications de la structure du génome.

Génétique des populations

Discipline étudiant le patrimoine génétique des populations, et en particulier la distribution et l'évolution des fréquences des génotypes en réponse aux forces évolutives que sont la sélection naturelle, la dérive, les mutations et les flux génétiques.

Génétique du paysage

Discipline qui étudie l'influence de la structure du paysage sur la structure génétique des populations (Balkenhol *et al.*, 2016 ; Manel *et al.*, 2003 ; Storfer *et al.*, 2007). Contrairement aux approches phylogéographiques, la génétique du paysage s'intéresse à la façon dont le paysage contemporain influence d'une part la sélection naturelle et d'autre part la variabilité génétique neutre (Manel *et al.*, 2003).

Graphe

Représentation d'un réseau formé par des entités en interaction sous la forme d'un ensemble de nœuds connectés par des liens.

Graphe génétique

Graphe dont les nœuds correspondent à des individus ou à des populations pour lesquels on dispose de données génétiques, et dont les liens représentent et/ou sont pondérés en fonction du degré de différenciation génétique entre ces nœuds (Dyer et Nason, 2004 ; Greenbaum et Fefferman, 2017).

Graphe paysager

Graphe dont les nœuds correspondent aux taches d'habitat d'une ou plusieurs espèces et dont les liens représentent et/ou sont pondérés en fonction de la distance paysagère séparant ces nœuds (Urban et Keitt, 2001).

Habitat

Ensemble des zones qui constituent la niche fondamentale d'une espèce, au sens d'Hutchinson (1957), c'est-à-dire celles qui réunissent l'ensemble des conditions abiotiques (climatiques, pédologiques, topographiques) et biotiques (trophiques, symbiotiques, parasitaires) nécessaires à l'accomplissement du cycle de vie de cette espèce. Ces zones ne sont pas nécessairement occupées par l'espèce considérée, différant alors de sa niche réalisée.

Hétérogénéité du paysage

En écologie du paysage, l'hétérogénéité du paysage intègre la nature et la proportion des différentes taches de ressources présentes dans le paysage (composition) et la distribution spatiale de ces taches (configuration) (Fahrig *et al.*, 2011).

Inférence

Le terme "inférence" désigne ici l'inférence statistique, c'est-à-dire le processus par lequel on estime des paramètres relatifs à la population étudiée à partir de données relatives à un échantillon de celle-ci.

Isolement par la distance

Phénomène par lequel la différenciation génétique entre des individus (ou populations) augmente avec la distance géodésique qui les sépare car les flux génétiques entre eux sont moins fréquents à longue distance (Slatkin, 1993 ; Wright, 1943). Souvent noté IBD pour *Isolation By Distance* en anglais.

Isolement par la résistance du paysage

Extension sémantique du phénomène d'IBD dans le cas où le flux génétique est mieux expliqué par des distances tenant compte de l'hétérogénéité de la matrice paysagère que par des distances géodésiques. Le terme IBR (*Isolation By Resistance*) désigne le cas particulier où la distance paysagère représentant l'influence du paysage sur le flux génétique est calculée en appliquant la théorie des circuits (McRae, 2006). Nous utilisons donc le terme IBLR (*Isolation By Landscape Resistance*) pour englober également les cas où la distance paysagère est calculée avec d'autres méthodes (p.ex. : basées sur des chemins de moindre coût).

Matrice paysagère

Élément dominant du paysage (Burel et Baudry, 2003), au sein duquel se trouvent les taches de ressources ou d’habitat de l’espèce étudiée.

Microsatellite

Locus (séquence d’ADN) au niveau duquel un motif de nucléotides comptant généralement deux ou trois paires de bases (et jusqu’à une douzaine) est répété un certain nombre de fois (5 à 50 fois). Les microsatellites sont fréquents dans le génome, ont une forte variabilité d’un individu à un autre au sein de la même espèce et sont rarement transcrits en ARN, les rendant le plus souvent insensibles à la sélection naturelle.

Modèle de distribution d’espèce

Un modèle de distribution d’espèce (*Species Distribution Model* ou SDM) relie la présence ou l’absence d’une espèce à des variables environnementales (cf. Annexe B3). Il est calibré à l’aide d’une méthode d’analyse statistique (par exemple une régression logistique ou l’algorithme Maxent (Phillips *et al.*, 2006)) permettant de relier des observations ponctuelles à des variables environnementales spatialisées (Guisan et Zimmermann, 2000). L’extrapolation du modèle permet de cartographier l’aire de distribution d’une espèce.

Modèle gravitaire

Méthode statistique utilisée pour modéliser les interactions entre les entités d’un réseau en se basant sur la formule de la loi de la gravitation universelle de Newton (cf. Annexe B5). Un modèle gravitaire modélise l’interaction entre deux entités en fonction des caractéristiques de chacune des entités et de la distance qui les sépare.

Package R

Le terme *package* R désigne un ensemble de fonctions, au sens informatique du terme, codé dans le langage du logiciel R (R Core Team, 2020). Cet ensemble de fonctions facilite la réalisation d’analyses, de traitements et de visualisations de données à l’aide de ce logiciel. On parle aussi d’une librairie de fonctions.

Paysage

Selon le Larousse, (i) étendue spatiale, naturelle ou transformée par l’homme, qui présente une certaine identité visuelle ou fonctionnelle, ou (ii) vue d’ensemble que l’on a d’un point donné. En écologie du paysage et dans cette thèse, le paysage est un espace hétérogène constitué de taches de ressources ayant une certaine configuration spatiale, formant des motifs ou patrons paysagers caractéristiques (adapté d’après Forman et Godron (1986)).

Population

Groupe d'individus vivant suffisamment proches les uns des autres pour que chaque individu puisse potentiellement se reproduire avec un individu de sexe opposé de ce même groupe (Frankham *et al.*, 2017).

Résistance du paysage

Influence des éléments paysagers sur le déplacement des espèces. Cette influence est souvent représentée sous la forme d'une surface de résistance, c'est-à-dire un *raster* dans un SIG dont la valeur des pixels est plus ou moins élevée selon la résistance qu'ils sont censés exercer sur le mouvement (Zeller *et al.*, 2012).

Structure génétique

Distribution de la variabilité génétique au sein d'un ensemble de populations, caractérisée par la diversité génétique au sein de chaque population d'une part et par la différenciation génétique entre populations d'autre part.

Tache

Élément paysager homogène défini par sa nature, sa taille et sa forme (Burel et Baudry, 2003). Il s'agit de l'unité élémentaire des paysages étudiés dans le cadre de l'écologie du paysage.

Télémetrie

Technique permettant de suivre le déplacement d'un objet à l'aide d'un GPS ou d'un émetteur radio (*radiotracking*). Le terme est utilisé dans cette thèse en référence à l'utilisation de cette technique pour le suivi d'animaux.

Théorie des graphes

Branche des mathématiques combinatoires dédiée à l'analyse des connexions entre des entités discrètes formant un réseau (Harary, 1969).

Bibliographie

- ADAMACK, A. T. et GRUBER, B. (2014). PopGenReport : simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, 5(4):384–387.
- ADRIAENSEN, F., CHARDON, J., DE BLUST, G., SWINNEN, E., VILLALBA, S., GULINCK, H. et MATTHYSEN, E. (2003). The application of least-cost modelling as a functional landscape model. *Landscape and Urban Planning*, 64(4):233–247.
- AFONSO, E., TOURNANT, P., FOLTÊTE, J.-C., GIRAUDOUX, P., BAURAND, P.-E., ROUÉ, S., CANELLA, V., VEY, D. et SCHEIFLER, R. (2016). Is the lesser horseshoe bat (*Rhinolophus hipposideros*) exposed to causes that may have contributed to its decline? A non-invasive approach. *Global ecology and conservation*, 8:123–137.
- ALBERT, C. H. et CHAURAND, J. (2018). How to select surrogate species to ensure the consistency of ecological networks across political and biological scales? *Sciences Eaux et Territoires*, 25:26–31.
- ALBERT, C. H., RAYFIELD, B., DUMITRU, M. et GONZALEZ, A. (2017). Applying network theory to prioritize multi-species habitat networks that are robust to climate and land-use change. *Conservation Biology*.
- ALBERT, E. M., FORTUNA, M. A., GODOY, J. A. et BASCOMPTE, J. (2013). Assessing the robustness of networks of spatial genetic variation. *Ecology Letters*, 16(1):86–93.
- ALLENDORF, F. W. et LUIKART, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.
- ALLENDORF, F. W., LUIKART, G. H. et AITKEN, S. N. (2012). *Conservation and the genetics of populations*. Wiley Blackwell, New York, 2 édition.
- ALLOUCHE, O., KALYUZHNY, M., MORENO-RUEDA, G., PIZARRO, M. et KADMON, R. (2012). Area–heterogeneity tradeoff and the diversity of ecological communities. *Proceedings of the National Academy of Sciences*, 109(43):17495–17500.
- ANDERSON, C. D., EPPERSON, B. K., FORTIN, M.-J., HOLDEREGGER, R., JAMES, P., ROSENBERG, M. S., SCRIBNER, K. T. et SPEAR, S. (2010). Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology*, 19(17):3565–3575.
- ANDERSON, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116.
- ANDREASEN, A. M., STEWART, K. M., LONGLAND, W. S., BECKMANN, J. P. et FORISTER, M. L. (2012). Identification of source-sink dynamics in mountain lions of the Great Basin. *Molecular Ecology*, 21(23):5689–5701.
- ANDRELLO, M., NOIROT, C., DÉBARRE, F. et MANEL, S. (2021). MetaPopGen 2.0 : A multilocus genetic simulator to model populations of large size. *Molecular Ecology Resources*, 21(2):596–608.
- ANDREWARTHA, H. G., BIRCH, L. C. et al. (1954). *The distribution and abundance of animals*. University of Chicago press.
- ANDREWS, K. R., GOOD, J. M., MILLER, M. R., LUIKART, G. et HOHENLOHE, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2):81–92.
- ANGELONE, S. et HOLDEREGGER, R. (2009). Population genetics suggests effectiveness of habitat connectivity measures for the European tree frog in Switzerland. *Journal of Applied Ecology*, 46(4):879–887.

- ANGELONE, S., KIENAST, F. et HOLDEREGGER, R. (2011). Where movement happens - scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography*, 34(5):714–722.
- ARCHAUX, F., LOREL, C. et VILLEMÉY, A. (2018). Landscape drivers of butterfly and burnet moth diversity in lowland rural areas. *Landscape Ecology*, 33(10):1725–1739.
- ARNAUD, J.-F. (2003). Metapopulation genetic structure and migration pathways in the land snail *Helix aspersa* : influence of landscape heterogeneity. *Landscape Ecology*, 18(3):333–346.
- ARRHENIUS, O. (1921). Species and area. *Journal of Ecology*, 9(1):95–99.
- AUERBACH, D. M., DARROW, W. W., JAFFE, H. W. et CURRAN, J. W. (1984). Cluster of cases of the acquired immune deficiency syndrome : Patients linked by sexual contact. *The American journal of medicine*, 76(3):487–492.
- AWADE, M., BOSCOLO, D. et METZGER, J. P. (2012). Using binary and probabilistic habitat availability indices derived from graph theory to model bird occurrence in fragmented forests. *Landscape Ecology*, 27(2):185–198.
- BAGUETTE, M., BLANCHET, S., LEGRAND, D., STEVENS, V. M. et TURLURE, C. (2013). Individual dispersal, landscape connectivity and ecological networks. *Biological Reviews*, 88(2):310–326.
- BALBI, M., ERNOULT, A., POLI, P., MADEC, L., GUILLER, A., MARTIN, M.-C., NABUCET, J., BEAUJOUAN, V. et PETIT, E. J. (2018). Functional connectivity in replicated urban landscapes in the land snail (*Cornu aspersum*). *Molecular Ecology*, 27(6):1357–1370.
- BALBI, M., PETIT, E. J., CROCI, S., NABUCET, J., GEORGES, R., MADEC, L. et ERNOULT, A. (2019). Ecological relevance of least cost path analysis : An easy implementation method for landscape urban planning. *Journal of Environmental Management*, 244:61–68.
- BALDI, P., BRUNAK, S., CHAUVIN, Y., ANDERSEN, C. A. et NIELSEN, H. (2000). Assessing the accuracy of prediction algorithms for classification : an overview. *Bioinformatics*, 16(5):412–424.
- BALKENHOL, N., CUSHMAN, S., STORFER, A. et WAITS, L. (2016). *Landscape genetics : concepts, methods, applications*. John Wiley & Sons.
- BALKENHOL, N., DUDANIEC, R. Y., KRUTOVSKY, K. V., JOHNSON, J. S., CAIRNS, D. M., SEGELBACHER, G., SELKOE, K. A., von der HEYDEN, S., WANG, I. J., SELMONI, O. et al. (2017). Landscape genomics : Understanding relationships between environmental heterogeneity and genomic characteristics of populations. In RAJORA, O. P., éditeur : *Population Genomics : Concepts, Approaches and Applications*. Springer.
- BALKENHOL, N. et FORTIN, M.-J. (2016). Basics of study design : Sampling landscape heterogeneity and genetic variation for landscape genetic studies. In BALKENHOL, N., CUSHMAN, S., STORFER, A. et WAITS, L., éditeurs : *Landscape genetics : Concepts, methods, applications*, pages 58–76. John Wiley & Sons, 1 édition.
- BALKENHOL, N., GUGERLI, F., CUSHMAN, S. A., WAITS, L. P., COULON, A., ARNTZEN, J., HOLDEREGGER, R., WAGNER, H. H. et al. (2009a). Identifying future research needs in landscape genetics : where to from here ? *Landscape Ecology*, 24(4):455–463.
- BALKENHOL, N., PARDINI, R., CORNELIUS, C., FERNANDES, F. et SOMMER, S. (2013). Landscape-level comparison of genetic diversity and differentiation in a small mammal inhabiting different fragmented landscapes of the Brazilian Atlantic Forest. *Conservation Genetics*, 14(2):355–367.
- BALKENHOL, N., WAITS, L. P. et DEZZANI, R. J. (2009b). Statistical approaches in landscape genetics : an evaluation of methods for linking landscape and genetic data. *Ecography*, 32(5):818–830.
- BALL, I. R., POSSINGHAM, H. P. et WATTS, M. (2009). Marxan and relatives : software for spatial conservation prioritisation. *Spatial conservation prioritisation : Quantitative methods and computational tools*, 14:185–196.
- BANVILLE, F., VISSAULT, S. et POISOT, T. (2021). Mangal. jl and EcologicalNetworks. jl : Two complementary packages for analyzing ecological networks in Julia. *Journal of Open Source Software*, 6(61):2721.

- BARANYI, G., SAURA, S., PODANI, J. et JORDÁN, F. (2011). Contribution of habitat patches to network connectivity : redundancy and uniqueness of topological indices. *Ecological Indicators*, 11(5):1301–1310.
- BARR, K. R., KUS, B. E., PRESTON, K. L., HOWELL, S., PERKINS, E. et VANDERGAST, A. G. (2015). Habitat fragmentation in coastal southern California disrupts genetic connectivity in the Cactus Wren (*Campylorhynchus brunneicapillus*). *Molecular Ecology*, 24(10):2349–2363.
- BARRAT, A., BARTHELEMY, M., PASTOR-SATORRAS, R. et VESPIGNANI, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752.
- BASCOMPTE, J. (2007). Networks in ecology. *Basic and Applied Ecology*, 8(6):485–490.
- BASTIAN, O. (2001). Landscape ecology—towards a unified discipline? *Landscape Ecology*, 16(8):757–766.
- BEIER, P., MAJKA, D. R. et NEWELL, S. L. (2009). Uncertainty analysis of least-cost modeling for designing wildlife linkages. *Ecological Applications*, 19(8):2067–2077.
- BEIER, P., MAJKA, D. R. et SPENCER, W. D. (2008). Forks in the road : choices in procedures for designing wildland linkages. *Conservation Biology*, 22(4):836–851.
- BELLARD, C., BERTELSMEIER, C., LEADLEY, P., THULLER, W. et COURCHAMP, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15(4):365–377.
- BEN-HUR, E. et KADMON, R. (2020). An experimental test of the area–heterogeneity tradeoff. *Proceedings of the National Academy of Sciences*, 117(9):4815–4822.
- BENJAMINI, Y. et HOCHBERG, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 57(1):289–300.
- BENNETT, A. F. (1999). *Linkages in the landscape : the role of corridors and connectivity in wildlife conservation*. IUCN.
- BERGEROT, B., TOURNANT, P., MOUSSUS, J.-P., STEVENS, V.-M., JULLIARD, R., BAGUETTE, M. et FOLTÊTE, J.-C. (2013). Coupling inter-patch movement models and landscape graph to assess functional connectivity. *Population Ecology*, 55(1):193–203.
- BERGÈS, L., AVON, C., BEZOMBES, L., CLAUZEL, C., DUFLLOT, R., FOLTÊTE, J.-C., GAUCHERAND, S., GIRARDET, X. et SPIEGELBERGER, T. (2019). Intégrer la connectivité paysagère dans la séquence ERC : une approche par la quantité d’habitat atteignable. *VertigO-la revue électronique en sciences de l’environnement*, 19(2).
- BERGÈS, L., AVON, C., BEZOMBES, L., CLAUZEL, C., DUFLLOT, R., FOLTÊTE, J.-C., GAUCHERAND, S., GIRARDET, X. et SPIEGELBERGER, T. (2020). Environmental mitigation hierarchy and biodiversity offsets revisited through habitat connectivity modelling. *Journal of Environmental Management*, 256:1–10.
- BERGSTEN, A. et ZETTERBERG, A. (2013). To model the landscape as a network : A practitioner’s perspective. *Landscape and Urban Planning*, 119:35–43.
- BERTIN, A., GOUIN, N., BAUMEL, A., GIANOLI, E., SERRATOSA, J., OSORIO, R. et MANEL, S. (2017). Genetic variation of loci potentially under selection confounds species–genetic diversity correlations in a fragmented habitat. *Molecular Ecology*, 26(2):431–443.
- BEZANSON, J., EDELMAN, A., KARPINSKI, S. et SHAH, V. B. (2017). Julia : A fresh approach to numerical computing. *SIAM review*, 59(1):65–98.
- BODIN, Ö. et NORBERG, J. (2007). A network approach for analyzing spatially structured populations in fragmented landscape. *Landscape Ecology*, 22(1):31–44.
- BOITANI, L., FALCUCCI, A., MAIORANO, L. et RONDININI, C. (2007). Ecological networks as conceptual frameworks or operational tools in conservation. *Conservation Biology*, 21(6):1414–1422.
- BÖNSEL, A. B. et SONNECK, A.-G. (2011). Habitat use and dispersal characteristic by *Stethophyma grossum* : the role of habitat isolation and stable habitat conditions towards low dispersal. *Journal of Insect Conservation*, 15(3):455–463.

- BORCARD, D. et LEGENDRE, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153(1-2):51–68.
- BORDES, F., MORAND, S., PILOSOFF, S., CLAUDE, J., KRASNOV, B. R., COSSON, J.-F., CHAVAL, Y., RIBAS, A., CHAISIRI, K., BLASDELL, K. *et al.* (2015). Habitat fragmentation alters the properties of a host-parasite network : rodents and their helminths in south-east asia. *Journal of Animal Ecology*, 84(5):1253–1263.
- BOSSENBROEK, J. M., JOHNSON, L. E., PETERS, B. et LODGE, D. M. (2007). Forecasting the expansion of zebra mussels in the United States. *Conservation Biology*, 21(3):800–810.
- BOSSENBROEK, J. M., KRAFT, C. E. et NEKOLA, J. C. (2001). Prediction of long-distance dispersal using gravity models : zebra mussel invasion of inland lakes. *Ecological Applications*, 11(6):1778–1788.
- BOULANGER, E., DALONGEVILLE, A., ANDRELO, M., MOUILLOT, D. et MANEL, S. (2020). Spatial graphs highlight how multi-generational dispersal shapes landscape genetic patterns. *Ecography*, 15(1):1–13.
- BOURDOUXHE, A., DUFLLOT, R., RADOUX, J. et DUFRÊNE, M. (2020). Comparison of methods to model species habitat networks for decision-making in nature conservation : The case of the Wildcat in southern Belgium. *Journal for Nature Conservation*, 58:125901.
- BOURGEOIS, M. (2015). *Impacts écologiques des formes d'urbanisation : modélisations urbaines et paysagères*. Thèse de doctorat, Université de Franche-Comté.
- BOWCOCK, A. M., RUIZ-LINARES, A., TOMFOHRDE, J., MINCH, E., KIDD, J. R. et CAVALLI-SFORZA, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, 368(6470):455–457.
- BOWMAN, J., ADEY, E., ANGOH, S. Y., BAICI, J. E., BROWN, M. G., CORDES, C., DUPUIS, A. E., NEWAR, S. L., SCOTT, L. M. et SOLMUNDSON, K. (2020). Effects of cost surface uncertainty on current density estimates from circuit theory. *PeerJ*, 8:e9617.
- BOWMAN, J., GREENHORN, J. E., MARROTTE, R. R., MCKAY, M. M., MORRIS, K. Y., PRENTICE, M. B. et WEHTJE, M. (2016). On applications of landscape genetics. *Conservation Genetics*, 17(4):753–760.
- BOWNE, D. R. et BOWERS, M. A. (2004). Interpatch movements in spatially structured populations : a literature review. *Landscape Ecology*, 19(1):1–20.
- BOYCE, M. S. (1992). Population viability analysis. *Annual review of Ecology and Systematics*, 23(1):481–497.
- BRADBURD, G. S., COOP, G. M. et RALPH, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, 210(1):33–52.
- BRADBURD, G. S., RALPH, P. L. et COOP, G. M. (2016). A spatial framework for understanding population structure and admixture. *PLoS Genetics*, 12(1):e1005703.
- BRADBURY, I. R. et BENTZEN, P. (2007). Non-linear genetic isolation by distance : implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series*, 340:245–257.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J. et OLSHEN, R. A. (1984). *Classification and regression trees*. CRC press.
- BROOKS, C. (2003). A scalar analysis of landscape connectivity. *Oikos*, 102(2):433–439.
- BROOKS, C. P. (2006). Quantifying population substructure : extending the graph-theoretic approach. *Ecology*, 87(4):864–872.
- BROWN, J. H. et KODRIC-BROWN, A. (1977). Turnover rates in insular biogeography : effect of immigration on extinction. *Ecology*, 58(2):445–449.
- BRUGGEMAN, D. J., WIEGAND, T. et FERNANDEZ, N. (2010). The relative effects of habitat loss and fragmentation on population genetic variation in the red-cockaded woodpecker (*Picoides borealis*). *Molecular Ecology*, 19(17):3679–3691.
- BRUVO, R., MICHIELS, N. K., D'SOUZA, T. G. et SCHULENBURG, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, 13(7):2101–2106.

- BULLOCK, J. M., BONTE, D., PUFAL, G., da SILVA CARVALHO, C., CHAPMAN, D. S., GARCÍA, C., GARCÍA, D., MATTHYSEN, E. et DELGADO, M. M. (2018). Human-mediated dispersal and the rewiring of spatial networks. *Trends in Ecology & Evolution*, 33(12):958–970.
- BUNN, A., URBAN, D. et KEITT, T. (2000). Landscape connectivity : a conservation application of graph theory. *Journal of Environmental Management*, 59(4):265–278.
- BUREL, F. et BAUDRY, J. (2003). *Landscape ecology : concepts methods and applications*. Science Publishers.
- BURNHAM, K. P. et ANDERSON, D. R. (2004). Multimodel inference : understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2):261–304.
- CALABRESE, J. M. et FAGAN, W. F. (2004). A comparison-shopper’s guide to connectivity metrics. *Frontiers in Ecology and the Environment*, 2(10):529–536.
- CANTWELL, M. D. et FORMAN, R. T. (1993). Landscape graphs : ecological modeling with graph theory to detect configurations common to diverse landscapes. *Landscape Ecology*, 8(4):239–255.
- CAPURUCHO, J. M. G., CORNELIUS, C., BORGES, S. H., COHN-HAFT, M., ALEIXO, A., METZGER, J. P. et RIBAS, C. C. (2013). Combining phylogeography and landscape genetics of *Xenopipo atronitens* (Aves : Pipridae), a white sand campina specialist, to understand Pleistocene landscape evolution in Amazonia. *Biological Journal of the Linnean Society*, 110(1):60–76.
- CARRASCO, L., PAPEŞ, M., SHELDON, K. S. et GIAM, X. (2021). Global progress in incorporating climate adaptation into land protection for biodiversity since Aichi targets. *Global Change Biology*, 27(9):1788–1801.
- CARROLL, C., McRAE, B. et BROOKES, A. (2012). Use of linkage mapping and centrality analysis across habitat gradients to conserve connectivity of gray wolf populations in western North America. *Conservation Biology*, 26(1):78–87.
- CARROLL, K. A., INMAN, R. M., HANSEN, A. J., LAWRENCE, R. L. et BARNETT, K. (2021). A framework for collaborative wolverine connectivity conservation. *Isience*, 24(8):102840.
- CASTEL, T., LECOMTE, C., RICHARD, Y., LEJEUNE-HÉNAUT, I. et LARMURE, A. (2017). Frost stress evolution and winter pea ideotype in the context of climate warming at a regional scale. *OCL*, 24(1):D106.
- CASTILLO, J. A., EPPS, C. W., JEFFRESS, M. R., RAY, C., RODHOUSE, T. J. et SCHWALM, D. (2016). Replicated landscape genetic and network analyses reveal wide variation in functional connectivity for American pikas. *Ecological Applications*, 26(6):1660–1676.
- CASTRIC, V. et BERNATCHEZ, L. (2003). The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). *Genetics*, 163(3):983–996.
- CAYUELA, H., BESNARD, A., COTE, J., LAPORTE, M., BONNAIRE, E., PICHENOT, J., SHTICKZELLE, N., BELLEC, A., JOLY, P. et LÉNA, J.-p. (2020). Anthropogenic disturbance drives dispersal syndromes, demography, and gene flow in amphibian populations. *Ecological monographs*, 90(2):e01406.
- CBD (2010). Decision adopted by the conference of the parties to the convention on biological diversity at its tenth meeting. *In Conference of the Parties to the Convention on Biological Diversity*. UNEP/CBD/COP.
- CE (1992). Council Directive 92/43/EEC of the conservation of natural habitats and of wild fauna and flora. Rapport technique, European Council.
- CEBALLOS, G., EHRLICH, P. R., BARNOSKY, A. D., GARCÍA, A., PRINGLE, R. M. et PALMER, T. M. (2015). Accelerated modern human-induced species losses : Entering the sixth mass extinction. *Science Advances*, 1(5):e1400253.
- CHASSÉ, P., BLATRIX, C. et FRASCARIA-LACOSTE, N. (2020). What is wrong between ecological science and policy? *Ecology Letters*, 23(12):1736–1738.
- CHASSÉ, P., BLATRIX, C. et FRASCARIA-LACOSTE, N. (2021). Determining the location of protected areas in France : Does ‘scientific interest’ matter? *Perspectives in Ecology and Conservation*, 19(3):379–386.

