



HAL
open science

Extraction de fragments forestiers et caractérisation de leurs évolutions spatio-temporelles pour évaluer l'effet de l'histoire sur la biodiversité : une approche multi-sources

Pierre-Alexis Herrault

► To cite this version:

Pierre-Alexis Herrault. Extraction de fragments forestiers et caractérisation de leurs évolutions spatio-temporelles pour évaluer l'effet de l'histoire sur la biodiversité : une approche multi-sources. Géographie. Université Toulouse le Mirail - Toulouse II, 2015. Français. NNT : 2015TOU20018 . tel-01276544

HAL Id: tel-01276544

<https://theses.hal.science/tel-01276544>

Submitted on 19 Feb 2016

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

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse - Jean Jaurès

Présentée et soutenue par :

Pierre-Alexis Herrault

le vendredi 19 juin 2015

Titre :

Extraction de fragments forestiers et caractérisation de leurs évolutions spatio-temporelles pour évaluer l'effet de l'histoire sur la biodiversité: une approche multi-sources

École doctorale et discipline ou spécialité :

ED TESC : Géographie et aménagement

Unité de recherche :

Laboratoire GEODE UMR 5602 CNRS / Laboratoire DYNAFOR UMR 1201 INRA

Directeur/trice(s) de Thèse :

Martin Paegelow (Professeur, GEODE, UT2J, Toulouse)

David Sheeren (Maitre de conférences, DYNAFOR, INP-ENSAT, Toulouse)

Jury :

Rapporteurs:

Anne Ruas (ICPEF, HDR, IFSTTAR, Marne-la-Vallée)

Vincent Godard (Professeur, LADYSS, Université Paris VIII Vincennes-St Denis)

Examineurs:

Anne Puissant (Maitre de conférences, HDR, LIVE, Strasbourg)

Marianne Cohen (Professeur, ENEC, Université Paris IV Panthéon-Sorbonne)

REMERCIEMENTS

Cette thèse a été une formidable aventure au cours de laquelle j'ai rencontré des gens passionnants et restera comme une période très heureuse de ma vie. Je tiens à dire ici le plaisir que j'ai eu à la réaliser et souhaite maintenant remercier toutes les personnes qui ont contribué de près ou de loin à cet aboutissement.

Je remercie très chaleureusement mon directeur de thèse Martin Paegelow (Professeur de Géographie) pour la confiance qu'il m'a accordée et pour m'avoir confié ce travail de thèse. Merci pour toutes ces discussions enrichissantes et pour toutes tes qualités humaines très rassurantes pendant ces trois années. Je remercie David Sheeren (maitre de conférences en Géographie), mon co-directeur de thèse, pour...pour TOUT en fait. Je crois bien que j'ai eu beaucoup de chance de travailler à tes côtés alors merci de m'avoir accordé tant d'attention, de m'avoir fait progresser et d'avoir fait en sorte que cette période ait été si stimulante. J'ai pris énormément de plaisir et suis déjà pressé de retravailler avec toi !

Je veux remercier également Laurence Hubert et Samuel Corgne, mes encadrants de stage de Master 1 et 2 effectués au Laboratoire COSTEL pour m'avoir transmis le virus de la recherche et d'avoir cru en moi à une époque où les questions étaient nombreuses et les réponses un peu moins.

Un très grand merci aux membres de mon jury de thèse pour m'avoir fait l'honneur d'évaluer ce travail et pour les discussions très enrichissantes durant ma soutenance. Merci à Marianne Cohen (Présidente du Jury, Examinatrice), Anne Ruas (Rapporteur), Vincent Godard (Rapporteur) et Anne Puissant (Examinatrice). Merci à toutes les personnes avec qui j'ai eu le plaisir de travailler. Les membres de mon comité de thèse : Thomas Houet, Claude Monteil, Annie Quin, Cyril de Runz et Sébastien Mustière. Je garde également un excellent souvenir de mon séjour au WSL (Swiss Federal Research Institute) à Zurich où j'ai eu le plaisir de collaborer avec Urs Gimmi et Thibaut Lachat : merci à vous les gars ! Merci à Mathieu Fauvel pour l'aide permanente qu'il m'a apporté, pour sa patience et pour les footings du midi aussi ! Enfin, merci à Laurent Larrieu pour tout ce qu'il m'a appris et son enthousiasme durant notre collaboration : j'ai pris énormément de plaisir.

Je remercie ensuite les laboratoires GEODE et DYNAFOR pour tous les moyens mis à disposition ayant participé au très bon déroulement de mes travaux. Evidemment, une grosse pensée à tous les membres de ces laboratoires que j'ai croisés au quotidien, en réunion ou autour d'un café. Une pensée particulière pour tous les doctorants et stagiaires que j'ai rencontrés et que j'ai sincèrement apprécié connaître : à Hugo, Lucie, Virginie, Romain, Eugénie, Laure, Nicolas, Martin, Mailys, Anthony, Carole, Marco, Léa, Guillaume, Camille... Une dédicace plus spéciale à mon Hugo avec qui j'ai partagé mon bureau pendant plus de deux ans : c'était un plaisir mecton ! Je veux également remercier Carole Planque pour son travail remarquable durant son stage et qui a considérablement fait avancer mes travaux de thèse → A toi de jouer maintenant !

Et puis il y a tous les amis qui ont gravité autour de cette thèse, qu'ils soient du bureau ou du labo d'à côté ou bien rencontrés dans les bars toulousains (et plus précisément Saint-Michel..) : Léo, Youen, Vincent, Adrien, Antoine, Juliette, Sophia, Joana, Xiaolin, Théo...à tous je vous dis un grand Merci pour votre grosse bonne humeur, pour toutes ces bières partagées et je suis très heureux de vous avoir rencontré !

Avant de conclure, je pense très fort à mes amis bretons qui, sans s'en rendre compte, ont été d'un soutien sans commune mesure et la soupape de décompression la plus efficace de la Terre. Alors avec le plus grand des plaisirs du monde : à mon Charlot (et Julie !), à mon Dude, à Julie, à mon Flo, à mon Rerel, à mon Babas, à Tyfn et Alex, à Gogue et Malo, à Beuf, au Vieux Claude et à tous les autres que j'ai vu un peu moins ces derniers temps mais à qui je pense aussi. Une grosse bise à vous tous les loulous !

Enfin, je vais conclure en remerciant ma famille pour tous leurs encouragements. Merci à mes parents pour leur soutien sans faille, pour avoir cru en moi et avoir toujours tout fait pour qu'on s'épanouisse dans nos vies. Une grosse bise à mon frère Simon à qui je pense bien fort également. Cette thèse est pour vous trois !

LISTE DE SIGLES ET ACRONYMES

BD : Base de Données

CDL : Change Description Language

EM : Expectation-Maximization

EMQ : Erreur Moyenne Quadratique

ES : Élément structurant

ESTDM: Event based Spatiotemporal Data Model

FAO: Food and Agriculture Organization of the United Nations

GCP: Ground Control Points

IFN: Inventaire Forestier National

IRC: Infra-Rouge Couleur

JRC: Join European Research

LETR: Long Term Ecosystem Research Network

MEDDE : Ministère de l'Environnement et du Développement Durable

MSS : Multi Spectral Sensor

MNT : Modèle Numérique de Terrain

ONU : Organisation des Nations Unies

ONF : Office National des Forêts

OS: Occupation du Sol

RMSE: Root Mean Square Error

RVB: Rouge Vert Bleu

STC : Space Time Composite

THRS : Très haute Résolution Spatiale

SOMMAIRE

REMERCIEMENTS	
LISTE DES SIGLES ET ACRONYMES	
SOMMAIRE	
INTRODUCTION GENERALE	1
PARTIE A. CADRE APPLICATIF DE LA THESE	7
INTRODUCTION DE LA PARTIE A	9
CHAPITRE 1. CADRE DE L'ETUDE : L'ECOLOGIE DU PAYSAGE	11
CHAPITRE 2. OBJET D'ETUDE : LA FORET PAYSANNE FRAGMENTEE ET LA BIODIVERSITE ASSOCIEE	31
CHAPITRE 3. SITE D'ETUDE ET DONNEES BIOLOGIQUES SELECTIONNEES	51
CONCLUSION DE LA PARTIE A	57
PARTIE B. EXTRAIRE LES ESPACES FORESTIERS ANCIENS ET ACTUELS A PARTIR DE SOURCES DE DONNEES HETEROGENES	61
INTRODUCTION DE LA PARTIE B	63
CHAPITRE 4. SOURCES DE DONNEES SPATIALES RETENUES ET SPECIFICITES	65
CHAPITRE 5. METHODE DE CORRECTION GEOMETRIQUE DES CARTES ANCIENNES	91
CHAPITRE 6. VECTORISATION MANUELLE DES ESPACES FORESTIERS	125
CHAPITRE 7. VECTORISATION AUTOMATIQUE DES ESPACES FORESTIERS	139
CONCLUSION DE LA PARTIE B	183
PARTIE C. APPARIER LES ESPACES FORESTIERS ET CARACTERISER LES EVOLUTIONS SPATIO-TEMPORELLES	187
INTRODUCTION DE LA PARTIE C	189
CHAPITRE 8. DEMARCHE METHODOLOGIQUE PROPOSEE	191
CHAPITRE 9. MISE EN ŒUVRE DE LA DEMARCHE METHODOLOGIQUE PROPOSEE	215
CONCLUSION DE LA PARTIE C	231
PARTIE D. MODELISER LES EFFETS DES EVOLUTIONS SPATIO-TEMPORELLES DES FORETS ET DE LEURS INCERTITUDES SUR LA BIODIVERSITE	235
INTRODUCTION DE LA PARTIE D	237
CHAPITRE 10. MODELISATION ESPECE-HABITAT : EFFET DE L'EVOLUTION DES FORETS SUR LA BIODIVERSITE	239
CHAPITRE 11. MODELISATION ESPECE-HABITAT : EFFET DES INCERTITUDES SPATIALES	275
CONCLUSION DE LA PARTIE D	289
CONCLUSION GENERALE	291
BIBLIOGRAPHIE	297
ANNEXE	333
LISTE DES FIGURES	381
LISTE DES TABLEAUX	389
TABLE DES MATIERES	393

INTRODUCTION GENERALE

Les paysages européens connaissent depuis de nombreuses années de profonds bouleversements liés pour la plupart aux activités anthropiques (Antrop and Van Eetvelde 2008). Les forêts sont parmi les écosystèmes les plus perturbés par le déploiement des civilisations puisque l'Homme a continuellement défriché ces milieux pour agrandir les zones de culture et de pâturage nécessaires à sa subsistance (Kirby and Watkins 1998). Par conséquent, les forêts européennes actuelles sont le résultat d'épisodes successifs de déforestation et de reforestation, parfois sur des sols autrefois occupés par l'agriculture (Dupouey et al. 2002). En outre, elles sont dans certaines régions, touchées par la fragmentation. Parallèlement à cela, une grande part des massifs forestiers a été conservée et exploitée pour les nombreuses ressources que les forêts fournissent. Le bois de chauffage, la construction navale ou les fortifications sont autant d'exemples d'activités anciennes ayant fortement exploité les ressources forestières au cours des siècles. Encore aujourd'hui, le développement de nouvelles énergies ou la fabrication de papier mobilisent ces ressources. Mis à part leur valeur économique, les massifs forestiers sont considérés comme les principaux réservoirs de biodiversité au sein de nos paysages agricoles et font l'objet d'une forte attention de la part de la communauté (Deconchat and Balent 1996).

Un des processus majeurs affectant la biodiversité concerne la fragmentation des paysages (Fahrig 2003). Elle modifie à la fois la configuration des patrons paysagers et leur composition entraînant par conséquent une évolution des populations d'espèces (présence, abondance, richesse) en réponse à ces changements (Turner 1989). C'est dans ce cadre que l'écologie du paysage s'est développée (Risser 1987, Turner et al. 2001). Cette discipline, à la croisée entre l'écologie et la géographie, place la structure spatiale des paysages au cœur des facteurs permettant de comprendre la diversité et la distribution des organismes sur de plus vastes étendues que dans l'écologie traditionnelle (McGarigal and Marks 1995). Elle s'est considérablement développée au fil des années laissant place à un grand nombre de fronts de recherche (Wu 2013). Elle s'appuie également sur différentes théories (notamment la théorie de la biogéographie insulaire et la théorie de la hiérarchie) qui formalisent et facilitent l'interprétation des effets des patrons spatiaux paysagers sur la biodiversité (Grinnell 1917; MacArthur 1967; Levins 1970; Wu 1999).

La majeure partie des études visant à expliquer la diversité et la distribution des organismes par le prisme du paysage s'est très longtemps contentée de mobiliser uniquement la structure actuelle du paysage. Or, il s'est avéré que cette information ne permettait pas toujours d'expliquer efficacement ces patrons de biodiversité. Cela s'explique notamment par la capacité de résilience des espèces en réponse à une perturbation de leurs habitats (Kuussaari et al. 2009). Certaines réagissent rapidement, voire instantanément (ex : les insectes, les oiseaux). Ces espèces disposent de caractéristiques leur permettant d'exploiter plus facilement des habitats annexes ou de se régénérer pour être en mesure de perdurer. Le paysage actuel est généralement suffisant pour expliquer leur présence. En revanche, d'autres organismes vont exprimer un délai de réponse à une modification de leurs habitats (ex : les plantes vasculaires, les champignons). De fait, puisqu'elles ne sont pas en mesure de répondre immédiatement, les événements de perturbation passés peuvent jouer un rôle à plus ou moins long terme sur leur présence actuelle. Les paysages passés et leurs dynamiques sont devenus, par conséquent, au fil des années, des facteurs d'explication crédibles pour étudier la distribution et la diversité de ces espèces dans nos paysages (Girel 2006). Ce dernier constat structure aujourd'hui le domaine de l'écologie historique.

L'écologie historique a rapidement porté son attention sur les écosystèmes forestiers puisque les premiers travaux dans ce domaine ont concerné la flore des forêts en Grande Bretagne au début

des années 1980 (Peterken and Game 1984). Des études similaires ont par la suite été conduites aux Etats-Unis (Foster 1988) et en Europe Occidentale, notamment dans les forêts de feuillus belges (Hermy et al. 1999). Ces travaux pionniers ont permis de démontrer que certaines espèces de plantes n'étaient présentes que dans les massifs forestiers existants depuis plusieurs décennies voire plusieurs siècles, sans discontinuité. Ils ont également participé à la diffusion de plusieurs concepts clés, aujourd'hui au cœur de l'écologie historique, avec parmi eux, l'ancienneté. Cette notion d'ancienneté fait référence à la continuité de présence d'un habitat au cours du temps. Elle ne traduit en rien la vieillesse des peuplements (son état de maturité) ou le type de gestion subi par un écosystème. Les forêts anciennes sont définies comme les surfaces forestières présentes depuis un ou plusieurs siècles sans discontinuité temporelle (Peterken 1977).

La continuité de présence des habitats est généralement déterminée sur la base d'un point de référence permettant de statuer sur la présence ou non de l'habitat suivi à une date donnée (Dupouey et al. 2007; Cousins 2009; Cristofoli et al. 2010). Dans cette logique, les données spatiales historiques sont aujourd'hui des données cruciales qui témoignent de l'état du paysage à des périodes passées et qui permettent de reconstruire l'évolution des milieux sur le long terme. La mise à disposition récente de ces données par les instituts géographiques nationaux facilite leur utilisation et donne la possibilité à de nombreux acteurs de s'appuyer sur elles pour mener des études d'écologie historique. En France, l'Institut Géographique National (IGN) distribue gratuitement plusieurs sources de données spatiales anciennes au format numérique telles que des cartes historiques, des plans cadastraux ou des photographies aériennes noir et blanc. Elles offrent l'opportunité de disposer d'informations spatialisées sur différents milieux depuis la fin du XVIII^e siècle (la carte de Cassini, Vallauri et al. 2012). Enfin, ces données comportent différents problèmes de qualité à la fois géométriques, sémantiques ou de complétude. Si les biais introduits par l'incertitude spatiale dans la caractérisation et le suivi des phénomènes spatiaux sont bien connus, ils ne sont que très rarement pris en compte dans les travaux d'écologie du paysage. (Lechner et al. 2012).

Les objectifs poursuivis dans cette thèse visent à répondre à ces différents problèmes méthodologiques dans le cadre d'une étude visant, à l'échelle de la région midi-pyrénéenne, à évaluer les effets de l'histoire des habitats forestiers sur la biodiversité actuelle qu'ils abritent. Les principales questions auxquelles nous souhaitons apporter des réponses sont les suivantes. Quelles sont les sources de données spatiales les plus adaptées aujourd'hui pour reconstruire l'évolution des forêts en France ? Sur quelles techniques s'appuyer pour corriger géométriquement ces données et permettre leur comparaison avec des données plus récentes ? Quelles approches employer pour extraire automatiquement les forêts de ces documents pour être en mesure de reproduire cette tâche sur de vastes étendues ? Comment reconstruire les évolutions spatio-temporelles des habitats forestiers en exploitant les données spatiales et modéliser leurs effets sur la biodiversité actuelle ? Les imperfections entraînées par ces documents sont-elles susceptibles d'impacter la réponse des modèles espèces-habitats ?

Afin de répondre à ces questions, nous proposons différentes approches méthodologiques qui s'appuient d'une part, sur des techniques employées en télédétection et en géomatique pour corriger géométriquement et extraire les éléments des données spatiales anciennes, et d'autre part, sur de la modélisation statistique. Nous faisons l'hypothèse que la biodiversité contenue actuellement dans les habitats forestiers des coteaux de Gascogne (Haute Garonne, 31) et en particulier, celle relative à un insecte (Diptera: Syrphidae), pourrait être expliquée, au moins pour une part, par la structure passée de ces habitats et leurs dynamiques. Ainsi, nous mobiliserons plusieurs sources de données spatiales (cartes historiques, photographies aériennes) permettant d'obtenir des informations sur les forêts à différents moments de temps. Nous mobiliserons aussi des relevés biologiques existants effectués à l'échelle d'un groupe de forêts présentes actuellement. Nous émettons également l'hypothèse que les techniques classiquement employées en écologie

pour traiter les données spatiales anciennes ne sont pas toujours adaptées et pourraient être améliorées. Enfin, nous nous attacherons à vérifier que les variables d'entrée produites à partir de données spatiales anciennes pourraient fortement impacter la réponse des modèles espèces-habitat selon leur qualité.

Cette thèse s'articule en quatre parties :

La première partie (Partie A) présente le cadre applicatif de cette thèse à savoir, l'écologie du paysage et l'écologie historique. Un rappel des grands fondements de l'écologie du paysage sera tout d'abord présenté en chapitre 1 pour être en mesure de mieux comprendre les enjeux liés à la variation de biodiversité dans les paysages ruraux. Puis, nous identifierons, à partir de la littérature, les principaux déterminants de la présence de biodiversité avant d'établir un rappel des grandes théories écologiques à l'origine de l'écologie du paysage. Le second volet de cette partie (chapitre 2) portera plus spécifiquement sur les spécificités des forêts paysannes fragmentées (notre objet d'étude) aujourd'hui reconnues comme de véritables réservoirs de biodiversité au sein de l'espace rural. La biodiversité caractéristique de ces milieux sera elle aussi abordée au travers des menaces dont elle fait aujourd'hui l'objet, des indicateurs qui sont classiquement utilisées pour la mesurer et des facteurs généraux susceptibles de l'affecter. Enfin, un troisième volet (chapitre 3) s'attachera à décrire le site d'étude adopté dans cette thèse ainsi que les données biologiques mobilisées.

La deuxième partie (Partie B) décrit la démarche développée pour extraire les espaces forestiers actuels et anciens issus de sources de données hétérogènes (Figure I-1). Nous décrirons dans un premier chapitre (chapitre 4) les sources de données spatiales disponibles pour reconstruire les évolutions spatio-temporelles des forêts sur le temps long et les critères de choix adoptés pour les sélectionner. Puis, nous aborderons dans une deuxième phase (chapitre 5) le problème de la correction géométrique des cartes anciennes. Ce second volet aura pour objectif de proposer une stratégie optimale de correction en s'appuyant sur les différents critères qui déterminent la qualité de la correction géométrique. Nous étudierons en chapitre 6, la démarche suivie pour extraire manuellement les espaces forestiers des différentes sources de données sélectionnées avant de proposer une démarche de vectorisation automatique pour les cartes anciennes et les données actuelles (chapitre 7).

La troisième partie (Partie C) porte sur la caractérisation des évolutions spatio-temporelles des forêts (Figure I-1). Nous expliciterons dans une première phase la démarche d'appariement adoptée pour identifier les forêts homologues sur les différentes sources et reconstruire la généalogie des forêts étudiées au cours du temps. Un second volet abordera ensuite la caractérisation de la structure spatiale des forêts étudiées. L'objectif de ce volet vise à sélectionner un certain nombre de descripteurs spatiaux qui permettront de traduire les différents facteurs susceptibles d'affecter la biodiversité forestière. Enfin, deux démarches seront proposées pour caractériser et quantifier les évolutions spatio-temporelles des forêts : l'une a pour objectif de mettre en évidence l'évolution de chaque forêt tandis que l'autre vise à identifier les grandes tendances d'évolution à l'échelle du groupe de forêts suivies. A noter que cette partie est constituée de deux chapitres (8 et 9). Le premier reprend la démarche méthodologique évoquée précédemment et le second détaillera sa mise en œuvre dans le contexte de notre étude.

Enfin, la dernière partie de la thèse (Partie D) aborde la question de la modélisation espèces-habitats (Figure I-1). L'objectif premier est d'évaluer les effets des évolutions spatio-temporelles des forêts sur la présence actuelle de biodiversité (chapitre 10). Nous décrirons d'abord la démarche développée et présenterons ensuite les résultats de cette étude sous la forme d'un article scientifique soumis à la revue *Landscape Ecology* (version révisée). Le second objectif vise quant à lui à évaluer les impacts potentiels de l'incertitude spatiale sur la réponse des modèles espèces-habitats (chapitre 11). Nous remobiliserons par conséquent les modèles définis en chapitre 10 pour vérifier cette hypothèse.

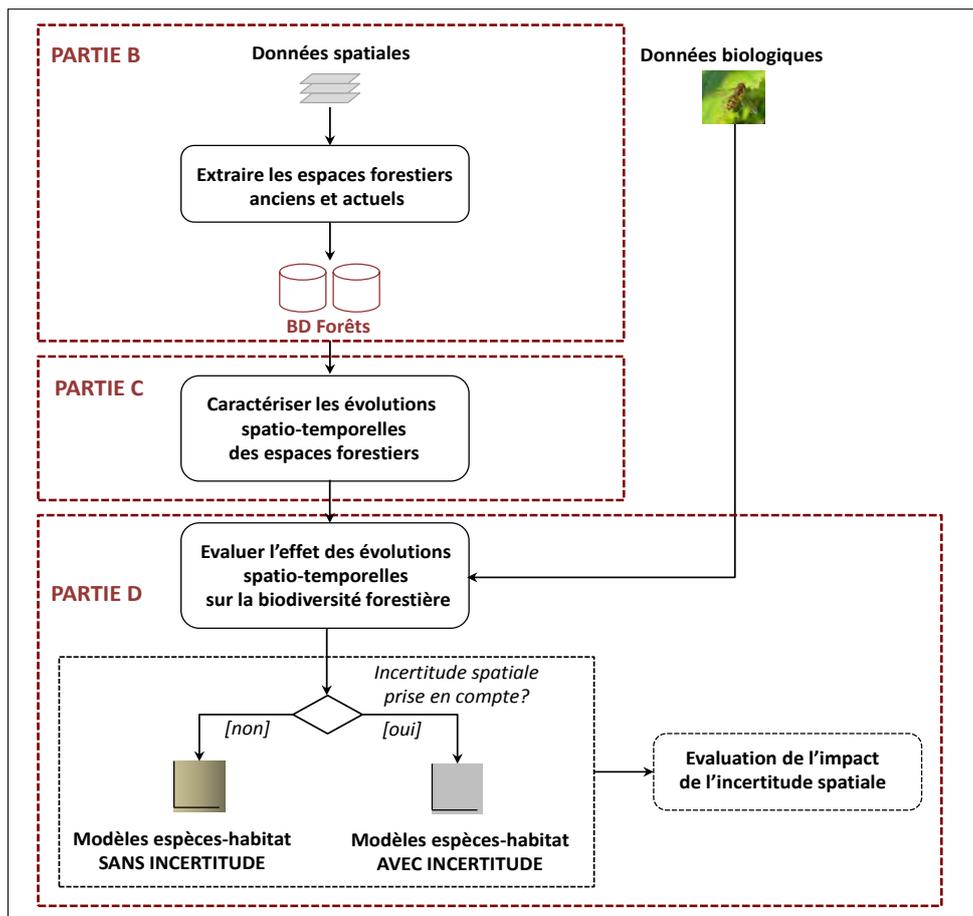


Figure I-1 : Démarche méthodologique suivie dans cette thèse

Cette thèse a été réalisée au sein du laboratoire DYNAFOR UMR INRA1201 et du laboratoire GEODE UMR CNRS 5602. Elle a bénéficié d'un cofinancement de la Région Midi-Pyrénées et de l'Université Fédérale Toulouse Midi-Pyrénées.

PARTIE A

Cadre applicatif de la thèse

INTRODUCTION DE LA PARTIE A

Cette première partie a pour objectif de présenter le cadre applicatif de cette thèse.

Pour mieux comprendre les enjeux liés à la variation de biodiversité forestière dans les paysages ruraux, un rappel des fondements de l'écologie du paysage est tout d'abord présenté. L'objectif est ensuite d'identifier les principaux déterminants de la présence de biodiversité et d'en expliquer les principes. Enfin, une présentation des grandes théories à l'origine de l'écologie du paysage permettra de définir les concepts établis pour analyser et interpréter les effets de ces déterminants sur la biodiversité.

Etant donné que ce travail de thèse s'intéresse aux forêts de coteaux de Gascogne, il est nécessaire de bien comprendre pourquoi la forêt paysanne est aujourd'hui un objet d'étude privilégié pour comprendre les relations paysage/biodiversité. Ce chapitre sera aussi l'occasion de définir les approches utilisées pour mesurer les indicateurs de biodiversité et de détailler ceux que nous étudierons dans ce travail de thèse. Un point sur les facteurs spécifiques susceptibles de faire varier la biodiversité forestière sera ensuite proposé avant de déterminer les grandes notions qui structurent les études d'écologie historique appliquées aux habitats forestiers.

Nous terminerons cette partie par la présentation du site d'étude adopté dans ce travail de thèse ainsi que les données biologiques sélectionnées pour évaluer les effets de l'histoire sur la présence de biodiversité actuelle.

CHAPITRE 1. CADRE DE L'ETUDE : L'ECOLOGIE DU PAYSAGE

SOMMAIRE

- 1.1. UNE DISCIPLINE A LA CROISEE ENTRE L'ECOLOGIE ET LA GEOGRAPHIE
- 1.2. L'HETEROGENEITE DU PAYSAGE COMME CAUSE DE LA BIODIVERSITE
- 1.3. LA FRAGMENTATION : UN PROCESSUS AFFECTANT POTENTIELLEMENT LA BIODIVERSITE
- 1.4. LES THEORIES ECOLOGIQUES FONDATRICES

1.1. UNE DISCIPLINE A LA CROISEE ENTRE L'ÉCOLOGIE ET LA GEOGRAPHIE

L'écologie du paysage a vu le jour à la fin des années 1930 avec l'émergence de la notion de paysage. Le terme allemand de *Landschaftsoekologie* s'impose aux biogéographes européens via les travaux du géographe-naturaliste Carl Troll dont les travaux consistent à exploiter des photographies aériennes pour étudier les interactions entre végétation et environnement à l'échelle des paysages (Troll 1939). C'est en revanche beaucoup plus tard que la notion d'écologie du paysage trouvera écho chez le plus grand nombre et précisément dans les années 1980 après la publication d'un manifeste résultant d'un séminaire situé dans l'Illinois (Risser 1987). Depuis, cette discipline ne cesse de progresser notamment grâce à l'exploitation d'images satellitaires et à l'usage de modèles informatiques. De nombreuses revues traitent exclusivement de cette discipline (*Landscape Ecology*), les universités l'enseignent tandis qu'une association internationale d'écologie du paysage (IALE : *International Association for Landscape Ecology*) organise de nombreuses rencontres et colloques à travers le monde.

Les définitions généralement apportées à la notion de paysage s'articulent autour de deux notions clés : l'espace et la perception. Le dictionnaire Littré (vers 1863-1873) définit le paysage comme une « *étendue du pays que l'on voit d'un seul aspect. Un paysage dont on aura vu toutes les parties l'une après l'autre n'a pourtant pas été vu ; il faut qu'il le soit d'un lieu assez élevé, où tous les objets auparavant dispersés se rassemblent sous un seul coup d'œil* ». Le Petit Larousse (1974) le détermine comme « *une étendue spatiale, naturelle ou transformée par l'homme, qui présente une certaine identité visuelle ou fonctionnelle* ». Avec le développement de la géographie de la perception, Roger Brunet et Robert Ferras (1993) verront le paysage « *comme une apparence et une représentation: un arrangement d'objets visibles perçu par un sujet à travers ses propres filtres, ses propres humeurs, ses propres fins* » (Brunet et al. 1993). Il faudra attendre le milieu des années 1970 pour voir la première définition du paysage de l'écologue apportée par Bertrand (1975). Pour ce dernier, « *le paysage est un médiateur entre la nature et la société ayant pour base une portion d'espace matériel qui existe en tant que structure et système écologique, donc indépendamment de la perception* ». La définition de Forman et Godron (1981), dans le premier livre de cours d'écologie du paysage est très proche de celle de Bertrand. C'est « *une portion de territoire hétérogène composée d'ensemble d'écosystèmes en interaction qui se répète de façon similaire dans l'espace* » (Forman et Godron 1981).

La définition que nous adopterons dans cette thèse est celle de Burel et Baudry (1999) qui est finalement la synthèse des deux définitions précédentes : « ***le paysage est un niveau d'organisation des systèmes écologiques, supérieur à l'écosystème ; il se caractérise essentiellement par son hétérogénéité et par sa dynamique gouvernée par les activités humaines. Il existe indépendamment de la perception*** » (Burel et Baudry 1999).

Selon Turner (1989), l'écologie du paysage est définie comme **l'étude des effets des « patrons spatiaux » sur les processus biologiques** où le patron fait référence spécifiquement à la structure du paysage. **En ce sens, l'écologie du paysage entretient d'étroites relations avec la géographie de par sa prise en compte de la dimension spatiale dans l'analyse des processus** (Turner 1989). Ces patrons spatiaux abritent un certain nombre d'éléments en interaction et susceptibles d'évoluer plus ou moins fortement à travers le temps (McGarigal et Marks 1995a). Ces changements modifiant de fait les effets de ces patrons sur les processus biologiques. Pour Fahrig (2003), l'écologie du paysage revient à étudier comment les paysages affectent l'abondance ou la distribution des organismes (Fahrig 2003a). En termes statistiques, les variables « réponse » en écologie du paysage sont des variables d'abondance, de présence ou de diversité d'espèces tandis que les variables « explicatives » sont généralement des variables qui décrivent la structure des paysages. Il est à noter également que la relation patron/processus ne s'analyse pas dans un sens unique. Si l'écologie du paysage s'attache

à comprendre comment la structure du paysage influence les processus écologiques, elle tente de comprendre également comment ces processus, en retour, peuvent impacter le paysage (Décamps et al. 2004). Mac Garigal et Marks (1995), considèrent l'écologie du paysage comme une discipline structurée par trois aspects fondamentaux : le rôle des humains, l'hétérogénéité spatiale, de vastes étendues d'analyse (Figure 1.1).

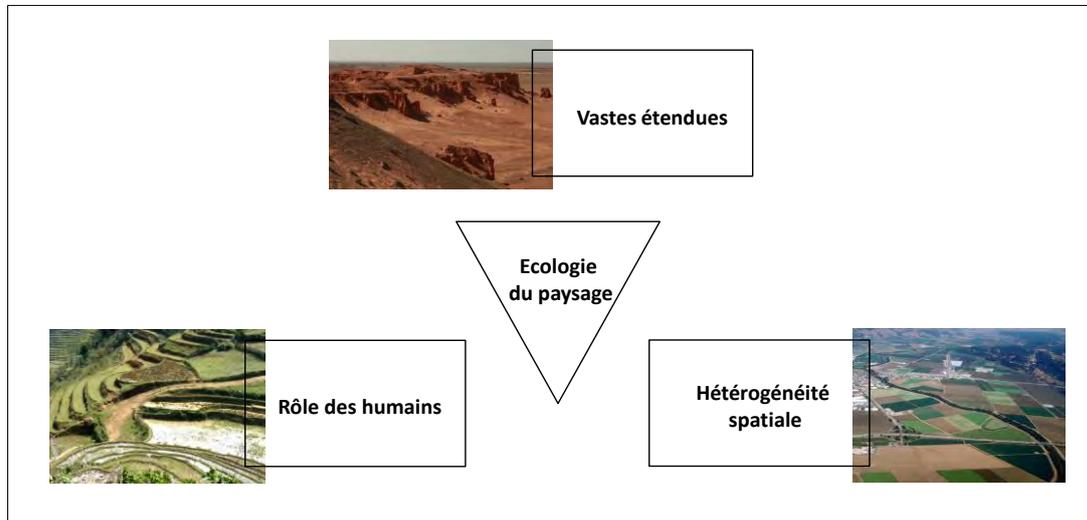


Figure 1-1. L'écologie du paysage structurée autour de trois aspects fondamentaux
(Source : McGarigal et Marks 1995)

La notion d'hétérogénéité spatiale est la notion qui caractérise le mieux l'essence de l'écologie du paysage (McGarigal et Marks 1995a; Wu 2013). Selon Burel et Baudry (1999), l'écologie du paysage est fondée sur l'hétérogénéité des espaces et des milieux et légitime même son rôle organisateur. Elle est surtout mise en avant pour différencier l'écologie du paysage des autres écologies traditionnelles telles que l'écologie des communautés ou des populations. L'écologie traditionnelle se contente d'étudier les interrelations qui existent entre les organismes et leur environnement sans tenir compte de la structure spatiale. Le but est de comprendre comment l'environnement, à savoir les processus biotiques et abiotiques, affecte l'abondance et la distribution des organismes. L'écologie du paysage fait intervenir les mêmes processus et patrons mais elle tente de comprendre comment la structure des paysages affecte l'abondance et la distribution des organismes (Figure 1-2). On cherche ainsi à caractériser l'hétérogénéité, à déterminer ses origines, à comprendre comment elle change à travers le temps et comment les humains la gèrent.

L'écologie du paysage étudie généralement les processus écologiques sur de plus vastes étendues que l'écologie traditionnelle. Cela s'explique par l'origine anthropocentrique de la discipline et particulièrement de sa **vision géographique aérienne** (l'étude des patrons paysagers à partir de photographies aériennes). L'écologie du paysage ne définit pas a priori des échelles spécifiques pouvant être universellement appliquées. On cherche plutôt à identifier des échelles adaptées qui sont en mesure de caractériser le mieux possible les relations entre l'hétérogénéité spatiale et les processus d'intérêt. Le développement de la **théorie de la hiérarchie** (Allen et Starr 1982) a notamment permis de fournir un cadre dans l'analyse des relations patrons/processus en fonction de différents niveaux d'organisation écologique (voir section 1.4.1). Depuis, il est devenu évident que les questions d'écologie du paysage doivent être appréhendées à travers de multiples et différentes échelles spatiales et temporelles (McGarigal et Marks 1995).

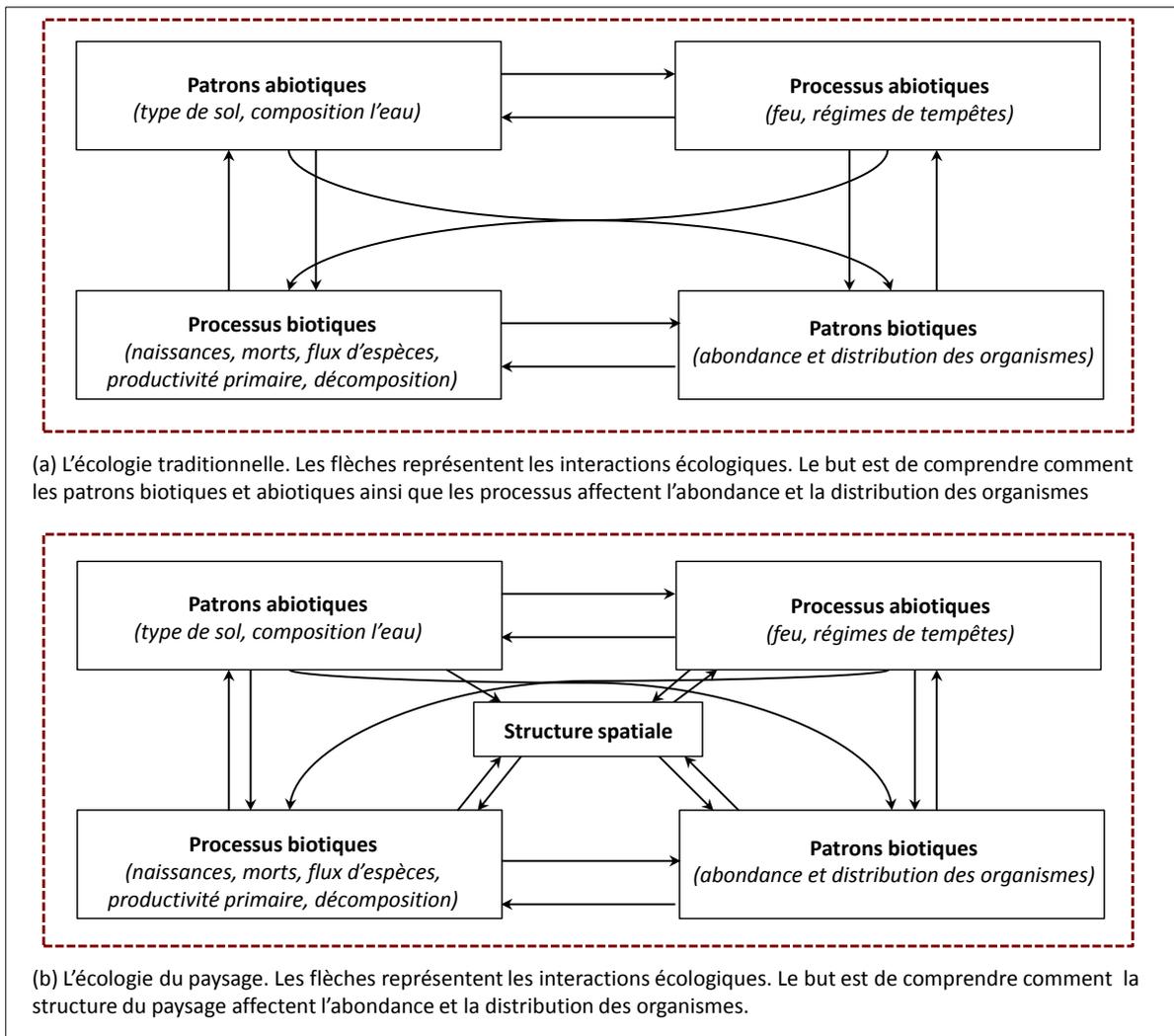


Figure 1-2. L'écologie traditionnelle et l'écologie du paysage
(Source: Fahrig 2003)

En affectant les processus et les patrons paysagers, les humains occupent une place centrale dans les études d'écologie du paysage. **Pour cette raison, l'écologie du paysage est une science interdisciplinaire avec des interrelations entre la société humaine et son environnement vital.** Une grande part de l'écologie du paysage s'effectue finalement à travers des environnements construits où les humains sont la force dominante des changements du paysage.

L'écologie moderne du paysage est donc à l'interface entre l'hétérogénéité spatiale et les processus écologiques tout en considérant les humains comme un des agents majeurs de la modification des paysages.

1.2. L'HÉTÉROGÉNÉITÉ COMME CAUSE DE LA BIODIVERSITÉ

En 1972, Smith définit un environnement comme hétérogène lorsque le taux d'un processus varie en fonction de la localisation dans l'espace (Smith 1972). Au contraire, l'environnement peut être homogène si ce processus a un taux constant entre ces différentes localisations (Dutilleul 1993; Deléglise et al. 2011). Plusieurs définitions de l'hétérogénéité existent (Dutilleul 1993; Wu et Loucks 1995; Dale 1999), néanmoins celle de Smith est applicable aux individus, aux espèces, aux facteurs environnementaux et à toute variables écologique. Nous considérerons par conséquent cette seule définition.

Il existe une relation entre l'hétérogénéité spatiale et les processus écologiques. Une variation de l'hétérogénéité spatiale produira des changements dans les fonctions et les processus écologiques qui apparaissent à travers la mosaïque paysagère (Li et Reynolds 1995; Pickett et Cadenasso 1995). La mosaïque paysagère peut être vue comme un ensemble d'unités spatiales contigües et distinctes (ou *patches*) (Pickett et Cadenasso 1995; Alard et Poudevigne 2002; Redon 2012). Chaque patch se différencie des autres par sa nature (ex : une forêt, une prairie) ou son apparence, sa taille, sa forme, sa composition ou sa persistance dans le temps. On peut alors définir la mosaïque paysagère plus généralement à travers deux grands aspects : **la composition et l'agencement spatial des patches (configuration)** (Fahrig 2005) (Figure 1-3). La quantité de forêts ou de zones humides, la longueur de lisière ou la densité de route sont des exemples relevant de la composition. La composition renseigne ainsi sur l'abondance d'une catégorie de patches dans un paysage donné. Leur configuration renseigne quant à elle sur l'organisation spatiale de ces patches dans le paysage (Pickett et Cadenasso 1995; Alard et Poudevigne 2002; Turner 2005). Cette hétérogénéité de la mosaïque paysagère est souvent appréhendée au travers du modèle **Tâche-Corridor-Matrice** (Burel et Baudry 1999; Décamps et al. 2004).

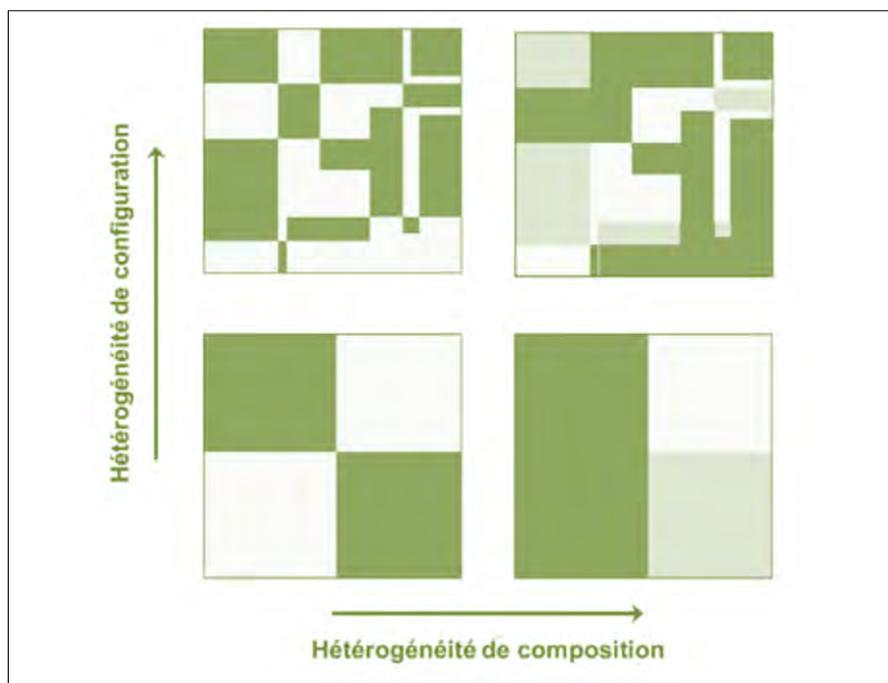


Figure 1-3. Hétérogénéité de composition et de configuration
(Source: Fahrig 2005)

Ce dernier modèle est la vision classique adoptée en écologie du paysage. Elle correspond à un espace organisé en tâches (correspondant à des patches d'habitat¹ de certaines espèces), reliées le plus souvent par des corridors (correspondant à des zones favorables aux déplacements des espèces d'un habitat à l'autre) et distribuées au sein d'une matrice qui elle permet peu les déplacements (Décamps et al. 2004). Les tâches peuvent être plus moins allongées, plus ou moins nombreuses et dispersées ; les corridors plus ou moins continus ; les matrices plus ou moins étendues (Figure 1-4). Bien que simplifié à l'extrême, ce cadre offre la possibilité de mettre en relation la configuration de l'espace avec les processus qui conditionnent le fonctionnement et la dynamique des populations,

¹ La notion d'habitat est encore controversée en écologie et fait l'objet de points de vue différents. Du point de vue de l'espèce, l'habitat est son aire de vie, là où elle accède à un ensemble de ressources permettant son développement et sa reproduction (Dennis et al. 2003)

communautés et écosystèmes ; il offre également l'opportunité d'étudier l'hétérogénéité des espaces qui nous entourent ainsi que leur évolution à travers l'espace et le temps.