- CHISHOLM, C., LINDO, Z. et GONZALEZ, A. (2011). Metacommunity diversity depends on connectivity and patch arrangement in heterogeneous habitat networks. *Ecography*, 34(3):415–424.
- CIOFI, C., BEAUMONT, M. A., SWINGLAND, I. R. et BRUFORD, M. W. (1999). Genetic divergence and units for conservation in the Komodo dragon *Varanus komodoensis*. *Proceedings of the Royal Society B*, 266(1435):2269–2274.
- CLARKE, R. T., ROTHERY, P. et RAYBOULD, A. F. (2002). Confidence limits for regression relationships between distance matrices : estimating gene flow with distance. *Journal of agricultural biological and environmental statistics*, 7(3):361–372.
- CLAUZEL, C. et BONNEVALLE, C. (2019). Apports de la modélisation spatiale pour la gestion de la Trame Verte et Bleue. *Cybergeo : European Journal of Geography*.
- CLAUZEL, C. et GODET, C. (2020). Combining spatial modeling tools and biological data for improved multispecies assessment in restoration areas. *Biological Conservation*, 250:1–15.
- CLEGG, S. M. et PHILLIMORE, A. B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of *Zosterops* in Vanuatu. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 365(1543):1077–1092.
- CLEVENGER, A. P., WIERZCHOWSKI, J., CHRUSZCZ, B. et GUNSON, K. (2002). GIS-generated, expert-based models for identifying wildlife habitat linkages and planning mitigation passages. *Conservation Biology*, 16(2):503–514.
- CLOBERT, J., BAGUETTE, M., BENTON, T. G. et BULLOCK, J. M. (2012). *Dispersal ecology and evolution*. Oxford University Press.
- CORNUET, J.-M., SANTOS, F., BEAUMONT, M. A., ROBERT, C. P., MARIN, J.-M., BALDING, D. J., GUILLEMAUD, T. et ESTOUP, A. (2008). Inferring population history with diy abc : a user-friendly approach to approximate bayesian computation. *Bioinformatics*, 24(23):2713–2719.
- CORREA AYRAM, C. A., MENDOZA, M. E., ETTER, A. et SALICRUP, D. R. P. (2016). Habitat connectivity in biodiversity conservation : a review of recent studies and applications. *Progress in Physical Geography*, 40(1):7–37.
- COSTANZA, R., D'ARGE, R., DE GROOT, R., FARBER, S., GRASSO, M., HANNON, B., LIMBURG, K., NAEEM, S., O'NEILL, R. V., PARUELO, J. et al. (1997). The value of the world's ecosystem services and natural capital. *nature*, 387(6630):253–260.
- COSTER, S. S., BABBITT, K. J., COOPER, A. et KOVACH, A. I. (2015). Limited influence of local and landscape factors on finescale gene flow in two pond-breeding amphibians. *Molecular Ecology*, 24(4):742–758.
- COULON, A., ABEN, J., PALMER, S., STEVENS, V., CALLENS, T., STRUBBE, D., LENS, L., MATTHYSEN, E., BAGUETTE, M. et TRAVIS, J. M. J. (2015). A stochastic movement simulator improves estimates of landscape connectivity. *Ecology*, 96(8):2203–2213.
- COULON, A., COSSON, J., ANGIBAULT, J., CARGNELUTTI, B., GALAN, M., MORELLET, N., PETIT, E., AULAGNIER, S. et HEWISON, A. (2004). Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape : an individual-based approach. *Molecular Ecology*, 13(9):2841–2850.
- CREECH, T. G., EPPS, C. W., MONELLO, R. J. et WEHAUSEN, J. D. (2014). Using network theory to prioritize management in a desert bighorn sheep metapopulation. *Landscape Ecology*, 29(4):605–619.
- CRISPO, E., MOORE, J.-S., LEE-YAW, J. A., GRAY, S. M. et HALLER, B. C. (2011). Broken barriers : human-induced changes to gene flow and introgression in animals : an examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays*, 33(7):508–518.
- CROOKS, K. R. et SANJAYAN, M. (2006). *Connectivity conservation*, volume 14. Cambridge University Press.
- CROSS, T. B., SCHWARTZ, M. K., NAUGLE, D. E., FEDY, B. C., ROW, J. R. et OYLER-McCANCE, S. J. (2018). The genetic network of greater sage-grouse : Range-wide identification of keystone hubs of connectivity. *Ecology and Evolution*, 8(11):1–19.

- CSARDI, G. et NEPUSZ, T. (2006). The igraph software package for complex network research. *International Journal of Complex Systems*, 1695(5):1–9.
- CURSON, J. (2014). *Plumbeous warbler (Dendroica plumbea)*. In del HOYO, J., ELLIOTT, A., SARGATAL, J., CHRISTIE, D. et de JUANA, E., éditeurs : *Handbook of the Birds of the World alive*. Barcelona - Lynx Edicions.
- CUSHMAN, S. A. et LEWIS, J. S. (2010). Movement behavior explains genetic differentiation in american black bears. *Landscape Ecology*, 25(10):1613–1625.
- CUSHMAN, S. A., MCKELVEY, K. S., HAYDEN, J. et SCHWARTZ, M. K. (2006). Gene flow in complex landscapes : testing multiple hypotheses with causal modeling. *The American Naturalist*, 168(4):486–499.
- CUSHMAN, S. A., SHIRK, A., HOWE, G. T., DYER, R. J., MURPHY, M. A. et JOOST, S. (2018). The least cost path from landscape genetics to landscape genomics : challenges and opportunities to explore NGS data in a spatially explicit context. *Frontiers in Genetics*, 9:215.
- CUSHMAN, S. A., SHIRK, A. et LANDGUTH, E. L. (2012). Separating the effects of habitat area, fragmentation and matrix resistance on genetic differentiation in complex landscapes. *Landscape Ecology*, 27(3):369–380.
- CUSHMAN, S. A., SHIRK, A. J. et LANDGUTH, E. L. (2013). Landscape genetics and limiting factors. *Conservation Genetics*, 14(2):263–274.
- DAIGLE, R. M., METAXAS, A., BALBAR, A. C., MCGOWAN, J., TREML, E. A., KUEMPEL, C. D., POSSINGHAM, H. P. et BEGER, M. (2020). Operationalizing ecological connectivity in spatial conservation planning with marxan connect. *Methods in Ecology and Evolution*, 11(4):570–579.
- DALE, M. (1977). Graph theoretical analysis of the phytosociological structure of plant communities : the theoretical basis. *Vegetatio*, 34(3):137–154.
- DALE, M. et FORTIN, M.-J. (2010). From graphs to spatial graphs. *Annual Review of Ecology, Evolution, and Systematics*, 41:21–38.
- DALE, M. R. (2017). *Applying Graph Theory in Ecological Research*. Cambridge University Press.
- DALE, M. R., DIXON, P., FORTIN, M.-J., LEGENDRE, P., MYERS, D. E. et ROSENBERG, M. S. (2002). Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25(5):558–577.
- DALE, M. R. et FORTIN, M.-J. (2014). *Spatial analysis : a guide for ecologists*. Cambridge University Press.
- DANON, L., DIAZ-GUILERA, A., DUCH, J. et ARENAS, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics : Theory and Experiment*, 2005(09):P09008.
- DARWIN, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. John Murray.
- DE CAMARGO, R. X., BOUCHER-LALONDE, V. et CURRIE, D. J. (2018). At the landscape level, birds respond strongly to habitat amount but weakly to fragmentation. *Diversity and Distributions*, 24(5):629–639.
- DECOCQ, G., GUILLER, A., KICHEY, T., Van de PITTE, K., GALLET-MORON, E., HONNAY, O. et CLOSSET-KOPP, D. (2021). Drivers of species and genetic diversity within forest metacommunities across agricultural landscapes of different permeability. *Landscape Ecology*, pages 1–18.
- DEN BOER, P. J. (1968). Spreading of risk and stabilization of animal numbers. *Acta biotheoretica*, 18(1-4):165–194.
- DIAMOND, J. M. (1975). The island dilemma : lessons of modern biogeographic studies for the design of natural reserves. *Biological Conservation*, 7(2):129–146.
- DÍAZ, S. M., SETTELE, J., BRONDÍZIO, E., NGO, H., GUÈZE, M., AGARD, J., ARNETH, A., BALVANERA, P., BRAUMAN, K., BUTCHART, S. et al. (2019). The global assessment report on biodiversity and ecosystem services : Summary for policy makers. Rapport technique, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

- DiBATTISTA, J. D. (2008). Patterns of genetic variation in anthropogenically impacted populations. *Conservation Genetics*, 9(1):141–156.
- DIDHAM, R. K., KAPOS, V. et EWERS, R. M. (2012). Rethinking the conceptual foundations of habitat fragmentation research. *Oikos*, 121(2):161–170.
- DIEKÖTTER, T., BILLETER, R. et CRIST, T. O. (2008). Effects of landscape connectivity on the spatial distribution of insect diversity in agricultural mosaic landscapes. *Basic and Applied Ecology*, 9(3):298–307.
- DIERINGER, D. et SCHLÖTTERER, C. (2003). Microsatellite analyser (MSA) : a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, 3(1):167–169.
- DIJKSTRA, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.
- DiLEO, M. F., SIU, J. C., RHODES, M. K., LÓPEZ-VILLALOBOS, A., REDWINE, A., KSIAZEK, K. et DYER, R. J. (2014). The gravity of pollination : integrating at-site features into spatial analysis of contemporary pollen movement. *Molecular Ecology*, 23(16):3973–3982.
- DiLEO, M. F. et WAGNER, H. H. (2016). A landscape ecologist’s agenda for landscape genetics. *Current Landscape Ecology Reports*, 1(3):115–126.
- DINIZ-FILHO, J. A. F., NABOUT, J. C., TELLES, M. P. d. C., SOARES, T. N. et RANGEL, T. F. L. (2009). A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genetics and molecular Biology*, 32(2):203–211.
- DIRZO, R. et RAVEN, P. H. (2003). Global state of biodiversity and loss. *Annual review of Environment and Resources*, 28(1):137–167.
- DIRZO, R., YOUNG, H. S., GALETTI, M., CEBALLOS, G., ISAAC, N. J. et COLLEN, B. (2014). Defaunation in the Anthropocene. *science*, 345(6195):401–406.
- DOAK, D. F., BAKKER, V. J., GOLDSTEIN, B. E. et HALE, B. (2015). What is the future of conservation ? *Protecting the wild*, pages 27–35.
- DOAK, D. F., MARINO, P. C. et KAREIVA, P. M. (1992). Spatial scale mediates the influence of habitat fragmentation on dispersal success : implications for conservation. *Theoretical population biology*, 41(3):315–336.
- DONADIEU, P. (2012). *Sciences du paysage - Entre théories et pratiques*. Lavoisier.
- DRAHEIM, H. M., MOORE, J. A., ETTER, D., WINTERSTEIN, S. R. et SCRIBNER, K. T. (2016). Detecting black bear source-sink dynamics using individual-based genetic graphs. *Proceedings of the Royal Society B*, 283(1835):1–9.
- DRAY, S. (2011). A New Perspective about Moran’s Coefficient : Spatial Autocorrelation as a Linear Regression Problem. *Geographical Analysis*, 43(2):127–141.
- DRAY, S., LEGENDRE, P. et PERES-NETO, P. R. (2006). Spatial modelling : a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196(3-4):483–493.
- DRIEZEN, K., ADRIAENSEN, F., RONDININI, C., DONCASTER, C. P. et MATTHYSEN, E. (2007). Evaluating least-cost model predictions with empirical dispersal data : a case-study using radiotracking data of hedgehogs (*Erinaceus europaeus*). *Ecological Modelling*, 209(2-4):314–322.
- DUFLOT, R., AVON, C., ROCHE, P. et BERGÈS, L. (2018). Combining habitat suitability models and spatial graphs for more effective landscape conservation planning : An applied methodological framework and a species case study. *Journal for Nature Conservation*, 46:38–47.
- DUNNING, J. B., DANIELSON, B. J. et PULLIAM, H. R. (1992). Ecological processes that affect populations in complex landscapes. *Oikos*, 65(1):169–175.
- DURAIAPPAH, A. K., NAEEM, S., AGARDY, T., ASH, N. J., COOPER, H. D., DIAZ, S., FAITH, D. P., MACE, G., McNEELY, J. A., MOONEY, H. A. et al. (2005). Ecosystems and human well-being : biodiversity synthesis ; a report of the millennium ecosystem assessment. Rapport technique, World Resources Institute.

- DYER, R. J. (2007). The evolution of genetic topologies. *Theoretical Population Biology*, 71(1):71–79.
- DYER, R. J. (2014). R package gstudio : analyses and functions related to the spatial analysis of genetic marker data. *R package version*, 1.
- DYER, R. J. (2015a). Is there such a thing as landscape genetics? *Molecular Ecology*, 24(14):3518–3528.
- DYER, R. J. (2015b). Population graphs and landscape genetics. *Annual Review of Ecology, Evolution, and Systematics*, 46:327–342.
- DYER, R. J. et NASON, J. D. (2004). Population graphs : the graph theoretic shape of genetic structure. *Molecular Ecology*, 13(7):1713–1727.
- DYER, R. J., NASON, J. D. et GARRICK, R. C. (2010). Landscape modelling of gene flow : improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, 19(17):3746–3759.
- ECONOMO, E. P. et KEITT, T. H. (2008). Species diversity in neutral metacommunities : a network approach. *Ecology Letters*, 11(1):52–62.
- ECONOMO, E. P. et KEITT, T. H. (2010). Network isolation and local diversity in neutral metacommunities. *Oikos*, 119(8):1355–1363.
- EDWARDS, L. J., MULLER, K. E., WOLFINGER, R. D., QAQISH, B. F. et SCHABENBERGER, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in medicine*, 27(29):6137–6157.
- ELTON, C. S. (1927). *Animal ecology*. Sedgwick and Jackson, London.
- EMARESI, G., PELLET, J., DUBEY, S., HIRZEL, A. H. et FUMAGALLI, L. (2011). Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. *Conservation Genetics*, 12(1):41–50.
- ÉRAUD, C., ARNOUX, É., LEVESQUE, A., VAN LAERE, G. et MAGNIN, H. (2012). Biologie des populations et statut de conservation des oiseaux endémiques des Antilles en Guadeloupe. Rapport d'étude, ONCFS-Parc National Guadeloupe.
- EVANS, J. S. et MURPHY, M. A. (2015). *GeNetIt*. R package version 0.1-0.
- EVERITT, B. et HOTHORN, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.
- EXCOFFIER, L., SMOUSE, P. E. et QUATTRO, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes : application to human mitochondrial DNA restriction data. *Genetics*, 131(2):479–491.
- FAGAN, W. et CALABRESE, J. (2006). Quantifying connectivity : balancing metric performance with data requirements. In CROOKS, K. et SANJAYAN, M., éditeurs : *Connectivity conservation*, pages 297–317. Cambridge University Press.
- FAHRIG, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 34(1):487–515.
- FAHRIG, L. (2013). Rethinking patch size and isolation effects : the habitat amount hypothesis. *Journal of Biogeography*, 40(9):1649–1663.
- FAHRIG, L. (2015). Just a hypothesis : a reply to Hanski. *Journal of Biogeography*, 42(5):993–994.
- FAHRIG, L. (2017). Ecological responses to habitat fragmentation per se. *Annual Review of Ecology, Evolution, and Systematics*, 48(1):1–23.
- FAHRIG, L. (2020). Why do several small patches hold more species than few large patches? *Global Ecology and Biogeography*, 29(4):615–628.
- FAHRIG, L. (2021). What the habitat amount hypothesis does and does not predict : A reply to Saura. *Journal of Biogeography*, 48(6):1530–1535.

- FAHRIG, L., BAUDRY, J., BROTONS, L., BUREL, F. G., CRIST, T. O., FULLER, R. J., SIRAMI, C., SIRIWARDENA, G. M. et MARTIN, J.-L. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. *Ecology Letters*, 14(2):101–112.
- FAHRIG, L. et PALOHEIMO, J. (1988). Determinants of local population size in patchy habitats. *Theoretical Population Biology*, 34(2):194–213.
- FAHRIG, L., WATLING, J. I., ARNILLAS, C. A., ARROYO-RODRÍGUEZ, V., JÖRGER-HICKFANG, T., MÜLLER, J., PEREIRA, H. M., RIVA, F., RÖSCH, V., SEIBOLD, S. *et al.* (2020). Resolving the sloss dilemma for biodiversity conservation : a research agenda. *Biological Reviews*, 00(1).
- FALL, A., FORTIN, M.-J., MANSEAU, M. et O'BRIEN, D. (2007). Spatial graphs : principles and applications for habitat connectivity. *Ecosystems*, 10(3):448–461.
- FARDILA, D., KELLY, L. T., MOORE, J. L. et MCCARTHY, M. A. (2017). A systematic review reveals changes in where and how we have studied habitat loss and fragmentation over 20 years. *Biological Conservation*, 212:130–138.
- FARINE, D. R. et WHITEHEAD, H. (2015). Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84(5):1144–1163.
- FAUBET, P. et GAGGIOTTI, O. E. (2008). A new Bayesian method to identify the environmental factors that influence recent migration. *Genetics*, 178(3):1491–1504.
- FERRARI, M. J., BJØRNSTAD, O. N., PARTAIN, J. L. et ANTONOVICS, J. (2006). A gravity model for the spread of a pollinator-borne plant pathogen. *The American Naturalist*, 168(3):294–303.
- FERRIER, S., MANION, G., ELITH, J. et RICHARDSON, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13(3):252–264.
- FISCHER, J. et LINDENMAYER, D. B. (2006). Beyond fragmentation : the continuum model for fauna research and conservation in human-modified landscapes. *Oikos*, 112(2):473–480.
- FLAVENOT, T., FELLOUS, S., ABDELKRIM, J., BAGUETTE, M. et COULON, A. (2015). Impact of quarrying on genetic diversity : an approach across landscapes and over time. *Conservation Genetics*, 16(1):181–194.
- FLETCHER, R. et FORTIN, M.-J. (2018). *Spatial ecology and conservation modeling*. Springer.
- FLETCHER, R. J., ACEVEDO, M. A., REICHERT, B. E., PIAS, K. E. et KITCHENS, W. M. (2011). Social network models predict movement and connectivity in ecological landscapes. *Proceedings of the National Academy of Sciences*, 108(48):19282–19287.
- FLETCHER, R. J., BURRELL, N. S., REICHERT, B. E., VASUDEV, D. et AUSTIN, J. D. (2016). Divergent perspectives on landscape connectivity reveal consistent effects from genes to communities. *Current Landscape Ecology Reports*, 1(2):67–79.
- FLETCHER, R. J., REVELL, A., REICHERT, B. E., KITCHENS, W. M., DIXON, J. D. et AUSTIN, J. D. (2013). Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature Communications*, 4(1):1–7.
- FLETCHER JR, R. J., DIDHAM, R. K., BANKS-LEITE, C., BARLOW, J., EWERS, R. M., ROSINDELL, J., HOLT, R. D., GONZALEZ, A., PARDINI, R., DAMSCHEN, E. I. *et al.* (2018). Is habitat fragmentation good for biodiversity? *Biological Conservation*, 226:9–15.
- FOLL, M. et GAGGIOTTI, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers : a Bayesian perspective. *Genetics*, 180(2):977–993.
- FOLTÊTE, J.-C. (2019). How ecological networks could benefit from landscape graphs : A response to the paper by Spartaco Gippoliti and Corrado Battisti. *Land Use Policy*, 80:391–394.
- FOLTÊTE, J.-C., CLAUZEL, C., GIRARDET, X., TOURNANT, P. et VUIDEL, G. (2012a). La modélisation des réseaux écologiques par les graphes paysagers. *Revue Internationale de Géomatique*, 22(4):641–658.

- FOLTÊTE, J.-C., CLAUZEL, C. et VUIDEL, G. (2012b). A software tool dedicated to the modelling of landscape networks. *Environmental Modelling & Software*, 38:316–327.
- FOLTÊTE, J.-C., CLAUZEL, C., VUIDEL, G. et TOURNANT, P. (2012c). Integrating graph-based connectivity metrics into species distribution models. *Landscape Ecology*, 27(4):557–569.
- FOLTÊTE, J.-C., GIRARDET, X. et CLAUZEL, C. (2014). A methodological framework for the use of landscape graphs in land-use planning. *Landscape and Urban Planning*, 124:140–150.
- FOLTÊTE, J.-C. et GIRAUDOUX, P. (2012). A graph-based approach to investigating the influence of the landscape on population spread processes. *Ecological Indicators*, 18:684–692.
- FOLTÊTE, J.-C., SAVARY, P., CLAUZEL, C., BOURGEOIS, M., GIRARDET, X., SAHRAOUI, Y., VUIDEL, G. et GARNIER, S. (2020). Coupling landscape graph modeling and biological data : a review. *Landscape Ecology*, 35(5):1035–1052.
- FOLTÊTE, J.-C. et VUIDEL, G. (2017). Using landscape graphs to delineate ecologically functional areas. *Landscape Ecology*, 32(2):249–263.
- FOLTÊTE, J.-C., VUIDEL, G., SAVARY, P., CLAUZEL, C., SAHRAOUI, Y., GIRARDET, X. et BOURGEOIS, M. (2021). Graphab : an application for modeling and managing ecological habitat networks. *Software Impacts*, 8:100065.
- FONTAINE, C., GUIMARÃES JR, P. R., KÉFI, S., LOEUILLE, N., MEMMOTT, J., van DER PUTTEN, W. H., VAN VEEN, F. J. et THÉBAULT, E. (2011). The ecological and evolutionary implications of merging different types of networks. *Ecology Letters*, 14(11):1170–1181.
- FORD, A. T., SUNTER, E. J., FAUVELLE, C., BRADSHAW, J. L., FORD, B., HUTCHEN, J., PHILLIPOW, N. et TEICHMAN, K. J. (2020). Effective corridor width : linking the spatial ecology of wildlife with land use policy. *European Journal of Wildlife Research*, 66(4):1–10.
- FORD, C. (2004). Nature, culture and conservation in France and her colonies 1840-1940. *Past & Present*, 183(1):173–198.
- FORMAN, R. T. (1995). Some general principles of landscape and regional ecology. *Landscape Ecology*, 10(3):133–142.
- FORMAN, R. T. et GODRON, M. (1986). *Landscape ecology*. John Wiley & Sons.
- FORMAN, R. T. T. (1990). The beginnings of landscape ecology in america. In ZONNEVELD, I. et FORMAN, R., éditeurs : *Changing landscapes : an ecological perspective*, pages 35–42. Springer.
- FORTIN, M.-J., JAMES, P. M., MACKENZIE, A., MELLES, S. J. et RAYFIELD, B. (2012). Spatial statistics - spatial regression, and graph theory in ecology. *Spatial Statistics*, 1:100–109.
- FORTUNA, M. A., ALBALADEJO, R. G., FERNÁNDEZ, L., APARICIO, A. et BASCOMPTE, J. (2009). Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, 106(45):19044–19049.
- FOTHERINGHAM, A. S. et O’KELLY, M. E. (1989). *Spatial interaction models : formulations and applications*, volume 1. Kluwer Academic Publishers Dordrecht.
- FRANKEL, O. H. et SOULÉ, M. E. (1986). *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.
- FRANKHAM, R. (1996a). Conservation genetics. *Annual Review of Genetics*, 29(1):305–327.
- FRANKHAM, R. (1996b). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10(6):1500–1508.
- FRANKHAM, R. (2006). Genetics and landscape connectivity. In CROOKS, K. R. et SANJAYAN, M., éditeurs : *Connectivity conservation*, chapitre 4, pages 72–96. Cambridge University Press.
- FRANKHAM, R. (2015). Genetic rescue of small inbred populations : Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, 24(11):2610–2618.
- FRANKHAM, R., BALLOU, J. D. et BRISCOE, D. A. (2004). *A primer of conservation genetics*. Cambridge University Press.

- FRANKHAM, R., BALLOU, J. D., RALLS, K., ELDRIDGE, M. D. B., DUDASH, M. R., FENSTER, C. B., LACY, R. C. et SUNNUCKS, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford University Press.
- FRANKHAM, R., BRADSHAW, C. J. et BROOK, B. W. (2014). Genetics in conservation management : revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, 170:56–63.
- FRANKLIN, I. R. (1980). Evolutionary change in small populations. In SOULÉ, M. et WILCOX, B., éditeurs : *Conservation biology : an evolutionary-ecological perspective*, volume 395, pages 135–149. Sunderland, MA : Sinauer Associates.
- FRUCHTERMAN, T. M. et REINGOLD, E. M. (1991). Graph drawing by force-directed placement. *Software : Practice and experience*, 21(11):1129–1164.
- GABRIEL, K. R. et SOKAL, R. R. (1969). A new statistical approach to geographic variation analysis. *Systematic zoology*, 18(3):259–278.
- GALPERN, P., MANSEAU, M. et FALL, A. (2011). Patch-based graphs of landscape connectivity : a guide to construction, analysis and application for conservation. *Biological Conservation*, 144(1):44–55.
- GALPERN, P., MANSEAU, M. et WILSON, P. (2012). Grains of connectivity : analysis at multiple spatial scales in landscape genetics. *Molecular Ecology*, 21(16):3996–4009.
- GALPERN, P., PERES-NETO, P. R., POLFUS, J. et MANSEAU, M. (2014). MEMGENE : Spatial pattern detection in genetic distance data. *Methods in Ecology and Evolution*, 5(10):1116–1120.
- GARNER, A., RACHLOW, J. L. et HICKS, J. F. (2005). Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology*, 19(4):1215–1221.
- GARNETT, S. et LINDENMAYER, D. (2011). Conservation science must engender hope to succeed. *Trends in Ecology & Evolution*, 26(2):59–59.
- GARNIER, S., ALIBERT, P., AUDIOT, P., PRIEUR, B. et RASPLUS, J.-Y. (2004). Isolation by distance and sharp discontinuities in gene frequencies : implications for the phylogeography of an alpine insect species, *Carabus solieri*. *Molecular Ecology*, 13(7):1883–1897.
- GARROWAY, C. J., BOWMAN, J., CARR, D. et WILSON, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, 1(4):620–630.
- GARROWAY, C. J., BOWMAN, J. et WILSON, P. J. (2011). Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology*, 20(19):3978–3988.
- GASTON, K. J. et FULLER, R. A. (2008). Commonness, population depletion and conservation biology. *Trends in Ecology & Evolution*, 23(1):14–19.
- GAUFFRE, B., MALLEZ, S., CHAPUIS, M.-P., LEBLOIS, R., LITRICO, I., DELAUNAY, S. et BADENHAUSSER, I. (2015). Spatial heterogeneity in landscape structure influences dispersal and genetic structure : empirical evidence from a grasshopper in an agricultural landscape. *Molecular Ecology*, 24(8):1713–1728.
- GILBERT-NORTON, L., WILSON, R., STEVENS, J. R. et BEARD, K. H. (2010). A meta-analytic review of corridor effectiveness. *Conservation Biology*, 24(3):660–668.
- GILLESPIE, J. H. (2010). *Population genetics : a concise guide*. JHU Press.
- GILPIN, M. (1991). The genetic effective size of a metapopulation. *Biological Journal of the Linnean Society*, 42(1-2):165–175.
- GILPIN, M. et SOULÉ, M. E. (1986). Minimum viable populations : processes of extinction. In SOULÉ, M. E., éditeur : *Conservation biology : the science of scarcity and diversity*, pages 19–34. Sinauer, Sunderland.
- GINI, C. (1912). Variabilità e mutabilità. *Memorie di metodologica statistica*, 10.
- GIPPOLITI, S. et BATTISTI, C. (2017). More cool than tool : Equivoques, conceptual traps and weaknesses of ecological networks in environmental planning and conservation. *Land Use Policy*, 68:686–691.

- GIRARDET, X., FOLTÊTE, J.-C. et CLAUZEL, C. (2013). Designing a graph-based approach to landscape ecological assessment of linear infrastructures. *Environmental Impact Assessment Review*, 42:10–17.
- GODET, C. et CLAUZEL, C. (2021). Comparison of landscape graph modelling methods for analysing pond network connectivity. *Landscape Ecology*, 36(3):735–748.
- GODET, L. (2010). La "nature ordinaire" dans le monde occidental. *L'Espace géographique*, 39(4):295–308.
- GODET, L. et DEVICTOR, V. (2018). What conservation does. *Trends in Ecology & Evolution*, 33(10):720–730.
- GONZALES, E. K. et GERGEL, S. E. (2007). Testing assumptions of cost surface analysis - a tool for invasive species management. *Landscape Ecology*, 22(8):1155–1168.
- GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338.
- GOWER, J. C. et LEGENDRE, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of classification*, 3(1):5–48.
- GRAVES, T. A., BEIER, P. et ROYLE, J. A. (2013). Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, 22(15):3888–3903.
- GRAVES, T. A., WASSERMAN, T. N., RIBEIRO, M. C., LANDGUTH, E. L., SPEAR, S. F., BALKENHOL, N., HIGGINS, C. B., FORTIN, M.-J., CUSHMAN, S. A. et WAITS, L. P. (2012). The influence of landscape characteristics and home-range size on the quantification of landscape-genetics relationships. *Landscape Ecology*, 27(2):253–266.
- GREENBAUM, G. et FEFFERMAN, N. H. (2017). Application of network methods for understanding evolutionary dynamics in discrete habitats. *Molecular Ecology*, 26(11):2850–2863.
- GREENBAUM, G., TEMPLETON, A. R. et BAR-DAVID, S. (2016). Inference and analysis of population structure using genetic data and network theory. *Genetics*, 202(4):1299–1312.
- GRIFFIOEN, R. (1996). Over het dispersievermogen van de moerassprinkhaan. *Nieuwsbrief Saltabel*, 15(1):39–41.
- GRINNELL, J. (1917). The niche-relationships of the California Thrasher. *The Auk*, 34(4):427–433.
- GUILLERA-ARROITA, G., LAHOZ-MONFORT, J. J. et ELITH, J. (2014). Maxent is not a presence-absence method : a comment on Thibaud et al. *Methods in Ecology and Evolution*, 5(11):1192–1197.
- GUILLOT, G., LEBLOIS, R., COULON, A. et FRANTZ, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology*, 18(23):4734–4756.
- GUISAN, A. et ZIMMERMANN, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3):147–186.
- GURRUTXAGA, M., LOZANO, P. J. et del BARRIO, G. (2010). GIS-based approach for incorporating the connectivity of ecological networks into regional planning. *Journal for Nature Conservation*, 18(4):318–326.
- HADDAD, N. M., BRUDVIG, L. A., CLOBERT, J., DAVIES, K. F., GONZALEZ, A., HOLT, R. D., LOVEJOY, T. E., SEXTON, J. O., AUSTIN, M. P., COLLINS, C. D. et al. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1(2):e1500052.
- HADDAD, N. M., GONZALEZ, A., BRUDVIG, L. A., BURT, M. A., LEVEY, D. J. et DAMSCHEN, E. I. (2017a). Experimental evidence does not support the Habitat Amount Hypothesis. *Ecography*, 40(1):48–55.
- HADDAD, N. M., HOLT, R. D., FLETCHER JR, R. J., LOREAU, M. et CLOBERT, J. (2017b). Connecting models, data, and concepts to understand fragmentation's ecosystem-wide effects. *Ecography*, 40(1):1–8.
- HADLEY, A. S. et BETTS, M. G. (2016). Refocusing habitat fragmentation research using lessons from the last decade. *Current Landscape Ecology Reports*, 1(2):55–66.

- HAHN, T., KETTLE, C. J., GHAZOUL, J., HENNIG, E. I. et PLUESS, A. R. (2013). Landscape composition has limited impact on local genetic structure in mountain clover, *Trifolium montanum* L. *Journal of Heredity*, 104(6):842–852.
- HAILA, Y. (2002). A conceptual genealogy of fragmentation research : from island biogeography to landscape ecology. *Ecological Applications*, 12(2):321–334.
- HALLEY, J. M., MONOKROUSOS, N., MAZARIS, A. D., NEWMARK, W. D. et VOKOU, D. (2016). Dynamics of extinction debt across five taxonomic groups. *Nature Communications*, 7(1):1–6.
- HAMILTON, M. (2011). *Population genetics*. John Wiley & Sons.
- HÄNFLING, B. et WEETMAN, D. (2006). Concordant genetic estimators of migration reveal anthropogenically enhanced source-sink population structure in the river sculpin *Cottus gobio*. *Genetics*, 173(3):1487–1501.
- HANSKI, I. (1989). Metapopulation dynamics : does it help to have more of the same? *Trends in Ecology & Evolution*, 4(4):113–114.
- HANSKI, I. (1994). A practical model of metapopulation dynamics. *Journal of Animal Ecology*, pages 151–162.
- HANSKI, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41–49.
- HANSKI, I. (2005). Landscape fragmentation, biodiversity loss and the societal response : The longterm consequences of our use of natural resources may be surprising and unpleasant. *EMBO reports*, 6(5):388–392.
- HANSKI, I. (2015). Habitat fragmentation and species richness. *Journal of Biogeography*, 42(5):989–993.
- HANSKI, I. et GILPIN, M. (1991). Metapopulation dynamics : brief history and conceptual domain. *Biological journal of the Linnean Society*, 42(1-2):3–16.
- HANSKI, I. et OVASKAINEN, O. (2002). Extinction debt at extinction threshold. *Conservation Biology*, 16(3):666–673.
- HARARY, F. (1969). *Graph theory*, volume 1. Addison-Wesley, Reading, MA.
- HARDY, O. J. et VEKEMANS, X. (1999). Isolation by distance in a continuous population : reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, 83(2):145.
- HARPER, J. L. et HAWKSWORTH, D. L. (1994). Biodiversity : measurement and estimation. preface. *Proceedings of the Royal Society B*, 345:5–12.
- HARTL, D. L., CLARK, A. G. et CLARK, A. G. (1997). *Principles of population genetics*, volume 116. Sinauer associates Sunderland, MA.
- HEDRICK, P. W. (2005). A standardized genetic differentiation measure. *Evolution*, 59(8):1633–1638.
- HEIN, C., ABDEL MONIEM, H. E. et WAGNER, H. H. (2021). Can We Compare Effect Size of Spatial Genetic Structure Between Studies and Species Using Moran Eigenvector Maps? *Frontiers in Ecology and Evolution*, 9:107.
- HERNANDEZ, P. A., GRAHAM, C. H., MASTER, L. L. et ALBERT, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29(5):773–785.
- HERRERA-ARROYO, M. L., SORK, V. L., GONZÁLEZ-RODRÍGUEZ, A., ROCHA-RAMÍREZ, V., VEGA, E. et OYAMA, K. (2013). Seed-mediated connectivity among fragmented populations of *Quercus castanea* (Fagaceae) in a Mexican landscape. *American Journal of Botany*, 100(8):1663–1671.
- HOBAN, S., BERTORELLE, G. et GAGGIOTTI, O. E. (2012). Computer simulations : tools for population and evolutionary genetics. *Nature Reviews Genetics*, 13(2):110.
- HOBAN, S., BRUFORD, M., JACKSON, J. D., LOPES-FERNANDES, M., HEUERTZ, M., HOHENLOHE, P. A., PAZ-VINAS, I., SJÖGREN-GULVE, P., SEGELBACHER, G., VERNESI, C. et al. (2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, 248:108654.
- HOBBS, R. J. (1992). The role of corridors in conservation : solution or bandwagon? *Trends in Ecology & Evolution*, 7(11):389–392.

- HOFMAN, M. P., HAYWARD, M. W., KELLY, M. J. et BALKENHOL, N. (2018). Enhancing conservation network design with graph-theory and a measure of protected area effectiveness : Refining wildlife corridors in Belize, Central America. *Landscape and Urban Planning*, 178:51–59.
- HOLDEREGGER, R. et GUGERLI, F. (2012). Where do you come from, where do you go? directional migration rates in landscape genetics. *Molecular Ecology*, 21(23):5640–5642.
- HOLLAND, J. D. et YANG, S. (2016). Multi-scale studies and the ecological neighborhood. *Current Landscape Ecology Reports*, 1(4):135–145.
- HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70.
- HOLZHAUER, S. I., EKSCHMITT, K., SANDER, A.-C., DAUBER, J. et WOLTERS, V. (2006). Effect of historic landscape change on the genetic structure of the bush-cricket *Metrioptera roeseli*. *Landscape Ecology*, 21(6):891–899.
- HUBERT, L. et ARABIE, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193–218.
- HUGHES, J. B., DAILY, G. C. et EHRLICH, P. R. (1997). Population diversity : its extent and extinction. *Science*, 278(5338):689–692.
- HUNTER JR, M. L. et GIBBS, J. P. (2006). *Fundamentals of conservation biology*. John Wiley & Sons.
- HUTCHINSON, G. E. (1957). Concluding remarks. Population studies : animal ecology and demography. *Cold Spring Harb Symp Quant Biol*, 22:415–427.
- HUTCHINSON, M. C., BRAMON MORA, B., PILOSOF, S., BARNER, A. K., KÉFI, S., THÉBAULT, E., JORDANO, P. et STOUFFER, D. B. (2019). Seeing the forest for the trees : Putting multilayer networks to work for community ecology. *Functional Ecology*, 33(2):206–217.
- HUTCHISON, D. W. et TEMPLETON, A. R. (1999). Correlation of pairwise genetic and geographic distance measures : inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, 53(6):1898–1914.
- INGLADA, J., VINCENT, A., ARIAS, M., TARDY, B., MORIN, D. et RODES, I. (2017). Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, 9(1):95.
- INGVARSSON, P. K. (2001). Restoration of genetic variation lost—the genetic rescue hypothesis. *Trends in Ecology & Evolution*, 16(2):62–63.
- JACKSON, H. B. et FAHRIG, L. (2012). What size is a biologically relevant landscape? *Landscape Ecology*, 27(7):929–941.
- JACKSON, H. B. et FAHRIG, L. (2015a). Are ecologists conducting research at the optimal scale? *Global Ecology and Biogeography*, 24(1):52–63.
- JACKSON, N. D. et FAHRIG, L. (2015b). Habitat amount - not habitat configuration - best predicts population genetic structure in fragmented landscapes. *Landscape Ecology*, 31(5):951–968.
- JACOBS, J. (1974). Quantitative measurement of food selection. *Oecologia*, 14(4):413–417.
- JAQUIÉRY, J., BROQUET, T., HIRZEL, A. H., YEARSLEY, J. et PERRIN, N. (2011). Inferring landscape effects on dispersal from genetic distances : how far can we go? *Molecular Ecology*, 20(4):692–705.
- JOLY, D., BOIS, B. et ZAKŠEK, K. (2012). Rank-ordering of topographic variables correlated with temperature. *Atmospheric and Climate Sciences*, 2(2):139–147.
- JOMBART, T. (2008). adegenet : a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405.
- JOST, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, 17(18):4015–4026.
- JULES, E. S., KAUFFMAN, M. J., RITTS, W. D. et CARROLL, A. L. (2002). Spread of an invasive pathogen over a variable landscape : a nonnative root rot on Port Orford cedar. *Ecology*, 83(11):3167–3181.

- KADOYA, T. (2009). Assessing functional connectivity using empirical data. *Population ecology*, 51(1):5–15.
- KAMADA, T., KAWAI, S. *et al.* (1989). An algorithm for drawing general undirected graphs. *Information processing letters*, 31(1):7–15.
- KEELEY, A. T., BASSON, G., CAMERON, D. R., HELLER, N. E., HUBER, P. R., SCHLOSS, C. A., THORNE, J. H. *et MERENLENDER, A. M.* (2018). Making habitat connectivity a reality. *Conservation Biology*, 32(6):1221–1232.
- KEELEY, A. T., BEIER, P. *et* GAGNON, J. W. (2016). Estimating landscape resistance from habitat suitability : effects of data source and nonlinearities. *Landscape Ecology*, 31(9):2151–2162.
- KEELEY, A. T., BEIER, P., KEELEY, B. W. *et* FAGAN, M. E. (2017). Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements. *Landscape and Urban Planning*, 161:90–102.
- KEIL, J. M. *et* GUTWIN, C. A. (1992). Classes of graphs which approximate the complete Euclidean graph. *Discrete & Computational Geometry*, 7(1):13–28.
- KEITT, T., URBAN, D. *et* MILNE, B. (1997). Detecting critical scales in fragmented landscapes. *Conservation Ecology*, 1(1):1–17.
- KELLER, D., HOLDEREGGER, R. *et* STRIEN, M. J. (2013). Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology*, 22(9):2467–2482.
- KELLER, D., HOLDEREGGER, R., *van* STRIEN, M. J. *et* BOLLIGER, J. (2015). How to make landscape genetics beneficial for conservation management? *Conservation Genetics*, 16(3):503–512.
- KEYGHOBADI, N. (2007). The genetic implications of habitat fragmentation for animals. *Canadian Journal of Zoology*, 85(10):1049–1064.
- KEYGHOBADI, N., ROLAND, J., MATTER, S. F. *et* STROBECK, C. (2005). Among- and within-patch components of genetic diversity respond at different rates to habitat fragmentation : an empirical demonstration. *Proceedings of the Royal Society B*, 272(1562):553–560.
- KHIMOUN, A., ARNOUX, E., MARTEL, G., POT, A., ERAUD, C., CONDÉ, B., LOUBON, M., THÉRON, F., COVAS, R., FAIVRE, B. *et al.* (2016a). Contrasted patterns of genetic differentiation across eight bird species in the Lesser Antilles. *Genetica*, 144(1):125–138.
- KHIMOUN, A., ERAUD, C., OLLIVIER, A., ARNOUX, E., ROCHETEAU, V., BELY, M., LEFOL, E., DELPUECH, M., CARPENTIER, M.-L., LEBLOND, G. *et al.* (2016b). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, 25(16):3831–3844.
- KHIMOUN, A., PETERMAN, W., ERAUD, C., FAIVRE, B., NAVARRO, N. *et* GARNIER, S. (2017). Landscape genetic analyses reveal fine-scale effects of forest fragmentation in an insular tropical bird. *Molecular Ecology*, 26(19):4906–4919.
- KIEREPKA, E. M., ANDERSON, S. J., SWIHART, R. K. *et* RHODES, O. E. (2020). Differing, multiscale landscape effects on genetic diversity and differentiation in eastern chipmunks. *Heredity*, 124(3):457–468.
- KIMURA, M. (1953). 'Stepping Stone' model of population. *Annual Report of the National Institute of Genetics Japan*, 3:62–63.
- KIMURA, M. *et al.* (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626.
- KIMURA, M. *et* WEISS, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561–576.
- KINDLMANN, P. *et* BUREL, F. (2008). Connectivity measures : a review. *Landscape Ecology*, 23(8):879–890.
- KININMONTH, S., *van* OPPEN, M. J. *et* POSSINGHAM, H. P. (2010). Determining the community structure of the coral *Seriatopora hystrix* from hydrodynamic and genetic networks. *Ecological Modelling*, 221(24):2870–2880.
- KOEN, E. L., BOWMAN, J., GARROWAY, C. J., MILLS, S. C. *et* WILSON, P. J. (2012). Landscape resistance and American marten gene flow. *Landscape Ecology*, 27(1):29–43.

- KOEN, E. L., BOWMAN, J., GARROWAY, C. J. et WILSON, P. J. (2013). The sensitivity of genetic connectivity measures to unsampled and under-sampled sites. *PLoS ONE*, 8(2):e56204.
- KOEN, E. L., BOWMAN, J. et WILSON, P. J. (2016). Node-based measures of connectivity in genetic networks. *Molecular Ecology Resources*, 16(1):69–79.
- KOEN, E. L., GARROWAY, C. J., WILSON, P. J. et BOWMAN, J. (2010). The effect of map boundary on estimates of landscape resistance to animal movement. *PLoS ONE*, 5(7):e11785.
- KOH, I., ROWE, H. I. et HOLLAND, J. D. (2013). Graph and circuit theory connectivity models of conservation biological control agents. *Ecological Applications*, 23(7):1554–1573.
- KOLACZYK, E. D. et CSÁRDI, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer.
- KONG, F., YIN, H., NAKAGOSHI, N. et ZONG, Y. (2010). Urban green space network development for biodiversity conservation : Identification based on graph theory and gravity modeling. *Landscape and Urban Planning*, 95(1-2):16–27.
- KOOL, J. T., MOILANEN, A. et TREML, E. A. (2013). Population connectivity : recent advances and new perspectives. *Landscape Ecology*, 28(2):165–185.
- KUEHN, R., SCHROEDER, W., PIRCHNER, F. et ROTTMANN, O. (2003). Genetic diversity - gene flow and drift in Bavarian red deer populations (*Cervus elaphus*). *Conservation Genetics*, 4(2):157–166.
- KUISMIN, M., SAATOGLU, D., NISKANEN, A. K., JENSEN, H. et SILLANPÄÄ, M. J. (2020). Genetic assignment of individuals to source populations using network estimation tools. *Methods in Ecology and Evolution*, 11(2):333–344.
- KUISMIN, M. O. et SILLANPÄÄ, M. J. (2017). Estimation of covariance and precision matrix, network structure, and a view toward systems biology. *Computational Statistics*, 9(6).
- LAFFERTY, K. D., DOBSON, A. P. et KURIS, A. M. (2006). Parasites dominate food web links. *Proceedings of the National Academy of Sciences*, 103(30):11211–11216.
- LAGESSE, C., BORDIN, P. et DOUADY, S. (2015). A spatial multi-scale object to analyze road networks. *Network Science*, 3(1):156–181.
- LAITA, A., KOTIAHO, J. et MÖNKKÖNEN, M. (2011). Graph-theoretic connectivity measures : what do they tell us about connectivity? *Landscape Ecology*, 26(7):951–967.
- LALIBERTÉ, J. et ST-LAURENT, M.-H. (2020). Validation of functional connectivity modeling : The Achilles' heel of landscape connectivity mapping. *Landscape and Urban Planning*, 202:103878.
- LANDGUTH, E., CUSHMAN, S., SCHWARTZ, M., MCKELVEY, K., MURPHY, M. et LUIKART, G. (2010). Quantifying the lag time to detect barriers in landscape genetics. *Molecular Ecology*, 19(19):4179–4191.
- LANDGUTH, E., CUSHMAN, S. A. et BALKENHOL, N. (2016). Simulation modeling in landscape genetics. In BALKENHOL, N., CUSHMAN, S., STORFER, A. et WAITS, L., éditeurs : *Landscape genetics : Concepts, methods, applications*, pages 101–113. John Wiley & Sons, 1 édition.
- LANDGUTH, E. L. et CUSHMAN, S. (2010). CDPOP : a spatially explicit cost distance population genetics program. *Molecular Ecology Resources*, 10(1):156–161.
- LAROCHE, F., BALBI, M., GRÉBERT, T., JABOT, F. et ARCHAUX, F. (2020). Three points of consideration before testing the effect of patch connectivity on local species richness : patch delineation, scaling and variability of metrics. *bioRxiv*, 640995, ver. 5 peer-reviewed and recommended by PCI Ecology, 1:1–21.
- LATOUCHE, P., BIRMELÉ, E., AMBROISE, C. et al. (2011). Overlapping stochastic block models with application to the french political blogosphere. *The Annals of Applied Statistics*, 5(1):309–336.
- LATTA, R. G. (2006). Integrating patterns across multiple genetic markers to infer spatial processes. *Landscape Ecology*, 21(6):809–820.

- LEADER-WILLIAMS, N., HARRISON, J. et GREEN, M. (1990). Designing protected areas to conserve natural resources. *Science Progress (1933-)*, pages 189–204.
- LEBLOND, G. (2008). Étude sur les structures de peuplement de l’avifaune du massif forestier du Parc national de Guadeloupe. Rapport technique, Parc National de Guadeloupe.
- LEGENDRE, P. et FORTIN, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, 10(5):831–844.
- LEGENDRE, P. et LEGENDRE, L. F. (2012). *Numerical ecology*. Elsevier.
- LEHNEN, L., JAN, P.-L., BESNARD, A.-L., FOURCY, D., KERTH, G., BIEDERMANN, M., NYSSSEN, P., SCHORCHT, W., PETIT, E. et PUECHMAILLE, S. (2021). Genetic diversity in a long-lived mammal is explained by the past’s demographic shadow and current connectivity. *Molecular Ecology*, 00(1).
- LENORMAND, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, 17(4):183–189.
- LEROUX, S. J., ALBERT, C. H., LAFUITE, A.-S., RAYFIELD, B., WANG, S. et GRAVEL, D. (2017). Structural uncertainty in models projecting the consequences of habitat loss and fragmentation on biodiversity. *Ecography*, 40(1):36–47.
- LEUNG, B., HARGREAVES, A. L., GREENBERG, D. A., MCGILL, B., DORNELAS, M. et FREEMAN, R. (2020). Clustered versus catastrophic global vertebrate declines. *Nature*, 588(7837):267–271.
- LEVINS, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, 15(3):237–240.
- LEVINS, R. (1970). Extinction. *Some mathematical questions in biology*.
- LINDENMAYER, D. B., BLANCHARD, W., FOSTER, C. N., SCHEELE, B. C., WESTGATE, M. J., STEIN, J., CRANE, M. et FLORANCE, D. (2020). Habitat amount versus connectivity : An empirical study of bird responses. *Biological Conservation*, 241:108377.
- LINDGREN, J. P. et COUSINS, S. A. (2017). Island biogeography theory outweighs habitat amount hypothesis in predicting plant species richness in small grassland remnants. *Landscape Ecology*, 32(9):1895–1906.
- LONG, F. H. (2013). Multivariate analysis for metabolomics and proteomics data. *In Proteomic and Metabolomic Approaches to Biomarker Discovery*, pages 299–311. Elsevier.
- LOOKINGBILL, T. R., GARDNER, R. H., FERRARI, J. R. et KELLER, C. E. (2010). Combining a dispersal model with network theory to assess habitat connectivity. *Ecological Applications*, 20(2):427–441.
- LOVETTE, I. J., BERMINGHAM, E., SEUTIN, G. et RICKLEFS, R. E. (1998). Evolutionary differentiation in three endemic West Indian warblers. *The Auk*, 115(4):890–903.
- LUQUE, S., SAURA, S. et FORTIN, M.-J. (2012). Landscape connectivity analysis for conservation : insights from combining new methods with ecological and genetic data. *Landscape Ecology*, 27(2):153–157.
- LUXIMON, N., PETIT, E. J. et BROQUET, T. (2014). Performance of individual vs. group sampling for inferring dispersal under isolation-by-distance. *Molecular Ecology Resources*, 14(4):745–752.
- MACARTHUR, R. et WILSON, E. (1967). *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- MAGWENE, P. M. (2001). New tools for studying integration and modularity. *Evolution*, 55(9):1734–1745.
- MAGWENE, P. M. et KIM, J. (2004). Estimating genomic coexpression networks using first-order conditional independence. *Genome biology*, 5(12):1–16.
- MAIR, L., MILL, A. C., ROBERTSON, P. A., RUSHTON, S. P., SHIRLEY, M. D., RODRIGUEZ, J. P. et MCGOWAN, P. J. (2018). The contribution of scientific research to conservation planning. *Biological Conservation*, 223:82–96.
- MALÉCOT, G. (1948). *Les mathématiques de l’hérédité*. Masson, Paris.

- MANEL, S., BERTHOUD, F., BELLEMMAIN, E., GAUDEUL, M., LUIKART, G., SWENSON, J., WAITS, L., TABERLET, P. et CONSORTIUM, I. (2007). A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology*, 16(10):2031–2043.
- MANEL, S. et HOLDEREGGER, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, 28(10):614–621.
- MANEL, S., SCHWARTZ, M. K., LUIKART, G. et TABERLET, P. (2003). Landscape genetics : combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18(4):189–197.
- MANNI, F., GUERARD, E. et HEYER, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation : how barriers can be detected by using Monmonier’s algorithm. *Human biology*, 76(2):173–190.
- MANTEL, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2):209–220.
- MARIS, V. (2014). *Nature à vendre : Les limites des services écosystémiques*. Éditions Quae.
- MARKO, P. B. et HART, M. W. (2011). The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution*, 26(9):448–456.
- MARROTTE, R. R. et BOWMAN, J. (2017). The relationship between least-cost and resistance distance. *PLoS ONE*, 12(3):e0174212.
- MARTENSEN, A. C., SAURA, S. et FORTIN, M.-J. (2017). Spatio-temporal connectivity : assessing the amount of reachable habitat in dynamic landscapes. *Methods in Ecology and Evolution*, 8(10):1253–1264.
- MARTIN, A. E. (2018a). The spatial scale of a species’ response to the landscape context depends on which biological response you measure. *Current Landscape Ecology Reports*, 3(1):23–33.
- MARTIN, C. A. (2018b). An early synthesis of the habitat amount hypothesis. *Landscape Ecology*, pages 1–5.
- MARTÍN-QUELLER, E. et SAURA, S. (2013). Landscape species pools and connectivity patterns influence tree species richness in both managed and unmanaged stands. *Forest Ecology and Management*, 289:123–132.
- MARZELLI, M. (1994). Ausbreitung von *mecostethus grossus* auf einer ausgleichs-und renaturierungsfläche. *Articulata*, 9(1):25–32.
- MATEO-SÁNCHEZ, M. C., BALKENHOL, N., CUSHMAN, S., PÉREZ, T., DOMÍNGUEZ, A. et SAURA, S. (2015). Estimating effective landscape distances and movement corridors : comparison of habitat and genetic data. *Ecosphere*, 6(4):1–16.
- MATTHEWS, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451.
- MAXWELL, S. L., FULLER, R. A., BROOKS, T. M. et WATSON, J. E. (2016). Biodiversity : The ravages of guns, nets and bulldozers. *Nature*, 536:143–145.
- MCCALLUM, H. et DOBSON, A. (2002). Disease, habitat fragmentation and conservation. *Proceedings of the Royal Society B*, 269(1504):2041–2049.
- MCPHERSON, J. M. et JETZ, W. (2007). Effects of species’ ecology on the accuracy of distribution models. *Ecography*, 30(1):135–151.
- MCRAE, B. H. (2006). Isolation by resistance. *Evolution*, 60(8):1551–1561.
- MCRAE, B. H. et BEIER, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, 104(50):19885–19890.
- MEDDE (2011). Stratégie nationale pour la biodiversité 2011-2020.
- MEIRMANS, P. G. (2014). Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources*, 14(4):726–733.