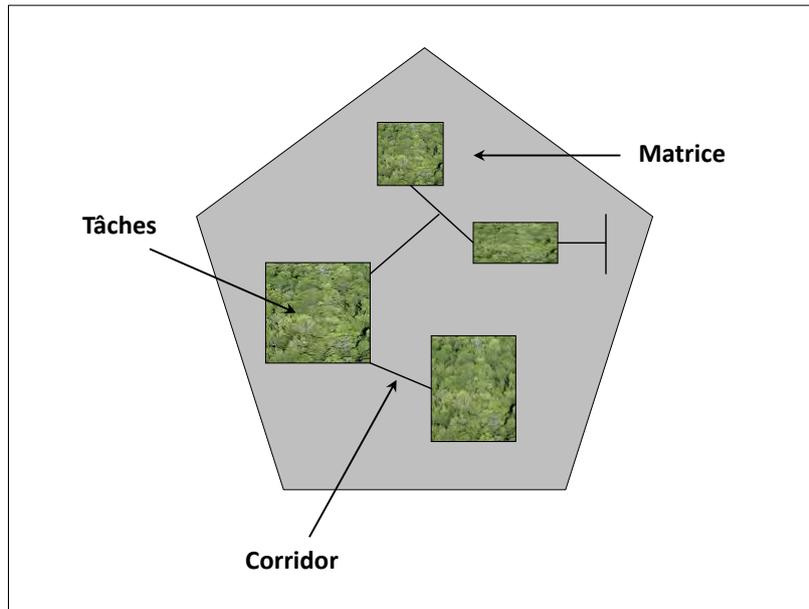


Figure 1-4. Le modèle Tache-Corridor-Matrice adopté en écologie du paysage

Concepts relatifs au modèle tâche-corridor-matrice :

Tâche : patches d'habitat de l'espèce qui apparaissent ponctuellement ou isolément dans la matrice

Corridor : éléments linéaires permettant la dispersion animale ou végétale entre deux tâches d'habitat

Matrice : espace dominant caractérisé par une certaine uniformité d'occupation du sol qui peut être défavorable à l'espèce (l'espèce ne s'y déplace pas ou peu car le milieu lui est plus hostile ou inadapté à ses besoins).

De nombreux travaux ont démontré les relations entre l'hétérogénéité spatiale du paysage évoquée précédemment et la biodiversité (Fraser 1998; Beese et Bryant 1999; Tamme et al. 2010). Cette hétérogénéité spatiale qui se traduit par une diversité d'habitats est à mettre en relation avec le concept de niche écologique que nous évoquerons plus loin.. Des effets positifs de l'hétérogénéité des paysages ont été démontré pour la flore à différentes échelles (Rescia et al. 1994; Burnett et al. 1998; Pausas et Lavorel 2003; Skov et Svenning 2003; Steiner et Köhler 2003; Dufour et al. 2006), pour les mammifères (Kerr et Packer 1997; Fraser 1998), les oiseaux (Freemark et Merriam 1986; Ewers et al. 2005) et avec la biodiversité en générale (Kati et al. 2010). La diversité des habitats permet notamment de faire cohabiter des espèces généralistes qui utilisent plusieurs types d'habitats et des espèces spécialistes qui utilisent des habitats spécifiques présents dans la mosaïque (Fahrig, et al. 2011 ; Redon, 2013). L'hétérogénéité réduit ainsi la compétition interspécifique (qui se déroule généralement le long des frontières de patches d'habitat car les individus interagissent surtout avec les voisins les plus proches) (Pacala et Levin 1997). Cela permet alors aux espèces faiblement compétitives de persister dans le paysage (Tilman et al. 1994). L'hétérogénéité spatiale influence également fortement le fonctionnement des communautés d'espèces, en modifiant les flux de nutriments et cycles de la matière (Rossignol 2006), la productivité et l'invasibilité des systèmes ainsi que la stabilité des populations d'espèces (De Boeck et al. 2006; Turnbull et al. 2007). Plusieurs

études ont notamment montré qu'en augmentant l'hétérogénéité spatiale d'un paysage, la production primaire ou la stabilité des populations d'espèces pouvaient être nettement améliorées (Turnbull et al. 2007).

1.3. LA FRAGMENTATION DES HABITATS : UN PROCESSUS AFFECTANT POTENTIELLEMENT LA BIODIVERSITÉ

Comme déjà évoqué brièvement, un habitat est un élément du paysage offrant les ressources nécessaires à la survie d'une espèce, à sa reproduction ou à ses déplacements (voir Dennis et al. 2003 pour différentes variantes de cette définition). La modification d'habitat peut s'effectuer différemment et peut avoir des conséquences très différentes sur la composition et la configuration spatiale du paysage (Figure 1-5). Selon Fahrig (2003), cette modification peut se traduire par : (A) diminuer le nombre de patches d'habitats avec pour conséquence une augmentation de l'isolement ; (B) augmenter le nombre de patches d'habitat par morcellement avec pour conséquence une diminution de la surface totale des habitats et une diminution de l'isolement ; (C) réduire le nombre de patches d'habitat de petite taille ce qui a pour effet d'augmenter la taille moyenne des habitats ; (D) réduire la taille des patches existants ce qui engendre aussi une augmentation de l'isolement ; (E) diminuer le nombre de patches d'habitats avec pour conséquence une réduction de l'isolement pour les patches restants. D'une manière générale, les habitats peuvent donc : diminuer ou augmenter en taille, diminuer ou augmenter en nombre (notamment par fragmentation mais aussi par apparition/disparition) et être plus isolé ou plus proche de leurs voisins.

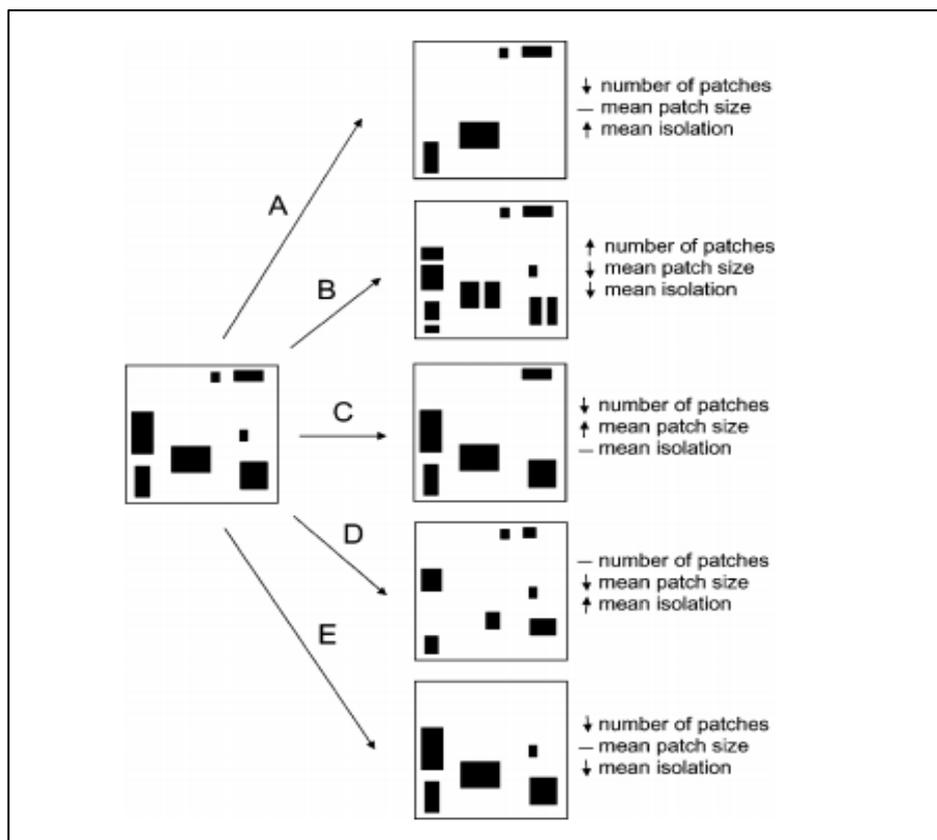


Figure 1-5 Les effets de la fragmentation sur la composition et configuration spatiale des patches d'habitats (source : Fahrig 2003)

De façon plus spécifique, la **fragmentation** est le processus qui transforme une zone continue homogène d'habitat en un nombre plus ou moins important de fragments de taille variable (figure 1-

5). Il en résulte un milieu hétérogène composé de fragments d'habitats discontinus noyés au sein de la matrice paysagère. La cause principale de la fragmentation des milieux naturels est l'activité de l'homme (construction d'infrastructures de transport, urbanisation, etc.) qui engendre plusieurs conséquences sur les habitats. La première conséquence de la fragmentation du paysage est la baisse de la superficie des habitats (effet *surface* de la fragmentation) qui à terme ne permettent plus de fournir les ressources suffisantes pour toute la population d'espèces ce qui entraîne une diminution du nombre d'espèces ou d'individus par espèces (Pimm 1998; Bolger et al. 2000). Nous y reviendrons plus loin en évoquant **la théorie de la biogéographie insulaire (section 1.4.2)**. La fragmentation engendre aussi une évolution du rapport périmètre/surface ce qui augmente la densité de lisières dans le paysage (effet *lisière* de la fragmentation) et entraîne une modification de la composition en espèces de chaque fragment d'habitat restant. Ceci s'effectue, en général, au détriment des espèces spécialistes. Une conséquence supplémentaire de la fragmentation des habitats est que chaque fragment résultant peut se retrouver progressivement isolé au sein de la matrice s'il n'est pas relié aux autres par des corridors ce qui a un impact sur les flux biologiques (effet *isolement* de la fragmentation).. La matrice est plus ou moins hostile aux espèces et n'est pas complètement perméable (on évoque généralement un niveau de porosité de la matrice (Burel et Baudry 1999; Ricketts 2001)). La composition et la configuration des corridors présents dans la matrice permettent d'abaisser ou d'augmenter ce niveau de perméabilité (Forman 1995).

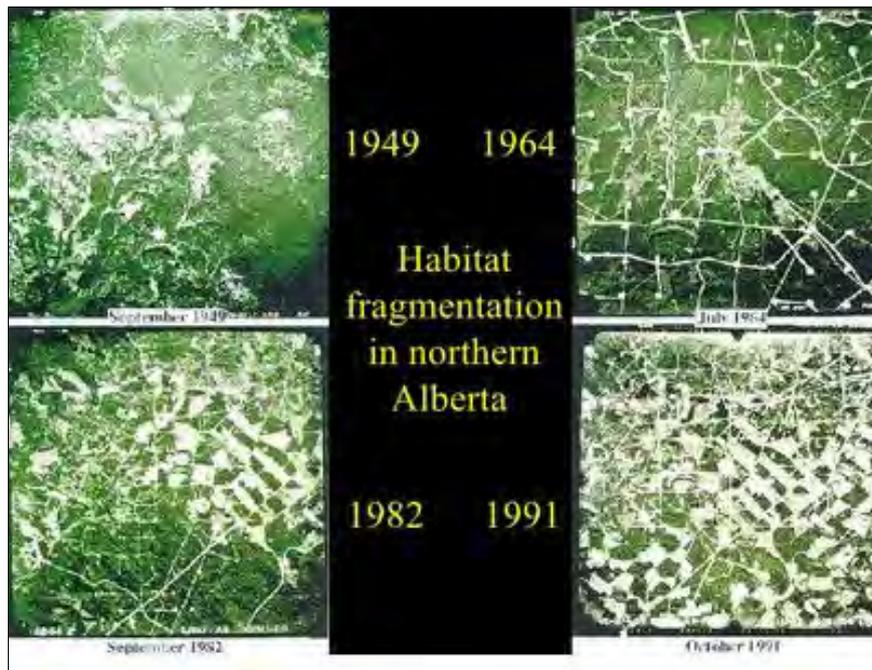


Figure 1-6. Fragmentation d'habitats forestiers au nord d'Alberta, Canada.
(Source : <http://coldcreek.ca/cool-stuff/biodiversity/>)

Même si sa superficie est peu amoindrie, l'habitat se retrouve morcelé en plusieurs fragments parfois inaccessibles, ceci en fonction de l'espèce et/ou de sa capacité de dispersion. Au contraire, une gestion forestière raisonnée peut ne pas entraîner de fragmentation si elle laisse un réseau de tâches forestières dans le paysage suffisamment proches pour qu'elles puissent encore être reliées par les espèces (Pereboom 2006). Autrement dit, deux habitats séparés physiquement dans l'espace (par fragmentation) ne sont pas forcément déconnectés. Ils sont déconnectés d'un point de vue structurel mais pas d'un point de vue fonctionnel. Tout dépend de la capacité de l'espèce à se déplacer. Le niveau de fragmentation et surtout l'effet que cela peut avoir sur les espèces dépend donc très fortement de l'espèce étudiée (Noss 1999).

Comme évoqué plus haut, un des principaux effets de la fragmentation sur la biodiversité concerne la modification des flux d'espèces liée au morcellement des habitats. Lorsque l'isolement est associé à une forte réduction des effectifs, il peut avoir des conséquences désastreuses sur la persistance des espèces dans le paysage. En effet, le risque d'extinction augmente avec la baisse des effectifs des populations lorsque la recolonisation par de nouvelles espèces est limitée en raison de l'isolement des fragments forestiers. Cette conséquence de la fragmentation a été décrite et conceptualisée par **la théorie des métapopulations** (voir section 1.4.2.2). Cette baisse de la diversité spécifique et génétique associée à l'isolement des populations peut avoir des conséquences néfastes sur la viabilité des populations : elle peut à court terme accentuer la consanguinité et faire décliner les performances des individus ; elle peut à moyen terme limiter les capacités des espèces à s'adapter dans un environnement changeant. On considère le plus souvent que le niveau de connectivité doit être égal en moyenne à l'immigration d'un individu par génération : ce niveau permettant ainsi d'atténuer les effets néfastes liés à la réduction des effectifs sans pour autant empêcher une adaptation locale (Lopez et al. 2009).

La fragmentation peut également augmenter l'hétérogénéité du paysage qui lui-même peut avoir des effets sur la diversité génétique et comportementale des populations. L'hétérogénéité encourage les adaptations locales et une meilleure utilisation des ressources. La fragmentation peut donc aussi avoir des effets positifs puisqu'elle augmente la diversité génétique inter-populations mais en faisant baisser la diversité intra-populations. La présence de couloirs de connexions (les corridors) peut alors avoir un double effet : soit ils permettent d'augmenter la variabilité génétique et d'accroître les chances de survie d'une population, soit ils diminuent les chances en faisant disparaître les adaptations locales. De plus, ces couloirs favorisent le plus souvent les espèces invasives.

Dans un contexte de fragmentation, pour qu'une espèce se maintienne, il est nécessaire qu'elle intègre la matrice dans son habitat ou qu'elle soit suffisamment mobile pour la traverser et ainsi retrouver ses habitats initiaux. Généralement, les espèces qui intègrent la matrice à leurs habitats disposent d'une grande flexibilité comportementale et sont nettement moins sujettes aux risques d'extinction. Il s'agit d'espèces généralistes (ou ubiquistes) que l'on retrouve dans de nombreux milieux. Toutefois, les espèces les plus mobiles n'ont pas nécessairement les capacités pour persister au sein de la matrice et l'utilisent généralement de façon temporaire. Elles utilisent des tâches d'habitats séparées dans le temps et dans l'espace (Pereboom 2006) ce qui renvoie aux notions de *complémentation* et de *supplémentation*. **La complémentation** est l'utilisation de tâches d'habitat de nature différente pour subvenir aux besoins nécessaires à la survie d'une espèce alors que **la supplémentation** est l'utilisation de plusieurs tâches de même nature lorsque la taille de chacun ne permet plus d'assurer les ressources nécessaires à une population (Kruuk et al. 1979; Virgos et Casanovas 1999). Enfin, les espèces qui n'utilisent pas la matrice sont les espèces spécialistes et représentent la majorité des espèces en voie de disparition comme certaines espèces d'oiseaux ou de mammifères (Caughley et al. 1996; Brooks et al. 1999).

1.4. LES THEORIES ECOLOGIQUES FONDATRICES

Le modèle tâche-corridor-matrice et plus généralement les notions d'écologie du paysage évoquées précédemment sont fondées sur des théories écologiques déjà bien établies depuis plusieurs décennies. Nous en donnons un aperçu ci-dessous. Le lecteur désireux d'approfondir le sujet pourra se reporter aux références citées.

1.4.1 UN CADRE THEORIQUE POUR ETUDIER LES EFFETS DE L'HETEROGENEITE SPATIALE

Les théories développées ci-après (celle de la hiérarchie et celle de la niche écologique) structurent les analyses concernant les effets de **l'hétérogénéité spatiale du paysage sur la biodiversité** et fournissent un cadre pour leur interprétation.

1.4.1.1. LA THEORIE DE LA HIERARCHIE

Au sein d'une mosaïque paysagère, l'hétérogénéité spatiale perçue est dépendante de plusieurs paramètres, notamment de l'étendue et du grain mais aussi, de la manière dont sont différenciés les patchs (Redon 2012). En écologie, **l'étendue** fait référence au cadre local ou régional d'un processus écologique alors qu'en géomatique, on parlera plutôt d'échelle d'analyse (ou échelle d'étude). Si on utilise plutôt les termes « locales » ou « régionales » pour qualifier respectivement une petite ou grande étendue, les géographes manipulent des « petites » ou des « grandes » échelles, **l'échelle** étant un rapport entre la taille d'un objet sur la carte et sa taille réelle. Par exemple, une carte au 1/50 000^e indique qu'un centimètre sur la carte représente 500 mètres sur le terrain tandis qu'une carte au 1/5 000 000^e signifie qu'un centimètre représente 50 kilomètres : plus le dénominateur est grand, plus l'échelle est petite. En écologie, l'échelle a un autre sens. Elle désigne la taille de l'unité d'observation, ce qui est aussi appelé le grain. Le **grain** chez les écologues désigne ainsi le niveau de détail chez les géomaticiens (Ruas 2004) alors qu'on parlera plutôt de résolution spatiale chez les spécialistes de la télédétection pour définir ce même terme. Lorsque le grain est fin, le niveau de détail est élevé alors qu'un grain grossier correspond à un faible niveau de détail. **L'hétérogénéité aura tendance à diminuer si l'on augmente la taille du grain (donc que l'on diminue le niveau de détail) alors qu'elle tendra plutôt à augmenter avec l'augmentation de l'étendue. De manière similaire, plus les patchs seront représentés avec une résolution spatiale élevée (ou un niveau de détail important) et plus l'hétérogénéité augmentera** (Figure 1-7).

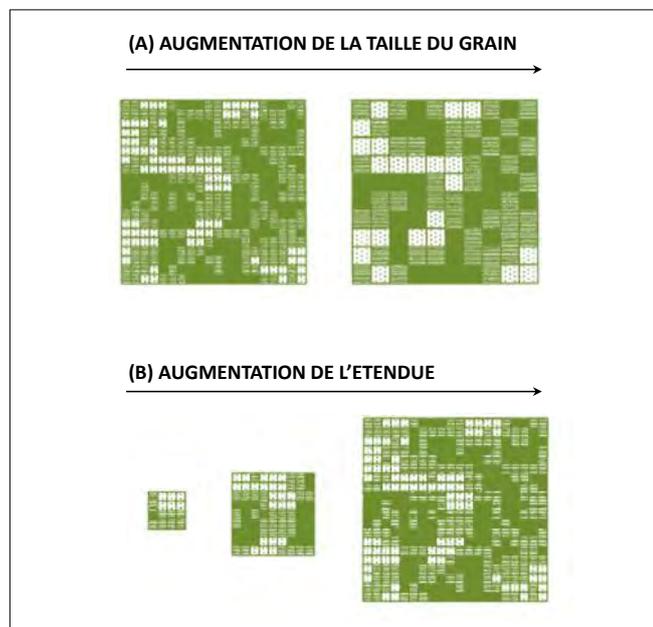


Figure 1-7. Evolution de l'hétérogénéité perçue en fonction du grain et de l'étendue : (A) l'hétérogénéité diminue avec la taille du grain ; (B) l'hétérogénéité augmente avec l'augmentation de l'étendue (Turner, 1989)

Nous venons d'évoquer les notions d'échelle, de grain et de niveau de détail dans la représentation de l'hétérogénéité paysagère. Plus généralement, la compréhension de la structure et

du fonctionnement des écosystèmes ne peut s'appréhender qu'en adoptant une vision multi-échelle, hiérarchique. Cette vision est nécessaire de par le caractère *complexe*² des systèmes écologiques constitués d'un grand nombre de composants dynamiques en interactions. Pour étudier ces systèmes, les écologues s'appuient sur la **théorie de la hiérarchie**. Une synthèse de cette théorie peut être trouvée dans l'ouvrage d'Allen et Starr (1982). Sans en donner une description exhaustive, mentionnons quelques éléments importants qui la caractérisent.

D'abord, un système hiérarchique est composé de plusieurs **niveaux d'organisation** structurés selon des relations d'ordre, verticales (Wu 1999). Ensuite, un même niveau d'organisation est généralement constitué de plusieurs entités en interactions selon des relations horizontales. Les interactions horizontales vont généralement faire émerger des phénomènes au niveau global (notion d'émergence caractéristique des systèmes complexes). Le niveau global agit lui aussi sur le niveau local (les niveaux d'organisation inférieurs). La structure du système se définit donc par le nombre de niveaux d'organisation, le nombre d'éléments par niveau, le nombre de relations entre les niveaux, et le nombre de relations entre les entités par niveau. Par ailleurs, les processus qui existent aux différents niveaux d'organisations opèrent à des **vitesse différentes**. Ainsi, les dynamiques qui se déroulent aux niveaux d'organisation inférieurs sont bien plus rapides que celles propres aux niveaux supérieurs. Autrement dit, les niveaux très englobant, d'étendue plus large, sont caractérisés par des évolutions lentes contrairement aux niveaux inférieurs plus spécifiques, d'étendue plus réduite, intégrant des processus plus rapides.

Pour illustrer ce principe, appuyons-nous sur la structure hiérarchique d'une communauté d'espèce au sein d'un paysage (Figure 1.8). Une communauté est installée dans un espace structuré par des interactions biotiques et abiotiques (l'environnement). Elle contient plusieurs populations (niveaux inférieurs à la communauté) qui elles-mêmes peuvent abriter des sous-populations d'individus (niveaux inférieurs aux populations). Certaines dynamiques, rapides, animent les sous-populations (compétition par exemple) tandis que des dynamiques plus lentes opèrent à l'échelle de la communauté (évolution de l'assemblage des espèces). La théorie de la hiérarchie suggère ainsi un système emboîté, dynamique composé de niveaux d'organisation et de processus agissant à des échelles spatiales et temporelles différentes.

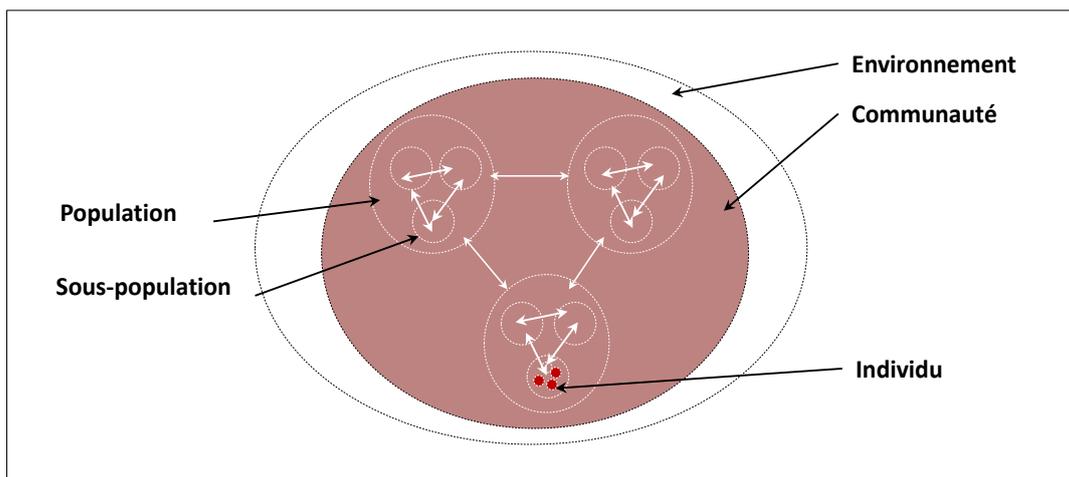


Figure 1-8. L'organisation hiérarchique d'une communauté d'espèces (source : <http://www.com.univ-mrs.fr/~poggiale/ens/M1/UE209/IntroModelM1.pdf>, adapté)

² Au sens de *complexus*, ce qui est tissé ensemble, en interactions, et qui ne peut être décomposé au risque de changer de nature (Morin et al. 1990)

Concepts relatifs à la structuration des organismes :

Individu : unité élémentaire. Être vivant fonctionnellement indépendant et autonome. Par exemple, un animal ou un arbre etc.

Population : ensemble des individus d'une même espèce, végétale ou animale, qui vivent au même instant dans un même écosystème (dans un milieu précis).

Communauté : Le concept de communauté en écologie fait débat depuis 1927 (Elton 1927). La définition la plus large est de considérer une communauté comme un assemblage d'espèces en interaction (directe ou indirecte) en un lieu donné.

Dans le cas de l'écosystème forestier, l'intensité d'une perturbation est plus élevée au sein d'une unité dite homogène que dans un peuplement forestier. Le fait qu'il y ait plus d'interactions au sein d'un seul peuplement qu'entre les différents peuplements feront que les dynamiques de feux par exemple qui occurred au niveau supérieur (au niveau du paysage forestier) seront moins fréquents (Figure 1-9). Dans une logique similaire, puisque le niveau supérieur (le paysage) accueille un niveau inférieur (le peuplement), il est de fait plus grand. Ainsi, des modifications de la composition et de la structure des peuplements induits par la mortalité naturelle des arbres à un niveau inférieur peuvent être propagés ou amplifiés dans les dynamiques à un niveau supérieur en modifiant par exemple la dynamique des feux à l'échelle du paysage (intensité, fréquence, configuration spatiale). En revanche, les événements des niveaux supérieurs tels que les incendies contrôlent ceux des niveaux inférieurs comme la composition en espèces, la structure d'âge et la disponibilité en bois morts par exemple. **Selon la théorie de la hiérarchie, si on observe un événement à une échelle donnée, on reconnaît qu'à d'autres échelles, il existe d'autres événements.**

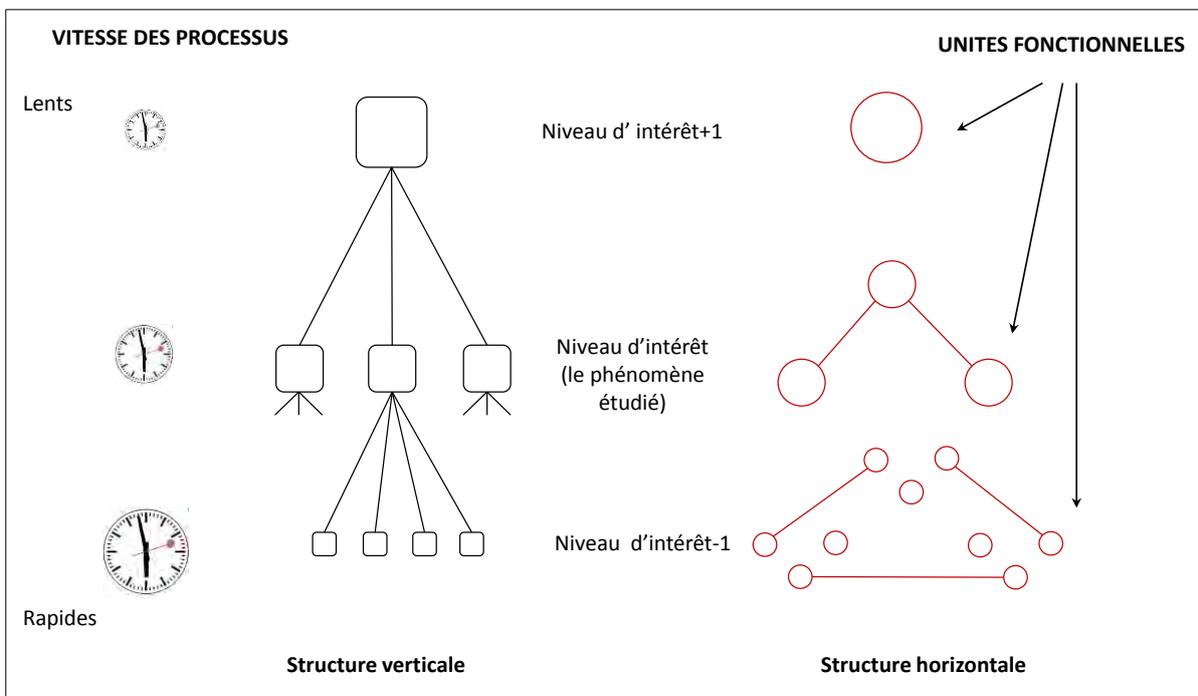


Figure 1-9. Représentation du principe général de la théorie de la hiérarchie (source : Redon 2012 d'après Wu, 1999)

1.4.1.2. LA THÉORIE DE LA NICHE ÉCOLOGIQUE

La théorie de la niche écologique apporte un second formalisme pour étudier les effets de l'hétérogénéité du paysage. La notion de niche permet d'associer à chaque espèce un environnement spécifique favorable à son développement permettant d'expliquer sa distribution spatiale. **De fait, en lien avec l'hétérogénéité spatiale, les espèces ne sont pas présentes partout mais uniquement dans les zones qui correspondent à leur niche** (Dufлот 2013).

Le concept de niche écologique a été introduit par le naturaliste Grinnell (Grinnell 1917) pour décrire les besoins et contraintes environnementales d'une espèce. Le concept opérationnel de la niche écologique est en revanche apparu plus tard avec les travaux du zoologiste Hutchinson (Hutchinson, 1957). Ce dernier décrit la niche comme « *un volume à n dimensions, dans lequel chaque point correspond à un état de l'environnement qui permettrait à une espèce d'exister indéfiniment* ». De fait, on comprend que l'identification de la niche écologique consiste à identifier le domaine de tolérance de l'espèce vis-à-vis des principaux facteurs des milieux (Frontier et Pichod-Viale 1995). Celle-ci est défini par deux grands types de paramètres : (1) les paramètres physico-chimiques caractérisant les milieux où vit l'organisme considéré ; (2) les paramètres biologiques décrivant les relations de compétition et de prédation en lien avec les espèces avoisinantes.

Le concept de niche est fondamental car il contribue à structurer les mécanismes de coexistence des espèces (Redon, 2012). Il est souvent reconnu que deux espèces ne peuvent coexister indéfiniment dans un milieu stable et homogène, la plus compétitive des deux espèces finissant à plus ou moins long terme par éliminer l'autre. Néanmoins, il a été observé que deux espèces voisines aux mêmes exigences alimentaires pouvaient cohabiter sans entrer en compétition. Un des exemples les plus connus est celui du cormoran noir et du cormoran huppé (Benamammar 2014). Ces deux espèces effectuent leur nidification sur les mêmes falaises en Grande Bretagne et pêchent dans les mêmes eaux. Ces espèces devraient alors rentrer en compétition mais elles ne capturent pas les mêmes proies : l'une pêche dans les eaux de surface et l'autre en profondeur. Ce type d'observation a conduit à la définition de la niche écologique proposée par Elton en 1927 (Elton 1927). Elle met l'accent sur les concepts de chaînes alimentaires ainsi que le niveau trophique d'une espèce et donc aux relations impliquées à des échelles plus locales (Guisan et Thuiller 2005).

Concept de la niche écologique :

Au sens d'Hutchinson (1957), la niche est donc un hyper volume où chaque dimension de cet espace représente une ressource ou une condition nécessaire de l'environnement. Sa définition s'oppose à celle d'Elton (Elton, 1927) puisqu'elle est utilisée pour étudier la répartition géographique des espèces à petites échelles (donc vastes étendues) tandis que celle d'Elton est reliée au concept de chaîne alimentaire et permet d'analyser des processus à une échelle beaucoup plus locale.

Hutchinson (1957) introduit également les concepts de niche fondamentale et de niche réalisée. La **niche fondamentale** représente le domaine de tolérance d'une espèce vis-à-vis des facteurs environnementaux. Il s'agit d'une niche théorique et rare sont les espèces occupant l'intégralité de leur niche (Pulliam 2000). Les processus de compétition, de prédation vont en effet contraindre l'espèce à occuper un espace différent de la niche fondamentale (Helaouët et Beaugrand 2009): c'est la **niche réalisée** (figure 1-10). Elle désigne la niche réelle à savoir, la zone réellement occupée par l'espèce même si les observations sont souvent partielles et qu'il est difficile de définir le contour exact de ces niches ou l'ensemble des facteurs qu'elle abrite.

Enfin, si la théorie de la niche est cruciale en écologie, c'est aussi parce qu'elle sert de base à l'établissement des modèles espèces/habitat (Guisan et Zimmermann 2000). La connaissance de la

niche d'une espèce permettra d'identifier les facteurs susceptibles de l'influencer dans le paysage et ainsi introduire des variables explicatives adéquates dans les modèles espèces/habitat.

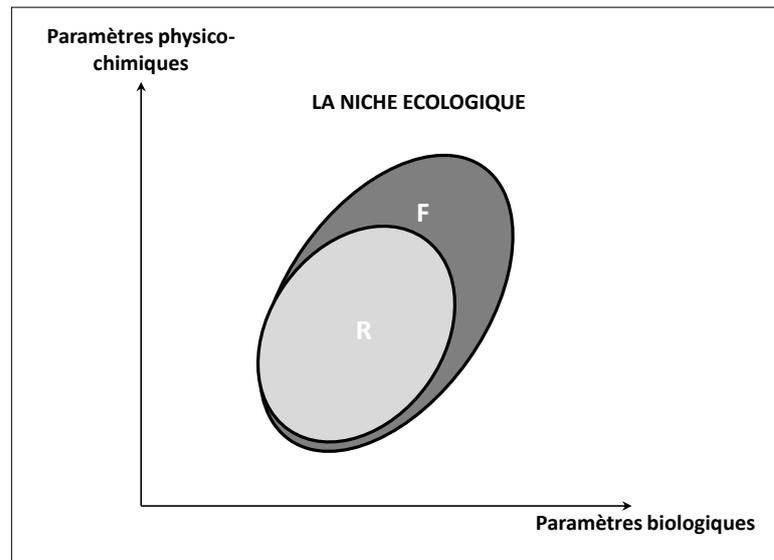


Figure 1-10. Le modèle de la niche écologique selon Hutchinson (1957). L'enveloppe F désigne la niche fondamentale tandis que l'enveloppe R désigne la niche réalisée

1.4.2 UN CADRE THEORIQUE POUR ANALYSER LES EFFETS DE LA FRAGMENTATION DES HABITATS

La fragmentation des paysages entraîne une réduction de la taille des habitats ainsi qu'un accroissement de leur isolement. La prédiction et l'évaluation de ces effets sur la biodiversité sont généralement appréhendées au travers de deux théories écologiques centrales : la théorie de la biogéographie insulaire et la théorie des métapopulations.

1.4.2.1. LA THEORIE DE LA BIOGEOGRAPHIE INSULAIRE

La théorie de la biogéographie insulaire est à l'origine d'un grand nombre de travaux de recherches en écologie du paysage au milieu des années 1960 suite à l'établissement d'un modèle superposable aux écosystèmes terrestres. Le modèle tâche-corrridor-matrice évoqué précédemment est directement issu de cette théorie. **En raison du morcellement du paysage, les fragments d'habitats naturels ou semi naturels (i.e. les tâches) sont vus comme des îles perdues au milieu d'un océan (la matrice).** Nous décrivons les principes de cette théorie ci-dessous en nous inspirant notamment de la synthèse proposée par Dufrêne (2003).

Cette théorie suggère en premier lieu qu'à surface égale, les patches d'habitat contiennent moins d'espèce sur une île que sur le continent. Pour illustrer ce principe, la relation Aire/espèce est le plus souvent mobilisée et permet de définir le lien direct qui existe entre la surface d'un habitat et le nombre d'espèces qui est observé en son sein. Cette relation Aire/espèces est une loi puissance, définie à l'origine par Arrhenius (Arrhenius 1921). Plus la surface est élevée, plus le nombre d'espèces différentes (i.e. la richesse spécifique) est grand. Cette relation a également été démontrée pour des « îles continentales » (c'est-à-dire pour des habitats présents au sein d'une matrice), comme illustré en figure 1-11, et en particulier pour des escargots d'îlots forestiers isolés. La relation se formalise sous la forme suivante :

$$S = C.A^z$$

Avec **S** = nombre d'espèces ; **C** = une constante propre au groupe biologique ; **A** = aire de la zone considérée et **z** la mesure de la pente de la droite qui varie généralement entre 0,2 et 0,3 avec une moyenne de 0,27 (Drakare et al. 2006).

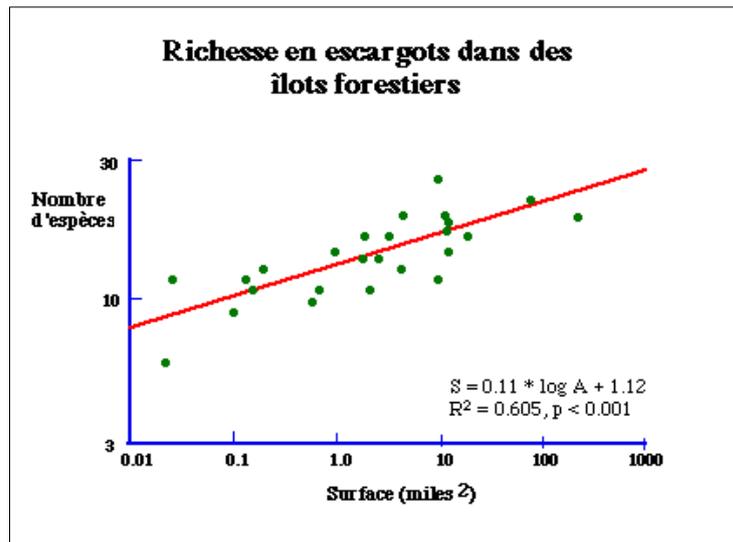


Figure 1-11. Relation aire espèce pour une population d'escargots forestiers. La relation est représentée en logarithmes ce qui en fait une droite (Dufrêne, 2003 d'après Blondel 1995)

Lorsque la relation obtenue pour les îles est comparée à celle obtenue pour des zones de surface équivalente sur le continent, les îles ont toujours moins d'espèces différentes mais cette différence s'amenuise avec l'augmentation de la surface (Figure 1-11). Pour un même accroissement de superficie, l'augmentation du nombre d'espèces s'effectue plus rapidement sur les îles que sur le continent.

C'est également sur les îles que se trouve la plus grande proportion d'espèces endémiques (donc spécifiques au lieu). Hawaï abrite par exemple 1200 plantes vasculaires qui pour la plupart sont endémiques à cette île (Clarke et al. 1984). Une autre particularité concerne la sédentarité des espèces présentes. Les espèces inféodées aux îles ont un pouvoir de dispersion nettement plus faible que celles vivant sur le continent. C'est notamment le cas de nombreux oiseaux ayant perdu leur capacité à voler sur les îles (comme les dodos de l'île d'Hawaï, (McNab 1994). C'est aussi le cas de certains insectes dotés d'ailes de plus petite taille (Darlington, 1971). Etant donné leur faible chance de survie en cas de migration vers une autre île ou continent, la sélection naturelle favorise au fil du temps les espèces plus sédentaires, ayant une faible capacité de dispersion.

En réalité, la richesse en espèces dans les milieux insulaires ne dépend pas que de la taille de ces milieux : le degré d'éloignement des îles par rapport à d'autres îles et surtout des continents (réservoirs d'espèces) joue également un rôle majeur. Considéré comme l'île la plus éloignée du monde (7 jours de mer depuis l'Afrique du Sud), l'île de Tristan Da Cunha est montrée une richesse particulièrement faible en espèces (Ryan et al. 2001). On observe également un gradient de richesse qui diminue au fur et à mesure que l'on s'éloigne du continent. L'isolement des habitats limite les flux de colonisation des espèces et conduit au déséquilibre écologique : le taux d'extinction est plus élevé que le taux de colonisation (cf. ci-dessous).

Le nombre plus limité d'habitats favorables est une autre explication de la pauvreté en espèces (cf. théorie de la niche écologique). Une faible superficie de l'île entraîne une réduction de la diversité des habitats et de leur taille. Ainsi, si une île est trop petite ou bien trop éloignée, elle ne

rassemble pas les conditions favorables à l'installation d'une population, ceci limitant d'autant plus les flux de colonisation des espèces.

Enfin, le processus de compétition entre espèces affecte lui aussi la richesse de biodiversité. Son influence est accentuée par la limitation des ressources disponibles et agit de deux manières possibles : soit une espèce expulse une espèce concurrente déjà présente, soit une espèce empêche une autre de s'installer et défend sa position. Généralement, les espèces présentes agrandissent leur niche écologique initiale pour accroître la disponibilité en ressources et empêchent d'autres espèces de s'installer en marge de leur niche initiale.

Mac Arthur et Wilson (1967) ont proposé un modèle qui suggère que la richesse spécifique (c'est-à-dire le nombre d'espèces différentes dans un habitat donné) d'une île est la résultante d'un équilibre entre deux dynamiques : l'extinction et la colonisation (Figure 1-12). L'évolution du taux de colonisation est représentée sous la forme d'une courbe décroissante. Le fait que les ressources soient limitées implique que la colonisation de l'île par une nouvelle espèce est d'autant plus faible que l'île est riche en espèces (niches écologiques déjà occupées). Parallèlement, le taux d'extinction est lui représenté par une courbe croissante du nombre d'espèces sur une île. Plus le nombre d'espèces est important, plus la compétition entre espèces est forte et provoque une hausse du taux d'extinction. **L'intersection des deux courbes (taux d'extinction est égal au taux de colonisation) conduit à une situation d'équilibre de la richesse spécifique : le nombre d'espèces reste constant. A ce point d'équilibre, le nombre d'espèces dépend uniquement de la surface de l'île et de son éloignement par rapport au continent (le réservoir d'espèces).**

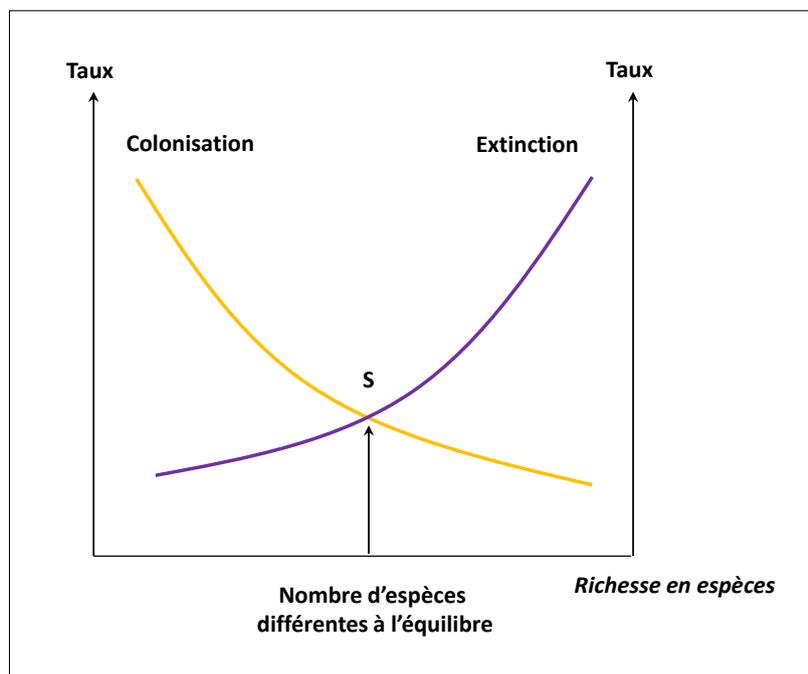


Figure 1-12. Le modèle de l'équilibre extinction/colonisation à la base de la théorie de la biogéographie insulaire (d'après MacArthur et Wilson, (1967). Lorsque le taux de colonisation est égal au taux d'extinction, la richesse en espèces reste constante.

Ce modèle de dynamique est en mesure de faire deux prédictions importantes (Figure 1-13):

(1) Pour un même taux de colonisation, le taux d'extinction diminue quand l'aire augmente (effet positif de la taille sur la richesse).

(2) A surface égale, pour un même taux d'extinction, le taux de colonisation diminue quand la distance au continent augmente (effet négatif de l'isolement).

Autrement dit, le taux d'extinction dépend de la surface de l'habitat tandis que la colonisation dépend de l'isolement.

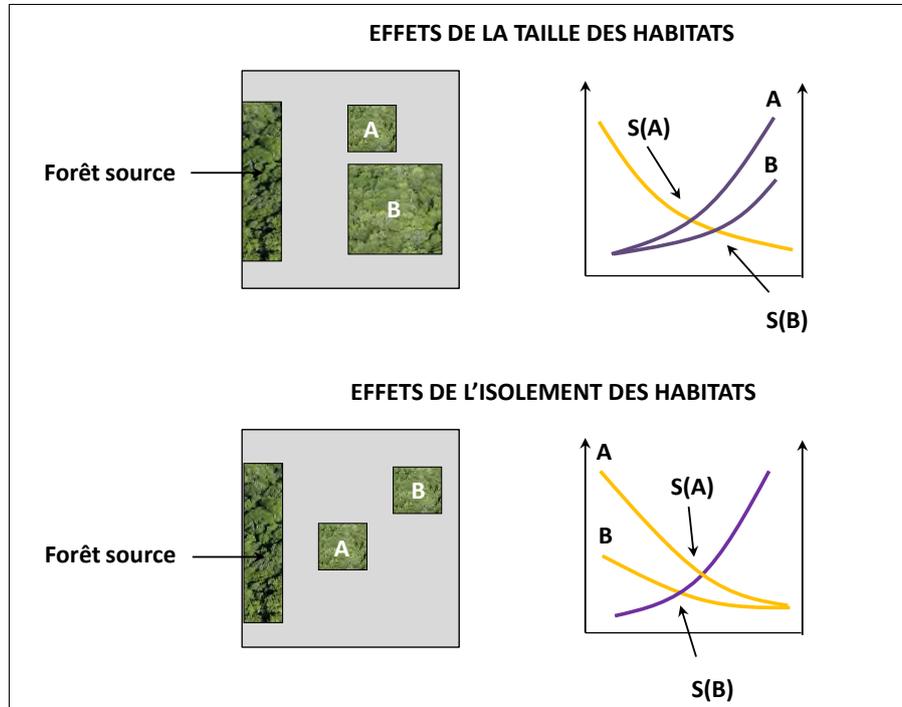


Figure 1-13. Illustration des effets de la taille et de l'isolement des habitats dans le modèle de la biogéographie insulaire (Mac Arthur et Wilson 1967). A taux d'immigration équivalent, le taux d'extinction diminue quand la taille de l'habitat augmente : l'équilibre des dynamiques colonisation/extinction montre que la richesse sera plus élevée dans l'habitat de plus grande taille. A surface égale, le taux de colonisation est plus faible quand l'habitat est plus éloigné de la source : la richesse sera donc plus élevée pour l'habitat le moins isolé.

Si ce modèle a servi de support à de nombreuses recherches, il présente aussi des simplifications importantes. Premièrement, à surface égale, les îles ne sont pas toujours comparables. Des facteurs tels que l'altitude, la diversité d'habitats ou leur complexité déterminent la richesse en espèces d'une île. Deuxièmement, le modèle ne traite que le nombre d'espèces et occulte les densités d'espèces. Or, cette densité d'espèces joue un rôle sur la résistance possible que peut avoir une espèce sur un immigrant. Troisièmement, les espèces sont toutes placées sur le même pied d'égalité alors qu'elles présentent des caractéristiques écologiques différentes. Leur capacité de dispersion va par exemple déterminer leur pouvoir de colonisation.

La reconnaissance de la relation Aire/Espèces et du modèle d'équilibre dynamique a fourni un certain nombre d'indices pour la mise en place de politiques de conservation. Par exemple, sur la base de la courbe Aire/Espèces, il serait logique de privilégier une seule surface de grande taille pour maintenir la biodiversité plutôt que d'installer des réserves de petite taille même si la surface est identique (Dufrêne 1992). Soulé et al (1986) se sont appuyé sur des populations mammifères africains pour démontrer que des parcs naturels de 10000m² perdront la moitié de leurs grands mammifères en moins de 5000 ans. Ainsi, deux grandes orientations ont été proposées sur la base de ces observations (Soulé et Simberloff 1986) :

(1) Les espaces dédiés à la biodiversité doivent être les plus grands possibles puisque le taux d'extinction est impacté par la surface.

(2) Les espaces dédiés à la biodiversité doivent être les plus proches possibles afin de maintenir un fort taux d'immigration et par conséquent diminuer le risque d'extinction.

Cette approche a toutefois été largement discutée et a fait l'objet d'une grande polémique dans la communauté scientifique : la polémique SLOSS « Single Large or Several Small » (McNeill et Fairweather 1993; Virolainen et al. 1998). L'approche contraire proposée par ces polémistes privilégie plutôt une dispersion des réserves dédiées à la biodiversité et ce pour plusieurs raisons. Premièrement, la relation Aire/Espèces ne donne aucune indication sur la similarité des espèces présentes sur des îles de petite taille. Deuxièmement, la dispersion d'habitats offrirait la possibilité de conserver certaines espèces occupant la même niche écologique et qui seraient en compétition. Troisièmement, des espèces typiques d'un seul stade de succession ont de grandes chances de disparaître sur une zone unique. Enfin, des catastrophes naturelles ou des épidémies impacteraient plus fortement les populations dans une réserve unique de grande taille alors que des patches dispersés favoriseraient la restauration de la population totale grâce aux noyaux réservés.

Nous verrons dans la suite du manuscrit que cette théorie de la biogéographie insulaire guidera nos choix pour caractériser la structure spatiale des habitats forestiers et modéliser leur relation avec nos indicateurs de biodiversité.

1.4.2.2. LA THÉORIE DES MÉTAPOPULATIONS

Une dernière théorie écologique mérite d'être présentée. Il s'agit de la théorie des métapopulations qui s'appuie sur des concepts similaires à la théorie de la biogéographie insulaire à savoir, l'extinction et la colonisation (on parle aussi de *turnover*). Elle offre la possibilité de prédire les habitats occupés en fonction de différentes relations que nous décrivons ci-après.

Levins, en 1970, introduit le concept de métapopulation comme une **population de populations locales** qui s'éteignent et recolonisent localement des habitats inoccupés (Levins 1970). La persistance d'une **métapopulation** dans une région n'est alors possible que si le taux d'extinction moyen est inférieur au taux de migration (Burel et Baudry 1999). Les populations locales occupent des **tâches** d'habitats qui peuvent être occupées ou libres à un instant donné. Les individus disperseurs peuvent ainsi quitter une tâche pour aller coloniser un élément vide ou renforcer une population de faible effectif. Ce modèle est particulièrement utile pour la biologie de la conservation puisque la plupart des habitats sont fragmentés ou dispose de plusieurs degrés de dispersion (Hanski et Gilpin 1991). Il permet d'estimer l'évolution de la proportion des habitats occupés au cours du temps en fonction de deux processus : la colonisation et l'extinction.

Le modèle de Levins est une équation différentielle prenant la forme suivante :

$$dP/dT = cP(1-P) - eP$$

où **c** est le patch inoccupé, **P** le nombre de sites occupés et **e** le patch disponible. La colonisation est alors exprimée par le produit de la probabilité de coloniser un patch inoccupé (**c**) et du nombre de sites inoccupés (**1-p**). L'extinction est quant à elle exprimée par le produit de la probabilité qu'un patch occupé devienne disponible (**e**) et du nombre de sites occupés (**p**). Pour que la population régionale soit à l'équilibre (extinction = colonisation), le taux de colonisation doit être plus grand que le taux d'extinction. L'équilibre est atteint pour le niveau **P* = 1-(e/c)**.

La différence majeure qu'entretient ce modèle avec le modèle de Mac Arthur et Wilson est que la colonisation ne s'effectue plus de manière constante à travers une seule source mais se fait à partir d'autres sites déjà occupés. Le taux de colonisation (**c**) est donc pondéré par le nombre sites

occupés (p). Le fait que le système fonctionne sans source extérieure implique que le taux de colonisation doit être plus grand que le taux d'extinction pour atteindre l'équilibre.

Concepts relatifs aux études sur les métapopulations (d'après Burel et Baudry, 1999)

Tâche : espace continu dans lequel une population locale trouve les ressources nécessaires à sa persistance

Population locale : ensemble d'individus vivant dans la même tâche et en interactions

Métapopulation : ensemble de populations locales séparées spatialement entre lesquelles un certain nombre d'échanges est possible

Turnover (ou taux de colonisation-extinction) : extinction de populations locales et établissement de nouvelles populations locales dans les habitats inoccupés par des individus provenant de populations locales déjà installées.

Une des grandes simplifications introduites par ce modèle, dans sa version initiale, concerne la non prise en compte de la dimension spatiale puisqu'il ne tient pas compte des coordonnées des sites. Ces derniers sont tous supposés avoir la même taille et équidistants les uns des autres. Cette simplification est cruciale puisqu'elle ne tient pas compte des effets potentiels de la fragmentation susceptibles de perturber fortement les dynamiques de populations locales.

Dufrène (2003) propose ainsi d'ajouter un terme symbolisant la perturbation (**D**) dans le modèle de Levin pour tenir compte de ces effets (Dufrène 2003). La nouvelle équation prend la forme suivante :

$$dP/dT = cP(1-P-D) - eP$$

Cette fois, le point d'équilibre théorique prend la valeur $P^* = 1 - D - (e/c)$. La destruction des habitats entraîne alors une disparition plus rapide des populations même si la présence d'habitats disponibles est encore assurée. Si cette valeur **D** est plus grande ou égale à la valeur $1 - (e/c)$, c'est-à-dire le point d'équilibre sans la présence d'une destruction, la population régionale s'éteindra.

D'autres modèles de métapopulations ont été développés pour améliorer celui de Levins et le rendre spatialement plus explicite (figure 1-14).

- Le modèle de Boorman et Levitt (Boorman et Levitt 1973)

Ce modèle suggère l'établissement d'un grand fragment central abritant une population permanente (qui ne s'éteint jamais) qui alimentera un jeu de populations locales avoisinant. Ce modèle se rapproche de la théorie de la biogéographie insulaire puisque ce grand fragment central joue le rôle de source d'individus disperseurs de la même façon que le continent pour les îles océaniques.

- Le modèle source-puits (Pulliam 1988)

Il prend en compte les différences de la qualité des parcelles d'habitat. Les parcelles dites sources contiennent des ressources abondantes qui permettent aux individus de produire plus de descendants que nécessaires pour les remplacer (croissance positive de la population). Le surplus de descendants produits est dispersé vers les d'autres parcelles. Dans les parcelles où les ressources sont moindres, les populations sont maintenues grâce au taux d'immigration d'autres individus : la croissance est alors négative s'il n'y a pas d'immigration. Dans ce modèle, les flux de migrants sont inexistants, les extinctions ne sont donc pas compensées par la

colonisation d'autres individus (Harrison et Taylor 1997). Les relations entre parcelles d'habitats ne sont conservées que par le biais des dynamiques locales puisque la dynamique régionale est inexistante.

- Le modèle de non-équilibre

Dans ce modèle, les flux de migrants sont inexistants, les extinctions ne sont donc pas compensées par la colonisation d'autres individus (Harrison et Taylor 1997). Les relations entre parcelles d'habitats ne sont conservées que par le biais des dynamiques locales puisque la dynamique régionale est inexistante.

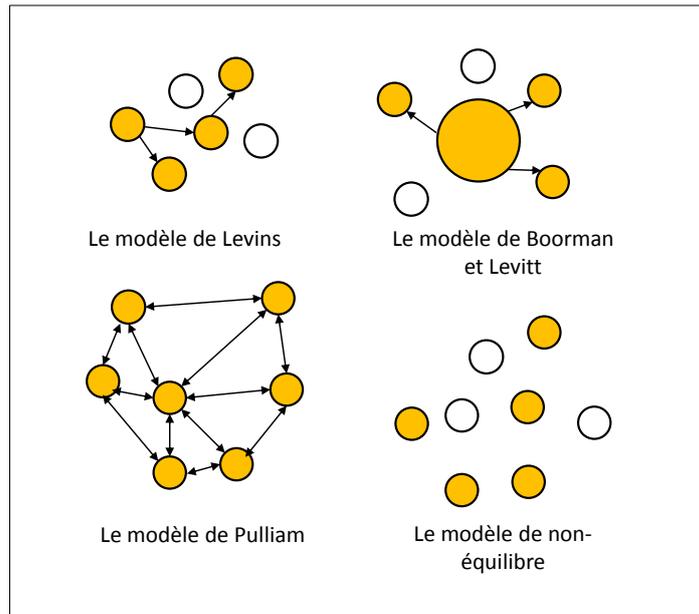


Figure 1-14. *Les modèles de métapopulation*

CHAPITRE 2. OBJET D'ETUDE : LA FORET PAYSANNE FRAGMENTEE ET LA BIODIVERSITE ASSOCIEE

SOMMAIRE

1.1. LA FORET PAYSANNE : UN RESERVOIR DE BIODIVERSITE DANS L'ESPACE RURAL

1.2. DE QUELLE BIODIVERSITE PARLE-T-ON ? COMMENT EST-ELLE MESUREE ?

1.3. LES FACTEURS GENERAUX AFFECTANT LA BIODIVERSITE FORESTIERE

1.4. L'ANCIENNETE DES FORETS : UN FACTEUR CLE POUR COMPRENDRE LEUR BIODIVERSITE

2.1. LA FORET PAYSANNE : RESERVOIR DE BIODIVERSITE DANS L'ESPACE RURAL

2.1.1 LA FORET PAYSANNE EN DECLIN

Occupant 30 % terres émergées, la forêt est le principal couvert végétal de la planète et se trouve en réduction sur plusieurs continents. Bien que cette baisse soit en légère diminution, on observe encore 8 millions d'ha/an de pertes nettes de forêts dans le monde (FAO¹, 2005). En Europe, les surfaces boisées représentent 4% de la surface forestière mondiale et sont en accroissement régulier. Entre 2000 et 2005, la forêt a gagné 660 000 ha/an.

La forêt française connaît une évolution similaire. Entre 1980 et 2010, la forêt a augmenté de 38000 ha/an (IFN 2011) et occupe actuellement 16,4 millions d'hectares. La surface a doublé depuis la fin du XIXe siècle puisqu'elle était comprise en 1830 entre 8,9 et 9,5 millions d'hectares (Cinotti 1996). Les cartes ci-dessous représentent respectivement le taux de boisement français en 2010 et le taux d'accroissement annuel de boisement entre 1980 et 2010 par département (Figure 2-1). Le taux de boisement varie de 0,5% dans la Manche à 67% en Corse du Sud. Le taux moyen français est à hauteur de 30 %. L'extension de la superficie forestière concerne majoritairement le gr et arc méditerranéen et la Corse ainsi que la Bretagne et les Pays de la Loire. Le sud-est profite des terres délaissées par le pastoralisme et l'agriculture tandis que des régions historiquement forestières connaissent une progression moins importante.

La forêt française est constituée à 75% de forêts privées soit près de 12 millions d'hectares sur les 16,4 millions de la forêt française. Elles sont possédées par 3,5 millions de forestiers privés dont les deux tiers ont au moins un hectare. Parmi eux, seulement 12% sont des agriculteurs. Ces forêts, à la charge des agriculteurs, sont appelées « forêt paysanne ». Elles peuvent être définies comme toutes les forêts, quel que soit le régime juridique ou la structure des peuplements forestiers, gérées par les exploitants agricoles (Normandin et Cinotti 2002). Nous verrons plus loin que les forêts que nous étudions dans cette thèse font partie de ces forêts paysannes.

En 30 ans, les surfaces boisées rattachées aux exploitants agricoles ont diminué de près de 60% passant de 3,1 millions d'hectares en 1970 à 1,3 millions en 2000. La séparation de l'agriculture et de la forêt est donc une tendance lourde (Normandin et Cinotti 2002). Cette baisse s'est sensiblement accélérée au cours des années 1990 passant de -55000ha/an dans les années 1980 à -74 000ha/an au milieu des années 1990 (Figure 2-2).

Le phénomène marquant des années 1990 est la nette diminution de la proportion des exploitants agricoles disposant de surfaces boisées. Si elle est restée stable jusqu'en 1988 (env. 43%), elle n'est plus que de 32% aujourd'hui. Finalement, la réduction du nombre d'exploitants agricoles possédant des surfaces boisées n'a fait qu'accompagner le mouvement général du recul du nombre d'exploitations même si la surface boisée s'est réduit proportionnellement beaucoup plus vite (-29%) que la surface agricole (-4,4%).

¹ Organisation des Nations Unies pour l'alimentation et l'agriculture

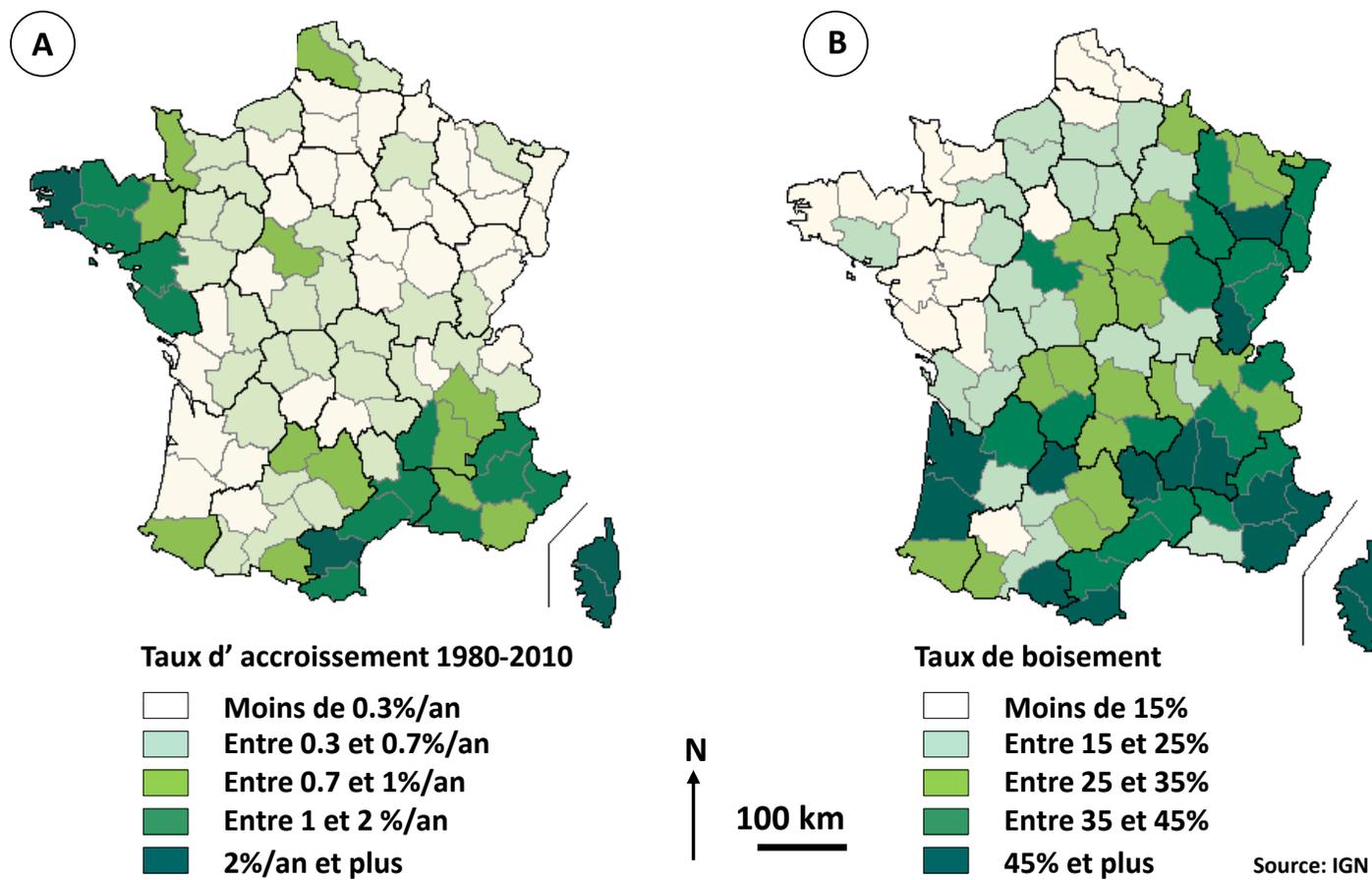


Figure 2-1 (A) Taux d'accroissement des surfaces boisées par département de 1980 à 2010 ; (B) Taux de boisement par département en 2010

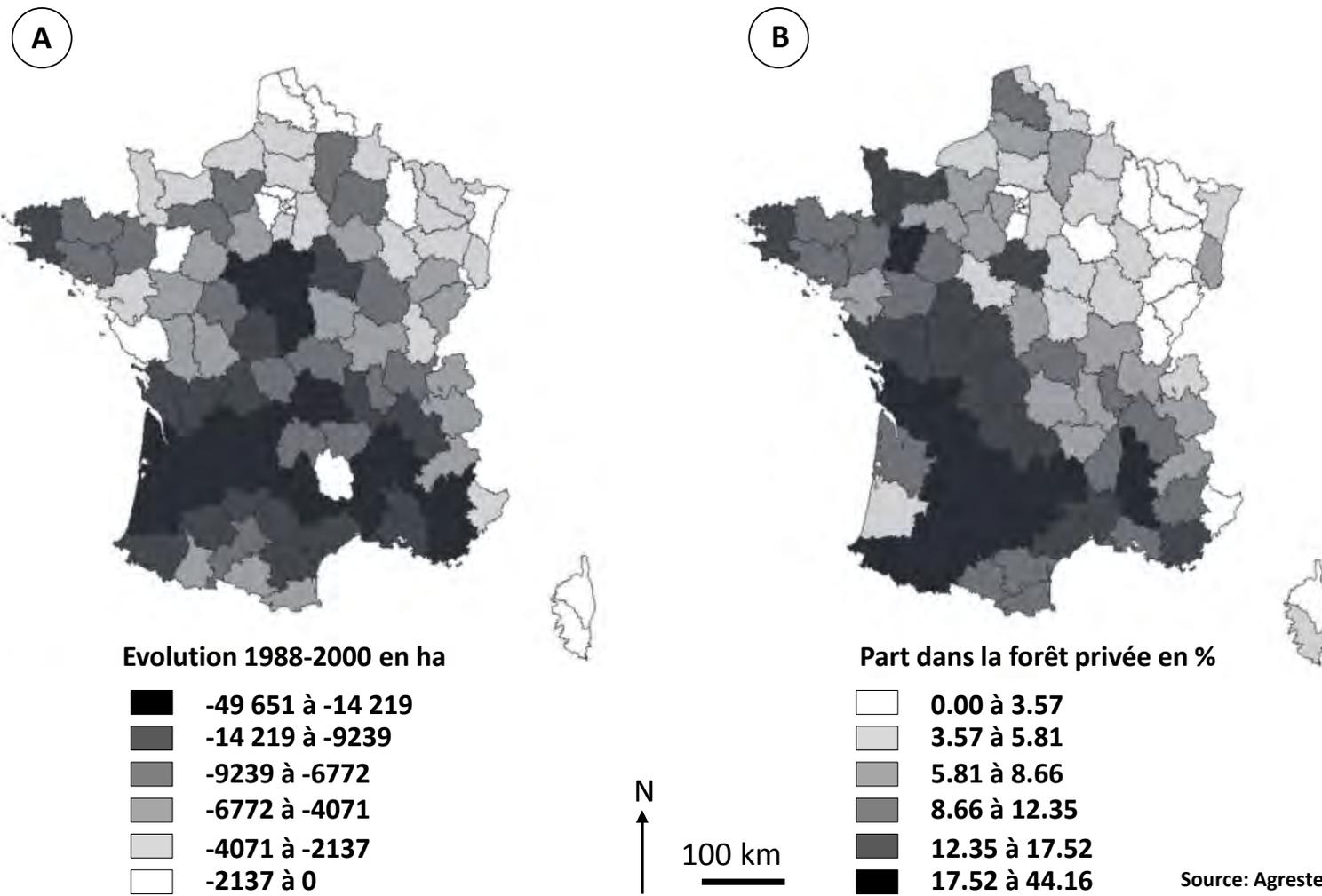


Figure 2-2. (A) Evolution de la surface des forêts paysannes par département de 1988 à 2000 (en ha) ; (B) Part de la forêt paysanne dans la forêt privée en 2000 (en %)

2.1.2 LA RECONNAISSANCE DE LA RICHESSE DES FORETS PAYSANNES

Si la forêt paysanne a connu un tel recul depuis 30 ans, c'est que la reconnaissance de sa valeur et des actions qui en découlent n'a eu lieu qu'au début des années 1990.

En 1995, la mise en place d'un Groupe National Agriculture Forêt (GNAF) sous l'égide du ministère de l'agriculture prévoit de s'intéresser aux forêts contenues dans les exploitations agricoles, aux forêts entretenues par les agriculteurs et aux boisements des terres agricoles dans une perspective d'aménagement du territoire. Dans ce cadre, tous les types de formations boisées sont pris en compte : les forêts anciennes, les accrues forestiers, les plantations sur des terres agricoles, les boisements linéaires et les différentes formes d'agroforesterie. Si la préoccupation d'aménagement du territoire est bien prise en compte via ce groupe, les questions environnementales ne sont encore que très peu abordées. Toutefois, la création d'un sous-groupe « forêt paysanne » au sein du GNAF témoigne de l'intérêt suscité et de l'évolution des mentalités. Il s'agit « *d'apporter des réponses concrètes et positives à la question du boisement des terres agricoles [...] dans le respect des activités agricoles de chacun et dans le souci de la protection et de valorisation de l'environnement* ». Plus largement, aujourd'hui, les forêts privées font l'objet d'un vaste réseau de protection et de conservation. Les ZNIEFF (Zone d'Intérêt Ecologique Faunistique et Floristique) ne font pas figure de protection juridique mais aident à la décision en matière d'aménagement du territoire tout en préservant le patrimoine naturel. La moitié de ces zones sont constituées de forêts dont les deux tiers sont des forêts privées. Par ailleurs, les deux tiers des forêts localisées dans les parcs naturels régionaux sont des forêts privées tout comme 14% des surfaces boisées membres du réseau NATURA 2000.

D'un point de vue scientifique, l'INRA (Institut National de la Recherche Agronomique) a pour sa part abordé la forêt paysanne en s'intéressant en premier lieu aux éléments linéaires et notamment aux fonctions écologiques des haies dans les exploitations agricoles (Forman et Baudry 1984; Baudry et Schreiber 1988). Dans la foulée et sous l'impulsion de la région Midi-Pyrénées, l'INRA s'est focalisée sur la forêt paysanne fragmentée si caractéristique des plaines de la région. Jusqu'à aujourd'hui, l'INRA² et l'IRSTEA³ mettent en œuvre des recherches sur les boisements des terres agricoles autour de plusieurs points majeurs tels que : le rôle des formations boisées sur la qualité de l'eau, les fonctions économiques et environnementales de la forêt paysanne à l'échelle des exploitations et des territoires ou encore les facteurs affectant la biodiversité forestière.

2.1.3 LA FORET PAYSANNE MORCELEE : UN ENVIRONNEMENT PARTICULIER

Dans les paysages ruraux, la forêt paysanne constitue souvent un des derniers refuges d'une vie sauvage et un des éléments clés du maintien de la biodiversité résiduelle (Deconchat et Balent 1996).

Il est important de faire la distinction entre la biodiversité d'un bois donné de celle d'un paysage. Dans le cas d'un paysage, le questionnement concerne l'influence d'un bois ou d'un groupe de bois sur la dynamique d'ensemble des populations d'espèces forestières ou agricoles dispersées dans l'espace. Une forêt disposant d'une richesse moindre peut ainsi jouer un rôle crucial dans la circulation d'espèces dans le paysage. Inversement, l'expansion de la forêt peut avoir des effets néfastes sur la diversité des espèces inféodées aux milieux ouverts.

L'une des caractéristiques majeures des formations boisées des paysages agricoles est leur fragmentation (cf. chapitre 1). Par conséquent, les conditions nécessaires à l'installation d'une population d'espèces dans un bois ne dépendent pas uniquement des caractéristiques internes du

² INRA : Institut National de la Recherche Agronomique

³ IRSTEA : Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture

bois mais aussi de sa morphologie (taille, forme, lisière) et de sa localisation dans le paysage (isolement). L'ensemble des connaissances acquises au fil des années sur ces différents points permet aujourd'hui d'alimenter le débat sur la gestion environnementale des paysages ruraux. Quels peuvent être les conséquences de l'introduction d'un bois dans le paysage ? Dans un cas, elle peut certes renforcer la qualité du corridor et faciliter les flux d'espèces mais dans un autre, elle peut n'avoir aucun effet sur des espèces circulant à une échelle différente. L'expansion de la forêt est-elle forcément un point positif pour la préservation de la biodiversité dans les espaces ruraux ? L'accroissement de la surface d'un bois permet en règle générale une diversification des ressources en son sein permettant de satisfaire un plus grand nombre d'espèces. Toutefois, elle peut aussi être source de barrière physique dans le paysage empêchant certaines espèces de rejoindre leur milieu privilégié ou bien affecter les espèces des milieux ouverts.

Ces deux exemples de questionnements montrent bien la complexité que représente l'implantation et la gestion durable des forêts dans les paysages agricoles morcelés. Le processus de fragmentation est difficile à appréhender : il modifie les mécanismes de dispersion des espèces qui doivent s'adapter. Il favorise certaines espèces tout en affectant d'autres et son caractère dynamique rend difficile la mise en place des politiques de conservation (Deconchat et Balent 1996).

2.2. DE QUELLE BIODIVERSITE PARLE-T-ON ?

La biodiversité peut se définir de différentes façons et fait référence à différentes notions en écologie. Le concept de biodiversité a été proposé par Walter Rozen en 1985 et a connu sa notoriété lors de la conférence de Rio en 1992. Au sens littéral, la biodiversité signifie la diversité du vivant. L'Organisation des Nations Unies en donne la définition suivante : "*the variability among living organisms from all sources, including, terrestrial, marine, and other aquatic ecosystems, and the ecological complexes of which they are part: this includes diversity within species, between species and of ecosystems*" ("Convention on Biological Diversity" des Nations Unies).

Ainsi, le terme biodiversité fait communément référence à la richesse en espèces d'un écosystème. Toutefois, cette vision est assez restrictive et les écologues incluent également d'autres formes de diversité. La diversité des espèces est une forme de biodiversité qui englobe la diversité du monde sous toutes ses formes (Larrieu et Gonin 2008) : (i) **la diversité taxonomique** (diversité des espèces), **la diversité génétique** (diversité au sein d'une espèce) et **la diversité des écosystèmes**. **La diversité fonctionnelle** est également prise en compte en considérant les différents services assurés par les différents groupes d'espèces au sein de l'écosystème.

La biodiversité est étudiée à différentes échelles écologiques : de l'individu à l'écosystème en passant par la population, l'espèce et la communauté (cf. section 1.4.1.1). Elle est aussi étudiée à différentes échelles spatiales : de la biodiversité d'un habitat particulier à celle d'une région et d'un continent.

2.2.1 UNE BIODIVERSITE FORESTIERE MENACEE

La France est réputée disposer d'une biodiversité forestière riche et menacée (Vallauri, 2009). Certaines espèces ont totalement disparues (le bison d'Europe, l'élan) alors que d'autres ont disparues avant d'être réintroduites. Les activités anthropiques sont les raisons qui permettent d'expliquer ces menaces et touchent la majorité des taxons même si les données sont parfois lacunaires pour certains groupes d'espèces.

La liste rouge des espèces forestières menacées (UICN⁴) donne un aperçu de l'ampleur des dégâts causés. La plupart des connaissances portent sur les espèces végétales. Ainsi, cette liste

⁴ Union Internationale de Conservation de la Nature

indique que 95 espèces végétales doivent faire l'objet d'une grande vigilance mais des données sur les groupes entiers d'espèces manquent et peuvent faire penser que ce constat est largement sous-évalué (Vallauri et Neyroumande 2009). Les oiseaux sont aussi fortement impactés. Selon l'ouvrage « Birds in Europe » publié par Birdlife, 226 espèces d'oiseaux ont un statut de conservation défavorable soit 43% de l'avifaune. La France qui héberge plus de la moitié de ces espèces menacées est largement concernée (40 des espèces d'oiseaux menacée) (Redon 2012). Par ailleurs, même si la majorité des espèces d'amphibiens ou de reptiles sont protégées au niveau national, peu de connaissances existent à leur sujet. C'est aussi le cas pour les insectes alors qu'ils sont indispensables à l'écosystème forestier. Certains groupes comme les coléoptères sont mieux documentés et rendent compte des effets provoqués puisque 40 % de ces derniers sont menacées d'extinction.

2.2.2 COMMENT EST-ELLE MESUREE ?

La biodiversité est le plus souvent mesurée au travers d'indicateurs (Gosselin et al. 2009; Collin et al. 2012; Nivet et al. 2012; Raymond 2012).

Selon Levrel (2007), « **un indicateur de biodiversité est un outil d'évaluation de la biodiversité construit à partir de données descriptives (qualitatives, quantitatives) mesurées périodiquement et qui permettent de faciliter le suivi des évolutions relatives à la biodiversité** ». Par conséquent, un indicateur de biodiversité forestière est la restriction d'indicateurs de biodiversité au domaine forestier. Il faut toutefois dissocier les indicateurs écologiques de la biodiversité forestière des indicateurs socio-économiques qui concernent la dimension sociale et la dimension économique de la biodiversité. Les indicateurs sont des outils essentiels pour plusieurs raisons. Ils permettent d'une part d'évaluer l'état de la biodiversité et d'autre part d'offrir un moyen de suivi de la préservation de la biodiversité. Ils facilitent la prise en compte des objectifs de conservation de la biodiversité à travers les politiques de planification de la gestion.

En forêt, les indicateurs traduisent la présence ou l'état de conservation d'espèces ou d'habitats. Ils peuvent être utiles pour la détection d'espèces d'intérêt, évaluer l'état d'habitats patrimoniaux⁵ ou menacées ou bien rendre un diagnostic sur l'état de santé de tel ou tel écosystème. Ces indicateurs peuvent être soumis à différentes contraintes qui rendent parfois leur construction laborieuse.

On distingue généralement deux types d'indicateurs pour la forêt: les indicateurs **directs** et **indirects**. Ces notions indiquent le fait que les données mesurées donnent directement ou non des informations sur celle que l'on veut caractériser, c'est-à-dire **la biodiversité ciblée** (GIP ECOFOR, 2012) (Figure 2-3).

Les indicateurs directs (ou taxonomiques) consistent à échantillonner directement les espèces à partir d'inventaires taxonomiques. Ils se focalisent sur le suivi des taxons (plantes, insectes, mammifères) et visent à renseigner sur l'état de la biodiversité directement. Ils peuvent prendre la forme d'indicateurs de richesse d'espèces (nombre d'espèces différentes dans une station ou un biotope donné), d'espèces indicatrices ou de diversité (Figure 2-3). Enfin, la construction de ces indicateurs est parfois difficile lorsque l'on souhaite étudier de vastes portions de territoires et disposer d'informations sur une ligne de temps relativement longue. L'accessibilité au terrain, les conditions météorologiques ou encore la difficulté d'observer certaines espèces (espèces nocturnes) peuvent par exemple constituer un frein au sein d'un protocole d'échantillonnage (Redon 2012)

Dans cette logique, les indicateurs indirects permettent de s'affranchir de ces difficultés d'échantillonnage. Ils s'appuient sur l'analyse de l'hétérogénéité de la structure, de la composition ou de l'organisation spatiale pour expliquer et prédire la diversité des espèces présentes dans les

⁵ habitat riche en espèces remarquables et protégées au niveau national ou international

forêts (Figure 2-3). On ne procède donc pas un à un à un échantillonnage direct des espèces mais on passe plutôt par un ensemble de proxys (les indicateurs de structure du paysage) qui vont permettre de prédire la présence de biodiversité. Par conséquent, une bonne connaissance des relations entre la structure du paysage et la biodiversité ciblée est primordiale (cf section 1.4.1.2 sur la niche écologique) pour être en mesure d'identifier les bons déterminants de la biodiversité que l'on souhaite étudier. Ces indicateurs ont donné lieu à un certain nombre de travaux concernant le suivi des écosystèmes forestiers. Un des plus récents est l'Indice de Biodiversité Potentielle (IBP) développé par Larrieu et Gonin (2008) dans le cadre des forêts privées françaises (Larrieu et Gonin 2008). L'idée initiale de cet outil est d'offrir aux gestionnaires un moyen simple et rapide d'évaluer la biodiversité potentielle à l'échelle du peuplement pour ensuite intégrer cette connaissance dans leur gestion courante. Ce dernier s'appuie sur dix facteurs reconnus comme les plus favorables à la diversité interne des peuplements forestiers : composition spécifique et structuration du peuplement, maturité et offres en micro-habitats, présence d'habitats liés à la forêt, continuité historique de l'état boisé. Plusieurs facteurs s'attachent à décrire le bois mort et les micro-habitats afin d'intégrer la diversité des espèces saproxyliques ⁶(25% de la diversité taxonomique forestière). La gestion et le peuplement sont pris en compte au moyen de sept facteurs alors que trois autres facteurs sont plutôt liés au contexte paysager. Un système de notation est ensuite utilisé pour caractériser chaque facteur. Enfin, l'essor des outils informatiques et des techniques de télédétection a aussi considérablement boosté ce type d'approches en permettant d'obtenir des données sur une vaste portion de territoire (Oindo et Skidmore 2002; Fairbanks et McGwire 2004; Gillespie et al. 2008). Palmer et al., (2007) ont par exemple utilisé l'indice SVH (Spectral Variations Analysis) faisant l'hypothèse que la variation de la réponse spectrale d'une image traduit en réalité la diversité des espèces du couvert étudié (Palmer et al. 2002). Rocchini et al. (2007) ont eux aussi démontré la potentialité des images multi spectrales et plus spécifiquement des canaux infrarouges pour prédire la richesse spécifique de plantes (Rocchini et al. 2007).

⁶ espèces inféodées aux bois morts

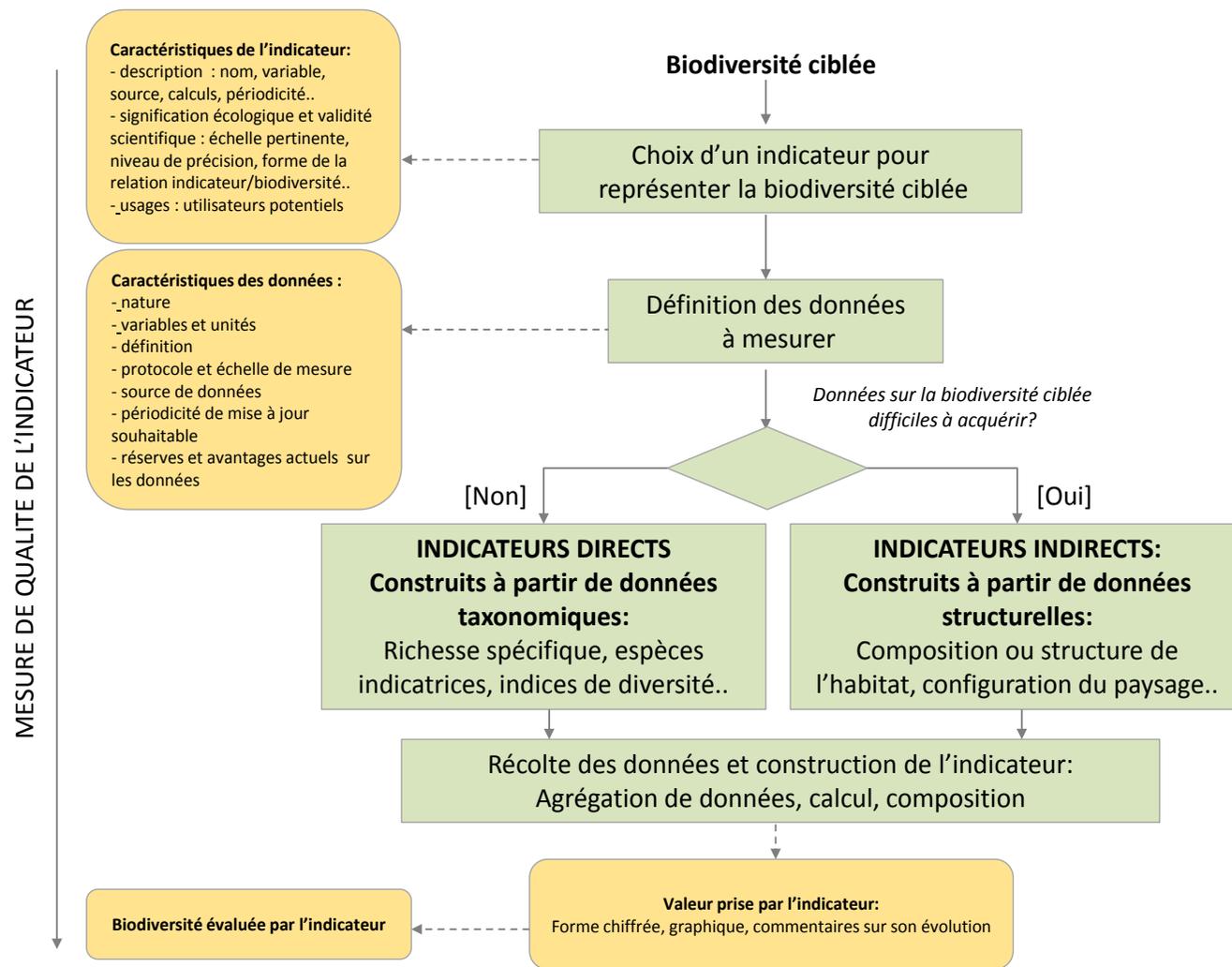


Figure 2-3. Comment mesurer la biodiversité ? (Source : GIP ECOFOR)

2.2.3 LA BIODIVERSITE FORESTIERE DANS CETTE THESE

La biodiversité étudiée dans ce travail de thèse s'intéresse à la biodiversité faunistique abritée par la forêt paysanne d'un paysage rural et non à la biodiversité en espèces ligneuses de ces forêts.

Dans cette logique, nous nous sommes concentrés plus spécifiquement sur une diversité taxonomique faunistique (un insecte et en particulier *le syrphe*, voir section 3.2) dont la présence dans ces paysages constitue un enjeu majeur pour la durabilité des écosystèmes. Cette diversité taxonomique est exprimée à l'aide d'un indicateur, la **richesse spécifique**, qui correspond au nombre d'espèces différentes dans un biotope ou une station donnée. Les diversités génétiques et fonctionnelles ne seront donc pas abordées dans ce travail.

2.3. LES FACTEURS GENERAUX AFFECTANT LA BIODIVERSITE FORESTIERE

Trois types de facteurs affectent potentiellement la biodiversité forestière, chacun agissant à une échelle spatiale différente : les facteurs stationnels, les facteurs morphologiques et les facteurs d'isolement (Figure 2.4).