- MEIRMANS, P. G. et HEDRICK, P. W. (2011). Assessing population structure : FST and related measures. *Molecular Ecology Resources*, 11(1):5–18.
- MELO, G. L., SPONCHIADO, J., CÁCERES, N. C. et FAHRIG, L. (2017). Testing the habitat amount hypothesis for South American small mammals. *Biological Conservation*, 209:304–314.
- MÉNDEZ, M., TELLA, J. L. et GODOY, J. A. (2011). Restricted gene flow and genetic drift in recently fragmented populations of an endangered steppe bird. *Biological Conservation*, 144(11):2615–2622.
- MERGEY, M., BARDONNET, C., QUINTAINE, T., GALAN, M., BODIN, C., HUBERT, P. et HELDER, R. (2017). Identifying environmental drivers of spatial genetic structure of the European pine marten (*Martes martes*). *Landscape Ecology*, 32(12):2261–2279.
- MERRIAM, G. (1984). Connectivity : a fundamental ecological characteristic of landscape pattern. In *Methodology in landscape ecological research and planning : Proceedings, 1st seminar, International Association of Landscape Ecology, Roskilde, Denmark, Oct 15-19, 1984*. Roskilde University Centre.
- MIELE, V. et MATIAS, C. (2017). Revealing the hidden structure of dynamic ecological networks. *Royal Society open science*, 4(6):170251.
- MIELE, V., MATIAS, C., ROBIN, S. et DRAY, S. (2019). Nine quick tips for analyzing network data. *PLoS Computational Biology*, 15(12):e1007434.
- MIGUET, P., FAHRIG, L. et LAVIGNE, C. (2017). How to quantify a distance-dependent landscape effect on a biological response. *Methods in Ecology and Evolution*, 8(12):1717–1724.
- MIGUET, P., JACKSON, H. B., JACKSON, N. D., MARTIN, A. E. et FAHRIG, L. (2016). What determines the spatial extent of landscape effects on species? *Landscape Ecology*, 31(6):1177–1194.
- MILLER-RUSHING, A. J., PRIMACK, R. B., DEVICTOR, V., CORLETT, R. T., CUMMING, G. S., LOYOLA, R., MAAS, B. et PEJCHAR, L. (2019). How does habitat fragmentation affect biodiversity? a controversial question at the core of conservation biology. *Biological Conservation*, 232:271–273.
- MILLETTE, K. L., FUGÈRE, V., DEBYSER, C., GREINER, A., CHAIN, F. J. et GONZALEZ, A. (2020). No consistent effects of humans on animal genetic diversity worldwide. *Ecology Letters*, 23(1):55–67.
- MILLETTE, K. L. et KEYGHOBADI, N. (2015). The relative influence of habitat amount and configuration on genetic structure across multiple spatial scales. *Ecology and Evolution*, 5(1):73–86.
- MILLIGAN, B. G. (in prep). Probabilistic graph models for landscape genetics. *PeerJ Preprints*.
- MILLIGAN, B. G., ARCHER, F. I., FERCHAUD, A.-L., HAND, B. K., KIEREPKA, E. M. et WAPLES, R. S. (2018). Disentangling genetic structure for genetic monitoring of complex populations. *Evolutionary Applications*, 11(7):1149–1161.
- MILLS, L. S. et ALLENDORF, F. W. (1996). The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 10(6):1509–1518.
- MIMET, A., CLAUZEL, C. et FOLTÊTE, J.-C. (2016). Locating wildlife crossings for multispecies connectivity across linear infrastructures. *Landscape Ecology*, 31(9):1955–1973.
- MINOR, E. S. et URBAN, D. L. (2008). A graph-theory framework for evaluating landscape connectivity and conservation planning. *Conservation Biology*, 22(2):297–307.
- MIRALDO, A., LI, S., BORREGAARD, M. K., FLÓREZ-RODRÍGUEZ, A., GOPALAKRISHNAN, S., RIZVANOVIC, M., WANG, Z., RAHBEK, C., MARSKE, K. A. et NOGUÉS-BRAVO, D. (2016). An anthropocene map of genetic diversity. *Science*, 353(6307):1532–1535.
- MOILANEN, A. (2011). On the limitations of graph-theoretic connectivity in spatial ecology and conservation. *Journal of Applied Ecology*, 48(6):1543–1547.
- MOILANEN, A., KUJALA, H. et LEATHWICK, J. R. (2009). The zonation framework and software for conservation prioritization. *Spatial conservation prioritisation : Quantitative methods and computational tools*, 135:196–210.

- MOILANEN, A. et NIEMINEN, M. (2002). Simple connectivity measures in spatial ecology. *Ecology*, 83(4):1131–1145.
- MONMONIER, M. S. (1973). Maximum-difference barriers : An alternative numerical regionalization method. *Geographical analysis*, 5(3):245–261.
- MONY, C., ABADIE, J., GIL-TENA, A., BUREL, F. et ERNOULT, A. (2018). Effects of connectivity on animal-dispersed forest plant communities in agriculture-dominated landscapes. *Journal of Vegetation Science*, 29(2):167–178.
- MORA, C., TITTENSOR, D. P., ADL, S., SIMPSON, A. G. et WORM, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biology*, 9(8):e1001127.
- MORAGA, A. D., MARTIN, A. E. et FAHRIG, L. (2019). The scale of effect of landscape context varies with the species' response variable measured. *Landscape Ecology*, 34(4):703–715.
- MORAN-LOPEZ, T., ROBLEDON-ARNUNCIANO, J., DIAZ, M., MORALES, J., LAZARO-NOGAL, A., LORENZO, Z. et VALLADARES, F. (2016). Determinants of functional connectivity of holm oak woodlands - fragment size and mouse foraging behavior. *Forest Ecology and Management*, 368:111–122.
- MORTELLITI, A., AMORI, G. et BOITANI, L. (2010). The role of habitat quality in fragmented landscapes : a conceptual overview and prospectus for future research. *Oecologia*, 163(2):535–547.
- MOUGENOT, C. et MELIIN, É. (2000). Entre science et action : le concept de réseau écologique. *Nature Sciences Sociétés*, 8(3):20–30.
- MOULHERAT, S. (2019). Méthodes et résultats du programme CIRFE. Rapport Livrable II, CIRFE - Cumul d'infrastructures linéaires de transports terrestres et relations écologiques fonctionnelles.
- MUNWES, I., GEFFEN, E., ROLL, U., FRIEDMANN, A., DAYA, A., TIKOCHINSKI, Y. et GAFNY, S. (2010). The change in genetic diversity down the core-edge gradient in the eastern spadefoot toad (*Pelobates syriacus*). *Molecular Ecology*, 19(13):2675–2689.
- MUREKATETE, R. M. et SHIRABE, T. (2018). A spatial and statistical analysis of the impact of transformation of raster cost surfaces on the variation of least-cost paths. *International Journal of Geographical Information Science*, 32(11):2169–2188.
- MURPHY, M., DYER, R. et CUSHMAN, S. A. (2016). Graph theory and network models in landscape genetics. In BALKENHOL, N., CUSHMAN, S., STORFER, A. et WAITS, L., éditeurs : *Landscape genetics : Concepts, methods, applications*, pages 165–180. John Wiley & Sons, 1 édition.
- MURPHY, M. A., DEZZANI, R., PILLIOD, D. S. et STORFER, A. (2010a). Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, 19(17):3634–3649.
- MURPHY, M. A., EVANS, J. S. et STORFER, A. (2010b). Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology*, 91(1):252–261.
- MYERS, N., MITTERMEIER, R. A., MITTERMEIER, C. G., DA FONSECA, G. A. et KENT, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772):853.
- NAIDOO, R., BALMFORD, A., FERRARO, P. J., POLASKY, S., RICKETTS, T. H. et ROUGET, M. (2006). Integrating economic costs into conservation planning. *Trends in Ecology & Evolution*, 21(12):681–687.
- NATHAN, R., GETZ, W. M., REVILLA, E., HOLYOAK, M., KADMON, R., SALTZ, D. et SMOUSE, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49):19052–19059.
- NATHAN, R., PERRY, G., CRONIN, J. T., STRAND, A. E. et CAIN, M. L. (2003). Methods for estimating long-distance dispersal. *Oikos*, 103(2):261–273.
- NAUJOKAITIS-LEWIS, I. R., RICO, Y., LOVELL, J., FORTIN, M.-J. et MURPHY, M. A. (2013). Implications of incomplete networks on estimation of landscape genetic connectivity. *Conservation Genetics*, 14(2):287–298.

- NEI, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3):583–590.
- NEIGEL, J. E. (2002). Is FST Obsolete? *Conservation Genetics*, 3(2):167–173.
- NEKOLA, J. C. et WHITE, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, 26(4):867–878.
- NEUDITSCHKO, M., KHATKAR, M. S. et RAADSMA, H. W. (2012). NETVIEW : a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS ONE*, 7(10):e48375.
- NEWMAN, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.
- NEWMAN, M. E. et GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.
- NIEMINEN, M., SINGER, M. C., FORTELIUS, W., SCHÖPS, K. et HANSKI, I. (2001). Experimental confirmation that inbreeding depression increases extinction risk in butterfly populations. *The American Naturalist*, 157(2):237–244.
- O'BRIEN, D., MANSEAU, M., FALL, A. et FORTIN, M.-J. (2006). Testing the importance of spatial configuration of winter habitat for woodland caribou : an application of graph theory. *Biological Conservation*, 130(1):70–83.
- OPDAM, P., STEINGRÖVER, E. et VAN ROOIJ, S. (2006). Ecological networks : A spatial concept for multi-actor planning of sustainable landscapes. *Landscape and Urban Planning*, 75(3-4):322–332.
- OPPEN, V., MADELEINE, J., PELOW, L. M., KININMONTH, S. et BERKELMANS, R. (2011). Historical and contemporary factors shape the population genetic structure of the broadcast spawning coral, *Acropora millepora*, on the Great Barrier Reef. *Molecular Ecology*, 20(23):4899–4914.
- PALMER, S. C., COULON, A. et TRAVIS, J. M. (2011). Introducing a "stochastic movement simulator" for estimating habitat connectivity. *Methods in Ecology and Evolution*, 2(3):258–268.
- PANZACCHI, M., VAN MOORTER, B., STRAND, O., SAERENS, M., KIVIMÄKI, I., ST CLAIR, C. C., HERFINDAL, I. et BOITANI, L. (2016). Predicting the continuum between corridors and barriers to animal movements using step selection functions and randomized shortest paths. *Journal of Animal Ecology*, 85(1):32–42.
- PARADIS, E. (2010). pegas : an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3):419–420.
- PASCHOU, P., DRINEAS, P., YANNAKI, E., RAZOU, A., KANAKI, K., TSETOS, F., PADMANABHUNI, S. S., MICHALODIMITRAKIS, M., RENDA, M. C., PAVLOVIC, S. et al. (2014). Maritime route of colonization of Europe. *Proceedings of the National Academy of Sciences*, 111(25):9211–9216.
- PASCUAL-HORTAL, L. et SAURA, S. (2006). Comparison and development of new graph-based landscape connectivity indices : towards the prioritization of habitat patches and corridors for conservation. *Landscape Ecology*, 21(7):959–967.
- PASCUAL-HORTAL, L. et SAURA, S. (2007). Impact of spatial scale on the identification of critical habitat patches for the maintenance of landscape connectivity. *Landscape and Urban Planning*, 83(2):176–186.
- PEDERSEN, A. B. et FENTON, A. (2007). Emphasizing the ecology in parasite community ecology. *Trends in Ecology & Evolution*, 22(3):133–139.
- PE'ER, G., HENLE, K., DISLICH, C. et FRANK, K. (2011). Breaking functional connectivity into components : A novel approach using an individual-based model, and first outcomes. *PLoS ONE*, 6(8):1–18.
- PENUELAS, J., POULTER, B., SARDANS, J., CIAIS, P., VAN DER VELDE, M., BOPP, L., BOUCHER, O., GODDERIS, Y., HINSINGER, P., LLUSIA, J. et al. (2013). Human-induced nitrogen–phosphorus imbalances alter natural and managed ecosystems across the globe. *Nature Communications*, 4(1):1–10.
- PEREIRA, H. M., NAVARRO, L. M. et MARTINS, I. S. (2012). Global biodiversity change : the bad, the good, and the unknown. *Annual Review of Environment and Resources*, 37:25–50.

- PEREIRA, M., SEGURADO, P. et NEVES, N. (2011). Using spatial network structure in landscape management and planning : a case study with pond turtles. *Landscape and Urban Planning*, 100(1):67–76.
- PÉREZ-ESPONA, S., PÉREZ-BARBERÍA, F., MCLEOD, J., JIGGINS, C., GORDON, I. et PEMBERTON, J. (2008). Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, 17(4):981–996.
- PÉREZ-MÉNDEZ, N., JORDANO, P. et VALIDO, A. (2018). Persisting in defaunated landscapes : reduced plant population connectivity after seed dispersal collapse. *Journal of Ecology*, 106(3):936–947.
- PÉREZ-RODRÍGUEZ, A., KHIMOUN, A., OLLIVIER, A., ERAUD, C., FAIVRE, B. et GARNIER, S. (2018). Habitat fragmentation, not habitat loss, drives the prevalence of blood parasites in a Caribbean passerine. *Ecography*, 41(11):1835–1849.
- PERRIN, A. (2021). *Fragmentation des habitats et interactions hôtes-parasites*. Thèse de doctorat, Université Bourgogne Franche-Comté.
- PERRIN, A., KHIMOUN, A., FAIVRE, B., OLLIVIER, A., de PRACONTAL, N., THÉRON, F., LOUBON, M., LEBLOND, G., DURON, O. et GARNIER, S. (2021). Habitat fragmentation differentially shapes neutral and immune gene variation in a tropical bird species. *Heredity*, 126(1):148–162.
- PETERMAN, W. E. (2018). ResistanceGA : An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, 9(6):1638–1647.
- PETERMAN, W. E., ANDERSON, T. L., OUSTERHOUT, B. H., DRAKE, D. L., SEMLITSCH, R. D. et EGGERT, L. S. (2015). Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conservation Genetics*, 16(1):59–69.
- PETERMAN, W. E., OUSTERHOUT, B. H., ANDERSON, T. L., DRAKE, D. L., SEMLITSCH, R. D. et EGGERT, L. S. (2016). Assessing modularity in genetic networks to manage spatially structured metapopulations. *Ecosphere*, 7(2):e01231.
- PETERMAN, W. E. et POPE, N. S. (2020). The use and misuse of regression models in landscape genetic analyses. *Molecular Ecology*, 30(1):37–47.
- PETERMAN, W. E., WINIARSKI, K. J., MOORE, C. E., da SILVA CARVALHO, C., GILBERT, A. L. et SPEAR, S. F. (2019). A comparison of popular approaches to optimize landscape resistance surfaces. *Landscape Ecology*, 34(9):2197–2208.
- PETERSON, E. E., HANKS, E. M., HOOTEN, M. B., VER HOEF, J. M. et FORTIN, M.-J. (2019). Spatially structured statistical network models for landscape genetics. *Ecological Monographs*, 89(2):e01355.
- PFLÜGER, F. J. et BALKENHOL, N. (2014). A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology*, 23(9):2146–2156.
- PHILLIPS, S. J., ANDERSON, R. P. et SCHAPIRE, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259.
- PILOSOV, S., FORTUNA, M. A., COSSON, J.-F., GALAN, M., KITTIPONG, C., RIBAS, A., SEGAL, E., KRASNOV, B. R., MORAND, S. et BASCOMPTE, J. (2014). Host–parasite network structure is associated with community-level immunogenetic diversity. *Nature Communications*, 5(1):1–9.
- PITHER, J. et TAYLOR, P. D. (1998). An experimental assessment of landscape connectivity. *Oikos*, 83(1):166–174.
- POLI, C., HIGHTOWER, J. et FLETCHER JR, R. J. (2020). Validating network connectivity with observed movement in experimental landscapes undergoing habitat destruction. *Journal of Applied Ecology*, 57(7):1426–1437.
- POOR, E. E., LOUCKS, C., JAKES, A. et URBAN, D. L. (2012). Comparing habitat suitability and connectivity modeling methods for conserving pronghorn migrations. *PLoS ONE*, 7(11):e49390.
- PRESSEY, R. (2004). Conservation planning and biodiversity : assembling the best data for the job. *Conservation Biology*, 18(6):1677–1681.
- PRESTON, F. W. (1962). The canonical distribution of commonness and rarity : Part i. *Ecology*, 43(2):185–215.

- PRITCHARD, J. K., STEPHENS, M. et DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.
- PROULX, S. R., PROMISLOW, D. E. et PHILLIPS, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, 20(6):345–353.
- PRUNIER, J., KAUFMANN, B., FENET, S., PICARD, D., POMPANON, F., JOLY, P. et LENA, J.-P. (2013). Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations : Towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology*, 22(22):5516–5530.
- PRUNIER, J. G., DUBUT, V., CHIKHI, L. et BLANCHET, S. (2017). Contribution of spatial heterogeneity in effective population sizes to the variance in pairwise measures of genetic differentiation. *Methods in Ecology and Evolution*, 8(12):1866–1877.
- PULLIAM, H. R. (1988). Sources, sinks, and population regulation. *The American Naturalist*, 132(5):652–661.
- PULLINGER, M. G. et JOHNSON, C. J. (2010). Maintaining or restoring connectivity of modified landscapes : evaluating the least-cost path model with multiple sources of ecological information. *Landscape Ecology*, 25(10):1547–1560.
- PUTMAN, A. I. et CARBONE, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, 4(22):4399–4428.
- R CORE TEAM (2020). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RALLS, K. et BALLOU, J. (1986). Captive breeding programs for populations with a small number of founders. *Trends in Ecology & Evolution*, 1(1):19–22.
- RAMAKRISHNAN, P. (1996). Conserving the sacred : from species to landscapes. *Nature & Resources*, 32(1):11–19.
- RAYFIELD, B., FORTIN, M.-J. et FALL, A. (2010). The sensitivity of least-cost habitat graphs to relative cost surface values. *Landscape Ecology*, 25(4):519–532.
- RAYFIELD, B., FORTIN, M.-J. et FALL, A. (2011). Connectivity for conservation : a framework to classify network measures. *Ecology*, 92(4):847–858.
- RAYMOND, M. et ROUSSET, F. (1995). GENEPOP : Population genetics software for exact tests and ecumenism. Vers. 1.2. *Journal of Heredity*, 86:248–249.
- REDDINGIUS, J. et DEN BOER, P. (1970). Simulation experiments illustrating stabilization of animal numbers by spreading of risk. *Oecologia*, 5(3):240–284.
- REED, G., LITVAITIS, J., CALLAHAN, C., CARROLL, R., LITVAITIS, M. et BROMAN, D. (2017). Modeling landscape connectivity for bobcats using expert-opinion and empirically derived models : how well do they work? *Animal Conservation*, 20(4):308–320.
- REICHERT, B. E., FLETCHER JR, R. J., CATTAN, C. E. et KITCHENS, W. M. (2016). Consistent scaling of population structure across landscapes despite intraspecific variation in movement and connectivity. *Journal of Animal Ecology*, 85(6):1563–1573.
- REINHARDT, K., KÖHLER, G., MAAS, S. et DETZEL, P. (2005). Low dispersal ability and habitat specificity promote extinctions in rare but not in widespread species : the Orthoptera of Germany. *Ecography*, 28(5):593–602.
- RESASCO, J. (2019). Meta-analysis on a decade of testing corridor efficacy : what new have we learned? *Current Landscape Ecology Reports*, 4(3):61–69.
- RIBEIRO, R., CARRETERO, M. A., SILLERO, N., ALARCOS, G., ORTIZ-SANTALIESTRA, M., LIZANA, M. et LLORENTE, G. A. (2011). The pond network : can structural connectivity reflect on (amphibian) biodiversity patterns? *Landscape Ecology*, 26(5):673–682.
- RICHARDSON, J. L., BRADY, S. P., WANG, I. J. et SPEAR, S. F. (2016). Navigating the pitfalls and promise of landscape genetics. *Molecular Ecology*, 25(4):849–863.

- RICKETTS, T. H. (2001). The matrix matters : effective isolation in fragmented landscapes. *The American Naturalist*, 158(1):87–99.
- RILEY, S. P., POLLINGER, J. P., SAUVAJOT, R. M., YORK, E. C., BROMLEY, C., FULLER, T. K. et WAYNE, R. K. (2006). FAST-TRACK : A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*, 15(7):1733–1741.
- RIPLEY, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(2):172–192.
- RISSER, P., KARR, J. et FORMAN, R. (1984). Landscape ecology : Directions and approaches. *Illinois Natural History Survey*, 2.
- ROBERTSON, E. P., FLETCHER, R. J., CATTANU, C. E., UDELL, B. J., REICHERT, B. E., AUSTIN, J. D. et VALLE, D. (2018a). Isolating the roles of movement and reproduction on effective connectivity alters conservation priorities for an endangered bird. *Proceedings of the National Academy of Sciences*, 115(34):8591–8596.
- ROBERTSON, J. M., MURPHY, M. A., PEARL, C. A., ADAMS, M. J., PÁEZ-VACAS, M. I., HAIG, S. M., PILLIOD, D. S., STORFER, A. et FUNK, W. C. (2018b). Regional variation in drivers of connectivity for two frog species (*Rana pretiosa* and *R. luteiventris*) from the US Pacific Northwest. *Molecular Ecology*, 27(16):3242–3256.
- ROBIN, V., GUPTA, P., THATTE, P. et RAMAKRISHNAN, U. (2015). Islands within islands : two montane palaeo-endemic birds impacted by recent anthropogenic fragmentation. *Molecular Ecology*, 24(14):3572–3584.
- RODGER, Y. S., GREENBAUM, G., SILVER, M., BAR-DAVID, S. et WINTERS, G. (2018). Detecting hierarchical levels of connectivity in a population of *Acacia tortilis* at the northern edge of the species' global distribution : Combining classical population genetics and network analyses. *PLoS ONE*, 13(4):e0194901.
- RONCE, O. (2007). How does it feel to be like a rolling stone? ten questions about dispersal evolution. *Annual Review of Ecology, Evolution, and Systematics*, 38:231–253.
- ROUSSET, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, 145(4):1219–1228.
- ROY, K., KAR, S. et DAS, R. N. (2015). Statistical methods in QSAR/QSPR. In *A primer on QSAR/QSPR modeling*, pages 37–59. Springer.
- ROZENFELD, A. F., ARNAUD-HAOND, S., HERNÁNDEZ-GARCÍA, E., EGUÍLUZ, V. M., SERRÃO, E. A. et DUARTE, C. M. (2008). Network analysis identifies weak and strong links in a metapopulation system. *Proceedings of the National Academy of Sciences*, 105(48):18824–18829.
- RUIZ-GONZALEZ, A., CUSHMAN, S. A., MADEIRA, M. J., RANDI, E. et GÓMEZ-MOLINER, B. J. (2015). Isolation by distance, resistance and/or clusters? lessons learned from a forest-dwelling carnivore inhabiting a heterogeneous landscape. *Molecular Ecology*, 24(20):5110–5129.
- RUIZ-GONZÁLEZ, A., GURRUTXAGA, M., CUSHMAN, S. A., MADEIRA, M. J., RANDI, E. et GÓMEZ-MOLINER, B. J. (2014). Landscape genetics for the empirical assessment of resistance surfaces : the European pine marten (*Martes martes*) as a target-species of a regional ecological network. *PLoS ONE*, 9(10):e110552.
- SACCHERI, I., KUUSSAARI, M., KANKARE, M., VIKMAN, P., FORTELIUS, W. et HANSKI, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, 392(6675):491–494.
- SAFNER, T., MILLER, M. P., MCRAE, B. H., FORTIN, M.-J. et MANEL, S. (2011). Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences*, 12(2):865–889.
- SAHRAOUI, Y. (2016). *Le paysage, entre esthétique et écologie - Modélisation rétrospective à partir de changements d'occupation du sol*. Thèse de doctorat, Université de Franche-Comté.
- SAHRAOUI, Y., FOLTÊTE, J.-C. et CLAUZEL, C. (2017). A multi-species approach for assessing the impact of land-cover changes on landscape connectivity. *Landscape Ecology*, 32(9):1819–1835.

- SAHRAOUI, Y., LESKI, C. D. G., BENOT, M.-L., REVERS, F., SALLES, D., van HALDER, I., BARNEIX, M. et CARASSOU, L. (2021). Integrating ecological networks modelling in a participatory approach for assessing impacts of planning scenarios on landscape connectivity. *Landscape and Urban Planning*, 209:104039.
- SALA, O. E., CHAPIN, F. S., ARMESTO, J. J., BERLOW, E., BLOOMFIELD, J., DIRZO, R., HUBER-SANWALD, E., HUENNEKE, L. F., JACKSON, R. B., KINZIG, A. *et al.* (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459): 1770–1774.
- SARKAR, S., PRESSEY, R. L., FAITH, D. P., MARGULES, C. R., FULLER, T., STOMS, D. M., MOFFETT, A., WILSON, K. A., WILLIAMS, K. J., WILLIAMS, P. H. *et al.* (2006). Biodiversity conservation planning tools : present status and challenges for the future. *Annu. Rev. Environ. Resour.*, 31:123–159.
- SAURA, S. (2018). The amount of reachable habitat - jointly measuring habitat amount and connectivity in space and time. In *International Conference of Ecological Sciences of the French Society for Ecology and Evolution*.
- SAURA, S. (2021). The Habitat Amount Hypothesis implies negative effects of habitat fragmentation on species richness. *Journal of Biogeography*, 48(1):11–22.
- SAURA, S., BASTIN, L., BATTISTELLA, L., MANDRICI, A. et DUBOIS, G. (2017). Protected areas in the world’s ecoregions : How well connected are they? *Ecological Indicators*, 76:144–158.
- SAURA, S., BERTZKY, B., BASTIN, L., BATTISTELLA, L., MANDRICI, A. et DUBOIS, G. (2018). Protected area connectivity : Shortfalls in global targets and country-level priorities. *Biological Conservation*, 219:53–67.
- SAURA, S., BERTZKY, B., BASTIN, L., BATTISTELLA, L., MANDRICI, A. et DUBOIS, G. (2019). Global trends in protected area connectivity from 2010 to 2018. *Biological Conservation*, 238:108–183.
- SAURA, S., BODIN, Ö. et FORTIN, M.-J. (2014). Stepping stones are crucial for species’ long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology*, 51(1):171–182.
- SAURA, S. et de la FUENTE, B. (2017). Connectivity as the amount of reachable habitat : conservation priorities and the roles of habitat patches in landscape networks. In GERGEL, S. E. et TURNER, M. G., éditeurs : *Learning landscape ecology : a practical guide to concepts and techniques*, pages 229–254. Springer.
- SAURA, S., ESTREGUIL, C., MOUTON, C. et RODRÍGUEZ-FREIRE, M. (2011). Network analysis to assess landscape connectivity trends : application to European forests (1990–2000). *Ecological Indicators*, 11(2):407–416.
- SAURA, S. et PASCUAL-HORTAL, L. (2007). A new habitat availability index to integrate connectivity in landscape conservation planning : comparison with existing indices and application to a case study. *Landscape and Urban Planning*, 83(2):91–103.
- SAURA, S. et RUBIO, L. (2010). A common currency for the different ways in which patches and links can contribute to habitat availability and connectivity in the landscape. *Ecography*, 33(3):523–537.
- SAURA, S. et TORNE, J. (2009). Conefor Sensinode 2.2 : a software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Modelling & Software*, 24(1):135–139.
- SAVARY, P., FOLTÊTE, J.-C., MOAL, H., VUIDEL, G. et GARNIER, S. (2021a). Analysing landscape effects on dispersal networks and gene flow with genetic graphs. *Molecular Ecology Resources*, 21(4):1167–1185.
- SAVARY, P., FOLTÊTE, J.-C., MOAL, H., VUIDEL, G. et GARNIER, S. (2021b). graph4lg : a package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, 12(3):539–547.
- SAWYER, S. C., EPPS, C. W. et BRASHARES, J. S. (2011). Placing linkages among fragmented habitats : do least-cost models reflect how animals use landscapes? *Journal of Applied Ecology*, 48(3):668–678.
- SCHADT, S., KNAUER, F., KACZENSKY, P., REVILLA, E., WIEGAND, T. et TREPL, L. (2002). Rule-based assessment of suitable habitat and patch connectivity for the Eurasian lynx. *Ecological Applications*, 12(5):1469–1483.
- SCHAEFFER, B., NICOLAS, V., AUSTERLITZ, F. et LARÉDO, C. (in prep). Population networks from dna sequences - methodological developments. *PeerJ Preprints*.