Les facteurs stationnels décrivent le milieu interne d'une forêt. Ils sont mesurés à une échelle très fine et traduisent la capacité d'accueil en espèces d'un peuplement forestier (Deconchat et Balent, 1996). Par exemple, certains types de peuplements comme les feuillus sont favorables à la présence d'une faune du sol riche et diversifiée alors que des peuplements de résineux seront plutôt susceptibles de favoriser la présence de fourmis mais nuiront à l'installation des communautés d'arthropodes ou de macro-faune. Les conditions d'accueil fournies par ces différents peuplements varient en fonctions des propriétés biologiques, physiques et chimiques portées par ces différentes essences. De fait, plus la structure interne du peuplement est diversifiée, plus la disponibilité en habitats différents sera grande et favorisera par conséquent la diversité en espèces. Typiquement, la présence de bois morts ou de très vieux arbres permet d'accroître la diversité en espèces en fournissant des conditions adéquates aux espèces saproxyliques par exemple (Bouget et al. 2013; Larrieu et al. 2014). Les stades jeunes sont, eux, majoritairement propices à l'installation de plantes ou d'oiseaux et apporteront des ressources différentes.

Les facteurs morphologiques concernent à la fois la superficie de la forêt, celle de son cœur mais également sa longueur de lisière. On retrouve par conséquent les propriétés générales qui influencent la biodiversité déjà évoqué en chapitre 1, notamment la surface de l'habitat (cf relation aire/espèce, section 1.4.2.1). Généralement, les lisières abritent une richesse importante (Laurance et Yensen 1991; Honnay et al. 1999). (Laurance et Yensen 1991) Au Royaume Uni, 50% des papillons forestiers et 33% des oiseaux sont localisés dans ce type de milieu (Fuller et Warren 1991). Néanmoins, ces espèces souvent très compétitives réduisent l'arrivée de migrants et disposent d'une grande capacité de dispersion, elles contribuent donc assez peu à la richesse des forêts. Les milieux intérieurs des forêts des Coteaux de Gascogne sont largement occupés par des espèces rares et fragiles d'une grande valeur patrimoniale (Deconchat et Balent 1996). Les petits bois de forme complexe sans milieu intérieur ne sont pas pour autant sans intérêt écologique (Fischer et Lindenmayer 2002). Ils peuvent abriter des espèces « reliques » (principalement végétales) qui s'y maintiennent malgré des conditions peu favorables après la disparition de gros bois. A une échelle plus large, ces petits bois peuvent aussi faciliter la circulation de certaines espèces entre des massifs de plus grande taille.

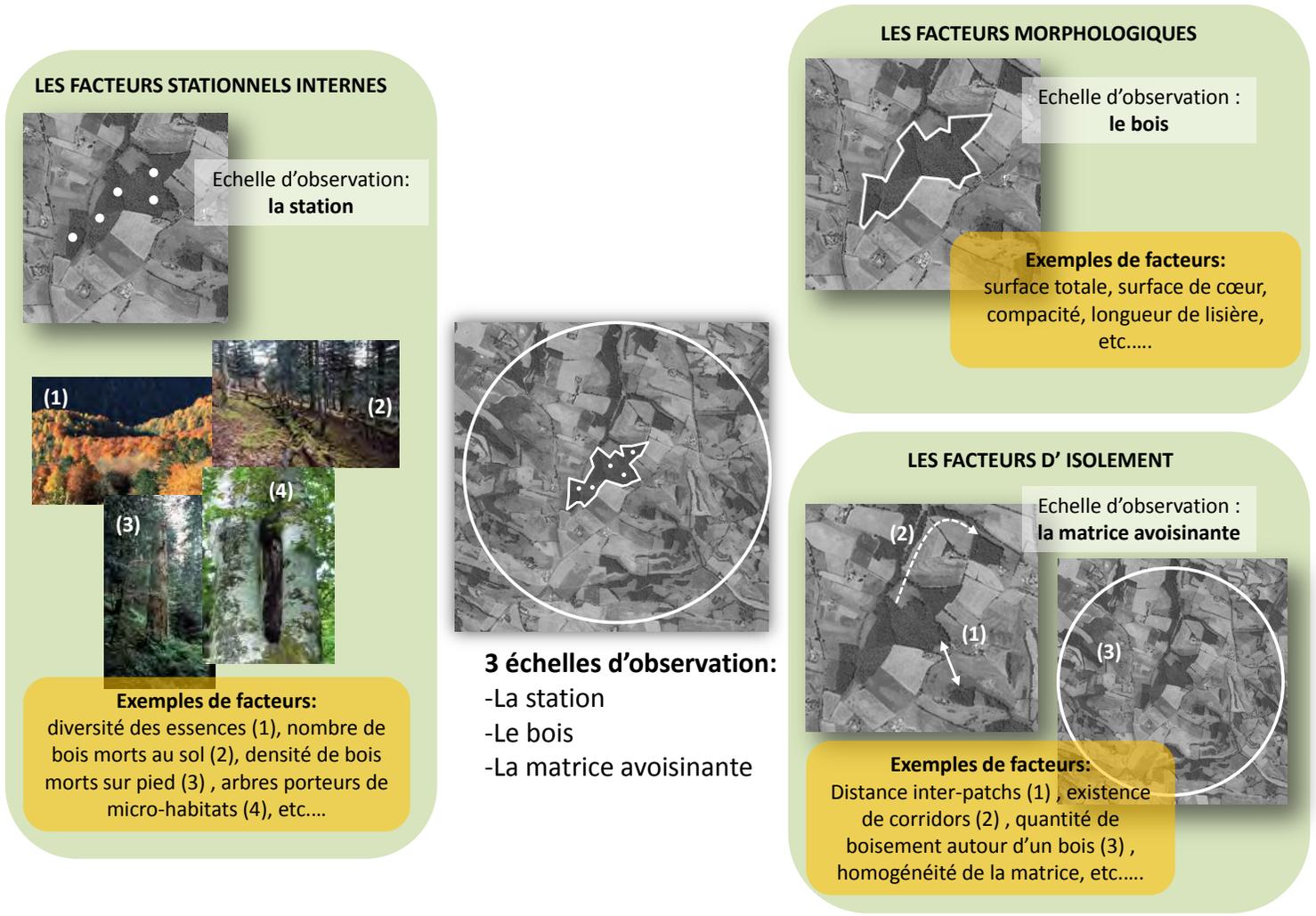


Figure 2-4. Les facteurs affectant la biodiversité forestière (clichés : Larrieu et Gonin, 2008)

Les facteurs d'isolement font référence à la connectivité des habitats forestiers (cf. 1.4.2.2). La faible distance entre les bois, l'existence de corridors (haies, ripisylves), l'absence de barrières (routes) ou l'homogénéité de la matrice paysagère sont autant de facteurs réduisant l'isolement d'un bois (Fahrig et Merriam 1985; Taylor et al. 1993). Les espèces exploitent plus ou moins bien ces éléments en fonction de leurs caractéristiques propres de dispersion et leur situation. Une distance de 100 m entre deux habitats constitue déjà un frein pour certaines espèces et celles qui dépassent ce stade sont considérées comme ayant une bonne capacité de dispersion (Saunders et al. 1991). Les caractéristiques des corridors peuvent aussi déterminer les espèces qui peuvent les emprunter : les espèces forestières sont par exemple plutôt favorisées par les patchs de grande largeur et de faible longueur (Sharitz et al. 1992). Inversement, ils peuvent jouer un rôle négatif en constituant une barrière dans le paysage pour d'autres espèces : les haies sont des voies de circulation pour les espèces forestières mais ne sont pas propices aux espèces des milieux ouverts (Opdam 1990). Une bonne connectivité favorise généralement l'arrivée continue de colonisateurs ou de propagules appartenant à de nombreuses espèces qui accroissent la valeur de la richesse du bois. Les grands bois sont aussi réputés avoir un impact sur les petits bois se trouvant à proximité favorisant les flux de dissémination.

2.4. L'ANCIENNETE DES FORETS : UN FACTEUR CLE POUR COMPRENDRE LEUR BIODIVERSITE

2.4.1 MATURITE VERSUS ANCIENNETE

Les travaux fondateurs de Peterken et son équipe au début des années 70 ont mis en lumière l'importance de prendre en compte l'histoire des écosystèmes forestiers dans l'analyse des relations qu'ils entretiennent avec la biodiversité qu'ils abritent (Peterken 1977; Peterken et Game 1984). Les auteurs font notamment remarquer que les concepts de forêt primaire ou secondaire de plus en plus utilisés dans la littérature scientifique ne disposent pas de définitions claires et précises. Ils en concluent qu'il est nécessaire de redéfinir ces concepts ou d'en proposer d'autres dans le but de faciliter les travaux concernant l'écologie historique.

Le concept d'ancienneté de l'état boisé a été proposé par Peterken et al. au début des années 1970 (Peterken, 1977). **Les forêts anciennes sont définies comme les surfaces forestières présentes depuis un ou plusieurs siècles sans discontinuité temporelle.** Cette continuité doit être déterminée à partir de données spatiales historiques qui constituent souvent le point de référence pour ce type d'analyse. Ce point de référence varie selon les pays et la disponibilité des sources historiques. La Grande-Bretagne est le pays qui dispose des cartes les plus anciennes (~1600) tandis que les Pays-Bas peuvent s'appuyer uniquement sur des documents plus récents produits au milieu du XIX^e siècle. En France, la carte exploitée la plus ancienne et disponible à large échelle est la carte de CASSINI datant du milieu du XIX^e siècle. Cette source a permis de caractériser l'état du minimum forestier en France et a fait l'objet d'un travail de cartographie des forêts anciennes à l'échelle nationale (Vallauri et al. 2012). **Par ailleurs, les déterminants de l'ancienneté des forêts sont uniquement liés au caractère continu de leur présence. Cela ne traduit en rien la vieillesse des peuplements et ne fait pas référence aux types de gestion subis par un bois, à ses stades sylviculturaux ou aux caractères de naturalité internes.**

Le concept de maturité des forêts fait lui référence aux critères de naturalité des forêts et fait abstraction de l'origine de la forêt et de son ancienneté (Peterken 1977). La présence de micro-habitats, l'âge des arbres, la présence de bois mort ou de très gros arbres sont tous des indicateurs de maturité des forêts. Le type de gestion et plus particulièrement l'intensité des cycles sylvicoles appliqués déterminent le stade de maturité en laissant se développer ou non les phases sénescentes. Une forêt peut donc être considérée comme ancienne car présente sans discontinuité depuis un siècle et demi ou plus mais pas nécessairement mature du fait que la gestion peut bloquer le

développement des stades de maturation. En revanche, les forêts matures sont nécessairement anciennes, on parle alors de vieilles forêts.

Plusieurs exemples en Europe permettent de comprendre un peu mieux les divergences qui existent entre maturité et ancienneté. La forêt de Balowieza en Pologne a été très largement étudiée en Europe en raison d'une présence continue depuis plusieurs siècles (Figure 2-5(A)). Il s'agit d'un vaste espace forestier situé à la frontière entre la Pologne et le Bélarus. En raison de plusieurs décennies de protection (depuis 1821), cette forêt a perduré en conservant son état naturel jusqu'à aujourd'hui. Elle abrite des espèces rares notamment le bison d'Europe alors que celui-ci a majoritairement disparu des forêts européennes (Vallauri, 2002). Elle se caractérise également par un volume important de bois mort (100 m³/ha) dont la présence est cruciale pour un grand nombre d'espèces (UNESCO World Heritage). La réserve de Zofinsky en République Tchèque est aussi une forêt ancienne dont les traces de maturité sont importantes en raison de son statut favorable de protection (Figure 2-5(B)). L'ancienneté de l'état boisé alliée à la maturité des peuplements a permis le développement de conditions pédologiques favorables à la richesse des espèces et plus spécifiquement à la distribution de la matière organique (Vallauri, 2002). La forêt de Tronçais en France et la forêt de Soignes en Belgique sont en revanche des exemples de forêts anciennes mais non matures (Figure 2-5(C-D)). La gestion intensive au cours du temps a bloqué le développement des phases de sénescence ainsi que les différents stades de maturation des arbres, ce qui explique qu'aujourd'hui, la présence de bois mort ou de très gros arbres est très faible voire nulle.



Figure 2-5. (A) *La forêt de Bialowieza (Pologne)* ; (B) *La forêt de Zofinsky (République Tchèque)* ; (C) *La forêt Tronçais (France)*; (D) *La Forêt de Soignes (Belgique)* (clichés : http://www.gip-ecofor.org/doc/drupal/liens_article/evenements/2011/ForetsAnciennes_VALLAURI_Dec2011.pdf)

2.4.2 LA SPECIFICITE DES FORETS ANCIENNES

Les travaux fondateurs de Peterken et al. en Grande-Bretagne au milieu des années 1970 ont mis en lumière les différences des communautés de plantes présentes au sein des forêts anciennes et des forêts récentes. Des résultats similaires ont ensuite été trouvés en Amérique du Nord à la fin des années 1980 et plus tard en Europe sous l'impulsion des recherches menées par Martin Hermy dans les forêts situées en Belgique (Foster 1988; Foster 1992; Honnay et al. 1998; Hermy et al. 1999). Ce papier fondateur publié par Hermy et son équipe (Hermy et al, 1999) permettra d'introduire le concept de forêts anciennes à travers toute l'Europe continentale. En France, l'INRA de Nancy et l'équipe de Jean-Luc Dupouey arriveront à des résultats similaires dans les forêts vosgiennes en démontrant les effets irréversibles des pratiques culturelles anciennes sur les communautés de plantes (Dupouey et al. 2002; Dambrine et al. 2007). Les travaux de Guillaume Decocq concernant l'impact des mottes castrales sur la biodiversité actuelle en Picardie contribueront également à alimenter les connaissances à propos de l'impact des activités anthropiques passées sur la présence actuelle de biodiversité (Decocq et al. 2002).

Les forêts anciennes se distinguent des autres forêts par deux caractéristiques majeures : leur pédologie et la biodiversité qu'elles abritent.

Le défrichement, l'utilisation des sols ou l'agriculture de manière générale sont susceptibles de modifier les équilibres chimiques et microbiens de sols. Selon Cateau et al. (2015), ces usages ont des impacts à plus ou moins long terme. A court terme, le labour peut par exemple provoquer des pertes de 50 à 75 % de la biomasse microbienne du sol, réduire la quantité d'azote et augmenter la nitrification nette (Compton et Boone 2000). L'occupation et les pratiques du sol peuvent également causer la destruction des horizons de surface du sol forestier sur long terme (Bellemare et al. 2002; Sciana et al. 2009). Il a notamment été montré que certains impacts sur le sol liés à l'activité agricole passée étaient visibles plus de 2000 ans après l'usage forestier (Dambrine et al. 2007).

La rupture de l'état boisé pour un usage agricole provoque des effets délétères sur la faune, la flore, la fonge forestière et les banques de graines au sol. Ces effets sont d'autant plus importants lorsque le sol a été labouré et la durée d'exploitation importante (Cateau et al. 2015). En effet, le retour à l'état forestier après une exploitation agricole du sol ne permet pas toujours la recolonisation du milieu par les espèces spécialistes, notamment quand leurs capacités de dispersion sont limitées. Il existe par exemple des différences significatives dans la composition spécifique des herbacées et des arthropodes entre les forêts anciennes et récentes (Gossner et al. 2013). Par ailleurs, les espèces davantage présentes en forêts anciennes qu'en forêts récentes sont nombreuses et servent généralement d'indicateurs d'ancienneté de l'état boisé. Les travaux de Martin Hermy dans les forêts de feuillus en Belgique ont notamment permis d'identifier 24 espèces de plantes uniquement inféodées aux forêts anciennes (Hermy et al. 1999). Certaines de ces espèces ont toutefois été trouvées uniquement à l'intérieur de boisements récents dans les études conduites par Foster et al. (Foster 1988; Foster 1992). Cette contradiction montre que la capacité de colonisation des espèces n'est pas liée uniquement aux caractéristiques propres des espèces mais également à d'autres facteurs.

Les raisons pour lesquelles certaines espèces ne parviennent pas à coloniser les boisements récents ont été précisées dans un contexte agricole (Flinn et Vellend 2005) (Figure 2-6). Les coupes forestières et l'abandon de l'agriculture tendent à créer de nouveaux paysages limitant la dispersion de graines. L'agriculture peut aussi limiter le recrutement via l'introduction de nouvelles relations biotiques et abiotiques. Par exemple, l'introduction d'espèces exotiques constitue un obstacle pour les espèces à plus faible statut de compétition.

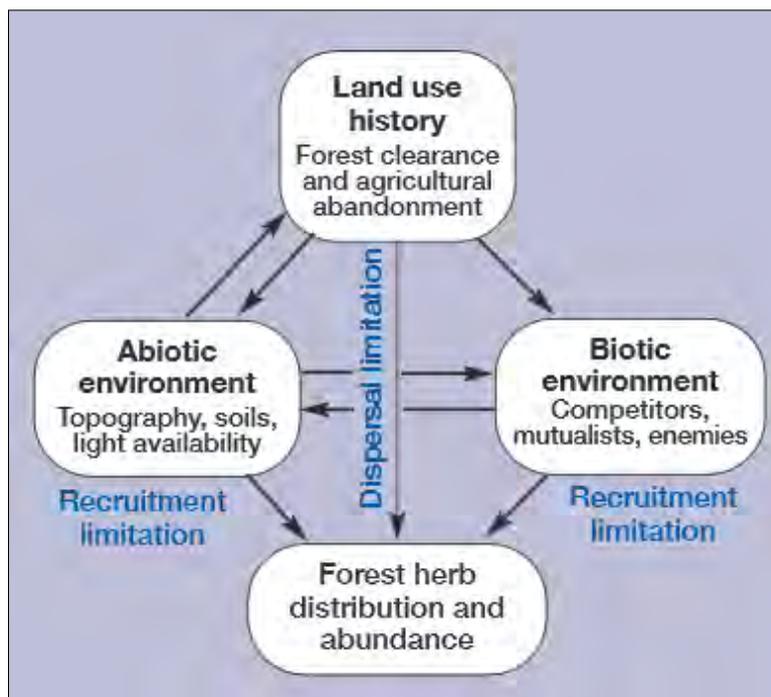


Figure 2-6. Les conséquences de l'usage passé sur la biodiversité (source : Flinn et Velend, 2005)

(Hermý et Verheyen 2007) parlent alors d'une course d'obstacles quand il s'agit d'évoquer la recolonisation d'espaces nouvellement boisées par les espèces caractéristiques des forêts anciennes. Dans le cas des plantes, il est nécessaire que les graines se dispersent, qu'elles s'intègrent à la nouvelle banque de graines, que ces graines germent, qu'elles deviennent des plantules pour arriver au stade juvénile avant d'atteindre le stade d'adulte où une nouvelle dispersion sera enfin possible. Toutes ces étapes sont marquées par des processus de mortalité, de prédation et de compétition entre espèces (Hermý et Verheyen 2007).

Plusieurs expérimentations ont été mises en place pour mesurer la contribution relative du recrutement et de la dispersion sur la colonisation. Pour le recrutement, les travaux ont consisté à implanter des espèces de forêts anciennes au sein de zones nouvelles boisées et d'en mesurer leur performance en termes de biomasse ou de fertilité (Donohue et al. 2000; Jamoneau et al. 2011). Les conclusions montrent que ces espèces sont au moins aussi performantes dans les forêts récentes : le recrutement ne semble donc pas être le problème majeur. Concernant la dispersion, les recherches ont portées sur l'analyse des lisières pouvant marquer une zone de transition entre une partie de forêt ancienne et une partie de forêt récente (Matlack 1994; Bossuyt et al. 1999). Les résultats de ces études ont montré que les espèces des fragments anciens partageant une frontière commune avec une zone nouvellement boisée pouvaient coloniser cet espace à la vitesse de 20 à 100 mètres par siècle. Pour les fragments isolés, la recolonisation est aléatoire voire nulle dès lors que la distance qui les sépare d'un autre habitat excède 200 mètres. Ceci s'explique en partie par le mode de dispersion des graines. La dispersion dans le temps semble être également un facteur déterminant. L'immense majorité des forêts anciennes n'a pas de banque de graines permanente, ce qui les rend très sensibles aux événements de mise en culture, même bref avant un retour à la forêt. On dénombre cependant assez peu de travaux dans ce domaine en raison de la lourdeur du processus d'échantillonnage nécessaire.

Bien que la majeure partie des travaux s'intéressant aux effets de l'ancienneté du paysage concerne des espèces de plantes, d'autres travaux, souvent plus récents, montrent qu'elles ne sont pas les seules impactées. Les oiseaux ont aussi montré qu'ils pouvaient être particulièrement

sensibles à l'ancienneté dans une étude conduite sur les populations d'oiseaux des forêts tropicales kényanes (Brooks et al, 1999). L'ancienneté des habitats tels que les mousses ou les lichens provoque des effets significatifs sur les communautés de micro-arthropodes. Enfin, les insectes peuvent aussi être concernés par ce type d'effets. (Bommarco et al. 2014) ont notamment démontré un impact significatif de l'ancienneté sur des populations d'abeille et de syrphes dans une zone prairiale du sud-est de la Suède.

2.4.3 LA SPECIFICITE DES VIEILLES FORETS

Les vieilles forêts peuvent être de deux sortes :

(1) les forêts non exploitées depuis plusieurs décennies voire plusieurs siècles qui évoluent de manière naturelle vers les stades ultimes du cycle sylvigénétique.

(2) les forêts exploitées selon les modes de gestion permettant de conserver des caractères de vieilles forêts

Ces forêts font l'objet d'une forte attention aujourd'hui (Vallauri, 2002). Vu leur stade de maturité, elles se rapprochent des dynamiques naturelles et permettent d'en apprendre davantage sur les phases de fin de cycle sylvigénétique. Les enjeux liés aux vieilles forêts peuvent être classés en trois sous-parties : la biodiversité, le stockage de carbone et la naturalité (Figure 2-7).

(1) La biodiversité

La biodiversité des forêts matures est riche en raison de la présence d'un grand nombre d'habitats aux substrats variés. Les vieux arbres présentent des « défauts » qui constituent des micro-habitats (Larrieu et al. 2012; Larrieu et al. 2014). Les cavités par exemple, causés par des pics ou des cassures de branches, sont des abris privilégiés pour de nombreuses espèces (ONF, 2010). Les cavités situées en hauteurs et causées par des pics profitent aux cavicoles⁷ secondaires (chouettes) ou à certains mammifères comme le loir et aux insectes. Les cavités causées par des branches, elles, sont des habitats propices à la présence de chauves-souris forestières et la présence de terreau dans ces cavités est favorable à certains insectes comme le Pique Prune (Savoie et Brustel, 2011) Enfin, comme déjà évoqué précédemment, les coléoptères saproxyliques se nourrissent du bois mort et contribuent à les faire évoluer.

⁷ espèces nichant majoritairement dans les cavités

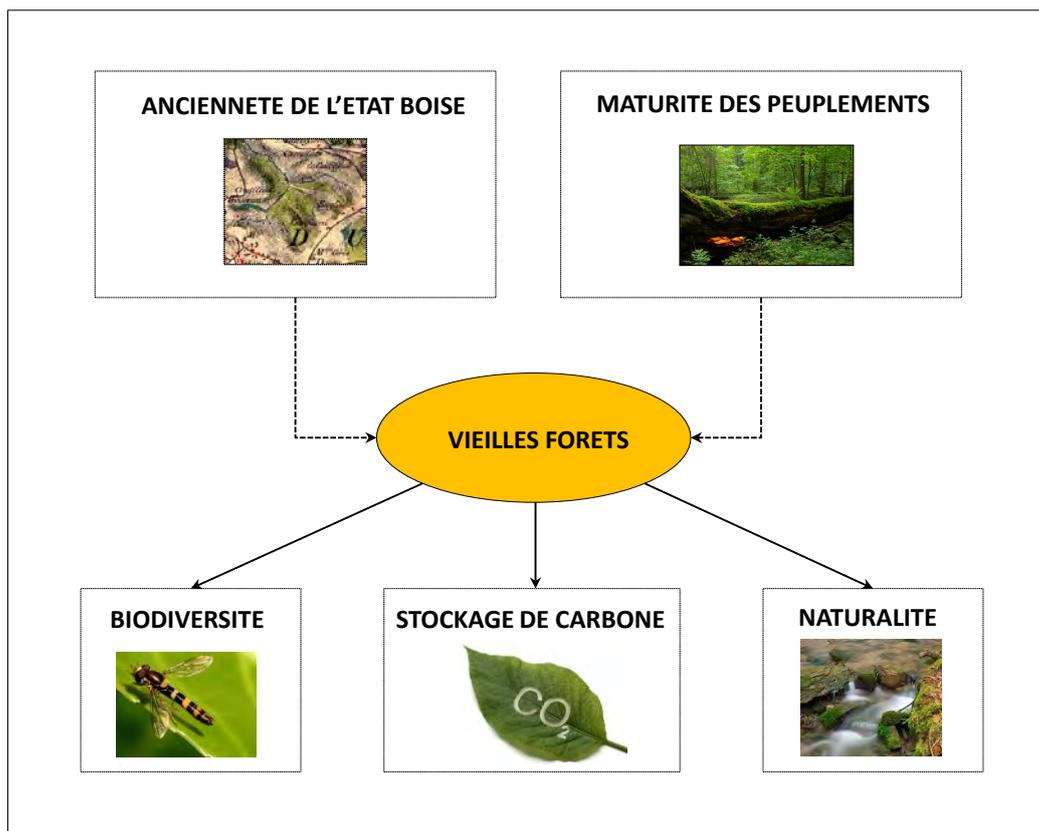


Figure 2-7. Les enjeux liés aux vieilles forêts

(2) Le stockage de carbone

Des études récentes montrent que les forêts matures continuent de stocker du carbone au fur et à mesure que leurs stades continuent de se développer. Luysaert et al. ont notamment démontré que les vieilles forêts stockent du carbone dans trois compartiments principaux (Luysaert et al. 2008) :

- Dans la biomasse aérienne ($0,4 \pm 0,1$ tC/ha/an)
- Dans les débris de bois ($0,7 \pm 0,2$ tC/ha/an)
- Dans le sol : systèmes racinaires et matières organiques ($1,3 \pm 0,8$ tC/ha/an)

Ainsi, il paraît évident que les pratiques d'exploitation intensives affectent la séquestration du carbone. Une forêt gérée de manière non intensive et combinant différentes classes d'âge stockera plus de carbone qu'une forêt gérée de façon intensive (Nunery et Keeton 2010).

(3) Une naturalité intéressante

Le degré de naturalité traduit l'impact de l'homme sur le milieu (GILG 2004; Paillet et Bergès 2010). Un fort degré de naturalité indique généralement une biodiversité riche bien que cela peut être discuté. Les vieilles forêts par définition sont peu perturbées par l'homme. On s'attend donc à trouver un fort degré de naturalité dans ces milieux.

Aujourd'hui, la volonté de préserver les vieilles forêts devenues très rares se heurtent à un contexte de pression sur la ressource en bois. Le manque de politiques communes concernant leur préservation explique en partie ces difficultés. Comment préserver la forêt lorsque seuls les usages économiques sont source de rémunération ? Les actions menées par le GIP ECOFOR tentent de

répondre à ces questions en réunissant différents acteurs du monde forestier. **Le colloque « Naturalité des eaux et des forêts » (17-20 septembre 2013)** via différentes présentations a en partie souligné l'importance de préserver ces vieilles forêts et les enjeux sous-jacents induits par ces milieux.

Synthèse de Chapitre :

La forêt paysanne apparaît comme un réservoir majeur de biodiversité au sein de l'espace rural. Bien qu'elles soient en déclin à l'échelle nationale et que sa richesse ait été reconnue plus tard, le morcellement des paysages qui touche nos régions en fait un environnement particulier, propice à de nombreuses espèces. La biodiversité des forêts françaises est par conséquent riche mais menacée (Vallauri, 2002). La plupart des groupes d'espèces sont fortement impactés et le manque de connaissances sur des groupes entiers d'espèces montrent que ces estimations sont probablement sous-évaluées. Différents indicateurs sont aujourd'hui employés pour mesurer la biodiversité. Les indicateurs directs s'appuient sur des inventaires taxonomiques pour rendre compte de l'état de la biodiversité à un endroit donné tandis que les indicateurs indirects s'appuient sur la composition et la configuration spatiale des paysages pour prédire cette biodiversité. Ces derniers offrent l'avantage de s'affranchir de contraintes d'échantillonnages lourdes et peu reproductibles. Ils impliquent aussi de connaître à priori les déterminants de la biodiversité ciblée afin de s'appuyer sur les bons critères pour la prédire efficacement. Trois grands facteurs affectent potentiellement la biodiversité forestière : (1) les facteurs stationnels qui décrivent les caractéristiques du milieu interne de la forêt (cf niche écologique, chapitre 1) ; (2) les facteurs morphologiques qui traduisent les effets de la surface des forêts ou de la longueur de leur lisière (cf relation aire/espèce, chapitre 1) ; (3) les facteurs d'isolement qui font référence à la connectivité des habitats forestiers (cf. théorie de la biogéographie insulaire, chapitre 1). Enfin, l'ancienneté des paysages est apparue ces trente dernières années comme un autre déterminant majeur de la variation de biodiversité au sein des espaces forestiers. La majeure partie des études dans ce domaine concerne la flore mais de récents travaux portant sur d'autres taxons (i.e. les insectes ou les oiseaux) montrent que l'ancienneté pourrait potentiellement jouer un rôle dans la présence de plusieurs groupes d'espèces.

CHAPITRE 3. SITE D'ETUDE ET DONNEES BIOLOGIQUES SELECTIONNEES

SOMMAIRE

3.1. LES FORETS DES COTEAUX DE GASCOGNE

3.2. LES SYRPHES : UN BIO-INDICATEUR DES PAYSAGES RURAUX

Ce chapitre présente le site d'étude dans lequel les bases de données taxonomiques ont été retenues pour notre étude ainsi que les grandes caractéristiques de ses paysages. Nous expliquerons ensuite pourquoi nous nous sommes focalisés sur les syrphes et quelles sont les grandes caractéristiques de ces espèces.

3.1. LES FORETS DES COTEAUX DE GASCOGNE

Les bases de données taxonomiques retenues dans cette thèse ont été constituées sur le site LTER (Long Term Ecological Research) « Vallées et Coteaux de Gascogne ».

3.1.1 LE SITE D'ETUDE

Toutes les expérimentations réalisées dans cette thèse ont été menées sur un territoire qui se situe dans le Sud-Ouest de la France, à environ 80 km de Toulouse dans les coteaux du Bas-Comminges en Haute-Garonne, canton d'Aurignac (43°16'28"N, 0°51'11" E, WGS-84) (Figure 3-1). Ce terrain fait l'objet de nombreuses observations *in situ* depuis le début des années 1980 : relevés faunistiques (oiseaux, insectes) et floristiques, suivi d'exploitations agricoles, suivi des pratiques de gestion des forêts, étude de l'évolution paysagère et de la biodiversité associée. Il constitue un des principaux terrains pilotes de l'UMR DYNAFOR, depuis la création de cette dernière. La profondeur temporelle des recherches qui sont développées et leur nature interdisciplinaire ont valu à ce site d'être intégré depuis 2007 dans le réseau international LTER (*Long Term Ecological Research*¹) (Deconchat et al. 2007). Ce site, « Vallées et Coteaux de Gascogne », couvre environ 30 000 hectares dans sa totalité. Les travaux engagés depuis l'origine permettent aujourd'hui de comparer de nombreuses données écologiques, agronomiques et sociologiques collectées à différentes périodes.

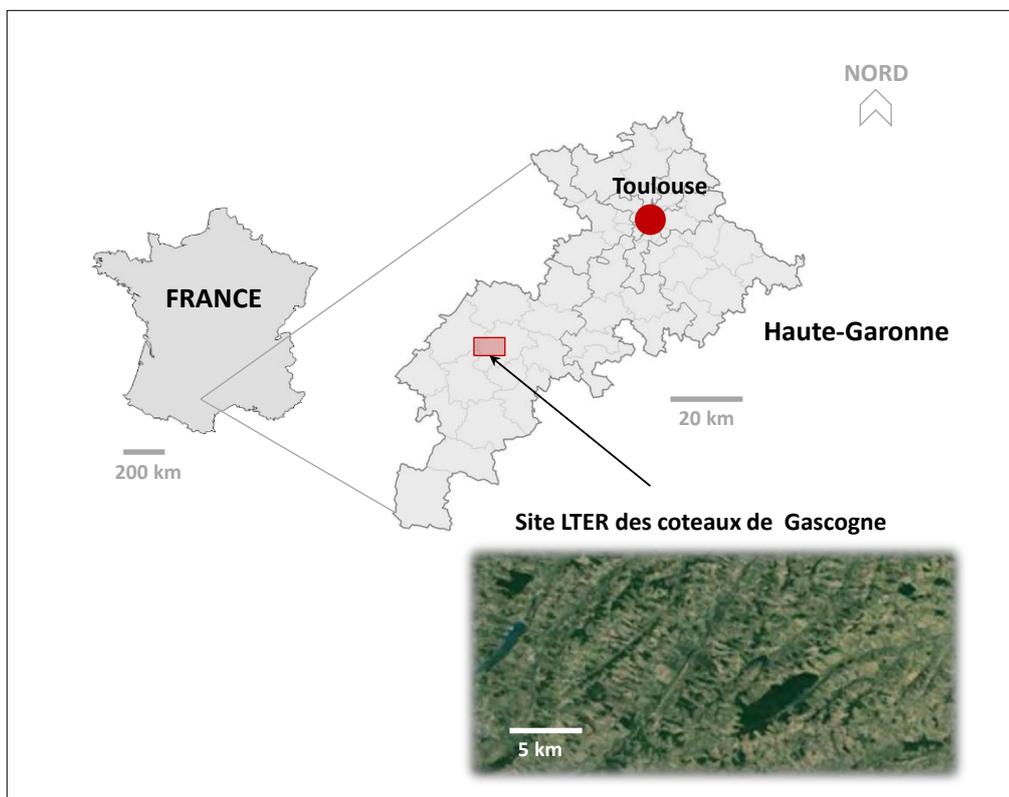


Figure 3-1 Carte de localisation du site d'étude

¹ <http://www.lter-europe.net/networks/france>

3.1.2 LE CLIMAT

Il est de type tempéré, caractéristique du bassin du Sud-Ouest, subissant les influences de la chaîne des Pyrénées, de la façade océanique ainsi que de la méditerranée (Joly et al. 2010). Il est caractérisé par des hivers plutôt doux et humides et des étés chauds et secs assez instables (perturbations orageuses). La température moyenne annuelle est élevée (entre 11 et 13°) avec une amplitude thermique annuelle entre l'hiver et l'été de 15 à 16°. Les précipitations en cumul annuel sont peu abondantes (moins de 800 mm).

3.1.3 LE PAYSAGE

Le paysage est marqué par la présence de nombreuses forêts paysannes de petites tailles, principalement privées, et gérées par les agriculteurs. Il intègre un vaste réseau hydrographique (dont la vallée de la Nère) responsable de la modulation du relief sous la forme d'une succession de vallons plus ou moins encaissés. Le relief est caractérisé par un ensemble collinaire dont les altitudes varient approximativement entre 250 et 400 m d'altitude (début du piémont Pyrénéen) (Figure 3-2).

Ce paysage est principalement composé de cultures en fond de vallées, de prairies, de haies et de bois (avec deux massifs forestiers importants, ceux de Lilhac-Fabas et Mauboussin) qui traduit une activité agricole de type polyculture-élevage. Jusqu'à la période de modernisation agricole, les activités dans cette région étaient très diversifiées et tournées vers l'autosubsistance : céréales, bovins, ovins, volailles, vignes, bois. Entre 1970 et 2000, le nombre d'exploitations du canton d'Aurignac a été divisé par 2,3. Ceci a conduit à un agrandissement de la taille moyenne des parcelles et une baisse de la quantité de prairies. La surface des éléments boisés, autrefois utilisés pour le bois de chauffage et le bois d'œuvre, est restée relativement stable (Bonthoux 2011).

Dans cette région, les petites forêts privées sont majoritaires et l'absence de plan de gestion pour plus de 90% de la surface forestière a conduit à une grande hétérogénéité spatiale entre les ces forêts. La couverture forestière est fragmentée et couvre 15 % de la surface totale. Les espèces d'arbres dominantes sont le chêne sessile, le chêne pédonculé et le chêne pubescent qui représentent à elles seules 80% des espèces (rapport IFN 2000). Les propriétaires forestiers ont généralement une connaissance lacunaire de la gestion forestière, ce qui peut fortement influencer les patrons de distribution de la biodiversité forestière (De Warnaffe et al. 2006; Andrieu 2013).



Figure 3-2 Les paysages des coteaux de Gascogne

3.2. LES SYRPHERS : UN BIO-INDICATEUR DES PAYSAGES RURAUX

Compte tenu du nombre de travaux réalisés par le laboratoire DYNAFOR sur les syrphes, un insecte de la famille des mouches (Diptera : Syrphidae), le choix de travailler sur ce taxon s'est imposé de manière évidente (Quin et al. 2006; Raymond 2013). Les syrphes constituent un taxon

privilegié en raison d'un grand nombre d'espèces présentes en France. On peut en avoir une idée en consultant la base de données nationale *Syrphid*² qui recense toutes les espèces observées en France métropolitaine d'après la littérature scientifique et qui cartographie celles-ci par département (Sarhou et al. 2010). Il existe également une base de données européenne, *Syrph the Net* (StN) qui renseigne les traits de vie de la majorité des espèces c'est-à-dire, leur biologie, leurs habitats, leur mode de vie, leur répartition, etc. (Speight 2010).

Comme les abeilles, les syrphes sont aujourd'hui considérés comme un bio-indicateur (Speight et Wainhouse 1989). Leur présence renseigne sur les caractéristiques ou l'état de conservation d'un écosystème. Ils occupent la plupart des milieux terrestres et sont facilement détectables. Ces insectes sont connus pour assurer un contrôle biologique des ravageurs des cultures : certaines espèces, au stade larvaire, se nourrissent des pucerons (Figure 3-3). Ils sont dits aphidiphages et de ce fait, sont très utiles comme moyen de lutte biologique (Chambers et Adams 1986; Tenhumberg et Poehling 1995). Ce sont des auxiliaires de culture. D'autres espèces sont spécifiques au milieu forestier. Certaines d'entre elles, au stade larvaire, ne se nourrissent que de matière en décomposition. Ils sont dits saprophages. Les saproxylophages quant à eux se nourrissent du bois mort. Ce groupe peut donc potentiellement être un indicateur des forêts anciennes à haut degré de naturalité, le développement des larves nécessitant de vieux arbres à cavités (Good et Speight 1996). A l'âge adulte, les insectes sont floricoles et jouent un rôle essentiel dans la pollinisation des cultures mais aussi des arbres et arbustes (Groot et Beyk, 2012).

Les syrphes sont également sensibles au paysage qui les entoure (Raymond et al., 2014). En effet, une partie d'entre eux, notamment les prédateurs des pucerons, doivent quitter les parcelles cultivées au cours de l'année pour survivre (Arrignon et al. 2007). Les éléments semi-naturels comme les haies, les lisières et les îlots boisés au sein de la matrice agricole jouent alors un rôle fondamental. Ils constituent autant de sites refuges pour se nourrir, se reproduire et passer l'hiver.

Malgré les connaissances déjà acquises sur ce taxon, les facteurs influençant les patrons de distribution de syrphes dans les paysages sont encore relativement mal connus (Keil et Konvicka 2005). En particulier, très peu d'études ont étudié l'influence de l'ancienneté des habitats (forestiers ou autre) sur la présence et l'abondance actuelle des syrphes généralistes ou spécialistes. Seule une étude très récente montre que les syrphes peuvent être sensibles aux perturbations de leur habitat et aux changements paysagers (Bommarco et al. 2014). D'une manière générale, les études en écologie historique mesurant l'impact du passé pour éclairer la biodiversité présente (pour reprendre le titre de l'article de (Girel 2006) n'en est qu'à ses débuts, en particulier pour les espèces faunistiques. Nous y reviendrons en partie 4. Ainsi, bien que notre travail s'inscrive dans le champ de la géomatique, il contribue aussi à alimenter les connaissances en écologie.



Figure 3-3 A gauche, un syrphé au stade adulte, à droite, un syrphé au stade larvaire (source : <http://www.naturablog.com>)

² <http://syrfid.ensat.fr/>

Un protocole d'échantillonnage conçu pour étudier l'effet du paysage

Dans ce travail, nous nous sommes focalisés sur les syrphes spécialistes des milieux forestiers. Nous avons bénéficié de relevés de syrphes déjà existants réalisés dans le cadre d'une étude précédente par des membres de l'UMR DYNAFOR (Ouin et al. 2006). Cette étude avait pour objectif de démontrer la relation aire/espèce entre des patchs forestiers et différents groupes d'espèces de syrphes.

Dans cette étude, l'échantillonnage des forêts s'est déroulé pendant deux mois, du 10 Mai au 10 Juin 2000 et du 13 Septembre au 13 octobre 2000. Les insectes ont été capturés à l'aide de tentes malaises et récoltés via des bouteilles à échantillonnages (Figure 3-4). Les bouteilles ont été remplacées chaque nuit au printemps et une fois le mois pour la période d'automne. Un total de 54 bois a été échantillonné pour cette expérimentation (Ouin et al 2006). Nous verrons en partie 4 que seuls 48 bois été retenus dans notre travail. Les changements forestiers survenus depuis 2000 expliquent en grande partie cette restriction.



Figure 3-4 Exemples de piégeages de syrphes : l'utilisation d'une tente malaise et d'une bouteille à échantillonnage

Les bois ont été sélectionnés selon un gradient de surface, d'isolement et de type de gestion. Ces bois diffèrent aussi en termes de structure et de composition. Le nombre de pièges par bois a été ajusté à la surface du patch dans le but d'obtenir une intensité de piégeage équivalente par bois (Tableau 3-1). Une large majorité de bois de petite taille (<5ha) contient seulement un seul piège. Pour les bois plus larges, davantage de pièges ont été installés. Un total de 101 tentes malaises a été installé dans les bois échantillonnés. Les espèces ont été classées selon la base de données *Syrph the Net* qui permet de différencier les groupes écologiques en se basant sur les habitats préférentiels de leur stade larvaire (Speight et al. 2000). Dans notre cas, nous nous sommes focalisés sur les espèces forestières pour étudier leur sensibilité aux dynamiques de leur habitat sur le temps long.

CLASSES DE SURFACE	NB DE PIÈGES	NB DE FORETS DANS LA CLASSE	MOYENNE DE RICHESSE EN ESPECES
S < 5ha	1	29	2.48 ± 1.68
5 < S < 10ha	2	9	3.77 ± 1.74
10 < S < 15ha	3	3	2.45 ± 1.37
15 < S < 25ha	4	4	2.48 ± 2.10
45 < S < 60ha	7	2	3.6 ± 0.98
150 < S < 200ha	14	1	5.67

Tableau 3-1 Pour chaque classe de surface de forêts : nombre de pièges, nombre de forêts dans la classe et moyenne de richesse en espèces par forêts

Nous reviendrons sur l'échantillonnage réalisé et les indicateurs de biodiversité estimés en partie 4 du manuscrit, lorsque nous présenterons nos modèles statistiques permettant de relier la diversité en syrphes aux caractéristiques de leur habitat. Retenons à ce stade que les forêts évoquées dans la suite du manuscrit sont celles pour lesquels des relevés biologiques ont été réalisés.

CONCLUSION DE LA PARTIE A

Cette première partie a permis de définir le cadre applicatif de cette thèse et de définir les grands enjeux relatifs à l'étude de la biodiversité forestière.

Nous avons tout d'abord rappelé en chapitre 1 les grands concepts de l'écologie du paysage. L'écologie du paysage se distingue de l'écologie traditionnelle notamment par sa prise en compte de la dimension spatiale où les effets de la structure du paysage (composition, configuration) sont au cœur de la discipline. Nous avons vu que l'effet du paysage sur la biodiversité était principalement fondé sur l'hétérogénéité spatiale. Nous avons aussi évoqué l'effet de la fragmentation des habitats sur les espèces. L'hétérogénéité spatiale et la fragmentation ont fait l'objet d'un grand nombre de travaux et s'appuient sur différentes théories écologiques fondatrices. La théorie de la hiérarchie et de la niche écologique sont ainsi mobilisées pour étudier les effets de l'hétérogénéité tandis que les théories de la biogéographie insulaire et des métapopulations offrent la possibilité d'analyser l'impact de la fragmentation des paysages.

Le chapitre 2 a permis de rappeler l'intérêt écologique des forêts paysannes dans l'espace rural mais que la reconnaissance de leur richesse n'a finalement eu lieu que très récemment. Ces forêts témoignent en réalité des effets que peut avoir la gestion anthropique des paysages sur la présence de biodiversité. Après avoir précisé la biodiversité que nous étudierons dans cette thèse, une description des méthodes et outils aujourd'hui employées pour la mesurer et la caractériser a été menée. Nous avons mentionné les différents facteurs affectant la biodiversité forestière en mettant en évidence l'importance de certaines propriétés spatiales et en particulier, la taille et l'isolement (ou connectivité) dont l'importance s'expliquent par les théories écologiques évoquées au chapitre 1. Enfin, l'ancienneté des forêts est aujourd'hui reconnue comme un facteur tout aussi important pour expliquer la présence actuelle de certaines espèces. Nous avons souligné la distinction entre les notions de maturité et d'ancienneté qui distinguent les vieilles forêts des forêts anciennes, et qui structurent les travaux menés en écologie historique.

Le chapitre 3 s'est attaché à décrire le site d'étude écologique à long terme « Vallées et Coteaux de Gascogne » sur lequel nous avons réalisé l'ensemble de nos expérimentations. Nous avons aussi présenté le modèle biologique retenu dans ce travail : les syrphes et en particulier les espèces spécialistes des forêts en rappelant leurs caractéristiques et l'intérêt que ce taxon présente pour évaluer l'état de conservation des écosystèmes forestiers. L'effet de l'histoire sur les syrphes a été peu étudié jusqu'ici ce qui constitue, sur le plan écologique, une des originalités de ce travail de thèse.

PARTIE B

**Extraire les espaces forestiers
anciens et actuels issus de
sources de données
hétérogènes : démarche
méthodologique**

INTRODUCTION DE LA PARTIE B

Cette partie constitue la première étape de la démarche méthodologique proposée pour reconstruire l'évolution des surfaces boisées sur le temps long. Elle est consacrée à la création de la base de données spatiale forestière historique intégrant les différents états passés et actuels, et destinée ensuite à mettre en lumière les évolutions (Partie C). L'approche méthodologique proposée se veut générique. Elle doit pouvoir être appliquée quel que soit la zone géographique du territoire métropolitain étudiée à condition que l'objet de la dynamique spatio-temporelle porte sur les espaces boisés.

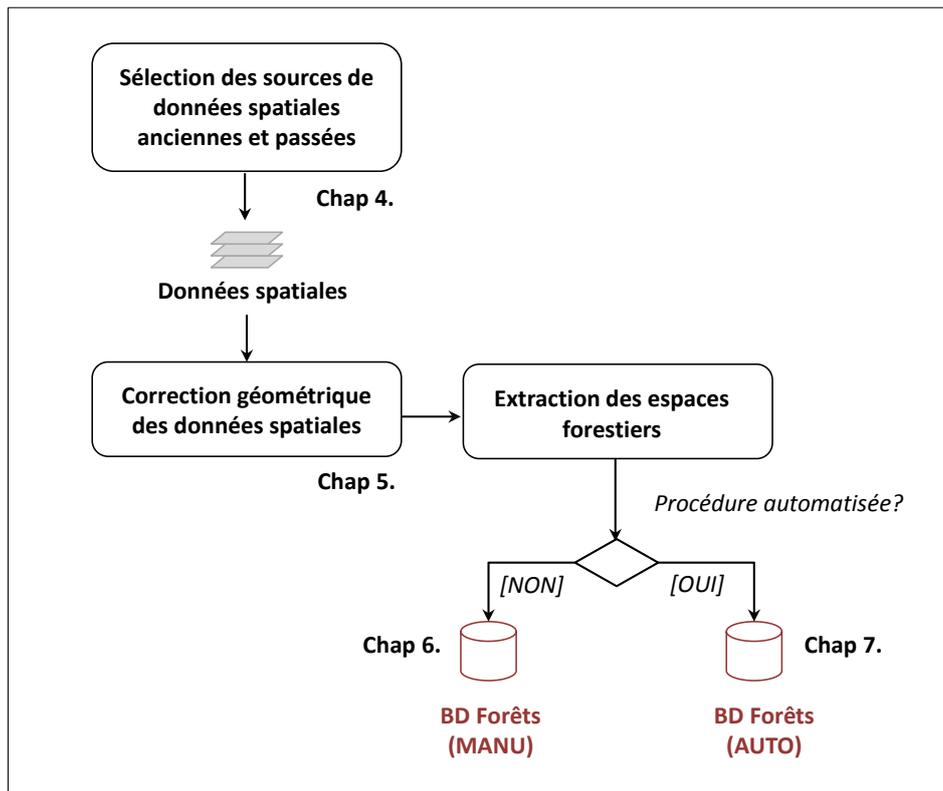


Figure B-1 Démarche générale appliquée pour la production d'une base de données spatiale forestière historique

La première tâche de cette démarche consiste à sélectionner les sources de données spatiales adaptées à notre objectif (Figure I-1). Ces sources doivent répondre à un certain nombre de critères qui seront détaillés en chapitre 4. L'objectif de ce premier travail vise à minimiser les différences entre les sources retenues tout en préservant la richesse de l'information et en mettant l'accent sur leur fiabilité.

Une fois ces sources de données spatiales retenues, une seconde étape vise à les intégrer dans un référentiel cartographique commun en procédant à un recalage géométrique (chapitre 5). L'enjeu est de pouvoir combiner les forêts aux différentes dates pour étudier leur évolution spatio-temporelle. Dans ce chapitre 5, notre contribution porte sur l'examen approfondi de différentes méthodes de recalage géométrique adaptées à des sources anciennes et à l'analyse de leur performance selon différentes valeurs de paramètres (type de méthode, nombre de points d'amer, répartition spatiale). Nous proposons également une nouvelle méthode de recalage global fondée sur la régression à noyau qui tient compte des déformations locales.

Enfin, la dernière étape de la démarche pour cette partie porte sur l'extraction des surfaces boisées dans chacune des sources retenues. Une première approche dite manuelle consiste à cartographier les forêts d'intérêt par digitalisation (chapitre 6). Elle permet d'élaborer un jeu de référence. La seconde approche est automatique et permet d'extraire les forêts en appliquant de nouvelles chaînes de traitement développées (chapitre 7). Pour celles-ci, notre contribution principale porte sur la définition d'une méthode de vectorisation automatique adaptée aux minutes de la carte d'Etat-Major du XIX^e siècle.

CHAPITRE 4.SOURCES DE DONNEES SPATIALES RETENUES ET SPECIFICITES

SOMMAIRE

- 4.1. CRITERES DE CHOIX DES DONNEES SPATIALES
- 4.2. LES MINUTES DE LA CARTE D'ETAT-MAJOR AU 1/40 000^e
- 4.3. LA REVISION TYPE 1889 DE LA CARTE D'ETAT-MAJOR AU 1/80 000^e
- 4.4. LES PHOTOGRAPHIES AERIENNES ANCIENNES NOIR ET BLANC NUMERISEES
- 4.5. LA BD ORTHO INFRA-ROUGE COULEUR DE L'IGN
- 4.6. BILAN : DES DIFFERENCES ENTRE SOURCES COMPATIBLES AVEC NOS OBJECTIFS

Dans ce chapitre, nous exposons les différentes données spatiales anciennes et actuelles retenues pour notre étude en justifiant les choix effectués quant à leur sélection. Ces choix sont fondés sur un certain nombre de critères qui sont d'abord précisés. Enfin, les spécificités de chacune des sources sont décrites avant de dresser un bilan des jeux de données mobilisés dans la suite du travail.

4.1. CRITERES DE CHOIX DES DONNEES SPATIALES

Nous décrivons ici les critères sur lesquelles nous nous sommes appuyés pour sélectionner les données spatiales anciennes et actuelles.

(1) Disponibilité géographique

Les sources sélectionnées devront être disponibles à l'échelle du territoire national afin que l'approche développée pour suivre les évolutions des forêts (Partie C) soit reproductible à cette échelle.

(2) Disponibilité temporelle

Dans la mesure où nous souhaitons évaluer les effets de l'histoire des forêts sur la biodiversité, nous devons pouvoir disposer de données spatiales aux alentours du minimum forestier français situé aux alentours de 1830 (Cinotti 1996; Koerner et al. 2000; Rudel et al. 2005). Cette date est particulièrement importante dans le choix de nos données puisqu'une forêt présente à cette date a plus de chances de faire partie du noyau forestier ancien français, sans que cela constitue toutefois une certitude. Par conséquent, si l'on veut avoir une chance d'identifier les forêts anciennes, nous devons être en mesure d'obtenir des données spatiales qui encadrent ce minimum forestier (Figure 4-1).

Depuis cette date, d'importants changements ont eu lieu et peuvent avoir eu un impact important sur la biodiversité. Nous avons par conséquent cherché à disposer d'informations spatialisées en remontant jusqu'à aujourd'hui avec des écarts entre sources relativement proches pour ne pas biaiser la quantification des évolutions.

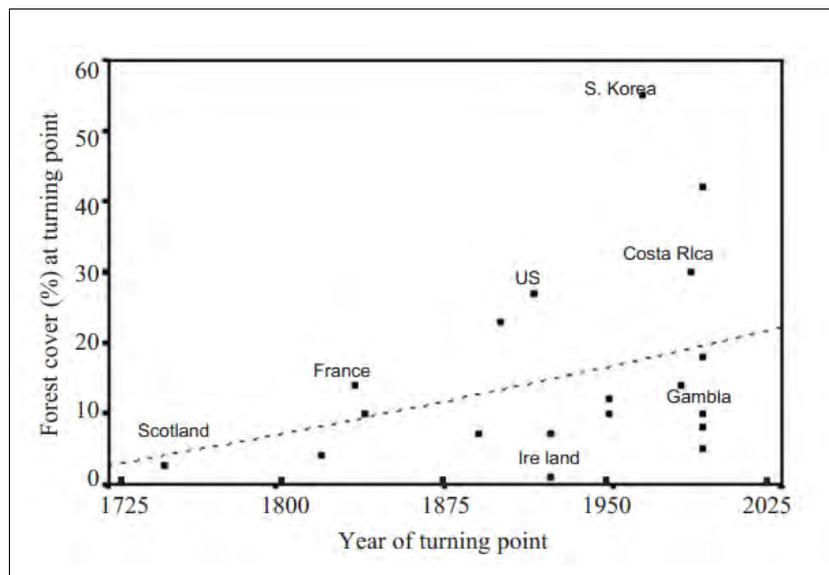


Figure 4-1. La date du minimum forestier en France (source: Rudel et al. 2005)

(3) Echelle, représentation, géométrie et complétude des données

Les sources sélectionnées devront disposer d'une échelle située dans une même gamme de valeur afin de comparer des informations qui ne présentent pas de différences importantes de ce point de vue. La représentation des informations dans ces sources doit permettre d'extraire aisément les espaces boisés. La complétude (ou exhaustivité) des données doit également être assurée pour comparer des paysages forestiers disposant d'une quantité d'informations similaire tandis que la qualité géométrique doit être suffisante pour caractériser précisément la structure spatiale des forêts (à l'échelle du patch) et étudier leurs évolutions (Partie C).

CHOIX EFFECTUES ET DONNEES SPATIALES RETENUES

Le premier choix effectué a consisté à sélectionner les **dessins-minutes d'Etat-Major au 1/40 000^e** pour disposer d'informations aux alentours du minimum forestier. Ces cartes en couleur ont été produites de 1818 à 1866, disposent d'une bonne exactitude géométrique (Dupouey et al. 2007) et sont mises à disposition par l'IGN¹ sur l'ensemble du territoire national (voir section 4.2 pour une description des dessins-minutes). D'autres sources existent à cette période et ont par conséquent été écartées (Figure 4-2). **La carte de l'Etat-major au 1/80 000^e** produite à la même période que les dessins minutes au 1/40 000^e n'a pas été retenue en raison de sa petite échelle et du type de représentation (noir et blanc). **La carte de Cassini au 1/86 400^e** produite de 1745 à 1790 a été écartée principalement pour la faible exactitude des contours géométriques des forêts ainsi que sa période de levé situé trop en amont du minimum forestier. **L'Atlas de Trudaine au 1/8640^e** constitué de 1749 à 1780 n'a pas été retenu pour les mêmes raisons. Enfin, **le cadastre napoléonien** est le document qui coïncide le mieux avec la période du minimum forestier car produit de 1807 à 1850. Il a été levé à une grande échelle allant du 1/5000^e au 1/500^e dans certaines zones géographiques et recense parcelle par parcelle l'usage des sols (Bret 1980; Fabre et Vernet 2006; Poirier 2006; Franchomme 2008; Vallauri et al. 2012; Pinoteau et Di Pietro, 2003). Toutefois, le cadastre est un système d'évaluation et de définition des propriétés foncières servant de base à la répartition de l'impôt à l'échelle nationale, son objectif initial n'est pas cartographique. Pour cette raison, les forêts ne sont pas toujours représentées ou bien les parcelles indiquées comme forêts ne correspondent pas toujours à cette catégorie d'objets sur le terrain (Dupouey et al. 2007). Il a donc été écarté lui aussi.

Nous avons ensuite opté pour **la Révision de la carte de l'Etat-major Type 1889 au 1/40 000^e** produite de 1889 à 1898 pour disposer d'informations aux alentours de 1900. Bien qu'en noir et blanc, cette carte s'est imposée naturellement compte tenu du faible nombre de sources cartographiques à cette date (Figure 4-2). Son échelle est en plus identique à celle des dessins-minutes d'Etat-major, ce qui est un point important dans l'optique d'une comparaison (voir section 4.3 pour une description de la Révision Type 1889). **La carte de France type 1922** constitue une alternative intéressante (en couleur et au 1/20 000^e) mais l'écart de temps qui la sépare des dessins-minutes de 1850 est apparu trop important pour caractériser ensuite les évolutions spatio-temporelles forestières.

Nous avons décidé de compléter notre jeu de données à l'aide de photographies aériennes pour s'affranchir du problème de représentation et de complétude des données. Par conséquent, quatre missions aériennes anciennes mises à disposition par l'IGN ont été sélectionnées. Deux premières missions planifiées en juillet 1953 et juin 1954 permettent de couvrir toute la zone d'étude au milieu du XX^e siècle. Puis, la grande répétitivité des missions aériennes après 1954 a offert la possibilité d'utiliser d'autres photos aériennes acquises en juin 1979 et permis de réduire l'écart temporel entre les sources. Enfin, le niveau de détail important offert par la BD ortho[®] infrarouge couleur (IRC) de l'IGN (résolution spatiale = 0.5m) permet d'obtenir une vision actuelle du paysage

¹ Institut Géographique National

forestier dans les coteaux de Gascogne tout en conservant un écart temporel très proche de celui qui existe entre les premières et les secondes missions aériennes anciennes. Elle a donc été retenue pour la période la plus récente.

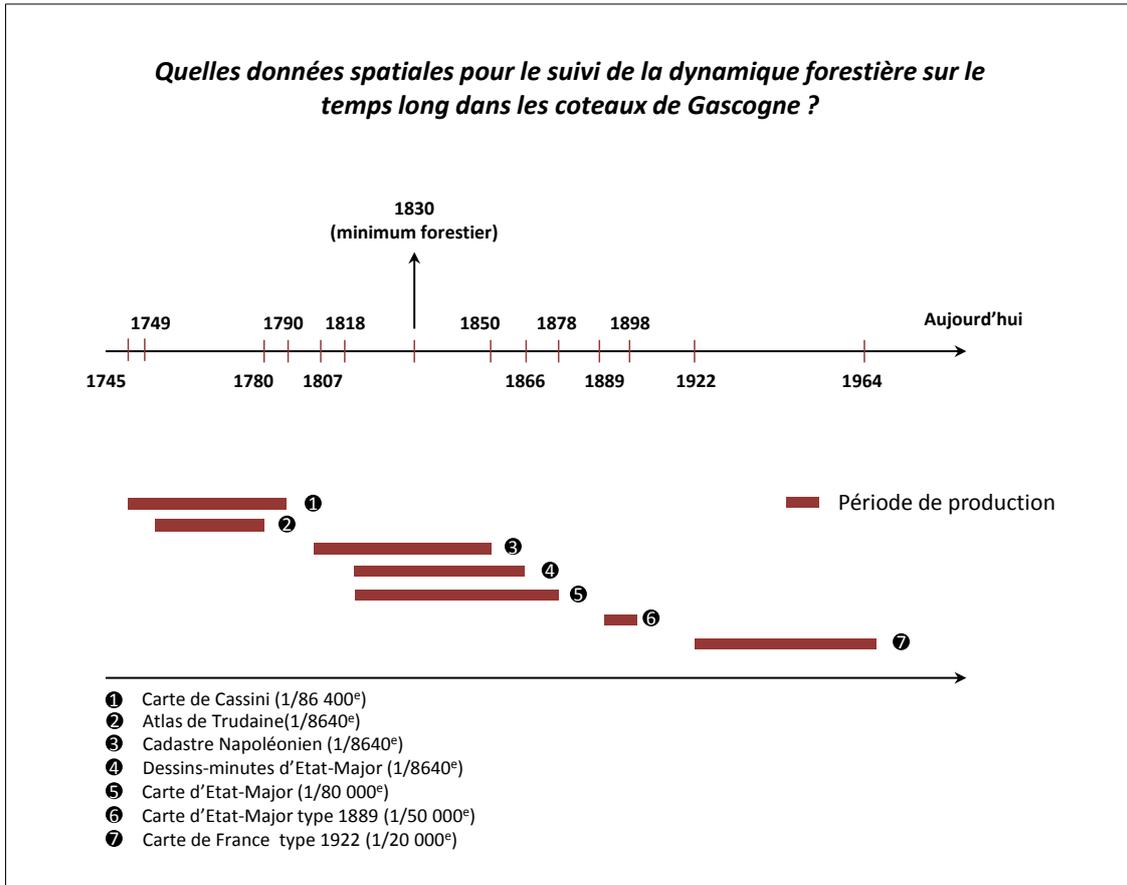


Figure 4-2. Disponibilité des données spatiales pour le suivi de la dynamique forestière sur le temps long dans les Coteaux de Gascogne

4.2. LES MINUTES DE LA CARTE D'ETAT-MAJOR AU 1/40 000^e

4.2.1 HISTORIQUE DE CONSTITUTION DE LA CARTE

La première carte topographique de la France, celle de Cassini, fut terminée aux alentours de 1789. C'est en tout cas à cette date que cessaient les relevés de terrain mais les dernières feuilles furent publiées jusqu'en 1815 (Berthaut 1898). Les levés à peine terminés, Napoléon 1er exprime déjà l'idée d'un nouveau projet cartographique de la France en partie à cause de la faible exactitude géométrique et l'absence de relief dans les cartes de Cassini. Toutefois, à cette époque, la France est en guerre : il faudra donc attendre 1815 et l'avènement de LOUIS XVIII pour que ce projet soit réellement initialisé.

Finalement, très peu de documents cartographiques seront établis entre la fin des levées de Cassini (~1789) et 1815. Cette période assez courte est toutefois fortement marquée par les progrès des sciences géodésiques, topographiques et cartographiques. Des instruments de relevés topographiques comme le théodolite ou le cercle répétiteur permettent par exemple de mesurer les angles horizontaux et verticaux. L'apparition des premiers éclimètres, machine à calculer inspirée de la calculatrice de Pascal, facilitent aussi considérablement le travail des ingénieurs dans le calcul des

cotes de niveau. A partir de 1830, des élimètres plus précis et supportables à dos d'homme amélioreront encore un peu plus ces tâches routinières de nivellement (Figure 4-3).

L'ordonnance royale du 11 juin 1817 présidé par le mathématicien Laplace réunit 14 membres et marque le vrai départ d'une nouvelle aventure cartographique. Elle prévoit notamment « *l'établissement d'une nouvelle carte topographique de la France, appropriée à tous les services publics, et combinée avec l'opération du cadastre général, ainsi que d'en poser les bases et le mode d'exécution* » (Berthaut 1898). Dans ce premier projet, l'échelle de levé est fixée au 1/10 000^e afin que la carte gravée soit éditée au 1/50 000^e. L'objectif est d'intégrer le cadastre dans un référentiel géodésique commun en s'appuyant sur les triangulations de 1^{er}, 2^{ème} et 3^{ème} ordre. La carte prévue est une collection de feuilles de 80*50cm qui découpent la France en quarts suivant un méridien qui passe par Paris, placé au milieu de la feuille de la bande centrale et sa perpendiculaire qui relie Brest à Strasbourg.

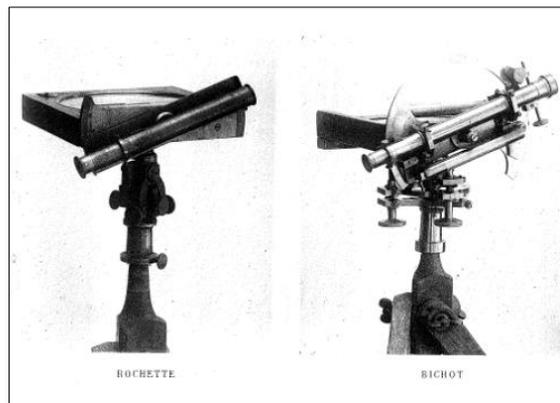


Figure 4-3. *Boussoles à élimètre de Rochette et Bichot (Source: Berthaut 1898)*

En Avril 1818, les premiers relevés de la carte d'Etat-major commencent autour de Paris, successivement dans les cantons de Vincennes et de Charenton. Le projet va alors dériver de ses objectifs initiaux. Les levés au 1/10.000^e prennent trop de temps. On décide donc d'entreprendre cette tâche au 1/20.000^e dans les zones frontalières ou côtières (risques potentiels liés aux ennemis de la France) puis le levé est finalement décidé au 1/40.000^e via l'instruction de 1823 (Berthaut 1898). Ces levés sont reportés sur ce qui est appelé les dessins-minutes. L'instruction de 1823 prévoit aussi de modifier l'échelle de la carte gravée. Si le 1/100 000^e est jugée pratique dans un premier temps car décimal, il n'est cependant pas possible de représenter les petites constructions. On finira par opter pour le 1/80 000^e, échelle suffisamment précise et qui a en outre l'avantage d'être proche de celle de Cassini au 1/86 400^e qu'elle doit remplacer (Berthaut 1898). La France est alors découpée en 273 feuilles couvrant chacune 64*40km. Finalement, cette instruction définira une charte graphique visant à représenter les éléments sur les dessins-minutes et qui aura pour objectif d'assurer l'homogénéité de la représentation sur ces documents.

Sur le terrain, des conflits entre ingénieurs civils et militaires apparaissent progressivement pendant la réalisation des levés. Les ingénieurs militaires reprochent notamment au service civil de ne pas transmettre leurs relevés et de ne pas être en mesure d'assurer correctement le rattachement du cadastre à la triangulation générale. Ainsi, à partir de 1827, une ordonnance royale confie la totalité des travaux aux ingénieurs militaires du service de l'Etat Major dont résultera la carte éponyme.

Malgré cette simplification administrative et la réduction des levés au 1/40 000^e, les opérations de terrain prennent beaucoup de temps. Ce n'est qu'en 1866 que ceux-ci seront entièrement terminés et en 1880 que la dernière des 273 feuilles sera achevée. Cet écart entre la

date de fin des relevés et la date de fin des dernières feuilles publiées s'explique à la fois par l'étape de gravure de la carte qui peut prendre parfois jusqu'à un an et par les différentes mises à jours effectuées sur les autres feuilles déjà imprimées (Costa et Robert 2009). Trois grandes périodes de mises à jour ont eu lieu : (i) 1861-1872 pour des mises à jour partielles visant à intégrer les chemins de fer, (ii) 1872-1889 pour des mises à jour plus systématiques afin de corriger certaines erreurs observées, (iii)-1889-1891 pour des mises à jour intégrant de nouveaux critères qui donneront d'ailleurs lieu à l'établissement d'une révision totale de toutes les feuilles au 1/80 000^e quelques années plus tard.

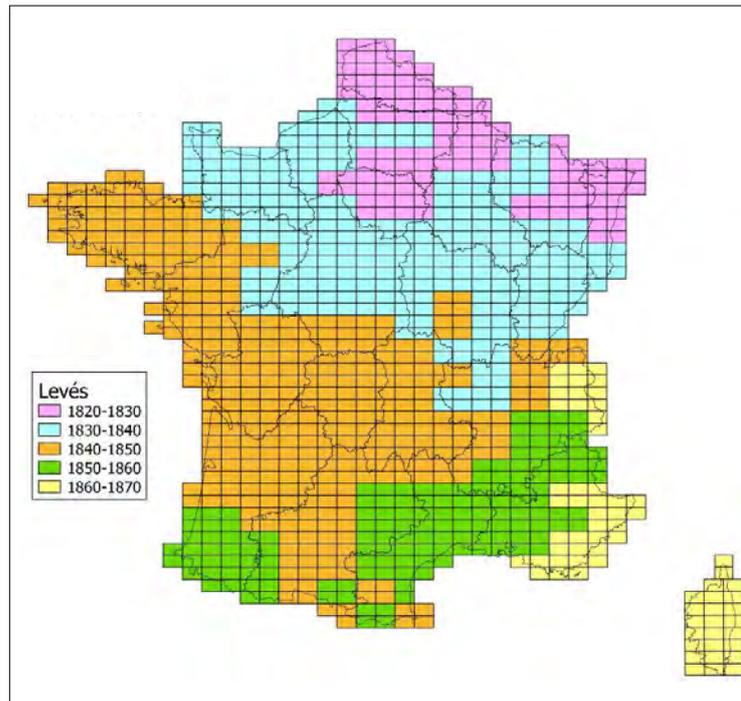


Figure 4-4. Tableau d'assemblage reprenant les dates de levés des feuilles d'Etat-majour au 1/80 000^e (Source : Dupouey, 2011)

4.2.2 SPECIFICATIONS DES DESSINS-MINUTES D'ETAT-MAJOR

Par souci de clarté, il est nécessaire dans un premier temps de distinguer les différents documents produits lors de la constitution de la carte d'Etat Major:

- les dessins-minutes, qui sont la traduction graphique des opérations de terrain, tracés à l'encre et au crayon, au 1/40 000^e
- la mise au net de ces dessins-minutes, à la même échelle et en couleur
- la carte gravée au 1/80 000^e en noir et blanc à partir des documents mis au net au 1/40 000^e

Dans ce travail, nous utiliserons uniquement **la mise au net des dessins-minutes au 1 :40.000e et en couleur.**

Nous décrivons ci-dessous les caractéristiques des dessins-minutes d'Etat-majour.

Paramètres géométriques :

- Projection et méridien :

La carte d'Etat-Major au 1/80 000^e a été levée dans la projection dite de Bonne, projection pseudo-conique équivalente permettant de conserver les surfaces (Figure 4-8). Celle de la carte de l'Etat Major est basée sur l'ellipsoïde de Plessis. Par conséquent, les coordonnées théoriques dans la projection de Bonne ont été associées aux points image du carroyage des dessins-minutes par l'IGN dans le but de pouvoir les reprojeter vers le système Lambert 93.

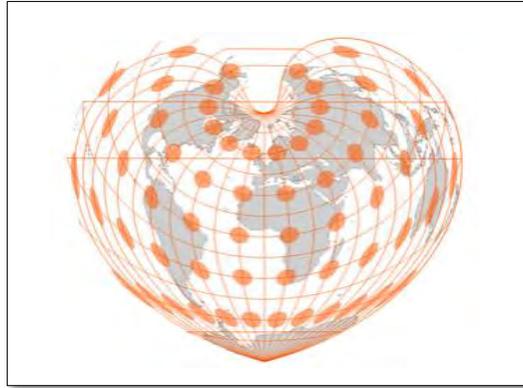


Figure 4-5. Exemple d'un canevas de Bonne

- Levés topographiques :

La triangulation utilisée pour produire les dessins-minutes est celle utilisée pour la carte d'Etat-Major au 1/80 000^e et en tout point identique à celle de Cassini. Néanmoins, cette fois-ci (pour la carte au 1/80 000^e et les dessins-minutes au 1/40 000^e), les mesures décimales sont chiffrés en mètres (Assemblée constituante de 1791) tandis que la méridienne (entre Dunkerque et Perpignan et qui passe par l'observatoire de Paris) mesure 551 584.7 toises² (Delambre et Legendre 1798). Un premier maillage a ainsi été défini puis les ingénieurs ont progressivement effectués les levés de détail au sein de ce premier maillage en s'appuyant sur le cadastre (Figure 4-6).

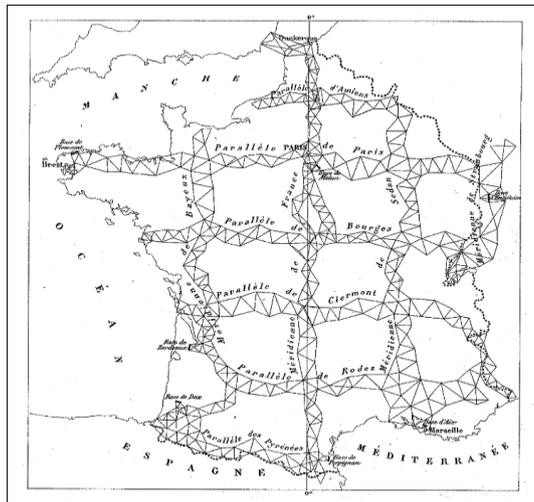


Figure 4-6. La triangulation du 1^{er} ordre de la France sur laquelle s'est appuyée la réalisation de la carte de Cassini et de la carte d'Etat-Major

L'instruction de 1823 (Berthaut 1898) qui mentionne le passage des minutes au 1/40 000^e précise que 20 à 25 points géodésiques devaient être suffisants pour l'établissement du squelette linéaire de ces minutes et que ces derniers doivent être correctement répartis sur la zone pour

² une toise = 2 mètres

couvrir toutes les irrégularités du terrain étudié. Néanmoins, il y est inscrit que « ... c'est à vous, Monsieur (*l'ingénieur chargé du levé*), à juger ce que vous reconnaissez indispensable, si vous deviez vous-même exécuter les opérations de détails.. » (Berthaut, 1898, Tome I, p 302). Ceci explique que les dessins-minutes d'Etat-Major comportent la plupart du temps un nombre plus important de points géodésiques dans les régions où le relief est très marqué. Ces points sont renseignés sur le cartouche située à gauche des feuilles des dessins-minutes (Figure 4-7).

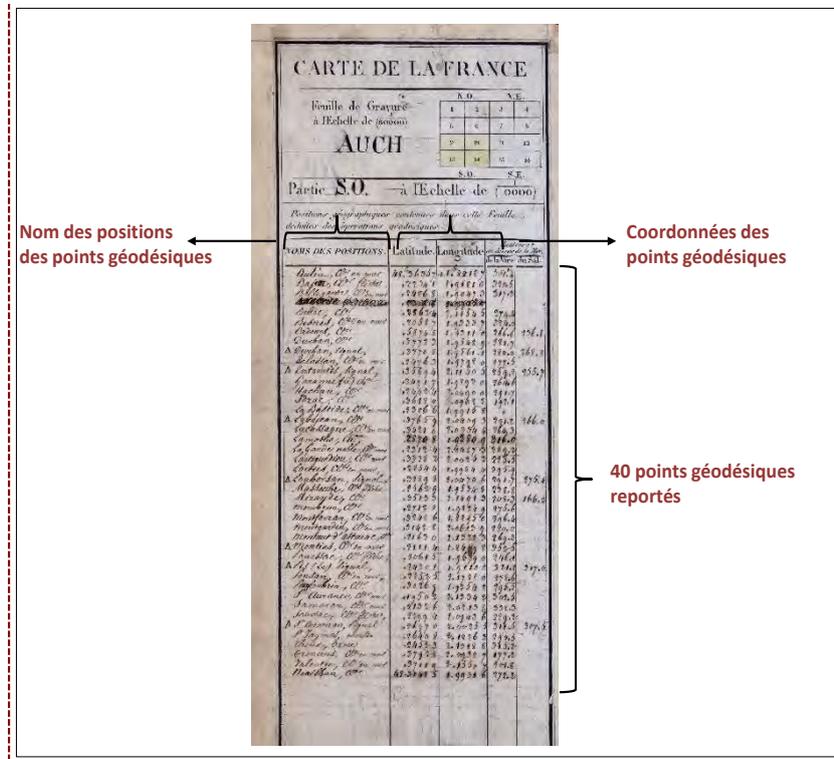


Figure 4-7. Exemple d'un cartouche présent sur un dessin-minute d'Etat-major (Feuille 241 NE, Saint Gaudens (31))

Echelle, découpage et emprise :

On dénombre au total 273 feuilles de la carte de l'Etat-Major au 1/80 000^e couvrant respectivement une surface de 64*40 km. Les dessins minutes au 1/40 000^e représentent le terrain par quart de ce que deviendra la feuille au 1/80 000^e (Figure 4-8).

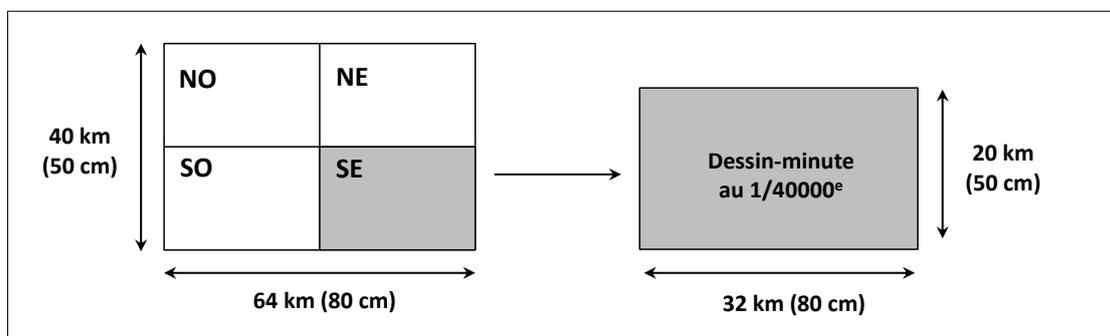


Figure 4-8. Etendue d'un dessin-minute au 1/40 000^e. La combinaison des quatre planches au 1:40 000e permet de constituer une feuille de la carte d'Etat-Major au 1/80 000^e (Source: Plumejeaud 2012, adapté)

Au total, 976 dessins-minutes existent à l'échelle de la France et portent le nom de la ville principale qu'ils contiennent. Les attributs NO, SO, NE et NO permettent en plus de localiser le quart de feuille correspondant à la division de la feuille au 1/80 000e (Figure 4.8)

Représentation et généralisation cartographique :

L'instruction de 1823 renseigne sur les indications apportées aux ingénieurs à propos des relevés et des mécanismes de représentation à adopter (Berthaut 1898). Le passage des minutes au 1/10 000e vers le 1/40.000e implique notamment de procéder à des choix de généralisation visant à homogénéiser les relevés et leur traduction. Cette charte n'aborde toutefois pas tous les thèmes figurant sur ces minutes. Une autre instruction de 1851 viendra compléter celle de 1823, notamment pour l'assemblage des communes et la réduction de l'échelle dans l'optique de la gravure au 1 :80.000e.

- Le relief (Tableau 4-1.a)

L'une des différences majeures entre la Carte de Cassini et la carte de l'Etat-major est la représentation du relief.

Dans les dessins minutes au 1/40.000^e, la représentation du relief repose sur le report des cotes de niveau et le tracé de *hachures* correspondant à des traits matérialisant les lignes de plus grande pente

- Les cotes de niveau sont déterminées sur les points culminants, les fonds, les bords ou les sources des eaux courantes et stationnaires ou encore sur les carrefours des grandes routes.
- Les hachures s'appuient sur des courbes de niveau non représentées.
- Les hachures doivent être faibles et espacées lorsque la pente est douce et doivent être plus serrées et fortifiées lorsque la pente devient plus raide. Leur longueur est inversement proportionnelle à la pente et leur épaisseur est proportionnelle à cette pente.
- La distance maximale entre deux hachures doit être de 2 mm dans les pentes les plus douces.
- La largeur minimale d'une hachure ne doit pas excéder un bon tiers de mm.
- Les hachures sont continues ou ne seront brisées que dans un but esthétique.

L'instruction de 1823 précise qu'il est laissé à l'ingénieur de faire le choix des détails. Parfois, certaines pentes faibles ne seront pas représentées dans « *l'intention de conserver l'ensemble et l'esprit des formes sur le terrain* » (Berthaut, 1898, Tome 1, p. 301). L'enjeu principal réside sur les bords des grandes rivières, sur les bords des routes où le figuré doit être plus soigné afin « *de servir avec succès aux opérations militaires d'attaque et de défense* » (Berthaut, 1898, Tome 1, p. 301).

L'instruction de 1851 reprendra ces spécifications liées au relief en les complétant :

- les hachures visant à représenter le relief seront parallèles aux lignes de plus grandes pentes
- les courbes de niveau ne sont plus prises en comptes dans le dessin des hachures
- l'hypothèse de la lumière oblique est abandonnée

- Les villes et les bourgs (Tableau 4-1.m)

- Seules les rues principales sont représentées.
- Les églises sont représentées par un rond accompagné d'une croix dont le centre sera celui du clocher.
- Les édifices remarquables sont représentés par un trait plein imitant strictement leurs contours.

- Les villages, hameaux et maisons isolés (Tableau 4-1.n)

- Toutes les rues sont représentées.

- La longueur minimale des parallélogrammes représentant les maisons est $\frac{1}{2}$ mm tandis que la largeur minimale ne doit pas être inférieure à $\frac{1}{3}$ mm.
- Les maisons sont individualisées.
- Les jardins sont délimités par un trait plein lorsqu'ils sont bordés par un mur et par un trait ponctué lorsqu'ils sont bordés par une haie.
- Les clochers sont représentés par un rond enfermant une croix ou un point lorsqu'ils ont été utilisés comme points géodésiques.
- Les maisons apparaissent en rouge lorsqu'elles sont en maçonnerie ou en noir lorsqu'elles sont en bois

- Les voies de communications (Tableau 4-1.p, q, r, s)

- Les routes principales sont représentées par un double trait dont l'écart entre ces deux traits est proportionnel aux routes de 1^{er}, 2nd et 3^e ordre.
- Les routes bordées d'arbres sont représentées avec des pointillés.
- Dans les grands bois percés par la chasse, on conserve tous les chemins et avenues qui sont représentés par un seul trait.

- Les rivières, ruissellement et canaux (Tableau 4-1.l)

- Si la largeur de la rivière excède 20 m, elle sera représentée par deux traits dont l'un des deux aura une largeur en proportion naturelle avec l'échelle.
- Si la largeur de la rivière oscille entre 10 et 20 m, elle sera représentée par deux traits égaux.
- Si la largeur de la rivière est inférieure à 10 m, elle sera représentée par un seul trait.

- Les forêts (Tableau 4-1.b)

- Les bois composés de plusieurs espèces se verront attribués les indications « C » pour le chêne et « H » pour le hêtre qui pourront être complétées par un « T » ou un « F » pour renseigner sur le taillis ou la futaie.

La représentation des forêts dans les dessins-minutes d'Etat-Major sera détaillée et discutée en section 4.2.3.

Le tableau 4-1 reprend les thèmes principaux observés dans les dessin-minutes. A noter toutefois que cette revue des thèmes n'est pas exhaustive mais elle devrait permettre au lecteur d'obtenir une vue d'ensemble des représentations existantes, notamment pour celles qui ne sont pas abordées dans les instructions de 1823 et de 1851. Pour une liste exhaustive des thèmes avec leur représentation, nous recommandons de consulter l'ouvrage publié par Costa et Robert 2009, le rapport de digitalisation des dessin-minutes (Favre et al. 2012) et le descriptif du produit Scan Etat Major (SCAN Historique®-SCAN Etat Major®-IGN).

NATURE	EXEMPLES	FIGURES/COULEURS	OBSERVATIONS
Le relief			
a) Modelé du relief		Hachures noires qui s'appuient sur les courbes de niveau	La longueur des hachures est inversement proportionnelle à la pente
L'occupation du sol			
b) Forêt		Aplats dont la teinte peut varier du jaune au vert (voir section 4.2.3)	Le plus souvent délimité par un contour noir ; parfois marqué « forêt » ou « bois »
c) Vigne		Aplat violet	Le plus souvent délimité par un contour noir
d) Prairie		Aplat bleu charron à vert sarcelle	Le plus souvent délimité par un contour noir
e) Prairie marécageuse		Aplat bleu azurin / cyan	Positionné le long d'un cours d'eau, parfois délimité par un contour noir
f) Pâquis, pâtures		Aplat dont la teinte est un mélange de bleu ciel passé (azur brume) et de rose	Absence de contour de délimitation
g) Marais		Aplat bleu azur clair	Le plus souvent marqué « marais » ; en bord de rivières, d'étang, de lacs
h) Glacier		Aplat blanc/bleu ciel + lignes de niveau bleu	Le plus souvent marqué « glacier »
i) Plage		Aplat orange	
Limites administratives			
j) Limite administrative		Trait épais rouge ou bleu/vert	Parfois marqués par des croisillons accolés aux traits

NATURE	EXEMPLES	FIGURES/COULEURS	OBSERVATIONS
L'hydrographie			
k) Etang et lac		Aplat bleu ciel irrégulier tâché de blanc	Le plus souvent marqué « lac » ou « étang » ; délimité par un contour bleu foncé
l) Rivière		Aplat bleu sarcelle délimitée par un double trait	Le plus souvent marqué « rivière » sur une portion du cours d'eau
L'urbanisme			
m) Villes ou villages		Aplat rouge dans les villes ou villages	Pas ou peu de détail sur les bâtiments, seules les rues principales apparaissent
n) Bâti en zone rurale		Rectangle rouge	Bâtiments individualisés en zone rurale
o) Détails planimétriques ponctuels		Symboles : croix, triangle, cercle englobant un point	Localisation de clochers, de points géodésiques, de calvaires (souvent très fiable géométriquement)
Les voies de communication			
p) Grande voie routière		Double trait noir	Présence de pointillés en bordure des traits en zone rurale pour signifier la présence d'arbres
q) Chemin de fer		Trait noir épais	Parfois indiqué « station » ou chemin de fer sur une portion de la voie »
r) Chemin		Double trait sans surcharge	Parfois marqué « chemin »
s) Sentier		Simple trait noir	

Tableau 4-1. Description des principaux thèmes observés sur les dessins-minutes d'Etat-Major au 1/40 000^e

4.2.3 REPRESENTATION DES ESPACES FORESTIERS DANS LES DESSINS-MINUTES

Comme d'autres occupations du sol, les forêts ne sont pas abordées dans l'instruction de 1823, ni dans l'instruction de 1851. Seuls les « bois » disposent de spécifications (Instruction de 1823, Berthaut 1898) mais leur définition ne permet pas d'attribuer une définition claire à la forêt représentée dans ces documents.

Cette absence de spécifications précises indique probablement qu'il était laissé à l'ingénieur le soin de trouver une représentation pour les forêts qui permettant de les distinguer des autres éléments sur ces cartes. Pour ces raisons, la représentation des forêts sur à l'échelle nationale n'est pas homogène et peut même varier fortement d'un endroit à l'autre du territoire, selon l'ingénieur cartographe en charge du levé.

Nous détaillons ci-après les différentes caractéristiques des forêts observées sur les dessins-minutes d'Etat-Major. Ceci a pour objectif d'avoir une vision globale des différentes représentations de ces objets à l'échelle du territoire français.

- La teinte

La teinte des forêts dans les dessins-minutes dépend directement de plusieurs critères :

- (1) le mélange de couleur choisi par l'ingénieur pour représenter l'objet
- (2) la surcharge d'éléments planimétriques et orographiques qui tend à assombrir la teinte
- (3) le niveau de qualité de conservation de la carte qui tend à éclaircir la teinte

La représentation des forêts sur les dessins minutes d'Etat-major varie du jaune pâle au vert foncé. Cela veut dire, d'un point de vue sémiologie graphique qu'il y a à la fois une variation de la teinte (jaune, vert) et une variation possible de la valeur (du clair au foncé). Entre ces deux extrêmes, il est possible d'observer toute une gamme de variations qui peuvent parfois se confondre avec d'autres occupations du sol (Tableau 4-2.j et a). Ces variations sont parfois observables au sein d'une même feuille puisque certaines de ces feuilles sont composées de plusieurs levés de terrain assemblés mais effectués par différents ingénieurs (Tableau 4-2.e et b).