- SCHLATHER, M., MALINOWSKI, A., MENCK, P. J., OESTING, M., STROKORB, K. *et al.* (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software*, 63(8):1–25.
- SCHMIDT, C., DRAY, S. *et* GARROWAY, C. J. (2021). Genetic and species-level biodiversity patterns are linked by demography and ecological opportunity. *bioRxiv*, pages 2020–06.
- SCHNEIDER, C. (2003). The influence of spatial scale on quantifying insect dispersal : an analysis of butterfly data. *Ecological Entomology*, 28(2):252–256.
- SCHNEIDER, D. W., ELLIS, C. D. *et* CUMMINGS, K. S. (1998). A transportation model assessment of the risk to native mussel communities from zebra mussel spread. *Conservation Biology*, 12(4):788–800.
- SCHOVILLE, S. D., BONIN, A., FRANÇOIS, O., LOBREAUX, S., MELODELIMA, C. *et* MANEL, S. (2012). Adaptive genetic variation on the landscape : methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, 43:23–43.
- SCHOVILLE, S. D., DALONGEVILLE, A., VIENNOIS, G., GUGERLI, F., TABERLET, P., LEQUETTE, B., ALVAREZ, N. *et* MANEL, S. (2018). Preserving genetic connectivity in the European Alps protected area network. *Biological Conservation*, 218:99–109.
- SCHREIBER, K. F. (1990). The history of landscape ecology in europe. *In* ZONNEVELD, I. *et* FORMAN, R., éditeurs : *Changing landscapes : an ecological perspective*, pages 21–34. Springer.
- SCHWARTZ, M. K. *et* MCKELVEY, K. S. (2009). Why sampling scheme matters : the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, 10(2):441–452.
- SEGELBACHER, G., CUSHMAN, S. A., EPPERSON, B. K., FORTIN, M.-J., FRANCOIS, O., HARDY, O. J., HOLDEREGGER, R., TABERLET, P., WAITS, L. P. *et* MANEL, S. (2010). Applications of landscape genetics in conservation biology : concepts and challenges. *Conservation Genetics*, 11(2):375–385.
- SERROUYA, R., PAETKAU, D., MCLELLAN, B. N., BOUTIN, S., CAMPBELL, M. *et* JENKINS, D. A. (2012). Population size and major valleys explain microsatellite variation better than taxonomic units for caribou in western Canada. *Molecular Ecology*, 21(11):2588–2601.
- SHIRK, A. *et* CUSHMAN, S. (2011). sGD : software for estimating spatially explicit indices of genetic diversity. *Molecular Ecology Resources*, 11(5):922–934.
- SHIRK, A., LANDGUTH, E. *et* CUSHMAN, S. (2017a). A comparison of individual-based genetic distance metrics for landscape genetics. *Molecular Ecology*, 17(6):1308–1317.
- SHIRK, A., WALLIN, D., CUSHMAN, S., RICE, C. *et* WARHEIT, K. (2010). Inferring landscape effects on gene flow : a new model selection framework. *Molecular Ecology*, 19(17):3603–3619.
- SHIRK, A. J., LANDGUTH, E. L. *et* CUSHMAN, S. A. (2017b). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, 18(1):55–67.
- SIMBERLOFF, D. (1988). The contribution of population and community biology to conservation science. *Annual Review of Ecology and Systematics*, 19(1):473–511.
- SIMBERLOFF, D. S. *et* ABELE, L. G. (1976). Island biogeography theory and conservation practice. *Science*, 191(4224):285–286.
- SIMPKINS, C. E., DENNIS, T. E., ETHERINGTON, T. R. *et* PERRY, G. L. (2017). Effects of uncertain cost-surface specification on landscape connectivity measures. *Ecological Informatics*, 38:1–11.
- SIMPKINS, C. E., DENNIS, T. E., ETHERINGTON, T. R. *et* PERRY, G. L. (2018). Assessing the performance of common landscape connectivity metrics using a virtual ecologist approach. *Ecological Modelling*, 367:13–23.
- SLATKIN, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803):787–792.
- SLATKIN, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 47(1):264–279.

- SMOUSE, P. E. et PEAKALL, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, 82(5):561–573.
- SONNECK, A.-G., BÖNSEL, A. et MATTHES, J. (2008). Der einfluss von landnutzung auf die habitate von stethophyma grossum (linnaeus, 1758) an beispielen aus mecklenburg-vorpommern. *Articulata*, 23:15–30.
- SORCI, G. et GARNIER, S. (2008). Parasitism. In FATH, S. J. . B., éditeur : *Encyclopedia of Ecology*, pages 2645–2650. Elsevier.
- SORK, V. L., SQUIRE, K., GUGGER, P. F., STEELE, S. E., LEVY, E. D. et ECKERT, A. J. (2016). Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American journal of botany*, 103(1):33–46.
- SOULÉ, M. E. (1985). What is conservation biology ? *BioScience*, 35(11):727–734.
- SOULÉ, M. E. (1986). *Conservation biology : the science of scarcity and diversity*. Sinauer, Sunderland.
- SOULÉ, M. E. et SIMBERLOFF, D. (1986). What do genetics and ecology tell us about the design of nature reserves ? *Biological Conservation*, 35(1):19–40.
- SPACKMAN, S. C. et HUGHES, J. W. (1995). Assessment of minimum stream corridor width for biological conservation : species richness and distribution along mid-order streams in Vermont, USA. *Biological Conservation*, 71(3):325–332.
- SPANOWICZ, A. G. et JAEGER, J. A. (2019). Measuring landscape connectivity : on the importance of within-patch connectivity. *Landscape Ecology*, 34(10):2261–2278.
- SPEAR, S. F., BALKENHOL, N., FORTIN, M.-J., MCRAE, B. H. et SCRIBNER, K. (2010). Use of resistance surfaces for landscape genetic studies : considerations for parameterization and analysis. *Molecular Ecology*, 19(17):3576–3591.
- SPIELMAN, D., BROOK, B. W. et FRANKHAM, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences*, 101(42):15261–15264.
- STEIN, A., GERSTNER, K. et KREFT, H. (2014). Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters*, 17(7):866–880.
- STEINIG, E. J., NEUDITSCHKO, M., KHATKAR, M. S., RAADSMA, H. W. et ZENGER, K. R. (2016). NETVIEW P : a network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources*, 16(1):216–227.
- STEVENSON-HOLT, C. D., WATTS, K., BELLAMY, C. C., NEVIN, O. T. et RAMSEY, A. D. (2014). Defining landscape resistance values in least-cost connectivity models for the invasive grey squirrel : a comparison of approaches using expert-opinion and habitat suitability modelling. *PLoS ONE*, 9(11):e112119.
- STORFER, A., MURPHY, M., EVANS, J., GOLDBERG, C., ROBINSON, S., SPEAR, S., DEZZANI, R., DELMELLE, E., VIERLING, L. et WAITS, L. (2007). Putting the "landscape" in landscape genetics. *Heredity*, 98(3):128–142.
- STORFER, A., MURPHY, M. A., SPEAR, S. F., HOLDEREGGER, R. et WAITS, L. P. (2010). Landscape genetics : where are we now ? *Molecular Ecology*, 19(17):3496–3514.
- STORFER, A., PATTON, A. et FRAIK, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, 9:68.
- STUBER, E. F. et GRUBER, L. F. (2020). Recent methodological solutions to identifying scales of effect in multi-scale modeling. *Current Landscape Ecology Reports*, 5(1):127–139.
- STUTCHBURY, B. J., MORTON, E. S. et al. (2001). *Behavioral ecology of tropical birds*. Academic Press.
- SUNDQVIST, L., KEENAN, K., ZACKRISSON, M., PRODÖHL, P. et KLEINHANS, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, 6(11):3461–3475.
- SZPIECH, Z. A., JAKOBSSON, M. et ROSENBERG, N. A. (2008). ADZE : a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, 24(21):2498–2504.

- TABERLET, P., FUMAGALLI, L., WUST-SAUCY, A.-G. et COSSON, J.-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular ecology*, 7(4):453–464.
- TANNIER, C., BOURGEOIS, M., HOUOT, H. et FOLTÊTE, J.-C. (2016). Impact of urban developments on the functional connectivity of forested habitats : a joint contribution of advanced urban models and landscape graphs. *Land Use Policy*, 52:76–91.
- TANSLEY, A. G. (1935). The use and abuse of vegetational concepts and terms. *Ecology*, 16(3):284–307.
- TANSLEY, A. G. (1939). *The British Islands and their vegetation*. Cambridge University Press.
- TARABON, S., BERGÈS, L., DUTOIT, T. et ISSELIN-NONDEDEU, F. (2019). Environmental impact assessment of development projects improved by merging species distribution and habitat connectivity modelling. *Journal of Environmental Management*, 241:439–449.
- TARABON, S., CALVET, C., DELBAR, V., DUTOIT, T. et ISSELIN-NONDEDEU, F. (2020a). Integrating a landscape connectivity approach into mitigation hierarchy planning by anticipating urban dynamics. *Landscape and Urban Planning*, 202:103871.
- TARABON, S., DUTOIT, T. et ISSELIN-NONDEDEU, F. (2020b). Pooling biodiversity offsets to improve habitat connectivity and species conservation. *Journal of Environmental Management*, 277:111425.
- TAYLOR, P. D., FAHRIG, L., HENEIN, K. et MERRIAM, G. (1993). Connectivity is a vital element of landscape structure. *Oikos*, 68(3):571–573.
- TAYLOR, P. D., FAHRIG, L. et WITH, K. A. (2006). Landscape connectivity : a return to the basics. In CROOKS, K. R. et SANJAYAN, M., éditeurs : *Connectivity conservation*, chapitre 2, pages 29–43. Cambridge University Press.
- TAYLOR, Z. et HOFFMAN, S. (2014). Landscape models for nuclear genetic diversity and genetic structure in white-footed mice (*Peromyscus leucopus*). *Heredity*, 112(6):588–595.
- TEIXEIRA, J. C. et HUBER, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences*, 118(10).
- TENENHAUS, M. et YOUNG, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50(1):91–119.
- TEWKSBRURY, J. J., LEVEY, D. J., HADDAD, N. M., SARGENT, S., ORROCK, J. L., WELDON, A., DANIELSON, B. J., BRINKERHOFF, J., DAMSCHEN, E. I. et TOWNSEND, P. (2002). Corridors affect plants, animals, and their interactions in fragmented landscapes. *Proceedings of the National Academy of Sciences*, 99(20):12923–12926.
- THÉBAULT, E. et FONTAINE, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856.
- THOMPSON, P. L., RAYFIELD, B. et GONZALEZ, A. (2014). Robustness of the spatial insurance effects of biodiversity to habitat loss. *Evolutionary Ecology Research*, 16(6):445–460.
- TILMAN, D., MAY, R. M., LEHMAN, C. L. et NOWAK, M. A. (1994). Habitat destruction and the extinction debt. *Nature*, 371(6492):65–66.
- TISCHENDORF, L. et FAHRIG, L. (2000). On the usage and measurement of landscape connectivity. *Oikos*, 90(1):7–19.
- TOBLER, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1):234–240.
- TOMA, Y., IMANISHI, J., YOKOGAWA, M., HASHIMOTO, H., IMANISHI, A., MORIMOTO, Y., HATANAKA, Y., ISAGI, Y. et SHIBATA, S. (2015). Factors affecting the genetic diversity of a perennial herb *Viola grypoceras* A. Gray var. *grypoceras* in urban fragmented forests. *Landscape Ecology*, 30(8):1435–1447.
- TORRENTA, R. et VILLARD, M.-A. (2017). A test of the habitat amount hypothesis as an explanation for the species richness of forest bird assemblages. *Journal of Biogeography*, 44(8):1791–1801.

- TOURNANT, P., AFONSO, E., ROUÉ, S., GIRAUDOUX, P. et FOLTÊTE, J.-C. (2013). Evaluating the effect of habitat connectivity on the distribution of lesser horseshoe bat maternity roosts using landscape graphs. *Biological Conservation*, 164:39–49.
- TRAUTNER, J. et HERMANN, G. (2008). Die Sumpfschrecke (*Stethophyma grossum* L., 1758) im Aufwind-Erkenntnis aus dem zentralen Baden-Württemberg. *Articulata*, 23(2):37–52.
- TROLL, C. (1966). *Landscape ecology*. ITC-UNESCO. Delft, The Netherlands.
- TUCKER, M. A., BÖHNING-GAESE, K., FAGAN, W. F., FRYXELL, J. M., VAN MOORTER, B., ALBERTS, S. C., ALI, A. H., ALLEN, A. M., ATTIAS, N., AVGAR, T. *et al.* (2018). Moving in the Anthropocene : Global reductions in terrestrial mammalian movements. *Science*, 359(6374):466–469.
- TURNER, M. G. (1989). Landscape ecology - the effect of pattern on process. *Annual Review of Ecology and Systematics*, 20(1):171–197.
- TYLIANAKIS, J. M., TSCHARNTKE, T. et LEWIS, O. T. (2007). Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature*, 445(7124):202–205.
- UENO, S., TOMARU, N., YOSHIMARU, H., MANABE, T. et YAMAMOTO, S. (2000). Genetic structure of *Camellia japonica* L. in an old-growth evergreen forest, Tsushima, Japan. *Molecular Ecology*, 9(6):647–656.
- URBAN, D. et KEITT, T. (2001). Landscape connectivity : a graph-theoretic perspective. *Ecology*, 82(5):1205–1218.
- URBAN, D. L., MINOR, E. S., TREML, E. A. et SCHICK, R. S. (2009). Graph models of habitat mosaics. *Ecology Letters*, 12(3):260–273.
- UROY, L., ALIGNIER, A., MONY, C., FOLTÊTE, J.-C. et ERNOULT, A. (2021). How to assess the temporal dynamics of landscape connectivity in ever-changing landscapes : a literature review. *Landscape Ecology*, 36:2487–2504.
- VAN DYCK, H. et BAGUETTE, M. (2005). Dispersal behaviour in fragmented landscapes : routine or special movements ? *Basic and Applied Ecology*, 6(6):535–545.
- VAN ETTEN, J. (2012). R package gdistance : distances and routes on geographical grids (version 1.1-4). *Journal of Statistical Software*, 76(1):1–13.
- VAN STRIEN, M. J. (2017). Consequences of population topology for studying gene flow using link-based landscape genetic methods. *Ecology and Evolution*, 7(14):5070–5081.
- VAN STRIEN, M. J., HOLDEREGGER, R. et VAN HECK, H. J. (2015). Isolation-by-distance in landscapes : considerations for landscape genetics. *Heredity*, 114(1):27–37.
- VAN STRIEN, M. J., KELLER, D. et HOLDEREGGER, R. (2012). A new analytical approach to landscape genetic modelling : least-cost transect analysis and linear mixed models. *Molecular Ecology*, 21(16):4010–4023.
- VAN STRIEN, M. J., KELLER, D., HOLDEREGGER, R., GHAZOUL, J., KIENAST, F. et BOLLIGER, J. (2014). Landscape genetics as a tool for conservation planning : predicting the effects of landscape change on gene flow. *Ecological Applications*, 24(2):327–339.
- VARVIO, S.-L., CHAKRABORTY, R. et NEI, M. (1986). Genetic variation in subdivided populations and conservation genetics. *Heredity*, 57(2):189–198.
- VELLEND, M. et GEBER, M. A. (2005). Connections between species diversity and genetic diversity. *Ecology Letters*, 8(7):767–781.
- VERITY, R. et NICHOLS, R. A. (2014). What is genetic differentiation, and how should we measure it - GST, D, neither or both ? *Molecular Ecology*, 23(17):4216–4225.
- VERSCHUUREN, B., WILD, R., OVIEDO, G. et MCNEELY, J. (2010). *Sacred natural sites : Conserving nature and culture*. Routledge.

- VIMAL, R. et MATHEVET, R. (2011). La carte et le territoire : le réseau écologique à l'épreuve de l'assemblée cartographique. *Cybergeo : European Journal of Geography*.
- VIMAL, R., MATHEVET, R. et THOMPSON, J. D. (2012). The changing landscape of ecological networks. *Journal for Nature Conservation*, 20(1):49–55.
- VITOUSEK, P. M., D'ANTONIO, C. M., LOOPE, L. L. et WESTBROOKS, R. (1996). Biological invasions as global environmental change. *American scientist*, 84(5):468–478.
- VITOUSEK, P. M., MOONEY, H. A., LUBCHENCO, J. et MELILLO, J. M. (1997). Human domination of Earth's ecosystems. *Science*, 277(5325):494–499.
- VOGT, P., RIITERS, K. H., IWANOWSKI, M., ESTREGUIL, C., KOZAK, J. et SOILLE, P. (2007). Mapping landscape corridors. *Ecological Indicators*, 7(2):481–488.
- WAGNER, H. H. et FORTIN, M.-J. (2013). A conceptual framework for the spatial analysis of landscape genetic data. *Conservation Genetics*, 14(2):253–261.
- WAGNER, H. H., HOLDEREGGER, R., WERTH, S., GUGERLI, F., HOEBEE, S. E. et SCHEIDEGGER, C. (2005). Variogram analysis of the spatial genetic structure of continuous populations using multilocus microsatellite data. *Genetics*, 169(3):1739–1752.
- WAITS, L. P. et STORFER, A. (2015). Basics of population genetics : quantifying neutral and adaptive genetic variation for landscape genetic studies. In BALKENHOL, N., CUSHMAN, S., STORFER, A. et WAITS, L., éditeurs : *Landscape genetics : Concepts, methods, applications*, pages 35–57. John Wiley & Sons.
- WANG, J. (2004). Application of the one-migrant-per-generation rule to conservation and management. *Conservation Biology*, 18(2):332–343.
- WANG, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B - Biological Sciences*, 360(1459):1395–1409.
- WANG, Y.-H., YANG, K.-C., BRIDGMAN, C. L. et LIN, L.-K. (2008). Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, 23(8):989–1000.
- WAPLES, R. S. et GAGGIOTTI, O. (2006). What is a population ? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6):1419–1439.
- WATTS, A. G., SCHLICHTING, P. E., BILLERMAN, S. M., JESMER, B. R., MICHELETTI, S., FORTIN, M.-J., FUNK, W. C., HAPEMAN, P., MUTHS, E. et MURPHY, M. A. (2015). How spatio-temporal habitat connectivity affects amphibian genetic structure? *Frontiers in Genetics*, 6:275.
- WATTS, D. J. et STROGATZ, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.
- WECKWORTH, B. V., MUSIANI, M., DECESARE, N. J., MCDEVITT, A. D., HEBBLEWHITE, M. et MARIANI, S. (2013). Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society B*, 280(1769):20131756.
- WEIR, B. S. et COCKERHAM, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6):1358–1370.
- WHITLOCK, M. C. et MCCAULEY, D. E. (1999). Indirect measures of gene flow and migration : $F_{st} \neq 1/(4nm + 1)$. *Heredity*, 82(2):117–125.
- WHITTAKER, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.
- WIENS, J. A. (1997). *Metapopulation dynamics and landscape ecology*. Academic Press.
- WILSON, E. O. (1988). *Biodiversity*. National Academy Press.

- WINIARSKI, K. J., PETERMAN, W. E., WHITELEY, A. R. et MCGARIGAL, K. (2019). Multi-scale resistant kernel surfaces derived from inferred gene flow : An application with vernal pool breeding salamanders. *Molecular Ecology Resources*, 20(1):97–113.
- WOLD, S., SJÖSTRÖM, M. et ERIKSSON, L. (2001). PLS-regression : a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130.
- WOMBLE, W. H. (1951). Differential systematics. *Science*, 114(2961):315–322.
- WRIGHT, D. H. (1983). Species-energy theory : an extension of species-area theory. *Oikos*, 41(3):496–506.
- WRIGHT, P., BELLAMY, C., HAMILTON, P., SCHOFIELD, H., FINCH, D. et MATHEWS, F. (2021). Characterising the relationship between suitable habitat and gene flow for *Myotis bechsteinii* and *Eptesicus serotinus* in Britain. *Landscape Ecology*, pages 1–10.
- WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2):97–159.
- WRIGHT, S. (1940). Breeding structure of populations in relation to speciation. *The American Naturalist*, 74(752):232–248.
- WRIGHT, S. (1943). Isolation by distance. *Genetics*, 28(2):114–138.
- XIA, Y., BJØRNSTAD, O. N. et GRENFELL, B. T. (2004). Measles metapopulation dynamics : a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281.
- ZELLER, K. A., CREECH, T. G., MILLETTE, K. L., CROWHURST, R. S., LONG, R. A., WAGNER, H. H., BALKENHOL, N. et LANDGUTH, E. L. (2016). Using simulations to evaluate Mantel-based methods for assessing landscape resistance to gene flow. *Ecology and Evolution*, 6(12):4115–4128.
- ZELLER, K. A., JENNINGS, M. K., VICKERS, T. W., ERNEST, H. B., CUSHMAN, S. A. et BOYCE, W. M. (2018). Are all data types and connectivity models created equal? validating common connectivity approaches with dispersal data. *Diversity and Distributions*, 24(7):868–879.
- ZELLER, K. A., MCGARIGAL, K. et WHITELEY, A. R. (2012). Estimating landscape resistance to movement : a review. *Landscape Ecology*, 27(6):777–797.
- ZERO, V. H., BAROCAS, A., JOCHIMSEN, D. M., PELLETIER, A., GIROUX-BOUGARD, X., TRUMBO, D. R., CASTILLO, J. A., EVANS MACK, D., LINNELL, M. A., PIGG, R. M. *et al.* (2017). Complementary network-based approaches for exploring genetic structure and functional connectivity in two vulnerable, endemic ground squirrels. *Frontiers in Genetics*, 8:81.
- ZETTERBERG, A., MÖRTBERG, U. M. et BALFORS, B. (2010). Making graph theory operational for landscape ecological assessments, planning, and design. *Landscape and Urban Planning*, 95(4):181–191.
- ZIÓŁKOWSKA, E., OSTAPOWICZ, K., KUEMMERLE, T., PERZANOWSKI, K., RADELOFF, V. C. et KOZAK, J. (2012). Potential habitat connectivity of European bison (*Bison bonasus*) in the Carpathians. *Biological Conservation*, 146(1):188–196.
- ZURELL, D., BERGER, U., CABRAL, J. S., JELTSCH, F., MEYNARD, C. N., MÜNKEMÜLLER, T., NEHRBASS, N., PAGEL, J., REINEKING, B., SCHRÖDER, B. *et al.* (2010). The virtual ecologist approach : simulating data and observers. *Oikos*, 119(4):622–635.



**Thèse de doctorat de l'établissement
Université Bourgogne Franche-Comté**

École Doctorale n° 554 Environnements - Santé

Spécialité : Biologie des populations et écologie

Par

Paul SAVARY

**Utilisation conjointe de graphes génétiques et paysagers
pour l'analyse de la connectivité écologique des habitats**

Annexes B - Méthodes d'analyse statistique

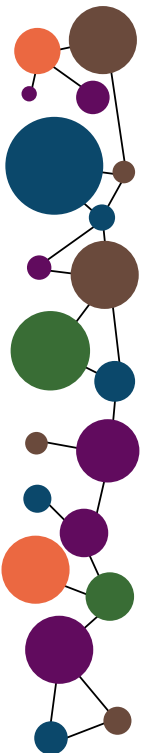
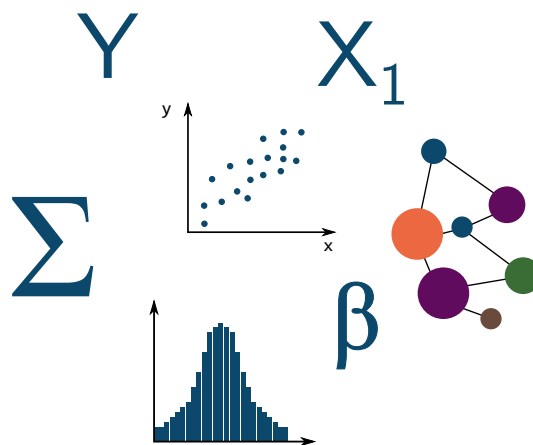


Table des matières

	Page
Table des matières	iii
Annexe B1 Utilisation du principe d'indépendance conditionnelle pour la construction de graphes génétiques	1
Annexe B2 La régression <i>Partial Least Squares</i> (PLS-R)	9
1 Introduction	9
2 Principe et expression du modèle	9
3 Évaluation de la qualité du modèle et choix du nombre de facteurs considérés	11
4 Évaluation de la significativité des effets des variables explicatives	12
Annexe B3 Les modèles de distribution d'espèces	15
1 Introduction	15
2 Création d'un SDM à partir d'une régression logistique	16
3 Qualité du modèle et choix d'un seuil de prédiction	17
Annexe B4 Les arbres de régression	21
1 Introduction	21
2 Principe de construction d'un arbre de régression	21
3 Élagage d'un arbre de régression	23
Annexe B5 Les modèles gravitaires	25
1 Introduction	25
2 Utilisation des modèles gravitaires en génétique du paysage	26
Bibliographie	29

Annexe B1

Utilisation du principe d'indépendance conditionnelle pour la construction de graphes génétiques

Le principe d'indépendance conditionnelle repose sur le fait que deux événements ou deux variables sont conditionnellement indépendants s'ils sont statistiquement indépendants après avoir pris en compte un troisième événement ou une troisième variable (Magwene, 2001). Un graphe d'indépendance est un graphe qui représente les relations d'indépendance conditionnelle entre un ensemble de variables (Magwene, 2001).

En génétique des populations, ce type de graphe a pour la première fois été utilisé par Dyer et Nason (2004). Dans ce cas, les "variables" sont des populations et la série des valeurs de chaque "variable" est la série de fréquences alléliques caractérisant la population. Créer un graphe d'indépendance génétique consiste à identifier les paires de populations qui peuvent être considérées indépendantes une fois que toutes leurs relations avec les autres populations ont été prises en compte. Nous présentons ici les bases mathématiques et les étapes de calcul permettant la construction de ce type de graphe (Figure 1).

Soit \mathbf{Y} un ensemble de p variables suivant une distribution normale multivariée : $\mathbf{Y} = \{y_1, y_2, \dots, y_p\}$. Les trois postulats suivants sont équivalents (Krzanowski et Marriott, 1995, in Magwene, 2001) :

- Les variables y_1 et y_2 sont indépendantes, conditionnellement à \mathbf{Y}_K , avec \mathbf{Y}_K tout sous-ensemble de \mathbf{Y} n'incluant ni y_1 ni y_2 .
- La corrélation partielle entre y_1 et y_2 est nulle : $\rho_{ij.\{K\}} = 0$
- Soit \mathbf{C} la matrice de covariance de l'ensemble des variables \mathbf{Y} , alors l'élément π_{ij} de la matrice de covariance inverse $\mathbf{\Pi} = \mathbf{C}^{-1}$ (aussi appelée matrice de précision) est nul.

Ainsi, pour évaluer les relations d'indépendance conditionnelle d'un ensemble de populations, il faut tout d'abord calculer une matrice de corrélation partielle ou de précision à partir de données génétiques. En génétique des populations, les génotypes pour plusieurs *loci* des individus des différentes populations sont codés sous la forme d'une matrice ayant autant de colonnes que d'allèles et de lignes que d'individus (Figure 1A). L'absence d'un allèle est codée par un 0. La présence d'une ou deux copies d'un allèle dans le génotype d'un individu est codée par un 0.5 ou un 1, respectivement. Si ces données sont codées avec les valeurs 0, 1 et 2, comme dans les travaux de Fortuna *et al.* (2009) et Smouse et Peakall (1999), cela n'affecte pas le calcul.

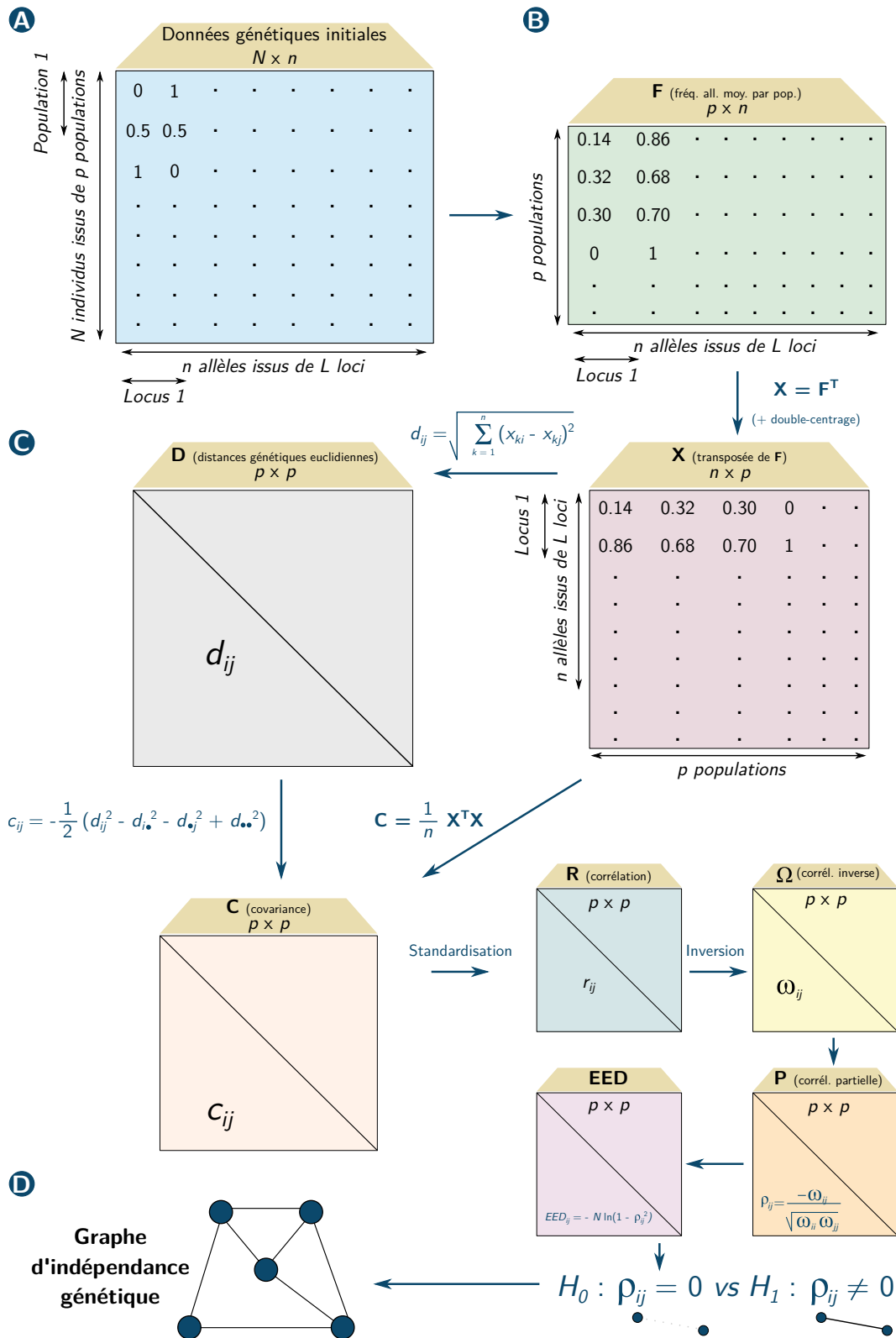


FIGURE 1 – Construction d'un graphe génétique à l'aide du principe d'indépendance conditionnelle. À partir des données génétiques initiales (A), on calcule les fréquences alléliques par population (B). (C) On calcule la matrice de covariance à partir de laquelle plusieurs étapes de calcul permettent d'aboutir au critère d'Edge Exclusion Deviance. (D) C'est à partir de ce critère qu'on teste la significativité de la corrélation partielle entre populations pour déterminer la présence ou non de chacun des liens du graphe d'indépendance génétique.

Dans un premier temps, les fréquences alléliques moyennes par population sont calculées. Ces fréquences sont les éléments d'une matrice \mathbf{F} comptant autant de colonnes qu'il y a d'allèles et autant de lignes qu'il y a de populations (Figure 1B). Ces fréquences alléliques forment les séries de valeurs caractérisant chaque population, considérées comme les variables lors de la construction du graphe d'indépendance génétique. L'étape suivante consiste à calculer la covariance entre les populations (entre les lignes de \mathbf{F} donc). Dyer et Nason (2004) calculent cette covariance en commençant par calculer une matrice de distance génétique euclidienne, suivant les travaux de Gower (1966) ayant démontré la dualité qui existe entre la distance et la covariance.

Pour cela, la matrice \mathbf{F} des fréquences alléliques moyennes par population doit être centrée à la fois par lignes et par colonnes pour que le calcul de la covariance à partir des distances génétiques lors des étapes ultérieures soit correct. Néanmoins, dans ce cas particulier, cette étape n'est pas obligatoire car (i) les sommes des valeurs par ligne sont égales au nombre de *loci* car la somme des fréquences alléliques vaut 1 pour chaque *locus*, et (ii) car le fait de centrer par colonnes n'affecte pas les distances euclidiennes entre les populations (lignes), compte tenu de la définition d'une distance euclidienne. Sans ce "double-centrage", la matrice de covariance entre populations calculée à partir des distances génétiques est équivalente à la matrice de covariance entre les colonnes de la transposée \mathbf{X} de la matrice doublement centrée \mathbf{F} des fréquences alléliques. Nous démontrons cela par la suite. Nous démontrons également pourquoi la covariance doit être calculée à partir de distances au carré et non pas de distances simples, d'un point de vue strictement mathématique, en nous appuyant sur les travaux d'Everitt et Hothorn (2011) (page 107), de Gower (1966) et Smouse et Peakall (1999) (équation 13) (Figure 1C).

La distance génétique euclidienne d_{ij} entre les populations i et j est calculée à partir de la transposée \mathbf{X} de la matrice \mathbf{F} des fréquences alléliques. \mathbf{X} est de dimension $n \times p$, avec n le nombre d'allèles et p le nombre de populations. La distance génétique est calculée avec la formule suivante :

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ki} - x_{kj})^2} \quad (1.1)$$

La covariance c_{ij} entre les variables/populations i et j se calcule avec la formule suivante :

$$c_{ij} = \frac{1}{n} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad (1.2)$$

Comme \mathbf{F} a été centrée à la fois par lignes et par colonnes, $\bar{x}_i = \bar{x}_j = 0$. Ainsi, la covariance entre les variables/populations i et j est simplement :

$$c_{ij} = \frac{1}{n} \sum_{k=1}^n x_{ki}x_{kj} \quad (1.3)$$

Et par conséquent :

$$\begin{aligned} c_{ii} &= \frac{1}{n} \sum_{k=1}^n x_{ki}^2 \\ c_{jj} &= \frac{1}{n} \sum_{k=1}^n x_{kj}^2 \end{aligned} \tag{1.4}$$

Il s'ensuit que la matrice de covariance \mathbf{C} est :

$$\mathbf{C} = \frac{1}{n} \mathbf{X}^T \mathbf{X} \tag{1.5}$$

tel que \mathbf{X}^T est de taille $p \times n$, \mathbf{X} de taille $n \times p$ et \mathbf{C} de taille $p \times p$.

La somme des éléments de chaque ligne de \mathbf{C} vaut ainsi :

$$\begin{aligned} \sum_{j=1}^p c_{ij} &= \sum_{j=1}^p \frac{1}{n} \sum_{k=1}^n x_{ki} x_{kj} \\ &= \frac{1}{n} \left[\left(\sum_{k=1}^n x_{ki} x_{k1} \right) + \left(\sum_{k=1}^n x_{ki} x_{k2} \right) + \dots + \left(\sum_{k=1}^n x_{ki} x_{kp} \right) \right] \\ &= \frac{1}{n} [(x_{1i} x_{11} + x_{2i} x_{21} + \dots + x_{ni} x_{n1}) + \dots + (x_{1i} x_{1p} + x_{2i} x_{2p} + \dots + x_{ni} x_{np})] \\ &= \frac{1}{n} \left[x_{1i} \times \left(\sum_{j=1}^p x_{1j} \right) + x_{2i} \times \left(\sum_{j=1}^p x_{2j} \right) + \dots + x_{ni} \times \left(\sum_{j=1}^p x_{nj} \right) \right] \\ &= \frac{1}{n} [x_{1i} \times 0 + x_{2i} \times 0 + \dots + x_{ni} \times 0] \\ &= 0 \end{aligned} \tag{1.6}$$

car les sommes des lignes de \mathbf{X} sont nulles étant donné que \mathbf{F} a été centrée par lignes et par colonnes.

La trace T de \mathbf{C} est :

$$T = \sum_{i=1}^p c_{ii} \tag{1.7}$$

Exprimons d_{ij}^2 en fonction des éléments de \mathbf{C} :

$$\begin{aligned} d_{ij}^2 &= \sum_{k=1}^n (x_{ki} - x_{kj})^2 \\ &= \sum_{k=1}^n (x_{ki}^2 - 2x_{ki}x_{kj} + x_{kj}^2) \\ &= \sum_{k=1}^n x_{ki}^2 + \sum_{k=1}^n x_{kj}^2 - 2 \sum_{k=1}^n x_{ki}x_{kj} \\ &= n \times (c_{ii} + c_{jj} - 2c_{ij}) \end{aligned} \tag{1.8}$$

On a alors :

$$\begin{aligned}
\sum_{i=1}^p d_{ij}^2 &= \sum_{i=1}^p n \times (c_{ii} + c_{jj} - 2c_{ij}) \\
&= n \times \left(\sum_{i=1}^p c_{ii} + \sum_{i=1}^p c_{jj} - 2 \sum_{i=1}^p c_{ij} \right)
\end{aligned} \tag{1.9}$$

Vu que $\sum_{j=1}^p c_{ij} = 0$ et que \mathbf{C} est une matrice symétrique, $\sum_{i=1}^p c_{ij} = 0$. On a alors :

$$\begin{aligned}
\sum_{i=1}^p d_{ij}^2 &= n \times (T + pc_{jj} - 2 \times 0) \\
&= n \times (T + pc_{jj}) \\
\sum_{j=1}^p d_{ij}^2 &= n \times (T + pc_{ii})
\end{aligned} \tag{1.10}$$

On peut alors calculer $\sum_{i=1}^p \sum_{j=1}^p d_{ij}^2$:

$$\begin{aligned}
\sum_{i=1}^p \sum_{j=1}^p d_{ij}^2 &= \sum_{i=1}^p \sum_{j=1}^p n \times (c_{ii} + c_{jj} - 2c_{ij}) \\
&= n \times \left(\sum_{i=1}^p \sum_{j=1}^p c_{ii} + \sum_{i=1}^p \sum_{j=1}^p c_{jj} - 2 \sum_{i=1}^p \sum_{j=1}^p c_{ij} \right) \\
&= n \times (pT + pT - 2 \times 0) \\
&= n \times 2pT
\end{aligned} \tag{1.11}$$

On calcule ensuite $d_{i\bullet}^2$, $d_{\bullet j}^2$ et $d_{\bullet\bullet}^2$:

$$\begin{aligned}
d_{i\bullet}^2 &= \frac{1}{p} \sum_{j=1}^p d_{ij}^2 \\
&= \frac{1}{p} \times n \times (T + pc_{ii}) \\
&= n \times \left(\frac{T}{p} + c_{ii} \right) \\
&= n \times \left(\frac{1}{p} \sum_{i=1}^p c_{ii} + c_{ii} \right) \\
d_{\bullet j}^2 &= n \times \left(\frac{1}{p} \sum_{i=1}^p c_{ii} + c_{jj} \right) \\
d_{\bullet\bullet}^2 &= \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^p d_{ij}^2 \\
&= \frac{n}{p^2} \times 2pT \\
&= 2 \times \frac{n}{p} \sum_{i=1}^p c_{ii}
\end{aligned} \tag{1.12}$$

Il découle de la formule utilisée pour calculer la distance euclidienne que :

$$\begin{aligned}
d_{ij}^2 &= n \times (c_{ii} + c_{jj} - 2c_{ij}) \\
c_{ij} &= -\frac{1}{2} \left(\frac{d_{ij}^2}{n} - c_{ii} - c_{jj} \right) \\
&= -\frac{1}{2n} (d_{ij}^2 - n \times c_{ii} - n \times c_{jj}) \\
&= -\frac{1}{2n} \left(d_{ij}^2 - n \times c_{ii} - n \times \frac{1}{p} \sum_{i=1}^p c_{ii} - n \times c_{jj} - n \times \frac{1}{p} \sum_{i=1}^p c_{ii} + 2n \times \frac{1}{p} \sum_{i=1}^p c_{ii} \right) \quad (1.13) \\
&= -\frac{1}{2n} \left[d_{ij}^2 - n \times \left(\frac{1}{p} \sum_{i=1}^p c_{ii} + c_{ii} \right) - n \times \left(\frac{1}{p} \sum_{i=1}^p c_{ii} + c_{jj} \right) + 2n \times \frac{1}{p} \sum_{i=1}^p c_{ii} \right] \\
&= -\frac{1}{2n} (d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2)
\end{aligned}$$

Ainsi, pour se conformer à la définition de la covariance, c_{ij} doit être calculée à partir de distances génétiques au carré bien que [Dyer et Nason \(2004\)](#) utilisent la formule suivante dans le package `popgraph` :

$$c_{ij} = -\frac{1}{2} (d_{ij} - d_{i\bullet} - d_{\bullet j} + d_{\bullet\bullet}) \quad (1.14)$$

La division par n dans l'équation (1.13) n'a aucune influence sur les étapes de calcul ultérieures car la matrice de covariance \mathbf{C} est ensuite standardisée pour obtenir la matrice de corrélation \mathbf{R} . Cette matrice de corrélation est inversée pour obtenir la matrice de corrélation inverse $\mathbf{\Omega}$, qui est standardisée à son tour. Les éléments non-diagonaux ω_{ij} de $\mathbf{\Omega}$ sont multipliés par -1 pour obtenir la matrice de corrélation partielle \mathbf{P} tel que ([Magwene, 2001](#)) :

$$\rho_{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \quad (1.15)$$

Enfin, pour déterminer si les populations i et j sont indépendantes conditionnellement à toutes les autres populations, il faut tester si chaque élément ρ_{ij} est significativement différent de 0. Pour cela, le critère d'*Edge Exclusion Deviance (EED)* est calculé selon la méthode de [Whittaker \(2009\)](#) tel que :

$$EED = -N \ln(1 - \rho_{ij}^2) \quad (1.16)$$

avec N le nombre total d'observations (nombre total d'individus ici, suivant [Dyer et Nason \(2004\)](#)).

Lorsque nous avons adapté cette méthode pour l'appliquer, nous avons considéré qu'un graphe d'indépendance génétique devait inclure des liens entre les populations qui sont corrélées positivement pour représenter des flux génétiques directs entre ces populations. Par conséquent, nous avons converti les éléments négatifs de \mathbf{P} en 0 avant de calculer l'*EED*, alors que ce n'était pas le cas dans la méthode originale de [Dyer et Nason \(2004\)](#). Sans cela, une corrélation partielle négative conduisait à la même valeur d'*EED* qu'une corrélation partielle positive de même valeur absolue.

L'*EED* a une distribution du χ^2 asymptotique à 1 degré de liberté ([Whittaker, 2009](#)). Cette propriété permet de tester la significativité de chaque valeur d'*EED* et donc de tester l'hypothèse nulle $H_0 : \rho_{ij} = 0$ contre l'hypothèse alternative $H_1 : \rho_{ij} \neq 0$ (Figure 1D). Si H_0 est rejetée, alors il y a un

lien entre les populations i et j sur le graphe d'indépendance génétique.

Un seuil de significativité de 0.05 est communément utilisé lors du test de la significativité de l'*EED*, sans que les p -valeurs ne soient ajustées dans la méthode originale de [Dyer et Nason \(2004\)](#). Cependant, lors de notre utilisation de cette méthode, nous avons ajusté les p -valeurs avec la méthode séquentielle d'[Holm \(1979\)](#) pour limiter le risque d'erreur de type I. En effet, $\frac{p(p-1)}{2}$ tests sont réalisés pour construire un graphe, ce qui augmente ce risque d'erreur et pourrait surestimer le nombre de liens dans le graphe.

N.B. : Bien que nous n'ayons pas inclus ces détails dans cette présentation des bases théoriques de la méthode de construction des graphes génétiques à partir du principe d'indépendance conditionnelle, la méthode `popgraph` telle qu'elle est mise en œuvre dans le package éponyme comporte d'autres différences importantes avec les étapes décrites ici. Par exemple, la matrice des fréquences alléliques initiales est soumise à une ACP et à plusieurs SVD (*Singular Value Decomposition*) sans que l'utilité de ces étapes n'ait pu être élucidée. Par ailleurs, la façon dont les étapes de calcul sont décrites dans l'article de [Dyer et Nason \(2004\)](#) ne coïncide pas avec l'ordre de ces étapes dans le code de la fonction `popgraph`. Une des conséquences majeures de ce détail est que la méthode mise en œuvre par [Fortuna et al. \(2009\)](#) en adaptant directement la méthode décrite dans l'article de [Dyer et Nason \(2004\)](#) dans un autre langage (MATLAB), avec quelques autres ajustements, diffère de façon très importante de la méthode `popgraph` et des bases théoriques décrites ici.

Annexe B2

La régression *Partial Least Squares* (PLS-R)

1 Introduction

La régression *Partial Least Squares* (PLS-R) est une généralisation de la régression linéaire multiple (Wold *et al.*, 2001) et constitue une alternative à la régression en composantes principales lorsque les variables explicatives \mathbf{X} sont fortement colinéaires ou corrélées. C'est également une extension de l'Analyse en Composantes Principales (ACP). En effet, elle est également basée sur une analyse factorielle mais contrairement à l'ACP, elle considère simultanément les variables issues de deux matrices \mathbf{X} et \mathbf{Y} pour réaliser cette factorisation (Long, 2013). Elle est souvent utilisée lorsqu'il y a davantage de variables explicatives que d'observations et permet aussi de modéliser simultanément plusieurs variables réponses \mathbf{Y} . Dans le chapitre 3 de cette thèse, nous n'avons pas présenté les résultats obtenus en modélisant plusieurs indices génétiques simultanément car ils étaient redondants avec ceux obtenus en les modélisant séparément. Ici, nous décrirons la façon dont la PLS-R s'applique à plusieurs variables réponses dans la mesure où le fait de n'en modéliser qu'une seule n'est qu'un cas particulier de cette approche plus globale.

2 Principe et expression du modèle

L'objectif de la PLS-R, comme de toute régression, est d'obtenir une expression de la forme suivante (Figure 2) :

$$\mathbf{Y} = \mathbf{XB} + \epsilon$$

telle que l'erreur ϵ soit minimale. La PLS-R permet cela de façon indirecte dans la mesure où les matrices \mathbf{X} et \mathbf{Y} sont toutes deux soumises à une factorisation dans un premier temps. Elles sont donc exprimées en fonction de facteurs, ou composantes. L'objectif de cette étape, qui fait l'intérêt de la PLS-R, est de maximiser la corrélation entre les facteurs exprimant \mathbf{X} et ceux exprimant \mathbf{Y} . C'est en quelque sorte comme si on réalisait une ACP "supervisée" des deux matrices de manière à ce que les composantes de l'une expliquent au mieux celles de l'autre. La solution optimale peut être obtenue en utilisant l'algorithme NIPALS, qui dépasse l'objet de ce document et est décrit par Tenenhaus (1998). Le plus souvent, les variables sont transformées avant les calculs, par exemple en les centrant et en les réduisant, ou en leur appliquant un logarithme.

Soit N le nombre d'individus, K le nombre de variables explicatives et M le nombre de variables réponses, les factorisations de la matrice \mathbf{X} de dimension $N \times K$ et de la matrice \mathbf{Y} de dimension $N \times M$ permettent d'obtenir les relations suivantes :

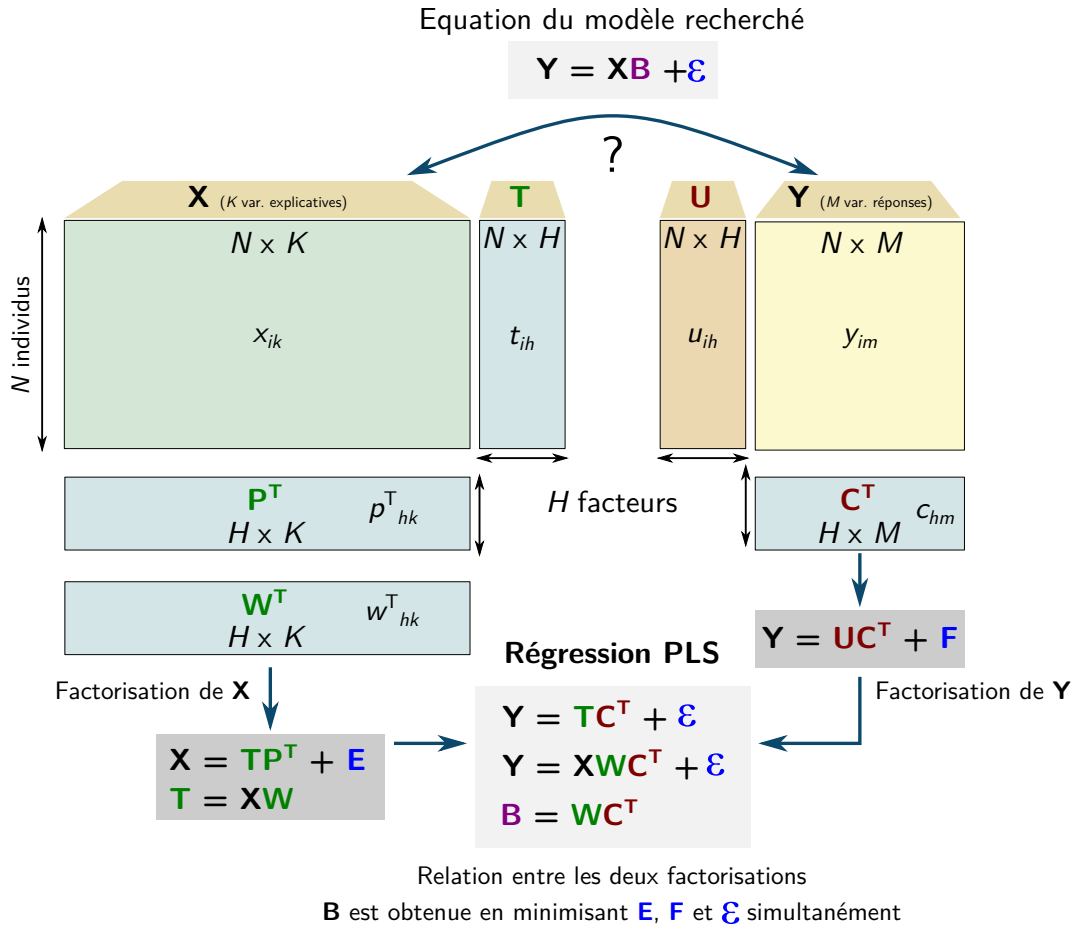


FIGURE 2 – Principe de la régression PLS, adapté à partir de Roy *et al.* (2015)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^T + \mathbf{F}$$

La matrice \mathbf{X} est donc liée aux variables factorielles de la matrice \mathbf{T} ($N \times H$) par les coefficients de la matrice \mathbf{P}^T ($H \times K$)¹ mais ces deux dernières ne permettent pas d'exprimer \mathbf{X} ($N \times K$) parfaitement dans la mesure où il y a un terme résiduel \mathbf{E} ($N \times K$) dans la relation. C'est à partir de la matrice \mathbf{W} ($K \times H$), qui tient compte des résidus, qu'on peut relier \mathbf{X} à \mathbf{T} :

$$\mathbf{T} = \mathbf{X}\mathbf{W}$$

Nous verrons par la suite comment le nombre H de facteurs considérés est déterminé.