La surcharge d'éléments planimétriques et orographiques tend aussi à assombrir la teinte des bois (Tableau 4-2.k). Il est possible d'observer sur l'exemple mentionné que la présence de hachure assombrit clairement la teinte verte initialement prévue pour caractériser l'objet « forêt ». Ces éléments peuvent aussi s'apparenter à du texte, des voies de communications, des limites administratives ou encore des symboles. La mauvaise qualité de conservation de certaines planches contribue en revanche à éclaircir et adoucir les couleurs appliquées initialement (Tableau 4-2.a).

- La dénomination

Si les forêts sur les dessins-minutes se distinguent par des aplats de couleur très variés, l'appellation qui leur est apportée est également hétérogène. En effet, on retrouve parfois la dénomination de « Bois » ou « Forêt » sur les objets sans qu'aucunes spécifications ne précisent la condition de leur présence (Tableau 4.2a, c, d, e, f, h). Une chose paraît plus certaine cependant, les forêts disposant de l'une ou l'autre dénomination sont tous des massifs de taille relativement importante : la présence explicite de ces dénominations est moins fréquente voire absente dès lors que les forêts sont de petite taille. (Tableau 4.2j, l).

- La délimitation

La majorité des forêts sur les dessins-minutes disposent d'un contour représenté par un trait noir épais permettant de délimiter correctement leur emprise spatiale. Néanmoins, dans les zones de forte pente, la surcharge des hachures est telle que le contour des bois y est parfois absent ou invisible (Tableau 4-2k). Le faible contraste qui existe entre les aplats de couleur dans ces zones ne permet pas de compenser l'absence de contour, ce qui rend leur délimitation plus complexe

Au regard des observations effectuées, nous ne sommes pas en mesure de donner une définition claire des forêts présentes sur les dessins-minutes d'Etat-major. Les conditions de représentation ne sont pas précisées tandis que le type de représentation et la dénomination varient d'une feuille à l'autre ou parfois même au sein d'une même feuille. De fait, cet aspect devra être pris en compte au moment d'extraire les surfaces boisées dans les différentes sources afin de ne pas biaiser la quantification des évolutions.

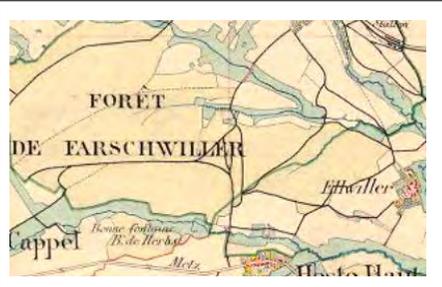
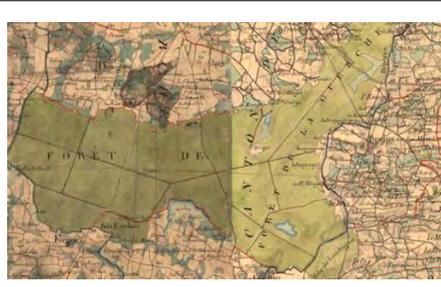
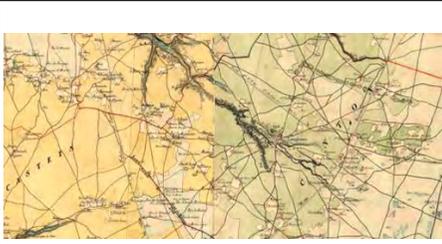
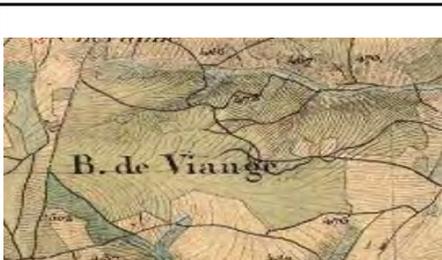
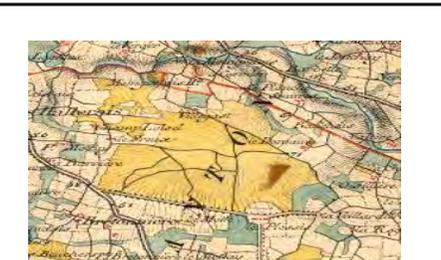
			
<p>a) Sarebourg (SO)</p>	<p>b) Château-Gontier (NO)</p>	<p>c) Loches (SE)</p>	<p>d) Saint-Gaudens (NE)</p>
			
<p>e) Mont de Marsan (NO)</p>	<p>f) Paris (NO)</p>	<p>g) Château-Chinon (SE)</p>	<p>h) Roanne (SO)</p>
			
<p>i) Mont de Marsan (SE)</p>	<p>j) Rennes (SE)</p>	<p>k) Castellane (NO)</p>	<p>l) St Omer (SE)</p>

Tableau 4-2. Exemples de représentation des surfaces boisées dans les dessins-minutes d'Etat-Major au 1/40.000^e

4.2.4 LES DESSINS-MINUTES DE LA ZONE D'ETUDE

Les dessins minutes utilisés dans cette étude sont au nombre de quatre et ont été téléchargés depuis le site du Geoportail de l'IGN. D'après les spécifications fournies par l'IGN, ces données ont été scannées à une résolution de 600 dpi ce qui signifie qu'en tenant compte de l'échelle des documents au 1/40 000^e, le côté d'un pixel dans l'image correspond à une distance de 1,7 mètres sur le terrain.

Sachant que chaque dessin-minute couvre une superficie de 32x20 km, l'ensemble de la zone retenue composée de 4 dessins-minutes couvrent au total une superficie de 64x40 km. Les détails des feuilles sélectionnées sont disponibles dans le tableau 4-3. Pour un aperçu des feuilles, voir la figure 4-9.

NOM DE LA FEUILLE	QUART DE FEUILLE	DEPARTEMENT	DIMENSIONS	FORMAT
Saint-Gaudens	241 NE	Haute Garonne (31)	32*20 km	Jpeg 2000
Saint-Gaudens	241 NO	Haute Garonne (31)	32*20 km	Jpeg 2000
Auch	229 SE	Gers (32)	32*20 km	Jpeg 2000
Auch	229 SO	Gers (32)	32*20 km	Jpeg 2000

Tableau 4-3. Caractéristiques des dessins-minutes sélectionnés dans la cadre de cette étude



Figure 4-9. Aperçu des dessins-minutes traités dans ce travail de thèse. 1) Feuille 229SO, 2) Feuille 229SE, 3) Feuille 241NO, d) 241NE

La section qui suit a pour objectif de présenter les spécificités de la seconde carte historique retenue dans ce travail à savoir, la Révision Type 1889 de la carte d'Etat-major.

4.3. LA REVISION TYPE 1889 DE LA CARTE D'ETAT-MAJOR AU 1/80 000^e

4.3.1 HISTORIQUE DE CONSTITUTION DE LA CARTE

Le développement accru des voies de communication et certaines erreurs observées dans la première version de la carte a naturellement conduit le service de l'Etat-Major à apporter des modifications et à entreprendre une révision totale de la carte initiale au 1/80 000^e (Berthaut 1898).

A partir de 1891, les révisions sont effectuées à partir d'un agrandissement par photographie au 1:40 000^e (puis au 1/50 000^e) de la carte gravée au 1/80 000^e. Cet agrandissement permettra d'améliorer la lecture des éléments inscrits en noir et blanc. D'un point de vue opérationnel, cet agrandissement nous permet également de pouvoir comparer deux sources cartographiques à environ cinquante ans d'intervalle, à la même échelle. Toutefois, ce format noir et blanc présente un défaut de lisibilité: les limites administratives sont difficilement visibles et les chemins se confondent facilement avec les ruisseaux. Elle souffre en outre d'une surcharge d'éléments planimétriques.

Chaque feuille au 1/80.000e est découpée en quatre planches et ces planches, au nombre de 965, couvrent une surface unitaire de 20 x 32 km (Figure 4-10). Cette subdivision des feuilles permet d'augmenter le nombre de graveurs employés à la révision de chacune des feuilles « ...assurant de la sorte, une plus gr ete rapidité dans les tâches d'impression. » (Nadal 2011).

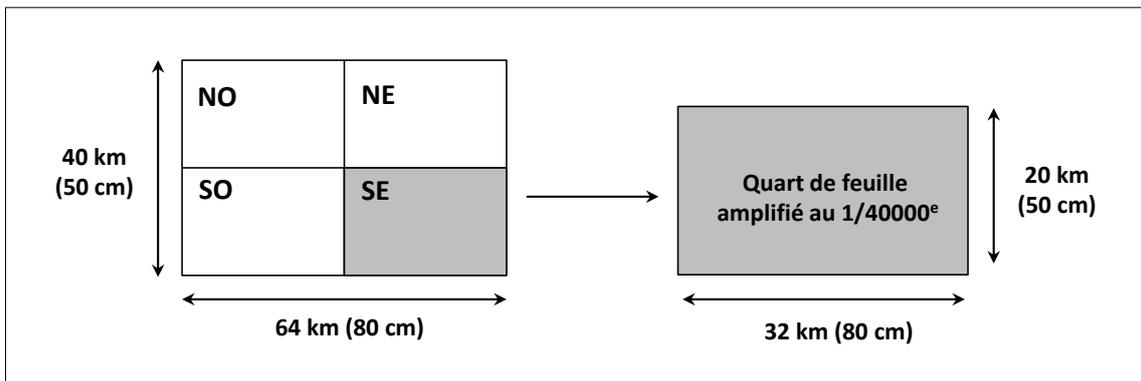


Figure 4-10. Structure d'un quart de feuille Type 1889 (Plumejeaud 2012, adapté)

A l'échelle nationale, les mécanismes de mises à jour opérées dans le cadre de cette révision semblent s'être déroulés de 1889 à 1898 (Costa et Robert 2009). Néanmoins, les travaux de Nadal (2011) sur la cartographie militaire des Pyrénées Françaises et Espagnols indiquent qu'en ce qui concerne notre zone d'étude, la révision s'est déroulée un peu plus tard, en 1900 (Figure 4-11)

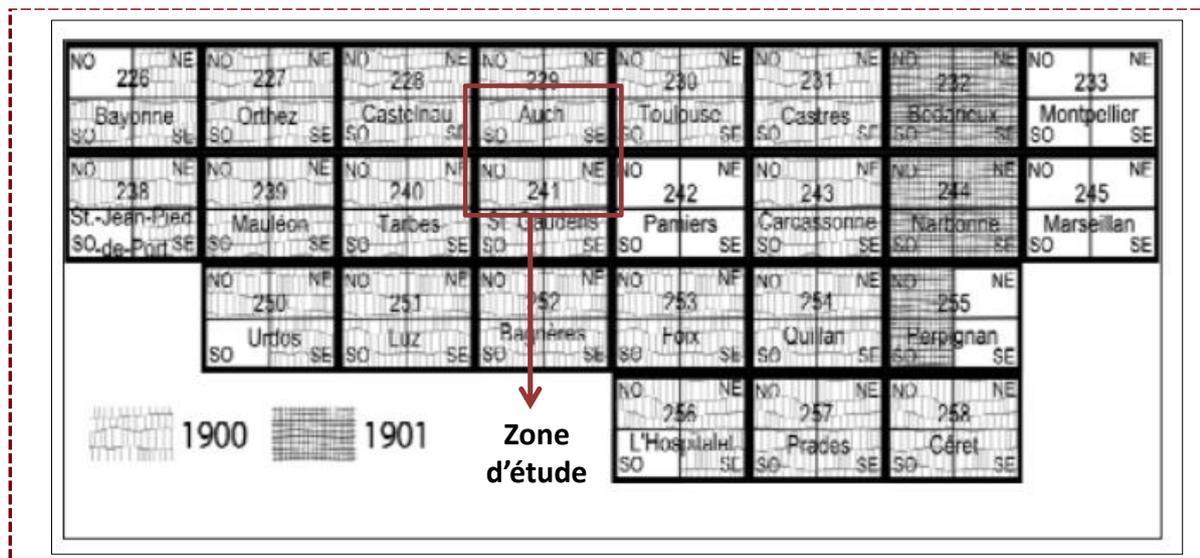


Figure 4-11. Révisions des feuilles de l'Etat-major au 1/80 000^e (Source : Nadal 2011)

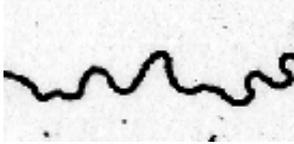
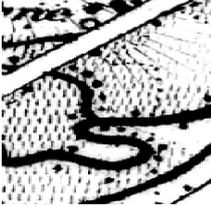
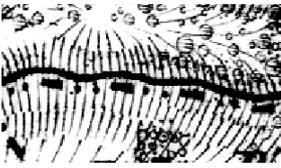
Les feuilles n°233 (Montpellier), 245 (Marseillan) et 242 (Pamiers) ont toutes été révisées de 1891 à 1893. Environ 10 années plus tard, toutes les autres feuilles présentées en figure 4-11 ont fait l'objet d'une mise à jour entre 1900 et 1901.

4.3.2 LES SPECIFICATIONS DE LA REVISION

Les spécifications de la révision type 1889 sont identiques à celles des planches-mères dont elle est issue à savoir les feuilles au 1/80 000^e, en noir et blanc.

La révision type 1889 va toutefois compléter plusieurs figurés et surtout généraliser la carte dans le but d'améliorer la lisibilité, critique souvent émise aux feuilles de l'Etat-major au 1/80 000^e (Tableau 4-12). Par exemple, les routes impériales et départementales sont devenues les routes « principales » (Tableau 4.4(m)) tandis que les chemins de moyenne et grande communication sont les routes « secondaires » (Tableau 4.4(o)).

Les principaux thèmes de cette carte sont répertoriés dans le tableau 4-4. Pour une revue complète des thèmes présents sur cette édition et des critères de spécifications, nous recommandons l'ouvrage de Costa et Robert (2009).

NATURE	EXEMPLES	FIGURES/COULEURS	OBSERVATIONS
L'hydrographie			
a) Fleuves, rivières, étang, plans d'eau		Aplats soulignés de lignes concentriques	Les étangs peuvent contenir des figurés complémentaires liés
b) Les rivières secondaires		Trait noir gras sinueux	
c) Les rivières de moins de 10 m TYPE 1889		Trait noir gras sinueux	Thème ajouté par le type 1889 → les rivières secondaires sont les rivières de moins de 10m
d) Les rivières à sec pendant l'été TYPE 1889		Trait noir gras+pointillés	Thème ajouté par le type 1889
L'urbanisme			
e) Bâti en zone rurale		Rectangles noirs	Bâtiments individualisés en zone rurale
f) Villes ou villages		Aplats grisés	
Le relief			
g) Pentes		Hachures noires	Dessinées au quart dans des courbes espacées de 10 mètres

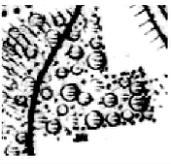
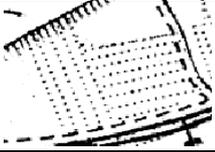
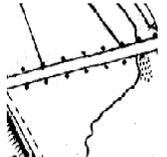
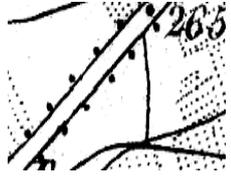
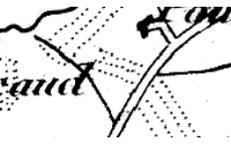
NATURE	EXEMPLES	FIGURES/COULEURS	OBSERVATIONS
L'occupation du sol			
h) Forêts		Trame de feuillus illustrée par des poncifs	Absence de contour géométrique
i) Les vignes		Lignes pointillés	Absence de contour mais délimitation possible par les lignes de pointillés
j) Prés et zones humides		Trame grisée	Absence de contour
k) Enclos bordés par une haie		Pointillés	Trait plein pour les enclos bordés de pierre
Les voies de communication			
l) Routes départementales et impériales		Trait double (0.1cm) bordé ou non de plantations	
m) Routes principales (TYPE 1889)		Trait double (0.1cm) bordé ou non de plantations	Thème ajouté par le type 1889 → routes départementales et impériales sont les « routes principales »
n) Chemins de gr ete communication		Trait double (0.05 cm)	Si présence de tirets sur un des deux traits alors c'est un chemin de moyenne communication
o) Routes secondaires TYPE 1889		Trait double (0.05 cm)	Thème ajouté par le type 1889 → Chemins de moyenne et gr ete communication sont les « routes secondaires »

Tableau 4-4. Description des principaux thèmes observés dans la Révision Type 1889. Les encadrés bordeaux indiquent les apports de la Révision Type 1889

4.3.3 REPRESENTATION DES ESPACES FORESTIERS DANS LA CARTE TYPE 1889

- Type de représentation et délimitation

Dans la révision Type 1889, les forêts sont représentées par une trame de feuillue relativement dense composée de poncifs en forme de cercles (Figure 4-12). Ces trames ne disposent pas d'un contour précis, ce qui rend plus difficile l'identification de l'étendue de la forêt, notamment dans les zones de fortes pentes (Figure 4-12(d et e)). Il est toutefois possible, dans certains cas, de s'appuyer sur les bords de la trame pour les délimiter (Figure 4-12(c)).

Un des avantages des feuilles type 1889 est son homogénéité dans la représentation des forêts. En effet, tous les espaces forestiers apparaissent avec la même symbologie.

- La dénomination

Comme dans les dessins-minutes au 1/40 000^e, il ne semble pas y avoir de spécifications particulières quant à la dénomination « bois » ou « forêts » bien que la présence de ces intitulés soit moins fréquente pour les forêts de petite taille (Figure 4.13(a et e)).

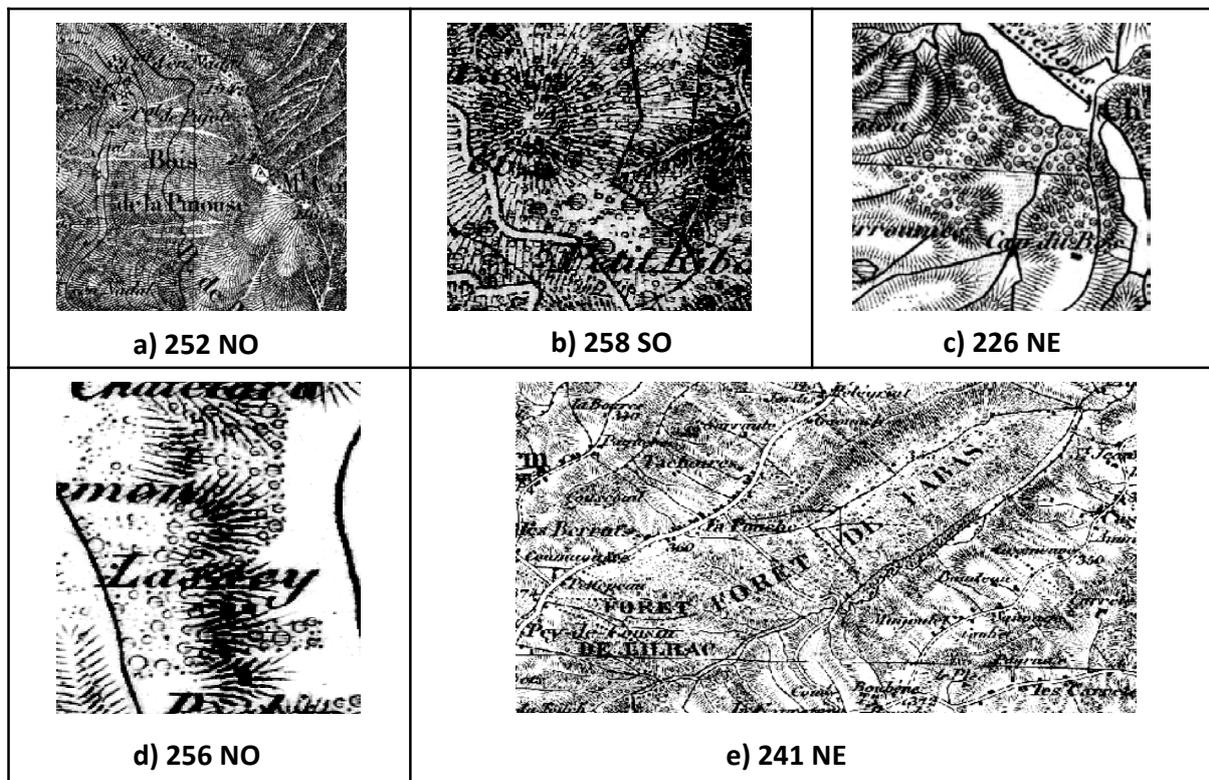


Figure 4-12. Exemples de représentation des surfaces boisées dans la révision Type 1889.

La révision Type 1889 souffre d'une surcharge d'éléments planimétriques importante. Elle est parfois difficilement lisible et la représentation des forêts rend complexe l'identification de leur contour. Toutefois, elle a l'avantage d'être homogène et bien qu'elle soit issue de la carte gravée au 1/80 000^e, son échelle (1/40 000^e) est compatible avec celle des dessins-minutes.

4.3.4 LES CARTES TYPE 1889 DE LA ZONE D'ETUDE

Au total, quatre feuilles ont été mobilisées pour cette période. Les données, scannées à 300 dpi et mises à disposition par l'IGN présentent une résolution spatiale de 3.89 mètres (Tableau 4-5). Pour un aperçu des feuilles, voir Figure 4-13.

Nom de la Feuille	Quart de feuille	Département	Dimensions	Format
Saint-Gaudens	241 NE	Haute Garonne (31)	32*20 km	Jpeg 2000
Saint-Gaudens	241 NO	Haute Garonne (31)	32*20 km	Jpeg 2000
Auch	229 SE	Gers (32)	32*20 km	Jpeg 2000
Auch	229 SO	Gers (32)	32*20 km	Jpeg 2000

Tableau 4-5. Caractéristiques des feuilles de la Révision Type 1889 sélectionnés dans la cadre de cette étude

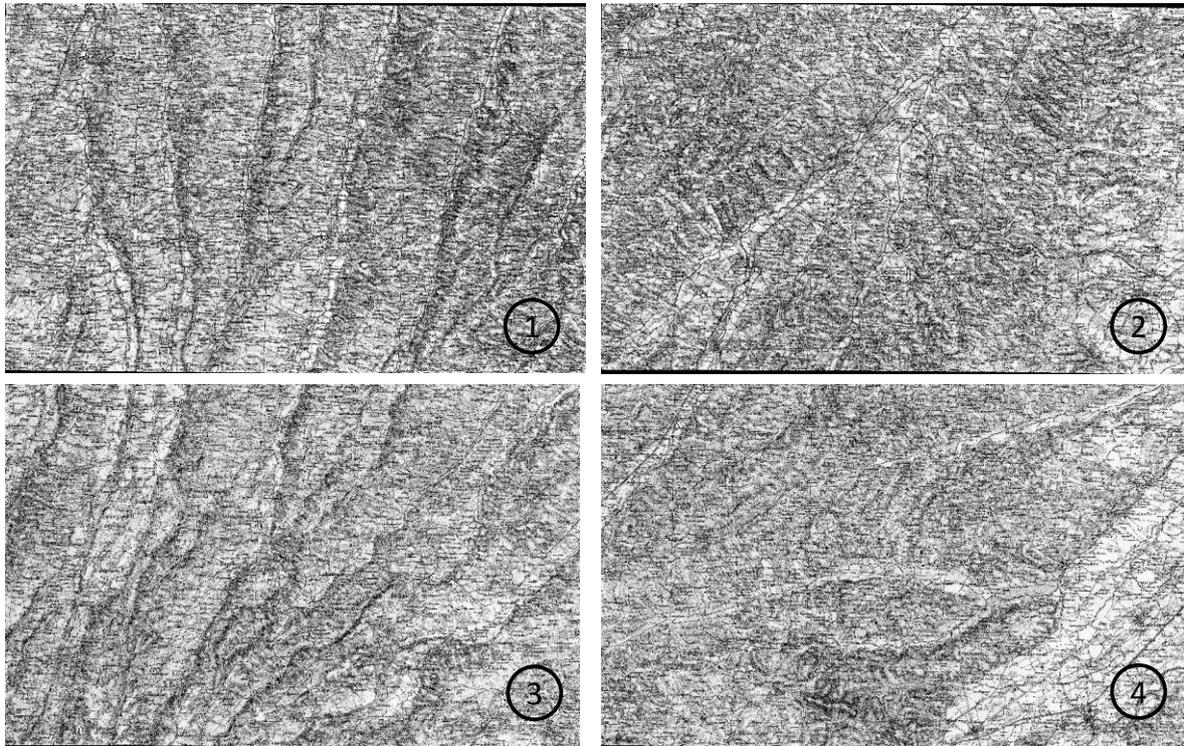


Figure 4-13. Aperçu des quatre feuilles sélectionnées. 1) Feuille 229SE ; 2) Feuille 229 SO ; 3) Feuille 241 NE ; 4) Feuille 241NO

4.4. LES PHOTOGRAPHIES AERIENNES NOIR ET BLANC NUMERISEES

Pour les périodes les plus récentes, nous choisissons d'utiliser des photographies aériennes pour s'affranchir d'un maximum d'artéfacts liés à la représentation cartographique tels que la généralisation ou le type de représentation.

Depuis plusieurs années, l'IGN met à disposition différentes missions aériennes anciennes accessibles gratuitement depuis le site du Géoportail. L'ancienneté de ces missions varie selon la zone géographique étudiée, la plus ancienne datant de 1930. Ces missions les plus anciennes sont toutefois très limitées et la qualité de prise de vue de certaines d'entre elles rendent parfois difficiles leur utilisation.

Pour couvrir notre zone d'étude, nous avons sélectionné quatre missions aériennes distinctes à deux périodes différentes.

4.4.1 LES MISSIONS AERIENNES DE 1953-1954

Les missions du 24 juillet 1953 et du 12 juin 1954 sont complémentaires et couvrent chacune une région distincte de notre zone d'étude. Nous considérerons dans la suite des travaux que ces photographies permettent d'obtenir des informations à une seule et même date. Elles ont été acquises au 1/25 000^e et couvrent respectivement une superficie de 4900x4900m. Il s'agit de clichés argentiques numérisés par l'IGN dont la résolution spatiale est de 0,9 mètres (Tableau 4-6, Figure 4-14). Au total, 43 clichés ont été nécessaires pour couvrir l'ensemble de notre zone d'étude.

MISSION	DATE	ECHELLE DE PRISE DE VUE	RESOLUTION	EMPRISE (1 CLICHE)	NB DE CLICHES SELECTIONNES	FORMAT
C1845-0021-F1845-2045	24/07/1953	1/25 000 ^e	0.9 m	4 900*4 900m	34	jpeg2000
C1744-0021-F1744-2144	12/06/1954	1/25 000 ^e	0.9 m	4 900*4 900m	9	jpeg2000

Tableau 4-6. Caractéristiques des missions retenues pour les années 1953-1954

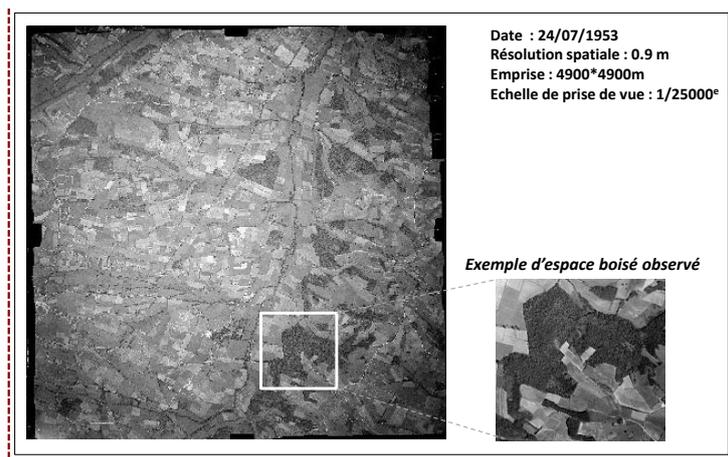


Figure 4-14. Exemple de cliché aérien issu de la mission aérienne de Juillet 1953

4.4.2 LES MISSIONS AERIENNES DE 1979

Les deux missions aériennes de 1979 sélectionnées sont répertoriées comme deux missions distinctes mais ont été pilotée le même jour à savoir, le 18 juin 1979 (Tableau 4-7). Durant ces missions, les clichés ont été photographiés au 1/29 000^e et ont été numérisées par l'IGN à une résolution de 0,6 mètres (Figure 4-15).

MISSION	DATE	ECHELLE DE PRISE DE VUE	RESOLUTION	EMPRISE (1 CLICHE)	NB DE CLICHES SELECTIONNES	FORMAT
C2445-0011-F1745-2445	18/06/1979	1/29 000 ^e	0.6 m	6 500*6 500m	20	jpeg2000
C2444-0051-F1744-2444	18/06/1979	1/29 000 ^e	0.6 m	6 500*6 500m	6	jpeg2000

Tableau 4-7. Caractéristiques des missions aériennes retenues pour l'année 1979

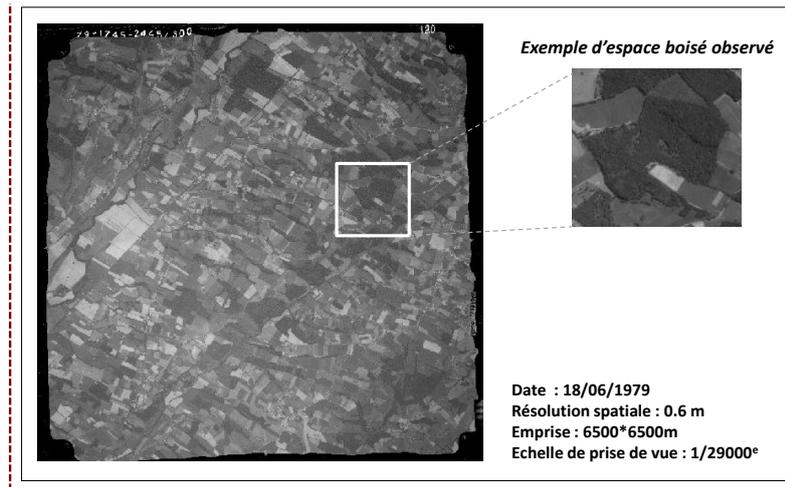


Figure 4-15. Exemple de cliché aérien issu des missions aériennes de Juin 1979

4.5. LA BD ORTHO® INFRAROUGE 2010 DE L'IGN

Pour la période la plus récente, nous avons utilisé la dernière version de la BD Ortho® de l'IGN disponible au moment du démarrage de notre travail, soit la version de 2010. La BD Ortho® fournit des photographies aériennes numériques en couleur, orthorectifiées (i.e. corrigées des déformations liées au relief). Etant donné que nous nous intéressons à la végétation (les forêts), nous avons opté pour les images en infra-rouge couleur plus adaptées. Elles sont constituées de trois canaux couvrant les longueurs d'onde du Vert (V), Rouge (R) et Proche Infra-rouge (PIR). La végétation présentant une forte réflectance dans le PIR, cette version est préférée aux images en couleurs naturelles (Figure 4-16).

Cette base de données permet aussi le suivi des éléments fins puisque la BD Ortho dispose d'une résolution spatiale de 0.5 mètre en couleur (mode multispectral) et de 1 mètre en noir et blanc (mode panchromatique).

Comme notre site d'étude se situe à cheval sur plusieurs départements, trois bases ont été retenues : (1) La Haute Garonne, (2) Le Gers et (3) les Hautes Pyrénées. Au total, 138 dalles ont été sélectionnées au sein de ces trois bases de données couvrant chacune une surface de 5000 m×5000m.



Figure 4-16. Exemple d'une dalle issue du produit BD ortho® IRC

4.6. BILAN : DES DONNEES DIFFERENTES QUI RESTENT COMPATIBLES

Nous avons finalement retenu cinq sources distinctes permettant de disposer de données spatiales forestières à cinq dates différentes sur une période de temps équivalente à un siècle et demie (Figure 4-17).

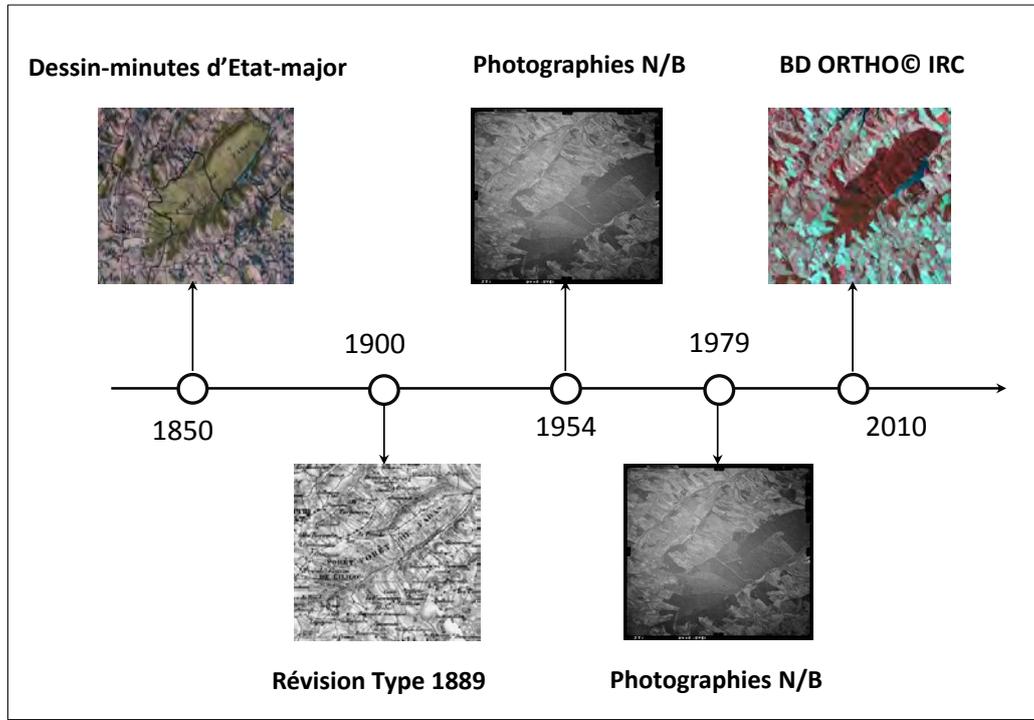


Figure 4-17. Récapitulatif des sources de données spatiales retenues dans le cadre de cette étude

Le suivi des dynamiques paysagères sur le temps long implique d'utiliser des données spatiales hétérogènes aux caractéristiques d'échelle, de représentation ou de qualité différentes. Dans notre cas, les différences observées entre sources restent acceptables et permettent de répondre à nos objectifs d'étude.

- Des différences d'échelle

Les données spatiales sélectionnées disposent d'une échelle située dans une gamme de valeur relativement réduite à savoir du 1/25 000^e au 1/40 000^e. De fait, comme nous avons pour objectif de retracer les évolutions spatio-temporelles des forêts et donc de comparer les informations issues de ces sources, ces différences d'échelles restent compatibles avec nos objectifs.

- Les différences de généralisation et de représentation

Une des différences observées concerne le type de représentation. Dans un cas (pour 1850 et 1900), nous avons retenu des cartes dont le contenu dépend d'un processus d'abstraction du monde réel et que les photographies aériennes sélectionnées pour les dates plus récentes permettent justement de s'affranchir de ce processus. Toutefois, malgré une diversité de représentation importante ou la présence de contours flous, le type de représentation adoptée dans les cartes sélectionnées permet aisément de différencier les forêts des autres éléments inscrits. Cela ne constitue donc pas un problème en soi pour les identifier.

Ensuite, il existe des différences entre cartes. S'il est difficile de connaître les règles précises de généralisation de ces documents, nous avons observé des différences de représentation, notamment

en ce qui concerne la taille minimale des objets (voir chapitre 6, section 6.4.3) ou leur forme (Figure 4-18). Les photographies aériennes disposent, elles, d'un niveau de détail très proche (résolution spatiale comprise entre 0.5 et 0.9 mètres) mais diffèrent d'un point de vue colorimétrique. Deux missions aériennes sont en noir et blanc tandis que les clichés de la BD Ortho® sont en couleur.

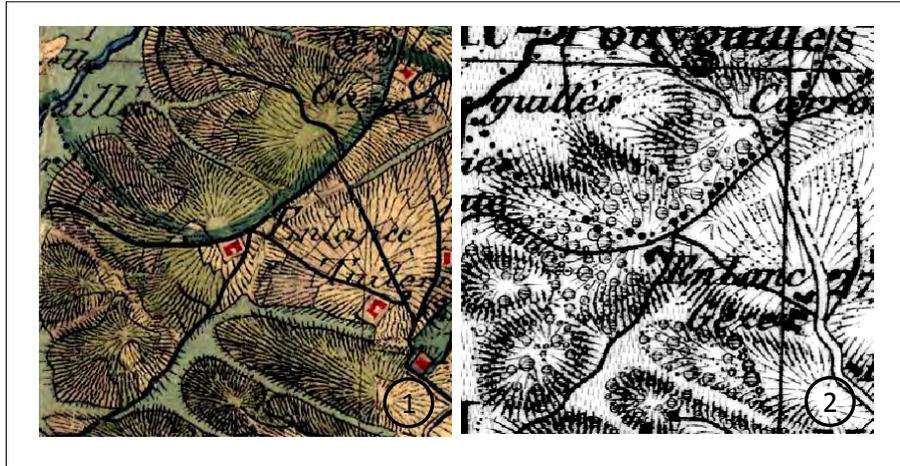


Figure 4-18. Exemple de la simplification de la forme d'une forêt causée par la présence de poncifs et l'absence de délimitation claire dans la Révision Type 1889. 1) Dessin-Minute d'Etat-Major, 2) Révision Type 1889

- Des différences de qualité

Les cartes sont concernées par des différences de qualité sémantique. Il est difficile de savoir si la représentation est conforme aux terrains, il peut donc y avoir potentiellement des erreurs de confusions sur la présence ou l'absence d'une forêt

Les sources sélectionnées sont également toutes concernées par des différences de qualité géométriques. Les différentes techniques utilisées pour lever les cartes ou bien capturer les clichés vont obligatoirement induire des degrés de précision variables tant du point de vue de la position des objets que dans la définition de leur contour. Cet aspect sera étudié par la suite (chapitre 5), dans la phase de correction géométrique des données spatiales visant à les intégrer dans un référentiel cartographique actuel et ainsi permettre leur comparaison.

CHAPITRE 5.METHODE DE CORRECTION GEOMETRIQUE DES CARTES ANCIENNES

SOMMAIRE :

- 5.1. RECALER, CORRIGER : UNE NECESSITE
- 5.2. PRINCIPE GENERAL DE LA CORRECTION GEOMETRIQUE
- 5.3. METHODES DE CORRECTION GEOMETRIQUE EXISTANTES
- 5.3. METHODES DE CORRECTION SPECIFIQUES AUX CARTES ANCIENNES
- 5.4. PROPOSITION D'UNE METHODE DE CORRECTION ALTERNATIVE : LA REGRESSION A NOYAU
- 5.5. SIMULER POUR CHOISIR UNE STRATEGIE OPTIMALE DE CORRECTION DES CARTES ANCIENNES
- 5.6. SIMULER POUR AIDER A DECIDER DANS UN CONTEXTE OPERATIONNEL

Ce chapitre présente la méthodologie développée pour corriger géométriquement les données spatiales anciennes. Après avoir décrit les différents déterminants de la qualité de la correction géométrique et méthodes existantes, nous présenterons une méthode alternative de correction basée sur la régression à noyau. Les résultats de cette approche seront comparés à ceux des méthodes conventionnelles puis discutés à la fin de ce chapitre.

5.1. CORRIGER, RECALER : UNE NECESSITE

La correction et le recalage géométrique des données spatiales sont des étapes indispensables à leur utilisation. Elles sont le plus souvent menées en amont de la chaîne de traitement dans la phase dite de « prétraitements des données ».

La correction géométrique est le processus de transformation d'un jeu de données spatiale vers un autre référentiel considéré comme la référence (plus juste). On parle alors de jeu de données à corriger et de jeu de données de référence (Billen et al. 1999).

Le recalage géométrique est l'étape qui assure le passage d'un jeu de données dans un référentiel initial vers un autre référentiel qui est celui du jeu de données cible. Au sens général, le jeu de données cible n'est pas obligatoirement la référence mais le choix d'adopter son référentiel vise à rendre compatible/superposable les données.

Dans cette thèse, nous parlerons uniquement de correction géométrique des données spatiales. Les cartes historiques doivent être intégrées dans un nouveau référentiel adapté permettant leur comparaison avec des données spatiales plus récentes. Elles seront ainsi transformées sur la base d'un jeu de données référence

Les champs d'application faisant intervenir le processus de correction géométrique sont nombreux compte tenu du caractère indispensable de cette tâche pour mesurer un phénomène à partir des données spatiales. Les données spatiales anciennes sont évidemment concernées par cette étape étant donné qu'elles ont été levées dans un système défini à l'époque de leur production (Leyk 2005; Jenny 2006; Jenny et Hurni 2011). En télédétection, les corrections géométriques consistent à intégrer les images brutes vers un système cartographique local ou global. Le processus de correction est alors calibré à partir des coordonnées pixellaires de l'image (Fonseca et Manjunath 1996; Zitova et Flusser 2003). Enfin, les bases de données vectorielles font aussi l'objet de corrections géométriques qu'il s'agit de les replacer dans un nouveau système de coordonnées pour les comparer à d'autres bases ou effectuer des mises à jour (Ubeda et Egenhofer 1997; Servigne et al. 2000). On s'appuie directement sur les coordonnées des objets pour les corriger.

5.2. PRINCIPE GENERAL DE LA CORRECTION GEOMETRIQUE

La problématique de la correction géométrique peut être réduite à déterminer la transformation spatiale permettant le passage d'un système de coordonnées à un système de référence (Billen et al. 1999). Les paramètres de la transformation sont calculés à partir de points d'amer (ou Ground Control Points = GCP) qui sont des points connus dans la BD à corriger et dans la BD référence (Figure 5-1).

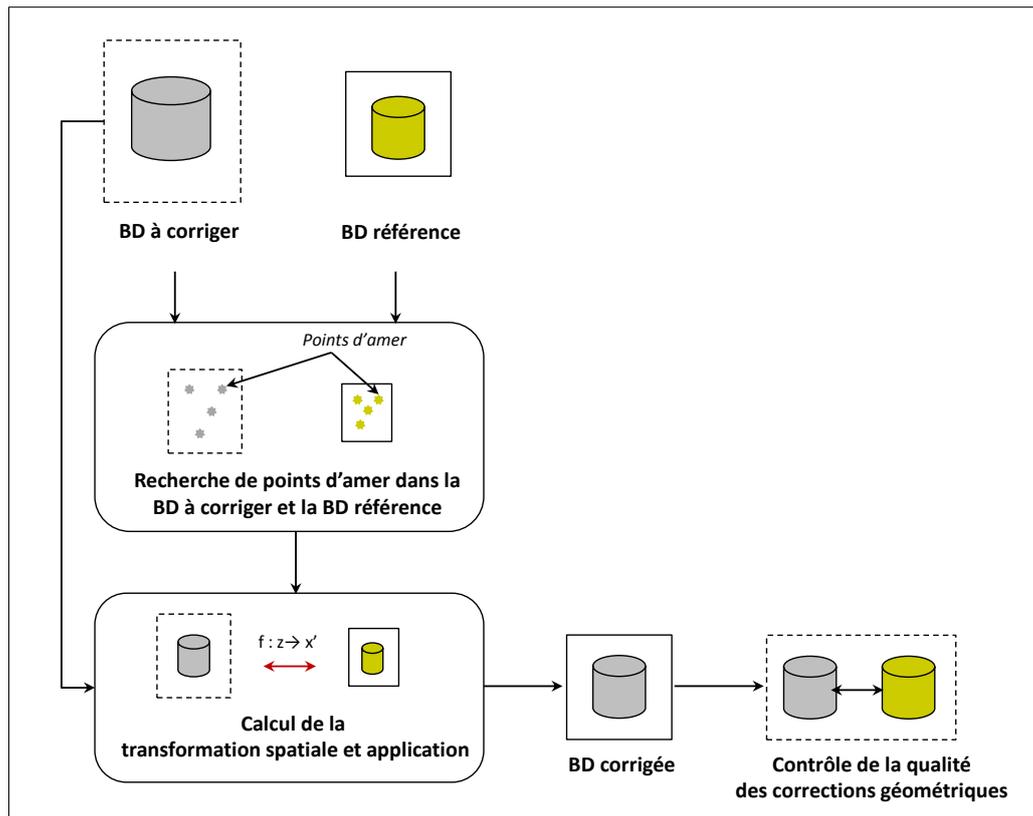


Figure 5-1. Principe de la correction géométrique des données spatiales

Le processus de correction géométrique se décompose en trois grandes étapes :

- (1) La recherche de points d'amer dans la BD référence et la BD à transformer. Ces points peuvent s'apparenter à des limites de régions, des intersections ou des points remarquables. On cherche le plus souvent à s'appuyer sur des objets stables d'un point de vue temporel pour ne pas biaiser le calcul de la transformation.
- (2) Le calcul et l'application de la fonction de transformation. La fonction de transformation est déterminée à partir des couples de points d'amer. A noter qu'il existe plusieurs type de transformation, certaines étant plus adaptées que d'autres au type de déformations rencontrées.
- (3) Le contrôle de la qualité des corrections qui consiste à comparer la BD corrigée à la BD référence pour en mesurer les écarts.

Ainsi, les travaux concernant les corrections géométriques portent sur divers aspects de la procédure incluant notamment l'impact du nombre de points d'amer, leur répartition spatiale, leur précision géométrique, la référence spatiale utilisée ainsi que la fonction mathématique choisie pour calculer la transformation géométrique (Ford et al. 1978; Orti 1981; Labovitz et Marvin 1986; Mather 1995; Buiten et Van Putten 1997).

Concernant le nombre de points d'amer (ou *Ground Control Points* = GCPs), Ford et al. (1978) ont montré que l'erreur quadratique moyenne de la position des GCPs était inversement proportionnelle au nombre de points sélectionnés après le calcul de plusieurs transformations polynomiales. La qualité de la transformation peut par conséquent être améliorée en augmentant le nombre de GCPs (Orti 1981; Moré et Pons 2012). Toutefois, d'autres effets doivent être pris en considération. Il a notamment été montré que la distribution spatiale des GCPs avait aussi une influence sur la qualité d'ajustement du modèle de transformation. Selon (Mather 1995) le nombre

optimal de GCPs dépend de leur patron de répartition spatiale. Ce dernier auteur préconise une répartition aléatoire des GCPs pour augmenter la qualité de la transformation. (Zhang et Cheng 2009) indiquent aussi que la qualité des corrections géométriques est plus sensible à la distribution spatiale des GCPs qu'à leur effectif (Zhang et Cheng 2009). En général, les auteurs ne sélectionnent pas leurs points d'amers de manière aléatoire (notamment pour contrôler la nature des points choisis) mais tendent vers une répartition spatiale uniforme pour optimiser la qualité des corrections géométriques (Ford et Zanelli 1985; Orun et Natarajan 1994; Zhang et Cheng 2009; Yang et Jiao 2011). Cette uniformité est souvent estimée de manière subjective (Zhang et al. 2006) mais peut aussi être conduite ou contrôlée à l'aide d'une méthode d'échantillonnage adaptée, en employant par exemple une tessellation de Voronoi ou des indices d'uniformité (Chen et Lee 1992; Zhang et Cheng 2009). Plus récemment, certains auteurs ont proposé des techniques d'optimisation de l'échantillonnage spatial pour optimiser la distribution des GCPs sans avoir à fixer le patron de distribution a priori. Wang et al. (2012) ont ainsi démontré l'efficacité de l'échantillonnage spatial basé sur le krigeage universel en comparant des échantillonnages aléatoires et uniformes.

En plus du nombre de GCPs et de leur répartition spatiale, l'exactitude géométrique de leur position influence aussi la qualité des corrections effectuées (Shlien 1979; Mather 1995; Billen et al. 1999). La qualité de positionnement absolu dépend de la qualité des points d'amer sélectionnés (ex : croisements de routes, coins de bâtiments, etc..) et de la qualité du jeu de référence (ex : relevés terrains au GPS, cartes topographiques, image). Le choix du jeu de référence doit naturellement être en adéquation avec le document à corriger et celui-ci est supposé être de plus grande qualité. Schlien (1979) a montré que les erreurs de correction d'une image Landsat MSS (Multi Spectral Sensor) à 30m de résolution spatiale augmentaient de 25% quand les GCP étaient repérés sur une carte au 1/250 000^e plutôt que sur une carte au 1/50 000^e. L'effet du jeu de référence peut donc être important. Mather (1995) a aussi démontré ces effets significatifs sur la qualité de correction d'une image panchromatique SPOT-1 en simulant des décalages de 0.25 à 0.5 pixels par rapport à la position actuelle des GCPs.

Le bilan de ces travaux montre que la problématique de la correction géométrique des données spatiale est bien connue ainsi que ses principaux déterminants. En plus de la méthode de transformation, le nombre, la répartition spatiale et l'exactitude géométrique des points d'amers seront les facteurs principaux dont nous devons tenir compte au moment de définir notre stratégie de correction.

5.3. METHODES DE CORRECTION GEOMETRIQUE EXISTANTES

Deux approches peuvent être utilisées pour corriger géométriquement une image ou une carte (Balletti 2006; Boutoura et Livieratos 2006). La première dite **globale** consiste à appliquer un seul et même modèle de transformation géométrique à l'ensemble du jeu de données à corriger dans le but de faire correspondre les objets et leurs homologues dans la carte de référence (Sprague et al. 2007). La seconde approche dite **locale** applique un modèle de transformation différent à chaque partition du jeu de données (White Jr et Griffin 1985; Fuse 1998). Le découpage du jeu de données peut se faire de différentes façons mais en général, il est réalisé par une triangulation de Delaunay¹. Nous détaillons ces deux approches ci-dessous.

¹ Pour rappel, un triangle au sens de Delaunay est tel qu'aucun point P (ici, aucun point d'amer) ne se situe à l'intérieur du cercle circonscrit du triangle formé. Elle présente la particularité de créer des triangles dont les angles ne sont pas supérieurs à 90° (Caloz et al. 2001)

5.3.1 METHODES DITES GLOBALES

Les approches les plus couramment utilisées sont les approches globales basées sur l'application d'une *transformation polynomiale* en cherchant à minimiser les écarts aux points de référence selon la méthode des moindres carrés.

Pour la suite, $z = [x, y]$ représente les coordonnées de l'image à transformer, $z' = [x', y']$ les coordonnées de la source de référence, et x et y les projections horizontales et verticales du système de coordonnées. Le jeu de points d'amers (GCP), noté S est défini tel que $(z_i, z'_i), i = 1, \dots, n$, où n est le nombre de GCPs. Par souci de clarté, seule la transformation de x' est discutée ici mais des conclusions identiques peuvent être tirées pour y' .

L'objectif de la correction géométrique basée sur la transformation polynomiale est de trouver la fonction $f: z \rightarrow x'$ qui minimise l'erreur moyenne quadratique sur S :

$$\hat{f} = \min_{f \in F} \left[\frac{1}{n} \sum_{i=1}^n (x'_i - f(z_i))^2 \right] \quad (1)$$

Où F est un jeu de fonctions polynomiales f d'ordre p défini par l'utilisateur. Si $p = 1$, nous sommes dans le cadre d'une transformation affine. L'augmentation de l'ordre du polynôme (e.g. $p = 2$ ou 3) tend à complexifier la transformation entraînant alors des déformations beaucoup plus importantes de l'image à corriger (Figure 5.2).

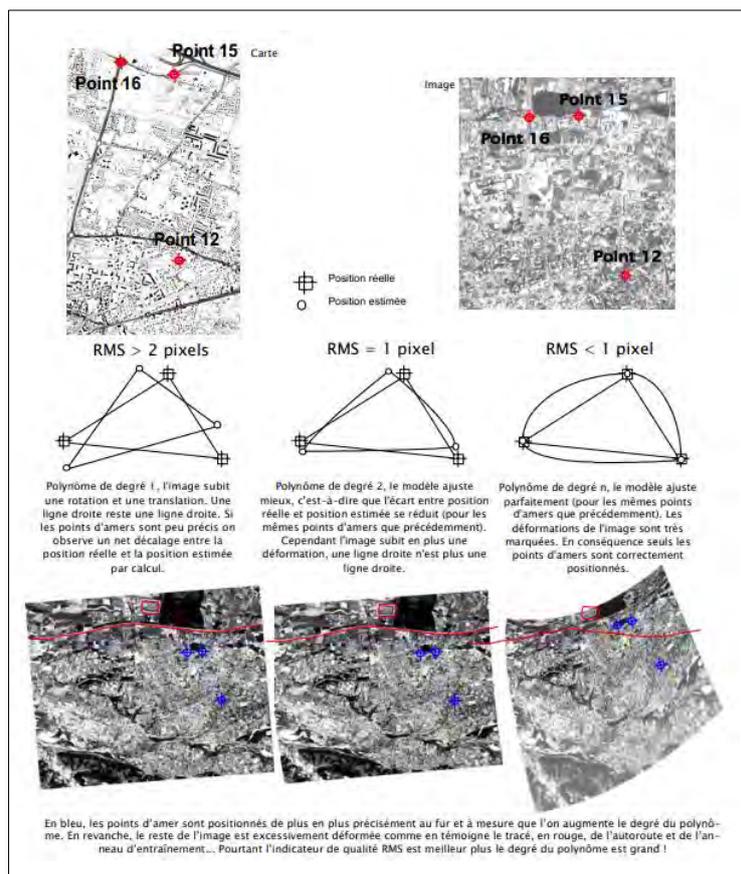


Figure 5-2. Conséquences de l'augmentation de l'ordre du polynôme sur l'image à corriger (source : Laffly, 2005). Le modèle ajuste de mieux en mieux les points d'amer mais entraîne des déformations de plus en plus marquées.

Afin de poursuivre la démonstration, prenons maintenant l'exemple de la transformation polynomiale d'ordre 2 ($p = 2$)

Si $n = 2$, la fonction de transformation $f: z \rightarrow x'$ peut être écrite telle que :

$$f(z) = w_0 + w_1x + w_2y + w_3xy + w_4x^2 + w_5y^2 \quad (2)$$

Les coefficients w_j sont trouvés en résolvant l'équation 1. En notant $\mathbf{w} = [w_0, \dots, w_5]^t$ et $(z) = [1, x, y, xy, x^2, y^2]^t$, l'équation 1 peut s'écrire comme ci-après:

$$\frac{1}{n} \sum_{i=1}^n (x'_i - \langle \mathbf{w}, \boldsymbol{\phi}(z_i) \rangle_{\mathbb{R}^6})^2 \quad (3)$$

De manière équivalente, on obtient l'équation 1 sous la forme matricielle suivante :

$$\frac{1}{n} \|\mathbf{x}' - \boldsymbol{\Phi}(z)^t \mathbf{w}\|^2 \quad (4)$$

Où \mathbf{x}' est le vecteur $[x'_1, \dots, x'_n]^t$ et $\boldsymbol{\Phi}(z)$ est la matrice des échantillons (les couples de points d'amers) définie comme $\boldsymbol{\Phi}(z) = [\phi(z_1), \dots, \phi(z_n)]$. Finalement, la solution optimale est trouvée en dérivant l'équation 4 et en annulant la dérivée :

$$\hat{\mathbf{w}} = (\boldsymbol{\Phi}(z)\boldsymbol{\Phi}(z)^t)^{-1}\boldsymbol{\Phi}(z)\mathbf{x}' \quad (5)$$

5.3.2 METHODES DITES LOCALES

Les méthodes locales sont aussi fondées, le plus souvent, sur les transformations polynomiales. La différence est qu'il n'existe pas qu'un seul modèle polynomial utilisé pour corriger l'ensemble de la zone à partir de l'ensemble des GCPs (i.e. une seule surface de tendance), mais autant de modèles qu'il y a de sous-ensembles de GCPs qui définissent chacun une partition de l'espace. Ainsi, quand la partition est basée sur une triangulation de Delaunay (cas usuel), un modèle polynomial de 1^{er} ordre est généralement estimé pour chaque triangle en s'appuyant sur les coordonnées des GCPs définissant les sommets (Saalfeld 1985; White Jr et Griffin 1985)(Figure 5-3). D'autres auteurs ont également proposé d'appliquer des transformations polynomiales d'un ordre plus élevé pour prendre en compte des déformations plus fortes (Akima 1978; Goshtasby 1986; Shimizu et Fuse 2003). Chaque modèle de transformation s'applique alors uniquement aux objets inclus dans le triangle correspondant et on suppose une variation linéaire à l'intérieur de chaque triangle (cf. section suivante).

Notons que la triangulation n'est pas la seule façon de partitionner l'espace pour opérer des transformations géométriques locales. On peut très bien procéder à un pavage de l'espace en utilisant par exemple une grille régulière de maille carrée. Néanmoins, cette façon de procéder est moins habituelle. Pour une revue complète des différentes méthodes de partition de l'espace et d'approximation locales, on pourra se reporter aux références suivantes (Goshtasby 1986; Goshtasby 1987).

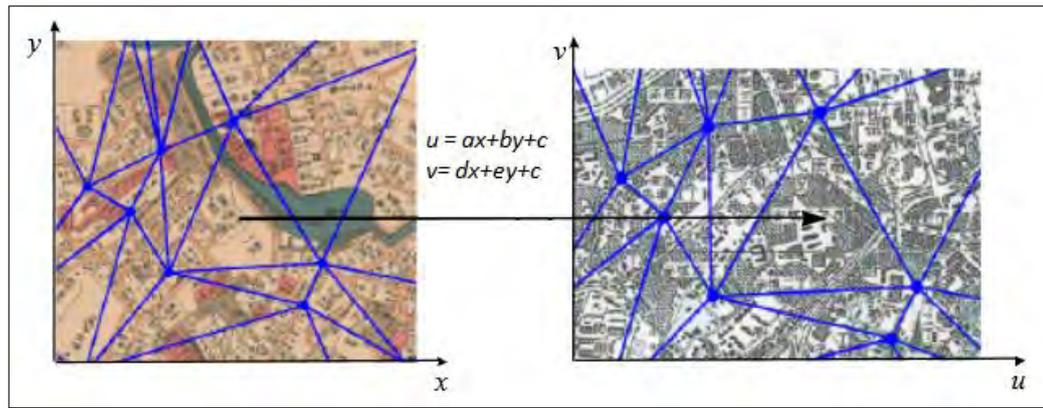


Figure 5-3. Transformation géométrique basée sur une partition de Delaunay et l'application d'une fonction polynomiale d'ordre 1 dans chacune des partitions (Shimizu et Fuse 2003)

5.4. METHODES DE CORRECTION SPECIFIQUES AUX CARTES ANCIENNES

Si le problème des corrections géométriques est déjà très ancien (Ford et al. 1978; Orti 1981; Ford et Zanelli 1985), la récente mise à disposition des données spatiales historiques a remis cette question au centre des considérations (Wieczorek et al. 2004; Sprague et al. 2007; Baily et al. 2011).

De manière générale, les difficultés rencontrées dans la correction géométrique des cartes anciennes résident principalement dans la prise en compte des déformations locales qui sont souvent très variables d'un endroit à l'autre de l'espace et non linéaires (Jenny 2006; Jenny et Hurni 2011). Ces dernières sont causées par divers paramètres comme la précision limitée des instruments de mesure de l'époque, une hétérogénéité des pratiques de relevés des ingénieurs cartographes, un manque de formalisation des spécifications cartographiques, une qualité limitée des systèmes géodésiques anciens, et une détérioration des cartes avec le temps (Leyk 2005). Les modèles globaux, qui représentent correctement la tendance générale des déformations s'avèrent alors moins adaptés à la prise en compte des déformations locales, même en augmentant le degré du polynôme. C'est la raison pour laquelle, les approches locales ou les transformations élastiques sont préférées aux approches globales pour corriger la géométrie des cartes anciennes (Fuse 1998; Shimizu et Fuse 2003; Balletti 2006).

5.4.1 CHOIX DES POINTS D'AMER

Pour la correction des cartes anciennes, il est souvent préconisé de récolter des points de contrôle dits *remarquables* sur les cartes anciennes qui peuvent être retrouvées sur les données spatiales actuelles, en l'occurrence des objets stables dans le temps (Benavides et Koster 2006; Balletti et Guerra 2009). Dupouey et al. (2007) proposent par exemple de s'appuyer sur des moulins, des clochers, des manoirs et des châteaux pour la correction des minutes d'Etat-major au 1/40 000^e. Si ces points sont a priori plus fiables car ayant servi aux relevés topographiques de l'époque, ils sont souvent irrégulièrement répartis sur la carte, présents en nombre limité, et plus éloignés des milieux naturels ce qui peut limiter la qualité de correction de ces objets et notamment celle des forêts qui nous préoccupent. De plus, les changements paysagers importants qui interviennent durant la période qui sépare la date de production de la carte ancienne de celle de la source de référence sont souvent importants et ne permettent pas toujours de constituer un jeu de points d'amer fiable et en nombre suffisant.

5.4.2 CHOIX D'UN MODELE DE TRANSFORMATION

La qualité de correction géométrique des cartes anciennes dépend notamment du choix de la fonction de transformation. Une revue complète de ces fonctions est donnée par (Boutoura et Livieratos 1986; Boutoura et Livieratos 2006). Il peut s'agir, dans le cas des approches globales, d'une transformation de type Helmert (Bitelli et al. 2009), d'une transformation affine (soit un polynôme d'ordre 1) (Baiocchi et Lelo 2005), de transformations polynomiales d'ordre plus élevé (Bitelli et Gatta 2011) ou encore, d'une transformation projective (Balletti 2006). Concernant les approches locales, de nombreux auteurs se sont appuyés sur une triangulation de Delaunay (Akima 1978; Fuse 1998; Shimizu et Fuse 2003; Balletti 2006; Dupouey et al. 2007; Balletti et Guerra 2009). D'autres procèdent à une transformation élastique lorsque les déformations observées sont locales et non-linéaires (Balletti 2006).

La plupart des études ayant pour objectif de recalculer géométriquement les données spatiales anciennes s'appuient sur des méthodes locales et le recalage élastique pour les raisons exposées précédemment (Akima 1978; Fuse 1998; Shimizu et Fuse 2003; Balletti 2006; Dupouey et al. 2007; Balletti et Guerra 2009).

Au regard des propos tenus précédemment, le principal avantage de cette approche concerne sa capacité à appliquer une multitude de modèles différents qui s'adaptent mieux aux déformations locales différentes aux multiples endroits de la carte.

Elle comporte également certaines limites qui portent sur deux points essentiels :

- Le besoin d'un nombre relativement élevé de points d'amer pour tirer les bénéfices de l'approche locale. Plus le nombre de partitions de l'espace est élevé et mieux les déformations locales seront prises en compte. Cela implique toutefois de sélectionner un grand nombre de points d'amer ce qui peut s'avérer difficile et gourmand en temps.

- La possibilité de déformer de différentes façons un même objet si celui-ci se situe à cheval entre deux partitions (i.e. deux triangles). La figure 5.4 illustre les conséquences d'une procédure de correction partitionnée sur la structure des objets inscrits dans une carte historique de la période Edo (Shimizu et Fuse 2003). L'image de gauche désigne l'image originale avant déformation: on constate que les rues qui traversent les différents triangles sont rectilignes. L'image au centre représente l'image transformée grâce à l'application d'une transformation affine dans les triangles définis. Les cercles dessinés par les auteurs permettent de constater que la rectitude des rues n'a pas été conservée puisque corrigées par deux modèles distincts. L'image de droite illustre les déformations engendrées par l'application d'un polynôme d'ordre 5 au sein des triangles. Les objets ont été grossièrement déformés au sein même des partitions définies mais les déformations entre partitions sont toutefois plus douces.

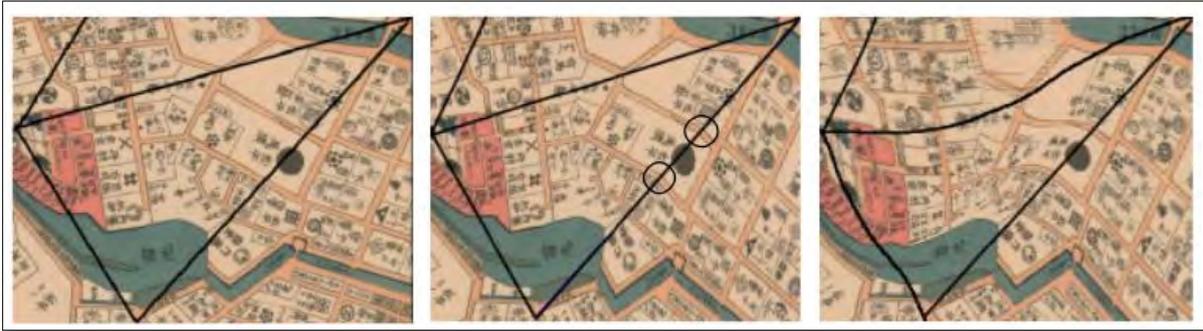


Figure 5-4. Exemples de déformations introduites par le modèle de Delaunay : l'image de droite est l'image originale, l'image au centre est l'image transformée par l'application d'une fonction polynomiale d'ordre 1 et l'image de droite est l'image transformée par l'application d'une fonction polynomiale d'ordre 5 (source : Shimizu et Fuse 2003)

5.4.3 BILAN

Bien qu'un nombre relativement important de travaux existe sur la correction géométrique des cartes anciennes, plusieurs questions subsistent à l'issue de l'examen de la littérature.

La première concerne le nombre de points d'amer. Combien de points sont nécessaires pour corriger efficacement une carte ancienne ? Il n'y a pas de consensus ou de recommandations concernant le nombre optimal de points pour corriger ces documents sachant qu'il est nécessaire de trouver un compromis entre cet effectif et le temps passé à rechercher les points d'amer. Ce compromis dépend fortement de la carte et de sa qualité qui vont conditionner la disponibilité de points d'amer sur la zone d'intérêt. Deuxièmement, on peut se poser la question de savoir quels sont les points d'amer à privilégier ? Faut-il se limiter aux points de relevés tels que les clochers et les moulins jugés a priori plus fiables mais en nombre limité et avec une répartition spatiale qui n'est pas toujours uniforme ? Peut-on s'appuyer sur d'autres catégories de points tels que les croisements de route ou les bâtiments a priori moins fiables car souvent généralisés ? L'effectif doit-il alors être plus important pour atteindre le même niveau de correction que ce qui est obtenu avec les points de relevés ? De manière plus générale, que faut-il privilégier ? La qualité des points d'amer, leur effectif ou leur distribution spatiale ? Concernant cette distribution spatiale, est-il préférable d'adopter une répartition aléatoire ou régulière ? Les réponses à ces questions sont-elles identiques quel que soit le type de modèle de transformation ?

Une autre question importante concerne l'approche de correction géométrique. Les modèles de transformation utilisés varient d'un auteur à l'autre en retenant tantôt l'approche globale, tantôt l'approche locale. Chacune des deux approches présentent des avantages et des inconvénients déjà évoqués. Dans ce cas, ne serait-il pas possible de définir une approche intermédiaire ? Est-il possible de définir une méthode globale qui prendrait mieux en compte les déformations locales ?

Nous proposons de répondre à ces différentes questions dans les sections suivantes.

5.5. PROPOSITION D'UNE METHODE DE CORRECTION ALTERNATIVE : LA REGRESSION A NOYAU

Au vu du bilan effectué dans la section précédente, nous proposons ici une nouvelle méthode de correction géométrique, adaptée aux cartes historiques, basée sur la régression à noyau (Shawe-Taylor et Cristianini 2004). Cette méthode utilise l'approche globale, c'est-à-dire en prenant en compte tous les points d'amer pour calculer la fonction de transformation mais peut également

considérer les déformations locales qui peuvent être différentes d'un endroit à l'autre (cf. noyau gaussien).

La régression à noyau est une méthode de régression non-paramétrique de plus en plus utilisées pour différentes applications (en bio-informatique, géostatistique, reconstruction 3D...). Les méthodes à noyau sont par exemple utilisées en télédétection ou en traitement d'images pour résoudre des problèmes de classification non-linéaire (Huang et al. 2002; Takeda et al. 2007; Fauvel et al. 2013). Elles commencent aussi à être utilisées pour prédire les données manquantes dans les séries temporelles de données spatiales, notamment celles qui visent à étudier les flux de transport (Haworth et al. 2014). Elles ont également été employées spécifiquement pour la correction d'images multimodales par (Ardizzone et al. 2009) ainsi que dans le cadre d'une série d'image vidéo pour le suivi de la criminalité par (Liu et al. 2011)). Les résultats montrent que la régression à noyau peut s'avérer très efficace lorsqu'il existe des problèmes liés à la super-résolution des images impliquant de fait des distorsions locales en forte proportion.

Les méthodes à noyau n'ont jamais été utilisées pour corriger géométriquement des cartes historiques alors que ces documents sont particulièrement entachés par les déformations locales. Elles sont donc potentiellement une alternative crédible aux problèmes de correction géométrique des cartes anciennes (cf. section 5.5.1)

5.5.1 LA REGRESSION RIDGE A NOYAU

Dans ce travail, la régression ridge à noyau (*Kernel Ridge Regression* - KRR – en anglais) est proposée comme une nouvelle méthode de correction géométrique des cartes anciennes.

Ce type de régression permet de résoudre un problème similaire à celui résolu par la régression polynomiale vue précédemment (*polynomial regression* -PR) mais introduit deux différences majeures :

- (1) la possibilité de régulariser la fonction de transformation en pénalisant les coefficients élevés de w (régularisation ridge).
- (2) les variables d'entrée ($\phi(z)$) ne sont pas restreintes aux monômes² du système de coordonnées grâce à l'utilisation de fonctions à noyau

5.5.1.1. L'INTRODUCTION DE LA PENALISATION RIDGE DANS LA REGRESSION POLYNOMIALE

Très souvent, le nombre de coefficients à estimer dans la régression polynomiale peut être trop important au regard du nombre d'échantillons disponibles pour les estimer. Dans ce cas, ces coefficients ne peuvent être estimés de manière unique: on parle alors de système sous-déterminé car il y a potentiellement une infinité de solutions au problème d'optimisation dont la plupart sont non-régulières.

Un moyen simple de gérer ce problème est donc d'introduire un terme de régularisation dans le problème d'optimisation qui pénalise les solutions non-régulières. Une méthode courante de régularisation est la régularisation ridge qui pénalise les valeurs élevées des coefficients de la fonction de transformation. On obtient par conséquent le problème d'optimisation suivant:

$$\frac{1}{n} \|\mathbf{x}' - \Phi(z)^t \mathbf{w}\|^2 + \|\mathbf{w}\|^2 \quad (6)$$

où λ est un paramètre de pondération entre la régularisation et l'erreur quadratique.

² Pour rappel, en algèbre, un monôme est une expression comportant un seul terme

En résolvant ce problème de la même façon que la PR, on obtient la solution suivante :

$$\hat{\mathbf{w}} = (\Phi(\mathbf{z})\Phi(\mathbf{z})^t + \mathbf{I})^{-1}\Phi(\mathbf{z})\mathbf{x}' \quad (7)$$

où \mathbf{I} est la matrice identité de taille $6*6$.

5.5.1.2. L'UTILISATION DES NOYAUX

L'utilisation de fonctions noyau k permet de travailler dans d'autres espaces de représentation que les espaces polynomiaux d'ordre p . Pour cela, le produit scalaire de la régression linéaire est remplacé par l'évaluation d'une fonction noyau. Cette fonction a la propriété d'être équivalente à un produit scalaire dans un espace de représentation permettant de mieux relier les données à corriger avec les données de référence.

Par exemple, la fonction noyau permettant une régression polynomiale d'ordre 2 s'écrit sous la forme :

$$k(z_i, z_j) = (1 + \langle z_i, z_j \rangle)^2 \quad (8)$$

L'équation (8) peut être réécrit tel que :

$$(1 + \langle z_i, z_j \rangle)^2 = [1 + x_i x_j + y_i y_j]^2 = 1 + x_i^2 x_j^2 + y_i^2 y_j^2 + 2x_i x_j y_i y_j + 2x_i x_j + 2y_i y_j \quad (9)$$

On retrouve alors le produit scalaire dans \mathbb{R}^6 défini dans (3).

Par rapport à la régression polynomiale (PR), la régression à noyau (KR) inclut toutes les solutions de la PR mais étend les possibilités d'apprentissage par l'utilisation d'autres fonctions noyau.

De façon similaire à la PR, elle tente de trouver une transformation entre z et x' ($g: z \rightarrow x'$) en utilisant la fonction de la forme:

$$g(z) = \sum_{i=1}^n \alpha_i k(z_i, z) \quad (10)$$

où les α_i sont les coefficients à estimer. Ces paramètres optimaux sont trouvés en minimisant une erreur quadratique régularisée similaire à (6):

$$\hat{\alpha} = \min_{\alpha} \left[\sum_{i=1}^n (x'_i - g(z_i))^2 \right] + \lambda \|g\|^2 \quad (11)$$

On résout ce problème d'optimisation de manière similaire à XX en dérivant et annulant la dérivée. On obtient :

$$\hat{\alpha} = (\mathbf{K}^{-1} + \mathbf{I})\mathbf{x}' \quad (12)$$

où \mathbf{I} est la matrice identité de taille $n*n$.

LES TYPES DE NOYAUX :

Il existe plusieurs types de fonctions noyau dont les plus courantes sont:

- Noyaux linéaires: $k(z_i, z_j) = \langle z_i, z_j \rangle$;
- Noyaux polynomiaux: $k(z_i, z_j) = (1 + \langle z_i, z_j \rangle)^p$;

- Noyaux gaussiens: $k(z_i, z_j) = \exp\left(-\frac{\|z_i - z_j\|^2}{2\sigma^2}\right)$.

Le noyau gaussien dépend de la distance spatiale entre deux points à corriger. On dit de ce noyau qu'il est local car il n'agit que dans un voisinage spatial défini dans l'apprentissage via le paramètre σ . Pour un z donné, en dehors de ce voisinage, la valeur du noyau est nulle. On voit alors à partir de (10) que les z_i d'apprentissage hors de ce voisinage n'auront pas d'influence sur $g(z)$. Les noyaux polynomiaux sont indépendants de la distance spatiale entre deux points à corriger. Les modèles construits à partir de ces noyaux sont considérés comme globaux.

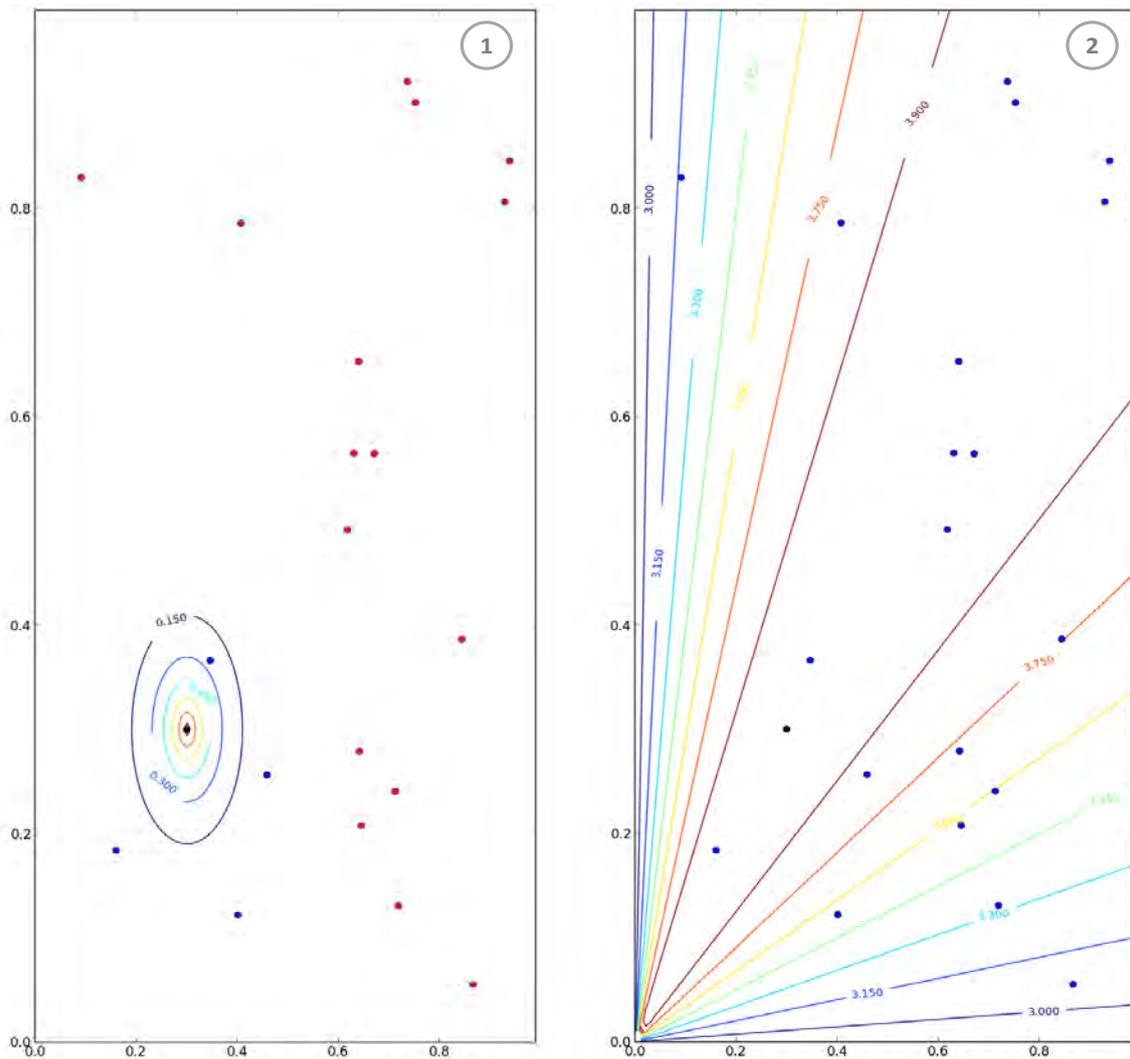


Figure 5-5. Illustration du principe du noyau gaussien (1) et du noyau polynomial (2).

Le principe de ces noyaux est explicité en figure 5-5. L'image de gauche illustre le principe du noyau gaussien. On observe que la valeur du noyau (représenté par les courbes de distance) décroît avec la distance au point à corriger noté P (point noir). Par conséquent, les points bleus sont les points du jeu d'exemple ayant un effet sur la prédiction de P. Les autres représentés en rouge sont trop éloignés et n'auront aucune incidence sur la prédiction de P. L'image de droite illustre le principe du noyau polynomial. On observe que les courbes de niveau ne sont pas fonction de la distance au point P (point noir), mais fonction de l'angle entre les points d'exemple (en bleu) et le point P.

5.6. SIMULER POUR CHOISIR UNE STRATEGIE OPTIMALE DE CORRECTION DES CARTES ANCIENNES

Dans les parties précédentes, nous avons présenté différentes méthodes de corrections géométriques classiques en discutant des avantages et des inconvénients de chacune d'elles a priori, d'après les conclusions de la littérature (sections 5.3 et 5.4). Nous avons alors proposé une approche alternative de correction fondée sur la régression ridge à noyau (section 5.5). Dans cette partie, nous confrontons les différentes approches de correction existantes ou proposée de manière à quantifier la performance de chaque méthode sur les données anciennes que nous mobilisons. De façon plus précise, nous explorons l'influence de différents paramètres de la démarche de correction géométrique selon un protocole bien établi. Cette étude se focalise sur la correction des minutes d'Etat-major au 1/40 000^e et vise à déterminer la meilleure stratégie de correction en répondant aux différentes questions laissées en suspens précédemment (section 5.4.3). Nous décrivons d'abord les données mobilisées pour l'étude avant de présenter le protocole et les résultats obtenus.

5.6.1 CARTES ANCIENNES A CORRIGER

Les cartes concernées par une étude approfondie de la procédure de correction géométrique sont les dessins minutes d'Etat-Major au 1/40 000^e. Comme indiqué précédemment (cf. chapitre 4), quatre quarts de feuille ont été sélectionnés (229 SE, 229 SO, 241 NE, 241NO) couvrant respectivement une superficie de 20 km*32km. A ce stade de notre étude, ces cartes disposent uniquement de coordonnées définies en ligne/colonnes qui ne permettent pas de les comparer aux données actuelles ou de quantifier les éléments inscrits dans ces documents. La référence choisie pour corriger géométriquement ces feuilles est la BD Ortho[®]IRC de 2010 alors intégrée dans le système de coordonnées Lambert 93. Les transformations spatiales définies consisteront par conséquent à intégrer les dessins-minutes d'Etat-Major dans ce même référentiel.

Au total, quatre types de paramètres entrent en jeu au moment de définir la stratégie de correction géométrique : (1) le choix de la méthode de transformation, (2) l'effectif de GCP, (3) la répartition spatiale des GCP, (4) la nature des GCP

(1) Le choix de la méthode de transformation

Si la majeure partie des travaux indique que les méthodes de transformations locales sont plus efficaces pour la correction des distorsions locales, ces dernières peuvent aussi conduire à des erreurs grossières de déformation. Dans cette logique, nous avons présenté une nouvelle méthode de transformation basée sur la régression à noyau.

Question(s) introduite(s) :

L'approche locale est-elle réellement plus performante que l'approche globale ?

La régression à noyau est-elle une alternative aux modèles de transformation conventionnels ?

(2) L'effectif de GCPs

Il est largement admis que les modèles de transformation géométriques sont plus efficaces dès lors que l'on augmente le nombre de GCP. Lorsqu'on utilise de surcroît une approche locale, il est souvent nécessaire de récolter un très gr et nombre de points afin d'optimiser la fonction de transformation, ce qui n'est pas toujours aisée dans le cas des cartes anciennes et très couteux en terme de temps d'une manière générale.

Question(s) introduite(s) :

Quelle que soit le type d'approche utilisé (globale ou locale), l'augmentation du nombre de GCPs conduit-il à faire baisser significativement l'erreur de correction ?

Existe-t-il un seuil à partir duquel l'augmentation du nombre de GCPs ne conduit plus à faire baisser significativement l'erreur de correction ?

(3) la répartition spatiale des GCPs

Il est souvent préconisé de sélectionner des GCPs en respectant une distribution spatiale uniforme mais celle-ci est souvent déterminée de manière subjective (Zhang et al. 2006). Néanmoins, les distorsions locales existantes dans les cartes anciennes sont liées à un certain nombre de paramètres qui conduisent à une répartition aléatoire de l'erreur sur la zone d'intérêt. Pour cette raison, nous faisons l'hypothèse qu'une répartition régulière des GCPs ne garantit pas obligatoirement une meilleure correction géométrique.

Question(s) introduite(s) :

Une répartition aléatoire des GCPs permet-elle de corriger plus efficacement les cartes anciennes qu'une répartition régulière de ces mêmes GCPs ?

(4) la nature des GCPs

Plusieurs travaux préconisent de s'appuyer sur les points remarquables stables dans le temps tels que les clochers ou les moulins ainsi que des points de levés potentiellement indiqués sur la feuille à transformer. Ces points sont toutefois présents en faible nombre et réparties non-uniformément sur la zone d'intérêt. Au contraire, des GCPs tels que les croisements de routes ou les bâtiments sont réputés moins précis mais sont généralement en plus gr et effectif, ce qui permet d'assurer une couverture quasi exclusive de la zone à transformer.

Question(s) introduite(s) :

La catégorie des GCPs sélectionnée a-t-elle un impact sur la qualité des corrections géométriques?

Dans le cas où un impact lié à ce paramètre existerait, serait-il plus judicieux de récolter un gr et nombre de points imprécis ou un plus petit nombre de points plus précis géométriquement?

5.6.2 PROTOCOLE D'EXPERIMENTATION

Nous présentons en figure 5-6 le protocole d'expérimentation mis au point pour étudier l'impact des paramètres de la correction géométrique sur la qualité des cartes anciennes. Il est composé de plusieurs étapes permettant de mesurer l'influence de : (1) la catégorie (ou nature) des points d'amer, (2) l'effectif de points d'amer, (3) la répartition spatiale des points d'amer et (4) la méthode de transformation.

Les effets de chacun des paramètres listés ont été mesurés indépendamment tout en maintenant la valeur des autres paramètres constante. Par exemple, l'impact du nombre de GCPs est mesuré en choisissant le même type de GCP pour un même type de distribution et la même méthode de transformation. Les différents paramètres ont par conséquent été évalués en évitant quelconque biais.

Ce protocole est appliqué à la feuille 241NE (voir chapitre 4), carte sélectionnée sur notre zone d'étude.

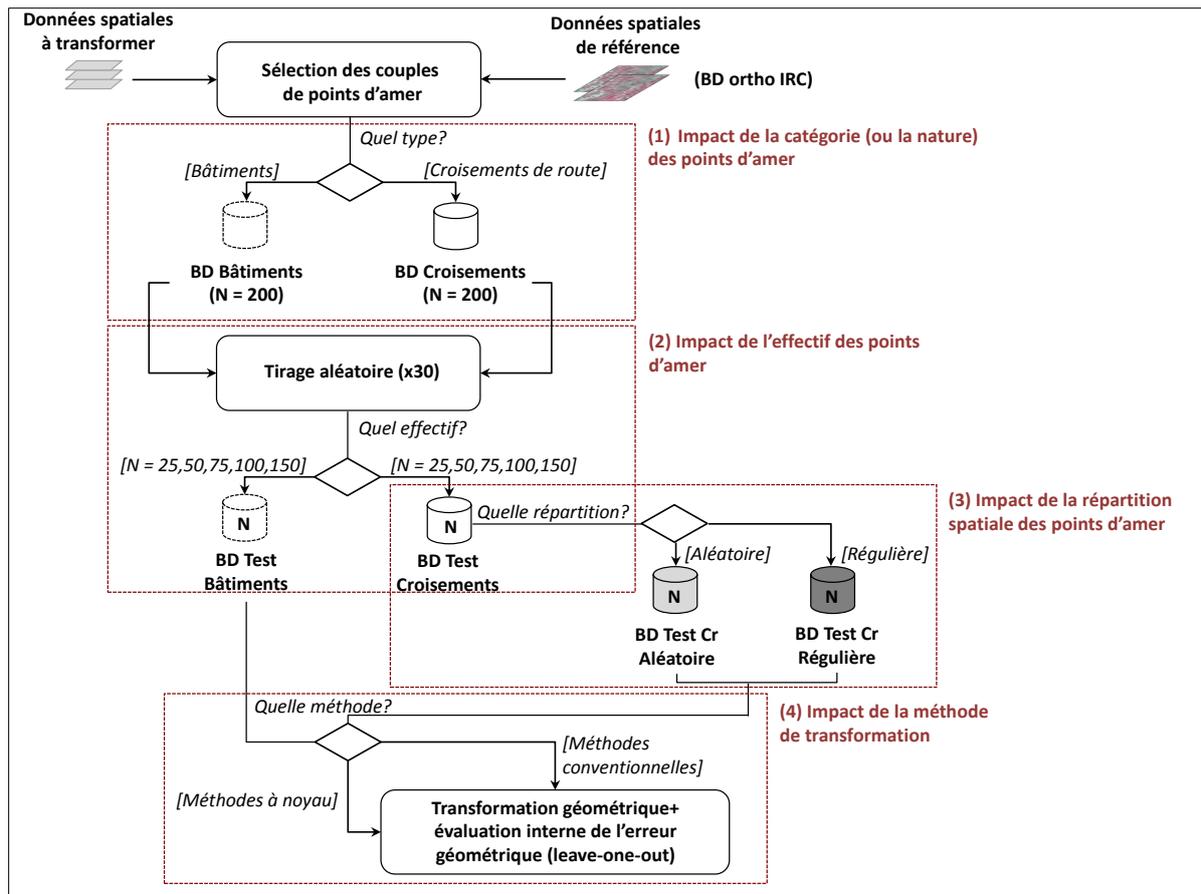


Figure 5-6. Protocole de simulation défini pour étudier l'impact des différents paramètres de la correction géométrique sur la qualité de transformation des cartes anciennes.