Ainsi, les variables factorielles t_1, t_2, \dots, t_h de \mathbf{T} sont des combinaisons linéaires des colonnes initiales de \mathbf{X} et des coefficients de la matrice \mathbf{W} (*weightings*) :

$$t_{ih} = \sum_{k=1}^K w_{kh} x_{ik}$$

Les facteurs de \mathbf{T} et \mathbf{U} sont choisis de façon à maximiser leurs corrélations. On a alors :

1. \mathbf{P}^T est la transposée de \mathbf{P} .

$$u_1 = r_1 t_1$$

$$u_2 = r_2 t_2$$

etc.

tels que r_1 et r_2 sont maximisés. De plus, les facteurs obtenus sont orthogonaux par construction (Tobias *et al.*, 1995). Cela explique pourquoi cette méthode constitue une solution lorsque des variables explicatives sont colinéaires ou corrélées.

Ainsi, les éléments des deux factorisations peuvent être mis en relation pour constituer la relation de régression PLS :

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^T + \epsilon$$

Par construction, cette relation permet de minimiser ϵ . On peut faire apparaître dans cette expression la matrice \mathbf{X} et ainsi comprendre comment la matrice \mathbf{B} a été obtenue :

$$\mathbf{Y} = \mathbf{X}\mathbf{W}\mathbf{C}^T + \epsilon$$

$$\mathbf{B} = \mathbf{W}\mathbf{C}^T$$

3 Évaluation de la qualité du modèle et choix du nombre de facteurs considérés

Nous nous placerons dans l'exemple où la matrice \mathbf{Y} ne contient qu'une seule variable y pour présenter le raisonnement adopté dans les PLS-R réalisées au chapitre 3 de cette thèse. D'après les équations précédentes, la variable y est exprimée comme une combinaison linéaire des variables t_1, t_2, \dots, t_h issues de la matrice \mathbf{T} , elle-même liée à la matrice \mathbf{X} par les coefficients de la matrice \mathbf{W} . On a donc :

$$\hat{y} = c_1 t_1 + c_2 t_2 + \dots + c_H t_H$$

avec c_1, c_2, \dots, c_H les coefficients de la régression issus de la matrice \mathbf{C} et t_1, t_2, \dots, t_H les facteurs issus de \mathbf{T} , tels que :

$$t_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p$$

On a donc :

$$\begin{aligned} y &= c_1 w_{11}x_1 + c_1 w_{12}x_2 + \dots + c_1 w_{1p}x_p + \\ & c_2 w_{21}x_1 + c_2 w_{22}x_2 + \dots + c_2 w_{2p}x_p + \\ & \dots \\ & c_H w_{H1}x_1 + c_H w_{H2}x_2 + \dots + c_H w_{Hp}x_p \end{aligned} \tag{2.1}$$

La question du nombre H de facteurs à considérer se pose alors. Pour le déterminer, pour chaque valeur h , un modèle à h facteurs est calculé, soit (i) à partir de toutes les observations, soit (ii) à partir

d'un sous-ensemble d'observations. Une observation peut être laissée de côté (*Leave One Out cross Validation*) ou un bloc entier représentant une proportion $\frac{1}{k}$ des observations (*k-fold cross validation*). À partir de ces modèles, les valeurs de y sont prédites. On obtient :

- \hat{y}_{hi} , prédiction de y_i à partir du modèle à h facteurs basé sur toutes les observations,
- $\hat{y}_{h(-i)}$, prédiction de y_i à partir du modèle à h facteurs calibré sans intégrer l'observation i .

La qualité des modèles est évaluée selon deux critères :

$$RSS_h = \sum_{i=1}^N (y_i - \hat{y}_{hi})^2$$

et

$$PRESS_h = \sum_{i=1}^N (y_i - \hat{y}_{h(-i)})^2$$

qu'on appelle respectivement la somme des carrés résiduels (*Residual Sum of Squares*, RSS) et la somme des carrés des erreurs de prédiction (*PRediction Error Sum of Squares*, PRESS). Considérer un facteur supplémentaire est pertinent si :

$$\sqrt{PRESS_h} \leq 0.95 \times \sqrt{RSS_{h-1}}$$

ce qui signifie qu'en ajoutant ce facteur, l'erreur de prédiction est inférieure à 90.25 % de la valeur de l'erreur du modèle obtenu sans ajouter ce facteur :

$$PRESS_h \leq 0.9025 \times RSS_{h-1}$$

On retrouve :

$$\frac{PRESS_h}{RSS_{h-1}} \leq 0.9025$$

et

$$1 - \frac{PRESS_h}{RSS_{h-1}} \geq 0.0975$$

Le critère Q^2 est égal à :

$$Q^2 = 1 - \frac{PRESS_h}{RSS_{h-1}}$$

La valeur de Q^2 est calculée pour chaque facteur h ajouté à un modèle. On considère qu'un facteur a un effet significatif sur le modèle s'il améliore la prédiction de y et donc d'après ce critère si $Q^2 \geq 0.0975$ (Tenenhaus, 1998).

4 Évaluation de la significativité des effets des variables explicatives

Les valeurs w_{kh} indiquent le poids qu'a chaque variable explicative k sur le facteur h . Elles permettent de comprendre quelles variables explicatives influencent les variables réponses et l'ampleur de cette influence. Ces contributions permettent de projeter les variables explicatives et les variables réponses dans l'espace constitué par les facteurs t_1, t_2, \dots, t_H pour comprendre les relations entre les variables. Par ailleurs, la significativité des poids w_{kh} des variables explicatives peut être validée par *bootstrap* (Pérez-Rodríguez *et al.*, 2018). Pour cela, le jeu de données est échantillonné aléatoirement et avec remise un grand nombre de fois et à chaque fois les poids w_{kh} sont estimés. Si l'intervalle

contenant 95 % des valeurs (2.5-97.5 %) n'inclut pas 0, alors le poids est significativement différent de 0 et on considère que la variable k a un effet significatif sur le facteur h . Si ce facteur a lui-même un effet significatif sur y , alors on pourra interpréter la relation entre la variable explicative k et la variable réponse y .

Annexe B3

Les modèles de distribution d'espèces

1 Introduction

De manière générale, une espèce s'observe dans un type de milieu particulier, qui lui est propre. C'est de ce constat que découle le concept de niche écologique. Si on confronte les observations de la présence ou de l'absence d'une espèce aux conditions environnementales dans lesquelles on a réalisé ces observations, on peut comprendre quelles sont les conditions nécessaires à la présence de cette espèce. Ces conditions sont identifiées d'une façon d'autant plus fiable et précise que l'on a réalisé un grand nombre d'observations, aussi bien de présences que d'absences, dans des conditions variées et qu'on dispose d'un nombre important de variables environnementales pouvant expliquer la présence de l'espèce. Si on dispose d'une carte de ces conditions environnementales à l'échelle de tout un territoire, on peut cartographier les zones au niveau desquelles la présence de l'espèce sera la plus probable à partir des relations identifiées.

Un modèle de distribution d'espèce (*Species Distribution Model*¹, noté SDM par la suite) est précisément un modèle qui relie la présence ou l'absence d'une espèce à des variables environnementales. Il est fréquemment utilisé aujourd'hui car les données d'observation sont de plus en plus nombreuses, les données environnementales (spatiales) de plus en plus accessibles et de nombreuses méthodes de modélisation permettent de s'adapter à chaque contexte (Guisan et Zimmermann, 2000). Leur utilisation permet de prédire l'effet du changement climatique sur la distribution des espèces ou l'expansion d'espèces invasives, entre autres (Fletcher et Fortin, 2018).

Le choix de la méthode de modélisation dépend essentiellement (i) des données d'observation dont on dispose (présence *vs* présence/absence *vs* abondance) et de la forme de la relation existant entre les variables environnementales et les observations (linéaire, non-linéaire, etc.).

Les données de présence sont très nombreuses aujourd'hui, notamment grâce aux programmes de science participative. On peut se contenter de ces observations pour modéliser la distribution d'une espèce avec les méthodes les plus simples. Néanmoins, il est préférable de confronter les conditions environnementales de ces points à d'autres conditions environnementales pour comprendre ce qui détermine la présence de l'espèce. Pour cela, on peut générer aléatoirement des points de *background* (ou

1. On parle aussi d'*Ecological Niche Model* (ENM) ou d'*Habitat Suitability Model* (HSM). Nous retenons ici le terme de SDM car c'est la distribution spatiale d'une espèce dans un contexte donné qui est modélisée. On ne peut pas savoir si cela coïncide véritablement avec sa niche écologique fondamentale ou réalisée, d'autant plus que le concept de niche en lui-même fait débat. Nous éviterons donc ce terme.

de "pseudo-absence"). L'algorithme Maxent est largement utilisé dans ce cas de figure (Elith *et al.*, 2011 ; Phillips *et al.*, 2006).

Sans données concernant l'absence de l'espèce, on ne peut pas connaître la prévalence de l'espèce. En d'autres termes, on ne sait pas quel est le ratio entre le nombre de points de présence et le nombre de points d'absence. Les SDM ne fournissent donc dans ce cas qu'une probabilité de présence relative, la véritable probabilité de présence de l'espèce étant proportionnelle à sa prévalence. De plus, lorsqu'on ne dispose pas de points d'absence, les potentiels biais liés à la distribution spatiale des observations sont difficilement évitables. Prenons l'exemple où les observateurs ne s'éloignent jamais de plus de 100 m des routes pour faciliter leurs déplacements sur le terrain. Si les points de présence sont confrontés à des points de pseudo-absence choisis aléatoirement dans toute la zone d'étude, il est probable que la proximité à la route soit identifiée comme un facteur favorable à la présence de l'espèce. Si à la fois les points de présence et d'absence sont situés à moins de 100 m des routes, ce qui est le cas si les observateurs relèvent également l'absence de l'espèce, ce biais sera évité.

Pour ces différentes raisons, lorsqu'on dispose de données de présence/absence, leur analyse à l'aide d'un modèle approprié est préférable à l'utilisation de l'algorithme Maxent à partir des points de présence uniquement (Guillera-Arroita *et al.*, 2014). Lorsqu'on dispose de ces données et qu'il y a une relation monotone entre les variables environnementales et la présence de l'espèce, la régression logistique est une des méthodes les plus utilisées. C'est celle que nous avons choisie pour réaliser le SDM de la Paruline caféïette dans le cadre de cette thèse et que nous décrivons par la suite. Nous mentionnons à titre indicatif l'existence des modèles suivants :

- *Envelope models* (BIOCLIM par exemple)
- *Generalised Additive Models* (GAM) : adaptés aux relations non-linéaires et non-monotones
- *Classification Trees* ou *Random Forests* : adaptés aux relations non-linéaires et non-monotones et à la prise en compte de variables quantitatives et qualitatives simultanément
- Voir Fletcher et Fortin (2018) et Guisan *et al.* (2017) pour une description de ces méthodes et d'autres.

2 Création d'un SDM à partir d'une régression logistique

La régression logistique fait partie de la grande famille des *Generalised Linear Models* (GLM). Elle permet de modéliser une variable réponse y binaire (0, 1) à partir d'un ensemble de p variables quantitatives et/ou qualitatives x_1, x_2, \dots, x_p renseignant sur les conditions environnementales au niveau de chaque point d'observation. Plus précisément, l'objectif est de modéliser la probabilité de présence sachant les conditions environnementales $P(y = 1|x_1, x_2, \dots, x_p) = p(x)$ à partir des conditions environnementales x_1, x_2, \dots, x_p . Par définition, $p(x)$ est bornée par 0 et 1. L'utilisation d'une relation de la forme $p(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ serait inadéquate car elle pourrait dépasser ces bornes. Une solution consiste à exprimer $p(x)$ de la façon suivante :

$$p(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}$$

Ce ratio est compris entre 0 et 1. La variation de $p(x)$ en réponse à l'augmentation de x_1 , si elle ne dépend que de x_1 , a une forme "en S". En transformant quelque peu cette relation, et en lui appliquant un logarithme, on tombe sur les expressions suivantes :

$$\frac{p(x)}{1-p(x)} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$$

$$\text{logit}(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

La fonction $\log\left(\frac{x}{1-x}\right)$ s'appelle la fonction "logit". $\text{logit}(p(x))$ prend des valeurs continues dépassant l'intervalle $[0, 1]$ quand $p(x)$ est comprise entre 0 et 1. On va modéliser $\text{logit}(p(x))$ en fonction des variables x_1, x_2, \dots, x_p . On retombe donc sur l'expression d'un modèle de régression linéaire multiple. C'est le principe des modèles linéaires généralisés (GLM). Ici, la "fonction de lien" permettant de retrouver un modèle de cette forme est la fonction "logit".

À l'aide d'une méthode le plus souvent basée sur la maximisation de la vraisemblance du modèle, on estime un ensemble de coefficients $\beta_0, \beta_1, \dots, \beta_p$ qui relient $\text{logit}(p(x))$ aux variables x_1, x_2, \dots, x_p . Suite à l'estimation du modèle, on peut calculer la valeur $\text{logit}(p(x))$ puis $p(x)$ pour l'ensemble des observations, et étendre ce calcul à l'ensemble des points pour lesquels on dispose des valeurs des variables x_1, x_2, \dots, x_p . C'est cela qui permet de cartographier la distribution de l'espèce.

Notons qu'en amont de la modélisation, il est particulièrement important de s'assurer que :

- Il n'y a pas de biais dans la distribution spatiale des points considérés.
- Les variables explicatives considérées ne sont pas fortement corrélées ou colinéaires.

3 Qualité du modèle et choix d'un seuil de prédiction

À l'issue de la modélisation, on peut se poser deux questions :

- Le modèle obtenu explique-t-il bien la distribution des observations de présence et d'absence ?
- À partir de quelle probabilité de présence peut-on prédire que l'espèce sera présente ?

Nous allons voir que ces deux questions sont liées et y répondre.

Observation/Prédiction	Présences prédites	Absences prédites
Présences observées	Vrais positifs (TP)	Faux négatifs (FN)
Absences observées	Faux positifs (FP)	Vrais négatifs (TN)

TABLE 1 – Matrice de confusion

On dispose pour chaque point d'une prédiction de probabilité de présence comprise entre 0 et 1 et suivant un gradient continu alors que la variable y initiale était binaire (0, 1). Pour confronter les prédictions aux observations et évaluer la qualité du modèle, il est donc nécessaire de les "binariser". On choisit pour cela une valeur seuil et si la probabilité de présence est supérieure à ce seuil, on prédit la présence de l'espèce ($\hat{y} = 1$). Pour un seuil donné, on obtient deux séries de 0 et de 1, l'une observée et l'autre prédite. Avec un modèle parfait, ces deux séries de valeurs seraient identiques. Mais, c'est très rare en pratique. Un modèle peut prédire la présence d'une espèce là où elle était absente, ou au

contraire son absence là où elle était présente. Il s'agit d'un faux positif et d'un faux négatif, respectivement. La qualité globale du modèle dépendra du nombre d'erreurs de chaque type auxquelles il conduit. Pour quantifier cela, on réalise une matrice de confusion qui contient les nombres de vrais positifs (TP), faux négatifs (FN), faux positifs (FP) et vrais négatifs (TN) (Table 1).

À partir de la matrice de confusion, on peut calculer différents indices pour caractériser la performance du modèle. Parmi eux :

$$\text{Spécificité} = \frac{TN}{TN + FP}$$

c'est-à-dire le taux de prédiction d'une absence lorsqu'on a observé une absence. On retrouve :

$$1 - \text{Spécificité} = \frac{FP}{TN + FP} = \text{Taux de faux positifs (FDR)}$$

On calcule également :

$$\text{Sensibilité} = \frac{TP}{FN + TP} = \text{Taux de vrais positifs (TPR)}$$

c'est-à-dire le taux de prédiction d'une présence lorsqu'on a observé une présence (taux de vrais positifs, TPR).

La valeur de ces indices calculés en "binarisant" les probabilités de présence avec une valeur seuil ne nous renseigne que sur la qualité de la prédiction pour ce seuil particulier. On ne connaît donc pas la qualité du modèle global ni le seuil optimal. Pour cela, on va répéter l'opération en choisissant itérativement des valeurs seuils comprises entre 0 et 1. On obtient une série de valeurs de FDR et de TPR, dont on représente la relation (Figure 3). La courbe obtenue s'appelle la courbe ROC (*Receiver Operating Characteristic*). Plus l'aire sous cette courbe est proche de 1, sa valeur maximale, meilleur est le modèle. Cette aire correspond à l'indice AUC (*Area Under the Curve*), souvent utilisé pour mesurer la qualité d'un SDM. Il est par construction indépendant du seuil choisi.

Pour identifier le seuil optimal au-delà duquel on peut prédire la présence de l'espèce, on utilise également l'indice de corrélation de Matthews (1975) car il tient compte de toutes les valeurs de la matrice de confusion :

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Il varie entre -1 et 1. Son utilisation est plus fréquente en informatique et en *machine learning* qu'en écologie (Baldi *et al.*, 2000). La valeur seuil maximisant cet indice (ou un autre) pourra être utilisée pour cartographier les zones d'habitat de l'espèce étudiée, comme nous l'avons fait pour la Paruline caféïette au chapitre 4 de cette thèse.

Comme tout modèle statistique, un SDM peut être sur-paramétré (*overfitting*). Il expliquera très bien les données utilisées pour le calibrer mais aura une faible valeur prédictive si on l'applique dans un autre contexte. Pour éviter cela, on peut calculer ces mêmes indices en considérant un ensemble d'observations non incluses lors de la calibration du modèle (données de validation). C'est en faisant cela que nous avons pu nous assurer que les coefficients associés aux variables environnementales n'étaient

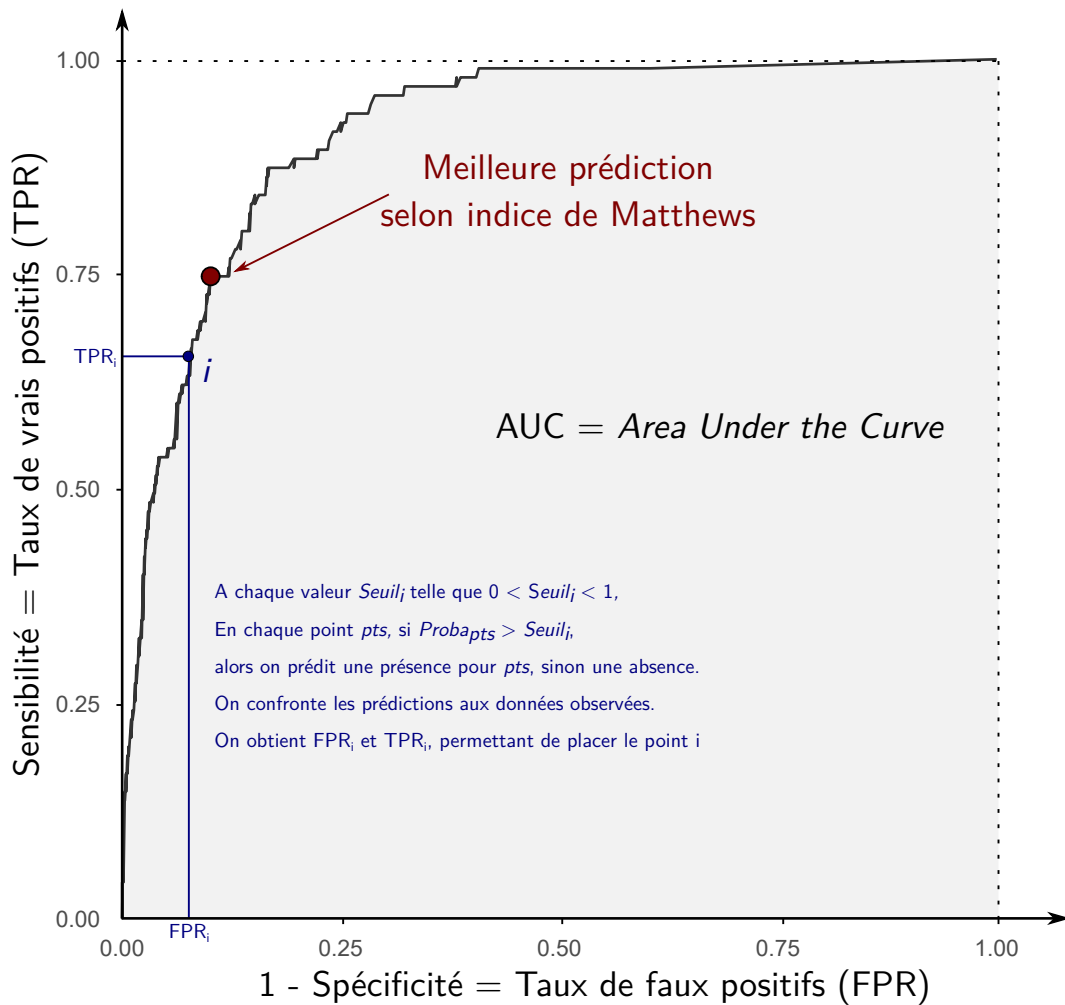


FIGURE 3 – Évaluation de la qualité d'un modèle de distribution d'espèce à l'aide de la courbe ROC (*Receiver Operating Characteristic*) et de l'AUC (*Area Under the Curve*). Le modèle fournit pour chaque point d'observation une estimation de probabilité de présence $Proba_{pts}$ entre 0 et 1. En choisissant itérativement des valeurs $Seuil_i$, on considère que si $Proba_{pts} > Seuil_i$, alors le point est classé comme un point de présence, et sinon comme une absence. On confronte ces résultats aux observations et on calcule les taux de vrais positifs (TPR) et de faux positifs (FPR) correspondants. Cela permet de placer le point i correspondant au $Seuil_i$. L'ensemble de ces points forment la courbe ROC. L'aire sous la courbe vaut au maximum 1 et plus elle se rapproche de 1, meilleur est le modèle.

pas biaisés par le jeu de données considéré dans le chapitre 4, et mettre en évidence l'importance du ré-échantillonnage des points d'observation de Basse-Terre.

Enfin, pour limiter le risque d'*overfitting* lors de la calibration du modèle, il est possible d'appliquer une régularisation aux coefficients du modèle. Pour ce faire, le package `glmnet` (Friedman *et al.*, 2021) permet d'utiliser la méthode *elastic net*, intermédiaire entre les méthodes *lasso* et *ridge* (Hastie *et al.*, 2017). Nous l'avons utilisée pour réaliser le SDM de la Paruline caféïette mais (i) l'amélioration des performances du modèle était négligeable et (ii) son utilisation ne reflétait pas la façon dont les SDM sont classiquement construits en amont de la construction d'un graphe paysager. Nous n'avons donc pas régularisé les coefficients de ce SDM.

Annexe B4

Les arbres de régression

1 Introduction

Au sein des méthodes statistiques basées sur les arbres, on distingue les arbres de régression et les arbres de classification. Tous deux sont parfois appelés "arbres de décision". Les premiers constituent des alternatives aux méthodes de régression, linéaires ou non, utilisées pour modéliser des variables réponses quantitatives. Les seconds constituent des alternatives aux méthodes de modélisation de variables qualitatives, telles que la régression logistique binaire ou multinomiale ou l'analyse discriminante, entre autres.

Nous avons modélisé des variables quantitatives avec des arbres de régression dans les chapitres 5 et 6 de cette thèse. Nous décrirons donc les bases de cette méthode. Nous n'aborderons pas le *bagging*, le *boosting* et les *random forests*, extensions de ces méthodes basées sur des approches de *machine learning*. Elles sont davantage adaptées à la prédiction des valeurs de la variable réponse qu'à l'interprétation de l'influence qu'ont des variables explicatives sur cette variable, ce qui ne correspondait pas aux objectifs poursuivis dans les chapitres 5 et 6. Ces méthodes sont décrites dans l'ouvrage de [James et al. \(2013\)](#), sur lequel nous nous basons dans les sections qui suivent.

2 Principe de construction d'un arbre de régression

Dans un modèle de régression linéaire, le lien entre la variable réponse Y et une variable explicative X_1 est de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

avec β_0 et β_1 deux constantes et ϵ les résidus non expliqués par le modèle, égaux à $Y - \hat{Y}$. Le modèle se base donc sur une relation de proportionnalité entre Y et X_1 .

Le principe d'un arbre de régression est très différent dans la mesure où il ne repose pas sur une relation de proportionnalité entre les variables. En effet, l'espace constitué par les variables explicatives X_1, X_2, \dots, X_p va être scindé en J régions (R_1, R_2, \dots, R_J) distinctes et non-chevauchantes en fonction des valeurs de ces différentes variables. Chaque région R_j constituera une feuille de l'arbre. La valeur \hat{Y}_{R_j} prédite par le modèle pour les observations appartenant à la région R_j sera simplement égale à la moyenne des valeurs de Y dans cette région. Les critères permettant de scinder l'espace des variables X_1, X_2, \dots, X_p en J régions ou feuilles seront représentés sous la forme d'un arbre, rendant

le modèle facilement interprétable (Figure 4).

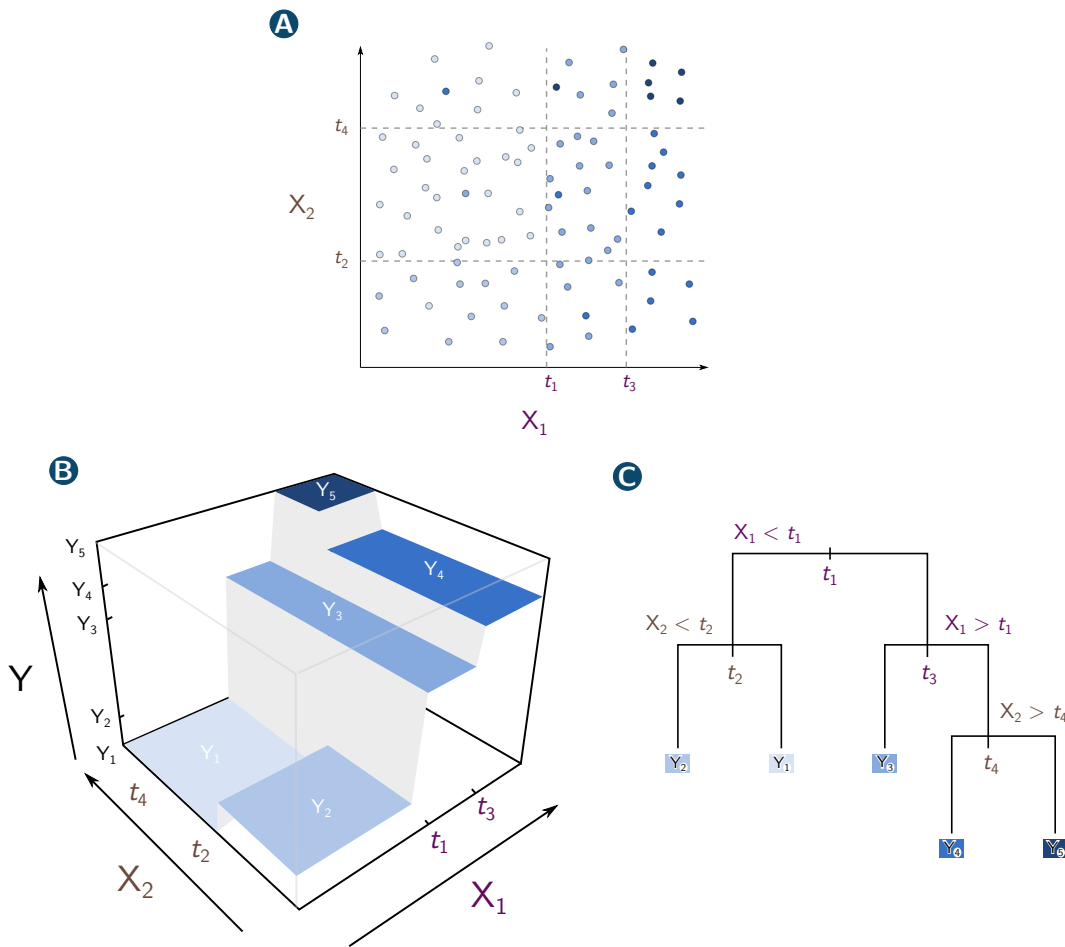


FIGURE 4 – Principe des arbres de régression, illustration adaptée à partir de [Hastie et al. \(2017\)](#). (A) Données d'entrée. Plus la couleur des points est foncée, plus la variable réponse Y prend une valeur forte. Le lien entre Y et les variables explicatives X_1 et X_2 n'est pas linéaire, ce qui justifie l'utilisation d'un arbre de régression pour modéliser Y . (B) Séparation optimale de l'espace formé par les variables Y , X_1 et X_2 en plusieurs régions (feuilles de l'arbre) au niveau desquelles Y prend des valeurs relativement homogènes en fonction des classes de valeurs respectives des variables X_1 et X_2 . (C) Arbre de régression représentant la séparation optimale en feuilles en fonction des classes de valeurs des variables X_1 et X_2 . En bas de l'arbre, on trouve pour chaque feuille la valeur moyenne de Y pour les observations appartenant à la feuille. Ces valeurs se retrouvent également en (B). Chaque séparation d'une branche de l'arbre en deux branches sépare les observations selon la classe à laquelle appartient la valeur qu'elles prennent pour les variables X_1 ou X_2 .

La qualité du modèle s'évalue à partir de la somme des carrés des erreurs (SCE), calculée de la façon suivante :

$$SCE = \sum_{j=1}^J \sum_{i \in R_j} (Y_i - \hat{Y}_{R_j})^2$$

avec Y_i la valeur de Y pour l'observation i , \hat{Y}_{R_j} la valeur de Y prédite pour l'observation i appartenant à la région R_j , c'est-à-dire la moyenne de Y au sein de cette région. Le meilleur modèle est identifié en cherchant les régions R_1, R_2, \dots, R_J qui permettent de minimiser la SCE .

Comme il y a une infinité de façons de scinder l'espace des p variables explicatives en J régions distinctes et non chevauchantes, ces régions vont être identifiées de façon récursive et binaire (*recursive binary splitting*). L'arbre est représenté la tête en bas (feuilles en bas) et construit de haut en bas,

donc du tronc (au niveau duquel toutes les observations appartiennent à la même région unique) vers les feuilles. Les branches partant du tronc se séparent au niveau de nœuds internes en d'autres branches jusqu'aux feuilles terminales. Chaque nœud interne donne naissance à deux régions selon un critère basé sur une des variables explicatives X . Par exemple, le critère donnant naissance aux deux premières régions R_1 et R_2 dépend de la variable X_j et d'une valeur seuil s :

$$R_1(j, s) = \{X|X_j < s\} \text{ et } R_2(j, s) = \{X|X_j \geq s\}$$

ce qui signifie qu'une observation i appartient à R_1 si $X_{ij} < s$ et à R_2 sinon. Pour construire l'arbre, on cherche les valeurs j et s qui minimisent la valeur suivante :

$$\sum_{i: X_i \in R_1(j, s)} (Y_i - \hat{Y}_{R_1})^2 + \sum_{i: X_i \in R_2(j, s)} (Y_i - \hat{Y}_{R_2})^2$$

Le résultat est obtenu assez rapidement. Cette étape est répétée, ce qui scinde à nouveau une région en deux nouvelles régions de façon à faire diminuer la valeur totale de SCE . L'arbre est construit ainsi. La figure 4 illustre le résultat que l'on peut obtenir à l'issue de ces étapes.

3 Élagage d'un arbre de régression

En pratique, il est toujours possible de créer de nouvelles régions et d'obtenir un arbre avec plus de branches et de feuilles terminales. Cela a plusieurs inconvénients. Tout d'abord, l'arbre obtenu est difficile à interpréter. De plus, les feuilles terminales incluent peu d'observations ce qui peut conduire à sur-interpréter les résultats obtenus. Enfin, l'arbre a une valeur prédictive faible s'il est trop complexe car dans ce cas, il est sur-paramétré (*overfitting*). Il existe plusieurs façons d'éviter cela.

Tout d'abord, on peut arrêter la construction de l'arbre lorsque la diminution de la SCE résultant de la création de nouvelles régions devient très faible. L'inconvénient est que même si un ajout donné fait peu diminuer la SCE , cela ne signifie pas que le suivant ne la fera pas diminuer davantage. Pour cette raison, on construit en général un très grand arbre T_0 qui est ensuite élagué pour réduire son nombre de branches et de feuilles. La construction de T_0 s'arrête lorsqu'une feuille contient un nombre d'observations inférieur à un certain seuil. Augmenter cette valeur seuil est d'ailleurs en soi une façon de réduire la taille de l'arbre. C'est ainsi que nous avons procédé au chapitre 6, car nous voulions qu'il y ait au minimum 40 observations par feuille pour pouvoir tester la significativité de la différence de ces valeurs par rapport à 0.

Pour élaguer T_0 , on pourrait faire une validation croisée en scindant les données en deux jeux de données, de calibration d'une part et de validation d'autre part. On estimerait alors l'erreur de prédiction des données de validation associée à chaque sous-arbre. Néanmoins, cette approche serait extrêmement longue. Il existe une alternative basée sur un principe de "coût complexité" (*cost complexity pruning*). Nous l'avons utilisée au chapitre 5. Elle consiste à évaluer la qualité des sous-arbres en calculant leur SCE et en la pénalisant par le nombre de feuilles qu'ils contiennent à l'aide d'un paramètre α . On cherche ainsi à minimiser pour chaque sous-arbre $T \subset T_0$:

$$\sum_{m=1}^{|T|} \sum_{i: x_i \in R_m} (y_i - \hat{y}_{R_m})^2 + \alpha |T|$$

avec $|T|$ le nombre de feuilles du sous-arbre T , R_m la région correspondant à la $m^{\text{ième}}$ feuille terminale et \hat{y}_{R_m} la valeur prédite par le sous-arbre pour la feuille R_m . On considère donc chaque observation i et la feuille R_m correspondante. Si $\alpha = 0$, c'est l'arbre T_0 qui permet de minimiser cette somme. À mesure que α augmente, ce sont des arbres plus petits qui permettent de la minimiser puisque le nombre de feuilles $|T|$ est pénalisé plus fortement. L'avantage de cette méthode est que lorsque α augmente, les branches sont élaguées d'une façon prévisible et les sous-arbres sont emboîtés. On peut donc facilement obtenir une séquence de sous-arbres en fonction de α . On se base alors sur une validation croisée pour déterminer la valeur α optimale et on identifie le sous-arbre correspondant.

Annexe B5

Les modèles gravitaires

1 Introduction

Les modèles gravitaires, traduction française des *gravity models*, s’inspirent de la loi de la gravitation universelle de Newton. Selon celle-ci, la force d’attraction F_{ij} qui s’exerce entre deux planètes i et j de masses m_i et m_j séparées par une distance d_{ij} a pour norme (en Newton) (Figure 5A) :

$$F_{ij} = F_{ji} = G \times \frac{m_i m_j}{d_{ij}^2}$$

tel que G est la constante universelle de gravitation.

Cette formule peut également s’exprimer de la façon suivante :

$$F_{ij} = F_{ji} = G \times m_i^o \times m_j^p \times d_{ij}^{2q}$$

avec $o = p = 1$ et $q = -1$.

Les modèles gravitaires s’inspirent de cette loi pour modéliser un flux ou l’intensité d’une interaction entre deux entités spatialement distinctes en fonction de la distance entre elles et des caractéristiques propres à chacune des entités. L’objectif est alors de déterminer les valeurs des exposants (o, p, q dans l’exemple). Ils ont été utilisés dans de nombreux domaines, en économie (marketing spatial, flux monétaires) ou en géographie (estimation de flux de migrants, réseaux de transport) en particulier. Ils n’ont été utilisés en écologie qu’après 2000, avec des applications à la dispersion de maladies (Ferrari *et al.*, 2006 ; Xia *et al.*, 2004) et à la connectivité (Bossenbroek *et al.*, 2001, 2007 ; Kong *et al.*, 2010).

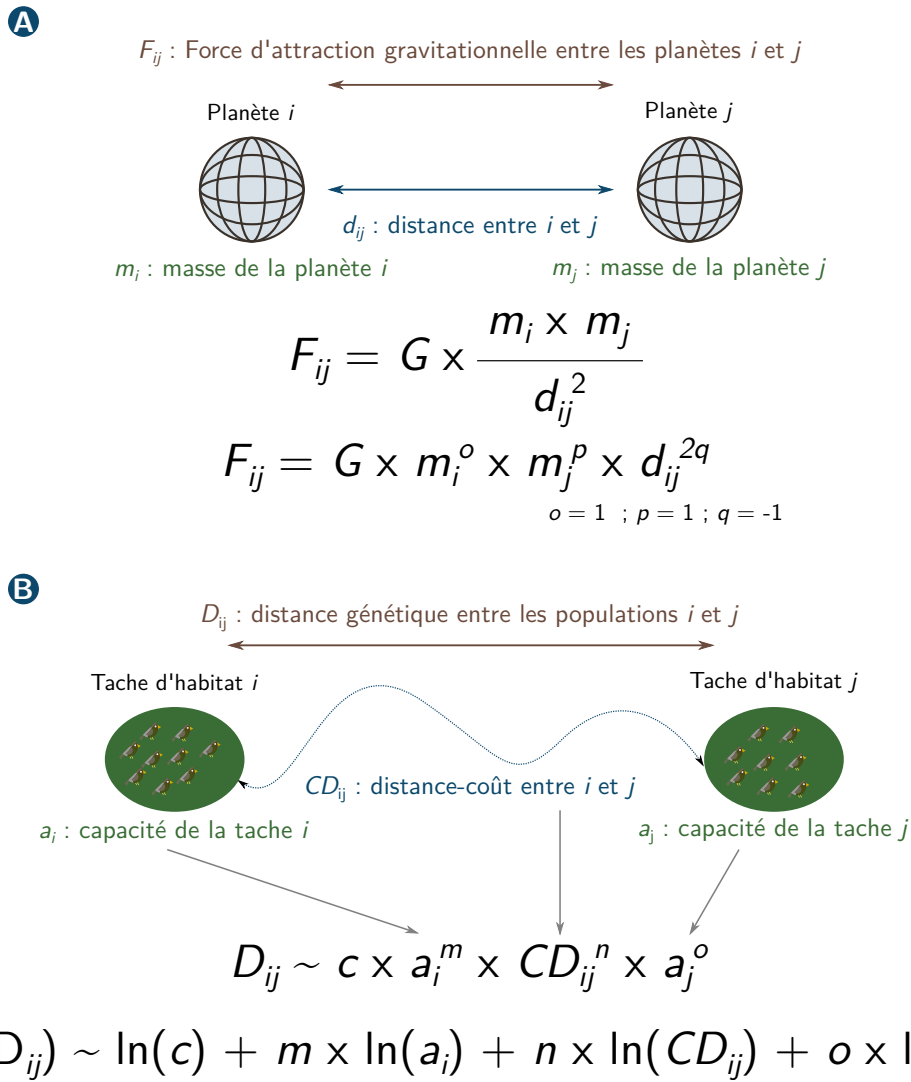


FIGURE 5 – Principe de l'utilisation des modèles gravitaires en génétique du paysage. (A) Loi de la gravitation universelle de Newton, permettant de calculer la force d'attraction gravitationnelle qui s'exerce entre deux planètes i et j en fonction de leurs masses respectives, de la distance qui les sépare et de la constante de gravitation G . (B) Application d'une loi d'interaction similaire à la loi de la gravitation pour expliquer la distance génétique entre deux populations i et j à partir des capacités de leurs taches d'habitat respectives et de la distance-coût qui les sépare. Le modèle est d'abord exprimé de façon similaire à la seconde notation de la loi de la gravitation universelle (A), puis en lui appliquant un logarithme, pour passer d'un modèle multiplicatif à un modèle additif ayant la forme d'une régression linéaire multiple.

2 Utilisation des modèles gravitaires en génétique du paysage

Les travaux de [Murphy et al. \(2010\)](#)

[Murphy et al. \(2010\)](#) ont introduit ce type de modèle en génétique du paysage dans le cadre d'une étude des déterminants de la structure génétique de populations de grenouilles (*Rana luteiventris*) occupant un réseau de lacs d'altitude. Ces auteurs ont montré grâce à ces modèles que la différenciation génétique entre les populations était influencée positivement par la distance entre les lacs, la présence de poissons dans les lacs et la topographie du paysage qui les sépare. Pour cela, ils ont réalisé un modèle de la forme suivante :

$$T_{ij} \sim kv_i^{\mu} w_j^{\alpha} c_{ij}^{-\beta}$$

tel que :

- T_{ij} est une estimation du flux génétique ($1 - D_{PS}$) entre les populations i et j .
- c_{ij} est une variable de résistance au flux entre les populations. Cette variable pouvait tenir compte de l'*elevation ratio relief*, du ratio entre surfaces de prairies et de forêts, entre autres. Elle est calculée dans une zone tampon (*buffer*) de largeur fixe autour du chemin reliant les populations en ligne droite.
- w_j est la distance géodésique euclidienne entre les sites.
- v_i est la variable caractérisant la population d'origine. Les variables prises en compte étaient notamment l'altitude ou la présence de poissons.
- k , α , μ et β sont les paramètres du modèle à estimer.

Cette expression du modèle n'est qu'un cas particulier car il en existe plusieurs ([Anderson, 1979](#) ; [Fotheringham et O'Kelly, 1989](#)).

Utilisations ultérieures et généralisation de l'approche

Lors des utilisations de modèles gravitaires qui ont fait suite à celle de [Murphy et al. \(2010\)](#) en génétique du paysage ([DiLeo et al., 2014](#) ; [Watts et al., 2015](#) ; [Zero et al., 2017](#)), d'autres expressions ont été utilisées. Ce qui les distingue en particulier est l'inclusion de variables relatives aux populations d'origine ou de destination, ou aux deux. Dans le cadre de cette thèse, nous avons inclus des variables relatives aux deux populations/taches d'habitat entre lesquelles nous avons modélisé la différenciation génétique, à l'instar de la loi de la gravitation universelle (Figure 5B). Notre modèle était donc de la forme suivante :

$$D_{ij} \sim c \times a_i^m \times CD_{ij}^n \times a_j^o$$

avec D_{ij} la distance génétique entre les populations i et j , c , m , n et o des constantes, a_i la capacité (ou l'effectif) de la population i et CD_{ij} la distance-coût entre les taches i et j occupées par les populations i et j . Notre modèle différait donc également de celui de [Murphy et al. \(2010\)](#) par le fait qu'au lieu d'inclure une variable relative à la distance géodésique euclidienne entre les populations et une autre relative à la résistance au flux entre ces populations, nous n'avons inclus qu'une seule variable représentant la distance-coût entre les populations. Cela se justifiait par le fait que nous utilisons les modèles gravitaires dans l'objectif d'optimiser des scénarios de coût. Par ailleurs, nous avons modélisé la distance D_{ij} estimée à l'aide du D_{PS} tandis que [Murphy et al. \(2010\)](#) modélisaient la variable T_{ij} égale à $1 - D_{PS}$.

L'estimation des paramètres du modèle ne peut pas se faire à partir des expressions des modèles multiplicatifs. Le produit est converti en une somme par transformation logarithmique. Dans le cas du modèle de [Murphy et al. \(2010\)](#) et dans notre cas, on obtient alors respectivement (Figure 5B) :

$$\ln T_{ij} \sim \ln k + \mu \ln v_i + \alpha \ln w_j - \beta \ln c_{ij}$$

et

$$\ln D_{ij} \sim \ln c + m \ln a_i + n \ln CD_{ij} + o \ln a_j$$

Ces modèles s'apparentent alors à des régressions linéaires multiples et peuvent être estimés à l'aide de la méthode des moindres carrés ordinaires (*Ordinary Least-Squares*) par exemple. Néanmoins, ils ne respectent pas la condition d'indépendance des observations car plusieurs flux et distances sont

relatifs à une même population. Une des solutions apportées à ce problème dans le cadre de l'utilisation de modèles gravitaires consiste à faire varier la constante k (ou c) en fonction de la population d'origine ou de destination. Le modèle est alors dit *singly constrained* (simplement contraint). C'est la solution utilisée par [Murphy et al. \(2010\)](#). Ces auteurs mentionnent le fait que le modèle est parfois non contraint (*unconstrained*) ou doublement contraint (*doubly constrained*), et les avantages et inconvénients associés à chaque option ([Murphy et al., 2010](#)). Ces différentes options permettent également de fixer la somme totale des flux dans le cadre du paramétrage du modèle gravitaire. Ce point dépasse notre application de ces modèles. Nous renvoyons vers [Bossenbroek et al. \(2001\)](#) pour une méthode plus détaillée intégrant un contrôle des flux totaux.

Bien que nous n'ayons pas utilisé cette terminologie, la méthode que nous avons utilisée permet également de tenir compte de la non-indépendance des observations en estimant une valeur c différente selon une des deux populations impliquées. En effet, nous avons calibré les modèles gravitaires en utilisant des modèles linéaires mixtes de type MLPE ([Clarke et al., 2002](#) ; [Van Strien et al., 2012](#)). Nous avons choisi ce type de modèle pour les raisons suivantes :

- Des travaux ont montré qu'ils étaient adaptés à la modélisation de distances génétiques ([Shirk et al., 2017](#)).
- Ils intègrent un effet aléatoire relatif à une des deux populations de chaque paire. Ainsi, une valeur c est calculée pour chaque population (*random intercept model*), comme dans l'application de [Murphy et al. \(2010\)](#).
- Ils rendent possible l'élagage des matrices de distances génétiques, à condition de faire permutation certaines populations pour que l'estimation de la covariance des résidus soit possible¹.

Contrairement à [Murphy et al. \(2010\)](#), nous n'avons pas comparé les modèles selon un critère d'AIC mais en utilisant le R^2_β développé par [Edwards et al. \(2008\)](#) pour l'évaluation de l'ajustement des modèles mixtes. Cela vient en partie du fait que cet indicateur permet de comparer des modèles n'incluant pas le même nombre d'observations, ce qui a justifié son choix initialement. Finalement, les modèles comparés incluaient le même nombre d'observations et de variables. Pour cette raison, l'utilisation de l'AIC, critère d'ajustement pénalisé par la complexité du modèle, plutôt que le R^2_β ne se serait pas justifiée.

Ainsi, les modifications de la méthode de [Murphy et al. \(2010\)](#) que nous avons réalisées ne sont pas censées affecter la validité de nos modèles. Dans les deux cas, les modèles prenaient la forme d'une régression linéaire multiple intégrant des effets aléatoires de type *random intercept*.

1. Nous avons discuté de ce point avec Maarten van Strien, à l'origine de l'utilisation de ce type de modèle en génétique du paysage. Il nous a fourni un script permettant leur utilisation pour calibrer des modèles gravitaires à partir de matrices élaguées.

Bibliographie

- ANDERSON, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, 69(1):106–116.
- BALDI, P., BRUNAK, S., CHAUVIN, Y., ANDERSEN, C. A. et NIELSEN, H. (2000). Assessing the accuracy of prediction algorithms for classification : an overview. *Bioinformatics*, 16(5):412–424.
- BOSSENBROEK, J. M., JOHNSON, L. E., PETERS, B. et LODGE, D. M. (2007). Forecasting the expansion of zebra mussels in the United States. *Conservation Biology*, 21(3):800–810.
- BOSSENBROEK, J. M., KRAFT, C. E. et NEKOLA, J. C. (2001). Prediction of long-distance dispersal using gravity models : zebra mussel invasion of inland lakes. *Ecological Applications*, 11(6):1778–1788.
- CLARKE, R. T., ROTHERY, P. et RAYBOULD, A. F. (2002). Confidence limits for regression relationships between distance matrices : estimating gene flow with distance. *Journal of agricultural biological and environmental statistics*, 7(3):361–372.
- DILEO, M. F., SIU, J. C., RHODES, M. K., LÓPEZ-VILLALOBOS, A., REDWINE, A., KSIAZEK, K. et DYER, R. J. (2014). The gravity of pollination : integrating at-site features into spatial analysis of contemporary pollen movement. *Molecular Ecology*, 23(16):3973–3982.
- DYER, R. J. et NASON, J. D. (2004). Population graphs : the graph theoretic shape of genetic structure. *Molecular Ecology*, 13(7):1713–1727.
- EDWARDS, L. J., MULLER, K. E., WOLFINGER, R. D., QAQISH, B. F. et SCHABENBERGER, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in medicine*, 27(29):6137–6157.
- ELITH, J., PHILLIPS, S. J., HASTIE, T., DUDÍK, M., CHEE, Y. E. et YATES, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1):43–57.
- EVERITT, B. et HOTHORN, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.
- FERRARI, M. J., BJØRNSTAD, O. N., PARTAIN, J. L. et ANTONOVICS, J. (2006). A gravity model for the spread of a pollinator-borne plant pathogen. *The American Naturalist*, 168(3):294–303.
- FLETCHER, R. et FORTIN, M.-J. (2018). *Spatial ecology and conservation modeling*. Springer.
- FORTUNA, M. A., ALBALADEJO, R. G., FERNÁNDEZ, L., APARICIO, A. et BASCOMPTE, J. (2009). Networks of spatial genetic variation across species. *Proceedings of the National Academy of Sciences*, 106(45):19044–19049.
- FOTHERINGHAM, A. S. et O’KELLY, M. E. (1989). *Spatial interaction models : formulations and applications*, volume 1. Kluwer Academic Publishers Dordrecht.
- FRIEDMAN, J., HASTIE, T., TIBSHIRANI, R., NARASIMHAN, B., TAY, K., SIMON, N. et QIAN, J. (2021). Package ‘glmnet’. *CRAN R Repository*.
- GOWER, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338.

- GUILLERA-ARROITA, G., LAHOZ-MONFORT, J. J. et ELITH, J. (2014). Maxent is not a presence–absence method : a comment on Thibaud et al. *Methods in Ecology and Evolution*, 5(11):1192–1197.
- GUISAN, A., THUILLER, W. et ZIMMERMANN, N. E. (2017). *Habitat suitability and distribution models : with applications in R*. Cambridge University Press.
- GUISAN, A. et ZIMMERMANN, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3):147–186.
- HASTIE, T., TIBSHIRANI, R. et FRIEDMAN, J. (2017). *The Elements of Statistical Learning : Data Mining, Inference and Prediction*, volume 2. Springer.
- HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70.
- JAMES, G., WITTEN, D., HASTIE, T. et TIBSHIRANI, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- KONG, F., YIN, H., NAKAGOSHI, N. et ZONG, Y. (2010). Urban green space network development for biodiversity conservation : Identification based on graph theory and gravity modeling. *Landscape and Urban Planning*, 95(1-2):16–27.
- KRZANOWSKI, W. et MARRIOTT, F. (1995). *Multivariate Analysis vol. 2 : Classification, Covariance Structures, and Repeated Measurements*. London : Arnold.
- LONG, F. H. (2013). Multivariate analysis for metabolomics and proteomics data. *In Proteomic and Metabolomic Approaches to Biomarker Discovery*, pages 299–311. Elsevier.
- MAGWENE, P. M. (2001). New tools for studying integration and modularity. *Evolution*, 55(9):1734–1745.
- MATTHEWS, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451.
- MURPHY, M. A., DEZZANI, R., PILLIOD, D. S. et STORFER, A. (2010). Landscape genetics of high mountain frog metapopulations. *Molecular Ecology*, 19(17):3634–3649.
- PÉREZ-RODRÍGUEZ, A., KHIMOUN, A., OLLIVIER, A., ERAUD, C., FAIVRE, B. et GARNIER, S. (2018). Habitat fragmentation, not habitat loss, drives the prevalence of blood parasites in a Caribbean passerine. *Ecography*, 41(11):1835–1849.
- PHILLIPS, S. J., ANDERSON, R. P. et SCHAPIRE, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259.
- ROY, K., KAR, S. et DAS, R. N. (2015). Statistical methods in QSAR/QSPR. *In A primer on QSAR/QSPR modeling*, pages 37–59. Springer.
- SHIRK, A. J., LANDGUTH, E. L. et CUSHMAN, S. A. (2017). A comparison of regression methods for model selection in individual-based landscape genetic analysis. *Molecular Ecology Resources*, 18(1):55–67.
- SMOUSE, P. E. et PEAKALL, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, 82(5):561–573.
- TENENHAUS, M. (1998). *La régression PLS : théorie et pratique*. Editions TECHNIP.
- TOBIAS, R. D. et al. (1995). An introduction to partial least squares regression. *In Proceedings of the twentieth annual SAS users group international conference*, pages 1250–1257. SAS Institute Inc Cary.
- VAN STRIEN, M. J., KELLER, D. et HOLDEREGGER, R. (2012). A new analytical approach to landscape genetic modelling : least-cost transect analysis and linear mixed models. *Molecular Ecology*, 21(16):4010–4023.
- WATTS, A. G., SCHLICHTING, P. E., BILLERMAN, S. M., JESMER, B. R., MICHELETTI, S., FORTIN, M.-J., FUNK, W. C., HAPEMAN, P., MUTHS, E. et MURPHY, M. A. (2015). How spatio-temporal habitat connectivity affects amphibian genetic structure? *Frontiers in Genetics*, 6:275.

- WHITTAKER, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.
- WOLD, S., SJÖSTRÖM, M. et ERIKSSON, L. (2001). PLS-regression : a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130.
- XIA, Y., BJØRNSTAD, O. N. et GRENFELL, B. T. (2004). Measles metapopulation dynamics : a gravity model for epidemiological coupling and dynamics. *The American Naturalist*, 164(2):267–281.
- ZERO, V. H., BAROCAS, A., JOCHIMSEN, D. M., PELLETIER, A., GIROUX-BOUGARD, X., TRUMBO, D. R., CASTILLO, J. A., EVANS MACK, D., LINNELL, M. A., PIGG, R. M. *et al.* (2017). Complementary network-based approaches for exploring genetic structure and functional connectivity in two vulnerable, endemic ground squirrels. *Frontiers in Genetics*, 8:81.