La première étape du protocole a consisté à sélectionner des GCPs de différentes catégories. A priori, la précision des GCPs varie selon leur appartenance à l'une ou l'autre catégorie. Pour comprendre l'effet de la catégorie de GCP sur la qualité des corrections, nous avons sélectionné des points de deux catégories différentes: les croisements de routes et les bâtiments. Ces objets sont largement présents dans l'espace et facilement identifiables pour être indépendant du jeu test utilisé. Nous avons fait l'hypothèse que les croisements de routes étaient plus fiables géométriquement que les bâtiments en raison d'une plus faible généralisation cartographique. Concernant ce choix de GCPs, nous avons volontairement exclu les bâtiments exceptionnels tels que les moulins, les églises ou les châteaux. En fait, comme déjà évoqué en section 5.4.1, ces objets disposent d'un statut particulier sur la carte puisqu'ils ont servi de points de repères aux levés topographiques effectués pour réaliser la carte. Leur position est par conséquent plus précise a priori que les bâtiments classiques et devrait faire l'objet d'une catégorie à part entière. Néanmoins, en raison de leur nombre limité et leur distribution hétérogène, ces points d'amer n'ont pas été utilisés dans l'analyse.

Pour la seconde étape du protocole, plusieurs tailles d'échantillon de GCPs ont été testés (25, 50, 75, 100 et 150) sur la base de chaque jeu complet de GCPs dans la même catégorie ($n = 200$). L'effectif à 25 points constitue un seuil minimal pour la calibration du polynôme d'ordre 3 tandis que l'effectif à 150 GCPs a été choisi car au-delà, le temps passé à la saisie des points d'amer a été jugé trop important. L'échantillonnage a été effectué par tirage aléatoire en conservant systématiquement quatre GCPs aux quatre coins de la carte à corriger dans le but d'assurer la

transformation de l'emprise totale de la zone d'intérêt. Ces quatre GCPs sont identiques pour tous les effectifs testés de la même catégorie de points mais différent d'une catégorie à une autre.

La troisième étape du protocole a consisté à choisir tous les GCPs sur la carte appartenant aux deux catégories sélectionnées selon un patron de distribution spatiale prédéterminé (Figure 5-7). Dans le cas des bâtiments, seule la distribution aléatoire a pu être maintenue. Dans le cas des croisements de route, deux patrons de distribution ont été comparés : une répartition aléatoire et une répartition régulière. La distribution régulière a été obtenue à partir d'une grille composée de cellules mesurant 1,6kmx1km (au total 200 unités sur la carte) et la répartition spatiale a été systématiquement vérifiée pour chaque échantillon établi, en utilisant la valeur moyenne de l'indice du Plus Proche Voisin (Clark et Evans 1954). Ce dernier est calculé en fonction de la distance moyenne de chaque point d'amer par rapport au point d'amer le plus proche. Il est exprimé en tant que rapport de la distance moyenne observée à la distance moyenne attendue. La distance attendue est la distance moyenne entre points d'amer voisins dans une distribution aléatoire hypothétique. Si l'indice est inférieur à 1, la répartition des GCPs est jugée régulière; si l'indice est supérieur à 1, la répartition est aléatoire.

La sélection de l'échantillon est répétée 30 fois pour chaque procédure de test de façon à éviter un biais lié aux points sélectionnés. Ces répétitions permettent ainsi d'estimer la variabilité de la qualité de l'ajustement en plus de sa valeur moyenne.

La quatrième étape du protocole consiste à appliquer différents modèles de transformations géométriques pour chaque échantillon de GCP défini. La méthode de régression à noyau est comparée aux trois modèles polynomiaux conventionnels (1^{er}, 2nd et 3^e ordre) ainsi que la méthode de correction basée sur la triangulation de Delaunay (polynôme de 1^{er} ordre pour chaque triangle). Plusieurs noyaux ont aussi été comparés : les noyaux polynomiaux (1^{er}, 2nd et 3^e ordre) et les noyaux gaussiens.

Finalement, la dernière étape a consisté à déterminer la qualité d'ajustement de chaque modèle testé. L'indicateur de performance sélectionné est l'Erreur Moyenne Quadratique (EMQ ou *RMSE* pour *Root Mean Square Error* en anglais) estimée par validation croisée. La validation croisée consiste à diviser le jeu de points d'amer S au hasard en, en k parties de taille équivalente S_1, \dots, S_k puis pour chaque partie S_i variant de 1 à k :

- (1) on calcule les paramètres de la fonction de transformation à partir des exemples $S-S_i$
- (2) on mesure les écarts géométriques à partir des exemples restants S_i
- (3) l'erreur réelle (EMQ) est estimée en effectuant la moyenne des écarts mesurés successivement

Dans notre cas, nous avons adopté la technique de validation croisée *leave-one-out* (LOO) dans laquelle le modèle est estimé sur les $n-1$ points d'amer (n étant le nombre total d'amers) et validé sur le point d'amer restant en répétant cette opération autant de fois qu'il y a d'amers (soit n fois).

Au total 120 transformations géométriques ont été appliquées et validées. La source de référence utilisée pour la correction dans le système de coordonnées Lambert 93 est la BDOrtho® de l'IGN à 0,5 m de résolution spatiale (cf. chapitre 4).

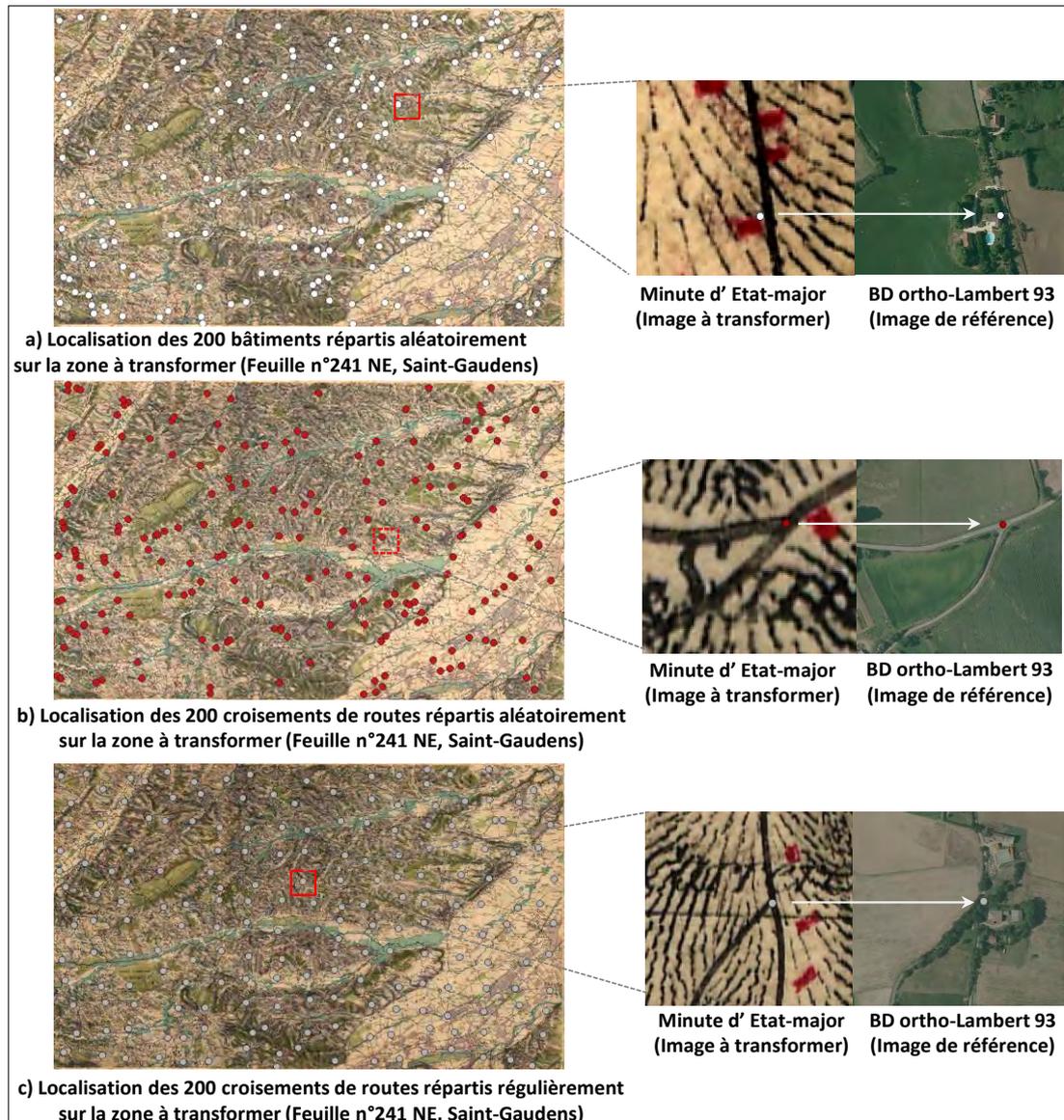


Figure 5-7. Illustration des trois jeux de points d'amer sélectionnés. (a) 200 bâtiments répartis aléatoirement sur la zone à transformer ; (b) 200 croisements de routes répartis aléatoirement sur la zone à transformer ; (c) 200 croisements de routes répartis régulièrement sur la zone à transformer

5.6.3 MISE EN ŒUVRE ET RESULTATS

Nous décrivons ci-dessous les résultats obtenus dans l'ordre des paramètres testés et de la mesure de leur influence.

(1) Impact de la catégorie des points d'amer

Les transformations géométriques effectuées sur la base des croisements de routes et des bâtiments pour le même effectif et la même distribution spatiale (aléatoire dans ce cas) révèlent peu de différences significatives entre les deux catégories de points (Figure 5-8). Ceci est vrai pour les modèles polynomiaux d'ordre 2 et 3, le modèle basé sur la triangulation de Delaunay ainsi que les modèles à noyau polynomiaux. Dans le cas de la régression à noyau gaussien, la qualité de l'ajustement est significativement meilleure avec les croisements de route qu'avec les bâtiments mais uniquement pour les faibles effectifs (25 et 50). On observe une différence respective de 7m et

5 m pour ces cas. Concernant la transformation polynomiale globale d'ordre 1, les résultats affichent une tendance inverse. Les bâtiments conduisent à obtenir une qualité de correction géométrique significativement meilleure avec un plus gr et effectif (configuration à 75 et 150 GCPs). On note des différences respectives de 1,09m et de 2,14 m pour ces cas (figure 5-8).

(2) Impact de l'effectif

Les relations entre le nombre de GCP et la qualité des transformations géométriques ont été étudiées sur la base des croisements de route répartis aléatoirement.

Comme attendu, la hausse de l'effectif conduit à améliorer la qualité de l'ajustement des modèles (cf. figure 5-9, barre d'erreur en bordeaux). Ceci est le cas pour la totalité des modèles testés avec une baisse moyenne de l'EMQ de 9,76 m entre les modèles basés sur 25 GCPs et ceux basés sur 150 GCPs.

En revanche, mis à part pour la triangulation de Delaunay, la hausse de l'effectif à partir de 75 GCPs conduit à faire baisser l'EMQ de manière beaucoup moins importante que l'accroissement du nombre de GCPs de 25 à 75. On enregistre une baisse moyenne de l'EMQ de 8.21 mètres lorsque l'effectif passe de 25 à 75 tandis que l'écart observé entre les modèles à 75 et 150 GCPs est de 1.78 mètres.

Pour la triangulation de Delaunay, la baisse de l'EMQ en fonction de la hausse de l'effectif est plus régulière. Augmenter l'effectif de 25 à 75 conduit à faire baisser l'EMQ moyenne de 3.23 mètres alors que l'on note une différence de 4.94 mètres entre les résultats obtenus à partir d'une configuration à 75 et 150 GCPs.

(3) Impact de la répartition spatiale

La sensibilité des transformations au patron de distribution spatiale des GCPs a été mesurée en utilisant les croisements de routes et en comparant deux distributions : aléatoire et régulière.

Pour l'ensemble des modèles de transformations, la distribution aléatoire a fourni de meilleurs résultats que la distribution régulière (figure 5-9). On observe une différence moyenne de 9,5 mètres entre les deux types de patrons de distribution quel que soit la taille de l'échantillon testé.

(4) Impact de la méthode de transformation

L'impact de la méthode de transformation sur la qualité des résultats de corrections géométriques a été analysé sur la base des croisements de routes répartis aléatoirement (Figure 5.9).

D'une manière générale, les méthodes de transformations locales (Delaunay) ou sensibles localement (méthode à noyau gaussien) sont plus efficaces que les approches globales conventionnelles, quel que soit la taille de l'effectif (figure 5-9). La triangulation de Delaunay basée sur 150 croisements répartis aléatoirement est le modèle le plus performant avec une EMQ moyenne de 46.44 ± 1.82 mètres et significativement meilleur (p value < 0.001) que les autres méthodes dans cette configuration. Le noyau gaussien permet également d'obtenir des résultats très performants mais légèrement en-dessous de la méthode par triangulation avec une EMQ moyenne de 50.03 ± 1.38 mètres.

Si on procède à une analyse plus détaillée des performances des différentes méthodes, on peut observer qu'à faible effectif, la méthode basée sur la triangulation de Delaunay et les méthodes à noyau en général sont plus efficaces que les méthodes polynomiales conventionnelles. Lorsque la taille de l'échantillon de GCP est fixée à 25, la triangulation de Delaunay est significativement meilleure (p value < 0.001) en fournissant une EMQ moyenne de $54,62 \pm 11,25$ mètres tandis qu'avec 50 GCPs, le noyau gaussien est la méthode la plus performante (EMQ moyenne= $51,62 \pm 4,84$ mètres).

On constate aussi qu'avec 25 GCPs, les noyaux polynomiaux d'ordre 2 et 3 permettent d'aboutir à une meilleure qualité d'ajustement que les transformations polynomiales conventionnelles avec des différences respectives de 9 et 14 mètres. Au-delà (c'est-à-dire à partir de 50 GCPs), les résultats observés avec les noyaux polynomiaux ne sont pas significativement différents (p value > 0.45) de ceux obtenus à partir des polynômes conventionnels. On constate aussi que l'accroissement de l'ordre des polynômes conventionnels et des noyaux polynomiaux permet d'améliorer légèrement la qualité de transformation quand le nombre de GCPs est important. Dans une configuration à 150 GCPs, on observe ainsi une différence de 4 mètres entre l'EMQ obtenue à partir d'un polynôme d'ordre 1 et 3. En revanche, à plus faible effectif, cela n'est pas le cas. Avec 25 croisements de routes, la transformation affine (polynôme d'ordre 1) est la plus efficace des méthodes polynomiales avec une EMQ moyenne de 64.01 ± 7.45 mètres. Concernant les méthodes à noyaux polynomiaux, les transformations d'ordre 2 et 3 sont toujours plus efficaces que l'ordre 1 quel que soit la taille de l'effectif de GCP. Néanmoins, le noyau polynomial d'ordre 2 fournit de meilleurs résultats que l'ordre 3 avec un écart moyen d'EMQ égal à 1.35 mètres.

Les résultats des tests de significativité sont disponibles en Annexe A.

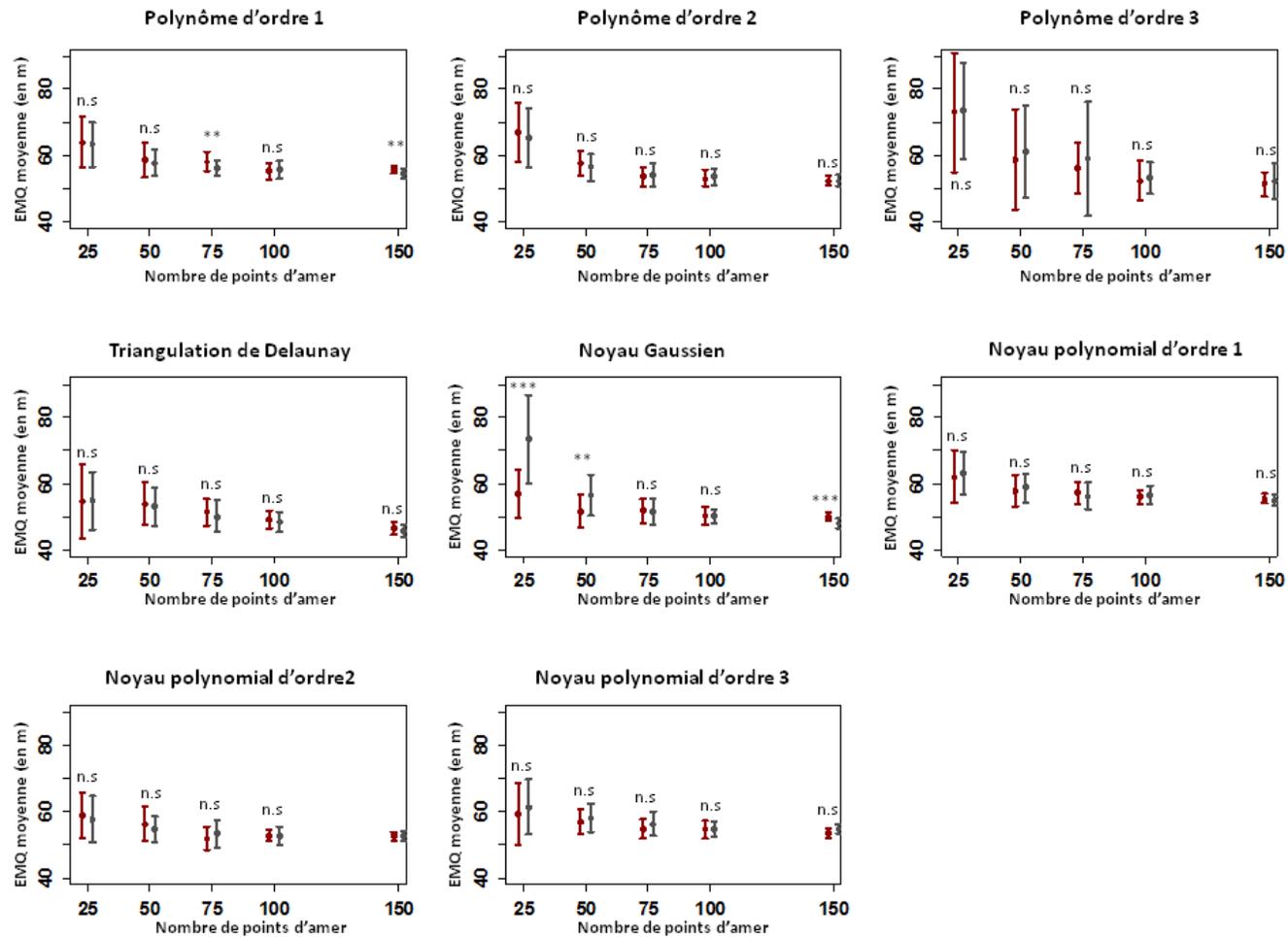


Figure 5-8. Résultats des simulations de transformations géométriques selon la nature des points d'amer. Les points et barres d'erreur représentent respectivement la moyenne et l'écart-type d'EMQ obtenue (avec en bordeaux les bâtiments et en gris avec les croisements de routes). Les résultats de significativité sont issus du test de Wilcoxon visant à comparer deux moyennes (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales.

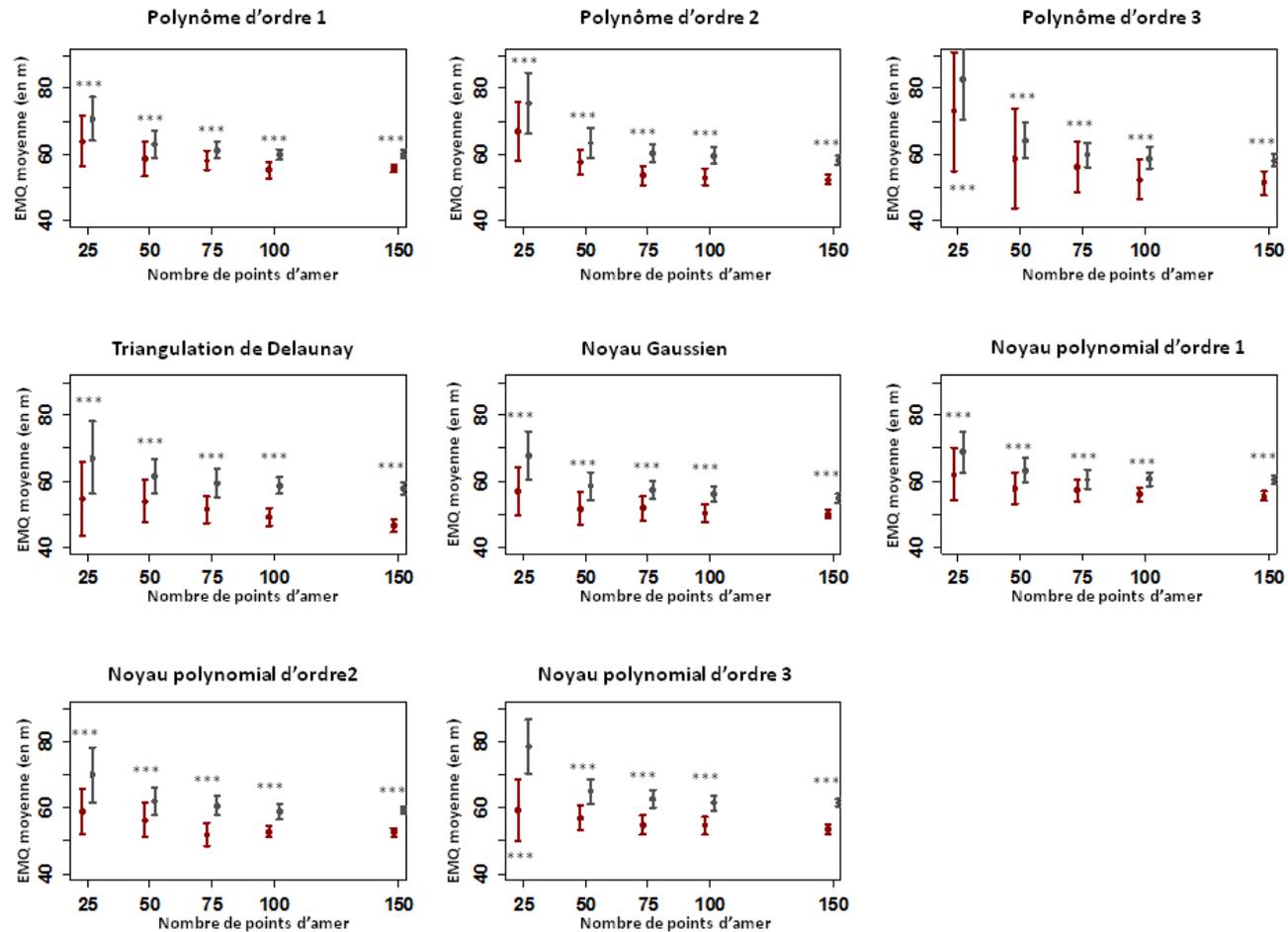


Figure 5-9. Résultats des simulations de transformations géométriques selon la répartition des points d'amer. Les points et barres d'erreur représentent respectivement la moyenne et l'écart-type d'EMQ obtenue (en bordeaux pour la répartition aléatoire, en gris pour la répartition régulière). Les résultats de significativité sont issus du test de Wilcoxon visant à comparer deux moyennes (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales

5.6.4 DISCUSSION

Le protocole de simulation appliqué dans cette étude permet d'apporter un certain nombre de réponses aux questions introduites en section 5.6.1. Plusieurs enseignements peuvent être tirés en testant la sensibilité des nouvelles méthodes proposées aux différents paramètres de correction et en comparant les performances des modèles couramment utilisés.

Premièrement, le choix de la catégorie de points a globalement peu d'impact sur la performance des modèles, du moins entre les croisements de route et les bâtiments pour les zones étudiées. Ceci indiquerait que le degré d'inexactitude de la position des objets est à peu près similaire. Ces résultats sont toutefois à nuancer puisque le noyau gaussien et le polynôme d'ordre 1 apparaissent plus sensibles au type de GCPs sous certaines configurations. Dans le cas du noyau gaussien, il est préférable d'utiliser des croisements de routes lorsque l'effectif est petit (25 ou 50). Il semble que cette catégorie de points permette de corriger plus efficacement les déformations locales tandis que les bâtiments tendent à améliorer la performance du polynôme d'ordre 1 dans une configuration à 75 et 150 GCPs.

Concernant la distribution spatiale des GCPs, nous pouvons observer que la distribution aléatoire permet d'obtenir une meilleure qualité de correction géométrique en général. Ceci se vérifie pour la totalité des modèles testés. La répartition de l'erreur de déformation dans les cartes historiques ne suivant généralement pas un patron régulier, une distribution aléatoire des GCPs conduit à mieux prendre en compte les distorsions existantes. De plus, la mise en place d'un maillage régulier dans la sélection des GCPs représente une contrainte forte pour l'utilisateur. Ceci peut le conduire à récolter des GCPs moins précis géométriquement car ces derniers sont parfois les seuls possibles présents au sein d'une unité de maillage régulier.

Dans le cas du nombre de GCPs, nous avons constaté qu'un accroissement de l'effectif garantissait un meilleur ajustement quel que soit le type de modèle de transformation géométrique. Ce résultat est identique et cohérent avec les travaux de Ford (1978) ou Dupouey (2011) qui montrent que l'EMQ est inversement proportionnelle au nombre de GCPs sélectionnés. Toutefois, l'évolution de l'EMQ n'est pas linéaire pour toutes les méthodes testées ou effectifs choisis.

Contrairement aux autres modèles de transformation, l'accroissement de l'effectif de GCPs pour la triangulation de Delaunay permet de faire baisser l'EMQ moyenne de façon quasi linéaire. La baisse enregistrée entre les modèles basés sur 25 et 75 GCPs ainsi qu'entre ceux basés sur 75 et 150 GCPs est pratiquement similaire. Cette observation est cohérente avec celle de Baletti (2006) qui observe la même tendance. L'accroissement du nombre de GCPs accroît le nombre de triangles divisant la zone d'intérêt. Par conséquent, la surface des triangles est plus petite, ceci permettant de mieux prendre en compte les déformations locales. Pour les autres méthodes, la baisse de l'EMQ est en moyenne quatre fois plus importante lorsque l'effectif passe de 25 à 75 que lorsqu'il passe de 75 à 150. Il semble par conséquent exister un seuil à partir duquel la baisse de l'EMQ peut être jugée insuffisante au regard du nombre de GCPs à sélectionner. D'après les résultats obtenus, nous estimons ce seuil à 75 GCPs. De plus l'écart moyen d'EMQ observé entre les configurations à 75 et 150 GCPs pour ces méthodes est deux fois moins important que celui constaté dans le cas de la triangulation de Delaunay pour ce même accroissement.

La comparaison des différentes méthodes de transformation est riche en enseignements.

Nous avons pu constater que la triangulation de Delaunay fournissait les meilleurs résultats de correction géométrique pour la quasi-totalité des effectifs testés. Cette observation est cohérente avec les travaux antérieurs (Shimizu et Fuse 2003; Balletti 2006; Balletti et Guerra 2009; Jenny et Hurni 2011) qui démontrent une plus grande efficacité des méthodes locales dans la correction de cartes historiques. Les distorsions locales dans les cartes anciennes étant nombreuses, l'application

d'un modèle différent pour recaler des partitions distinctes de la zone d'intérêt est plus efficace. Ce constat est d'autant plus renforcé au regard des performances affichées par le noyau gaussien. Après la triangulation de Delaunay, elle constitue la méthode la plus performante et permet même d'obtenir de meilleurs résultats que la triangulation dans une configuration à 50 GCPs. Sensible localement, contrairement aux autres méthodes à noyau, le noyau gaussien permet aussi de tenir compte efficacement des déformations particulières et non-linéaires existantes dans les cartes anciennes. De plus, il permet de s'affranchir des problèmes de partitionnement de la zone d'intérêt susceptibles d'entraîner des artéfacts puisqu'un objet à cheval sur deux régions peut être déformé par deux modèles distincts (section 5.4.2).

Quant aux résultats obtenus à partir des méthodes à noyaux polynomiaux, nous avons observé très peu de différences avec ceux issus des méthodes polynomiales conventionnelles. Ceci indique que la plupart des calculs des paramètres des fonctions polynomiales classiques n'a pas été biaisé par la présence de système sous déterminé. L'introduction du paramètre de pénalisation n'a par conséquent eu aucun effet sur l'erreur finale de correction géométrique. On observe toutefois une plus grande efficacité des noyaux polynomiaux d'ordre 2 et 3 que les polynômes conventionnels du même ordre dans une configuration à 25 GCPs. Pour ces deux cas, le nombre de GCPs utilisés est faible au regard du nombre de paramètres à estimer (respectivement 6 et 10). L'introduction du paramètre de régularisation dans la KRR permet ainsi de gérer efficacement cette situation de sous-faible effectif en pénalisant les solutions non-régulières.

Enfin, si l'accroissement de l'ordre du polynôme dans les méthodes conventionnelles permet globalement d'améliorer l'EMQ, ce constat n'a pas été vérifié avec les modèles à noyau d'ordre 2 et 3. Le noyau polynomial d'ordre 2 est systématiquement plus efficace que l'ordre 3. Ceci pourrait s'expliquer par l'introduction de déformations trop importantes induites par l'ordre 3 qui aurait tendance à moins bien s'ajuster en dehors des points d'amers.

Finalement, le choix de la stratégie à appliquer paraît assez évident mais peut être nuancé au regard de l'effort qui doit être fourni dans la sélection des GCPs.

Premièrement, le type de GCPs a peu d'importance tandis qu'il est préférable d'adopter une répartition aléatoire des GCPs pour minimiser l'EMQ.

Deuxièmement, la triangulation de Delaunay est la méthode la plus efficace en fournissant l'EMQ la plus basse pour la quasi-totalité des effectifs testés. Dans le cas de l'application de cette méthode, augmenter le nombre de points garantit une baisse de l'erreur de transformation géométrique mais cette tâche peut s'avérer très coûteuse en termes de temps. Dans ce sens, les performances des autres méthodes de transformation montrent qu'au-delà de 75 GCPs, la hausse de l'effectif conduit à améliorer très légèrement la qualité au regard de l'effort qui doit être fait pour sélectionner tous ces points. Parmi elles, le noyau gaussien paraît ainsi bien adapté à la correction des cartes historiques puisque dans une configuration à 75 GCPs, il fournit des résultats identiques à la triangulation de Delaunay sans toutefois souffrir d'éventuels biais liés au partitionnement de la zone d'intérêt (la triangulation de Delaunay).

Le choix de la stratégie à adopter est par conséquent fortement dépendante de l'effectif de GCPs et doit tenir compte de deux considérations des objectifs de l'utilisateur : combien de feuilles doivent être géoréférencées ? Quelle qualité est attendue ? Pour quel objectif ?

5.7. SIMULER POUR AIDER A DECIDER DANS UN CONTEXTE OPERATIONNEL

Les travaux développés dans la section précédente ont également été mobilisées dans un contexte opérationnel et plus particulièrement le projet OCS Ancien visant à produire une cartographie des anciens usages de sols à partir des dessins-minutes d'Etat-major au 1/40 000°.

5.7.1 LE PROJET OCS ANCIEN

Le Ministère de l'Écologie, du Développement Durable et de l'Énergie (MEDDE) a confié en 2013 à l'IGN la production d'une carte des forêts et des autres occupations anciennes des sols de France métropolitaine à partir des dessins-minutes d'Etat-Major au 1/40 000^e.

Le cadre général de la convention liant le MEDDE et l'IGN précise que cette cartographie des a pour objectif d'apporter des informations importantes sur le fonctionnement des écosystèmes, une meilleure connaissance de l'état de référence en matière de conservation des habitats et dans la mise en œuvre de leur gestion. Elle se destine principalement aux gestionnaires de l'environnement, forestiers, responsables de réserves naturelles, de parcs nationaux qui pourront ainsi s'appuyer sur ce document pour différents objectifs :

- (1) pour la gestion conservatoire de la nature en identifiant les sites d'espèces ou de milieux rares.
- (2) pour la compréhension des causes de leur apparition et de leur maintien
- (3) pour l'identification des réservoirs de biodiversité et les écosystèmes dégradés, opération qui nécessite de connaître leur état avant la perturbation.

Dans ce cadre, différents travaux préliminaires sont aujourd'hui menés pour évaluer les besoins nécessaires à la production de cette cartographie. Ces travaux portent principalement sur deux points :

- (1) la correction géométrique de la carte : cet aspect vise à déterminer une stratégie optimale de correction c'est-à-dire en fournissant une qualité géométrique importante tout en minimisant le temps de traitement et l'intervention humaine.
- (2) la numérisation des entités : ce volet consiste là aussi à définir une démarche de vectorisation optimale qui soit applicable à toutes les couches d'occupation du sol (OS) et reproductible à large échelle dans le but de l'inscrire dans une démarche de production (cf. chapitre 7).

Suite à l'étude que nous avons menée et les résultats obtenus (section 5.6), nous avons participé activement à la réflexion de mise en place de la démarche méthodologique en collaboration avec l'IGN et l'INRA de Nancy. Nous détaillons ci-dessous les expérimentations réalisées dans ce cadre qui concernent les corrections géométriques des minutes de la carte d'Etat-Major. Nous présenterons aussi en chapitre 7 les résultats obtenus sur la vectorisation automatique pour ce projet.

5.7.2 PROTOCOLE D'EXPERIMENTATION DEFINI DANS LE PROJET OCS ANCIEN

Au total, quatre minutes de l'Etat-Major ont été sélectionnées dans le cadre de ce protocole. Il s'agit des feuilles 239SE, 239SO, 253NO et 257SE (Figure 5-10). Ces quatre feuilles ont été retenues car elles disposent de caractéristiques distinctes qui permettent de faire l'hypothèse d'une structure de déformation différente (tableau 5-1). La première (239SO) est située au cœur de la chaîne pyrénéenne atlantique. Elle se caractérise surtout par un relief marqué. La seconde (239SE) est localisée dans le piémont des Pyrénées atlantique. Elle est surtout marquée par un relief modéré et la présence de bourgs (donc signaux, clochers) ou de voies de communication (donc croisements de routes) qui offrent de nombreuses possibilités pour la sélection de GCPs. La feuille 257SE est localisée au cœur de la chaîne pyrénéenne orientale et se caractérise par un relief marqué avec une prédominance de milieux naturels et semi-naturels. Enfin, le quart de feuille 253NO est située en Ariège, dans le piémont pyrénéen, et offre d'avantage de possibilités pour récolter des GCPs fiables géométriquement avec, là aussi, la présence de bourgs ou voies de communication.



Figure 5-10. Aperçu des quatre minutes d'Etat-major sélectionnées dans le cadre du protocole OCS ANCIEN. (1) 239SO ; (2) 239SE ;(3) 253NO. (4) 257SE

MINUTE D'ETAT-MAJOR 1850	SITUATION ET CONTEXTE GEOGRAPHIQUE
239SO (Mauléon-64)	<ul style="list-style-type: none"> • située à la frontière franco-espagnole au cœur de la chaîne pyrénéenne atlantique • relief très marqué • difficultés de récolter des points géométriquement fiables : prédominance des milieux naturels et semi-naturels (forêts, zones hydrographiques), peu d'aménagements urbains
239SE (Mauléon-64)	<ul style="list-style-type: none"> • située dans le piémont de la chaîne pyrénéenne atlantique • relief modéré • facilités de récolter des points géométriquement fiables : présence de bourgs, de voies de communications
257SE-(Prades-66)	<ul style="list-style-type: none"> • située à la frontière Franco-espagnole au cœur de la chaîne pyrénéenne orientale • relief très marqué • difficultés de récolter des points géométriquement fiables : prédominance des milieux naturels et semi-naturels (forêts, zones hydrographiques), peu d'aménagements urbains
253NO-(Foix-09)	<ul style="list-style-type: none"> • située dans le piémont de la chaîne pyrénéenne • relief modéré • facilités de récolter des points géométriquement fiables : présence de bourgs, de voies de communications

Tableau 5-1 Détails concernant la situation et le contexte géographique des quatre minutes d'Etat-major sélectionnées dans le cadre du protocole OCS ANCIEN

La première étape du protocole, défini avec l'IGN et l'INRA de Nancy, a consisté à scinder le jeu de points d'amer en deux sous-groupes pour chacune des feuilles: **les points d'amer d'entraînement** (avec lesquels nous estimons la fonction de transformation) et les **points d'amer de validation** (sur lesquels nous calculerons l'erreur de transformation et qui n'interviennent pas dans l'estimation des paramètres de la fonction de transformation) (figure 5-11). Cette distinction est effectuée à l'aide d'un tirage aléatoire et a été répétée 30 fois pour chaque effectif testé et pour chaque méthode.

Afin de tester l'impact de l'effectif des GCPs, plusieurs tailles d'échantillons ont été testées parmi la base de points d'amer d'entraînement (25, 50, 100, 250, 500 et 1000). L'échantillonnage a été effectué par tirage aléatoire sans conserver de points d'amer aux coins de la zone à transformer en raison du nombre important de répétitions effectuées pour chaque effectif (n=100).

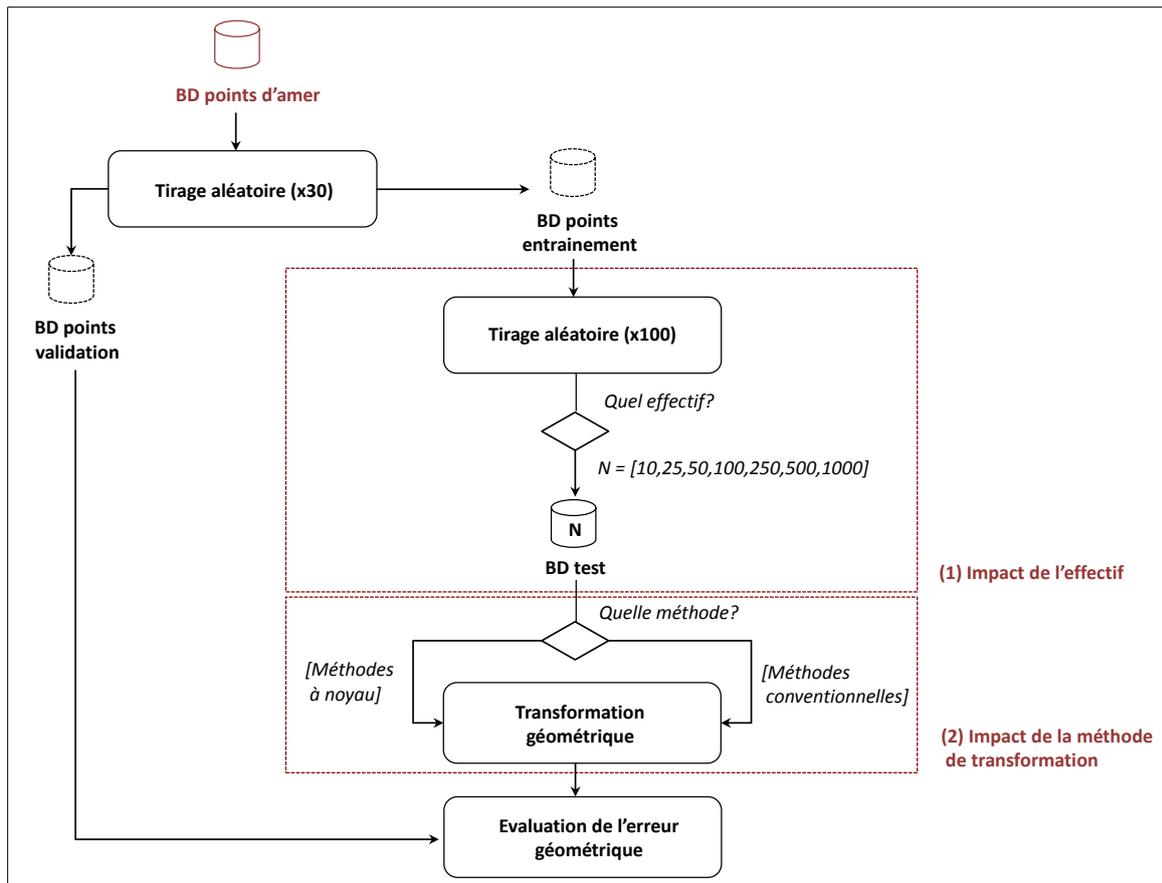


Figure 5-11. Protocole de simulation défini dans le projet OCS ANCIEN pour étudier l'impact des différents paramètres de la correction géométrique sur la qualité de transformation des cartes anciennes.

La troisième étape du protocole consiste à appliquer différents modèles de transformations géométriques pour chaque échantillon de GCP défini. La méthode de régression à noyau est comparée aux trois modèles polynomiaux conventionnels (1^{er}, 2nd et 3^e ordre) ainsi qu'à la méthode de triangulation de Delaunay (basée sur une transformation du 1er ordre pour chaque triangle). Plusieurs noyaux ont aussi été comparés : les noyaux polynomiaux et les noyaux gaussiens.

Un jeu très important de points d'amer pour chaque carte a été mis à disposition par l'équipe de l'INRA Nancy (Tableau 5-2). Ce jeu de points d'amer a été construit à la fois à partir de points remarquables (moulins, clochers, signaux), de croisements de routes, de bâtiments et de limites de régions telles que les bords des forêts.

N°FEUILLE	NB TOTAL D'AMERS DISPONIBLES	NB D'AMERS D'ENTRAINEMENT SELECTIONNES ALEATOIREMENT (X30)	NB D'AMERS DE VALIDATION SELECTIONNES ALEATOIREMENT (X30)
239SE	2444	1500	944
239SO	3574	1750	1824
253NO	1376	1000	376
257SE	1430	1000	430

Tableau 5-2. Détails des tirages aléatoires effectués sur chacun des jeux de points pour constituer la base de points d'entrainements et de validation à chaque échantillon testé

Enfin, la dernière étape a consisté à évaluer la qualité de chaque modèle testé à l'aide des échantillons de validation indépendants. L'indicateur de performance retenu est l'erreur moyenne

quadratique (EMQ) calculée pour chaque répétition, ce qui permet d'obtenir au final une valeur moyenne de l'EMQ avec son écart-type.

5.7.3 RESULTATS

Nous fournissons en figure 5-12, 13, 14 et 15 les résultats obtenus. Les tests réalisés permettent d'effectuer plusieurs constats.

Premièrement, on observe un même patron d'erreur pour tous les modèles de transformation. Deux groupes de cartes sont clairement identifiables : la carte 253 NO et les autres. Les différences entre groupes sont toutefois plus ou moins marquées et varient selon la méthode de transformation et l'effectif. Par exemple, les écarts sont moins importants pour la méthode de Delaunay que pour le polynôme d'ordre 1. Ils peuvent être aussi plus élevés pour l'effectif à 25 points comme dans le cas du polynôme d'ordre 3. Les deux cartes disposant d'un relief très marqué (257SE et 239SO) sont celles qui affichent généralement les plus grands scores d'EMQ. Ceci est vrai pour les polynômes conventionnels et les noyaux polynomiaux tandis que cette observation est plus nuancée pour le noyau gaussien et la méthode de Delaunay.

Deuxièmement, l'EMQ a tendance à baisser avec l'accroissement de l'effectif. Les écarts sont plus élevés cependant entre les effectifs à 25 et 50 et 50 et 75 que les autres. Seules deux méthodes affichent des baisses d'EMQ quasi linéaires entre tous les effectifs : la méthode de Delaunay et la méthode du noyau gaussien. Pour celles-ci, la hausse de l'effectif conduit systématiquement à une meilleure qualité de correction. Elles sont aussi celles qui permettent d'aboutir aux meilleurs scores d'EMQ avec 1000 points parmi tous les modèles testés (respectivement 26.92m en moyenne et 43.44m en moyenne sur les quatre cartes). Enfin, la baisse de l'effectif conduit à une plus grande baisse d'EMQ pour le polynôme et noyau polynomial d'ordre 3 que les autres méthodes de la même famille, notamment entre 25 et 50 points d'amers.

Les résultats des tests de significativité sont fournis en Annexe B.

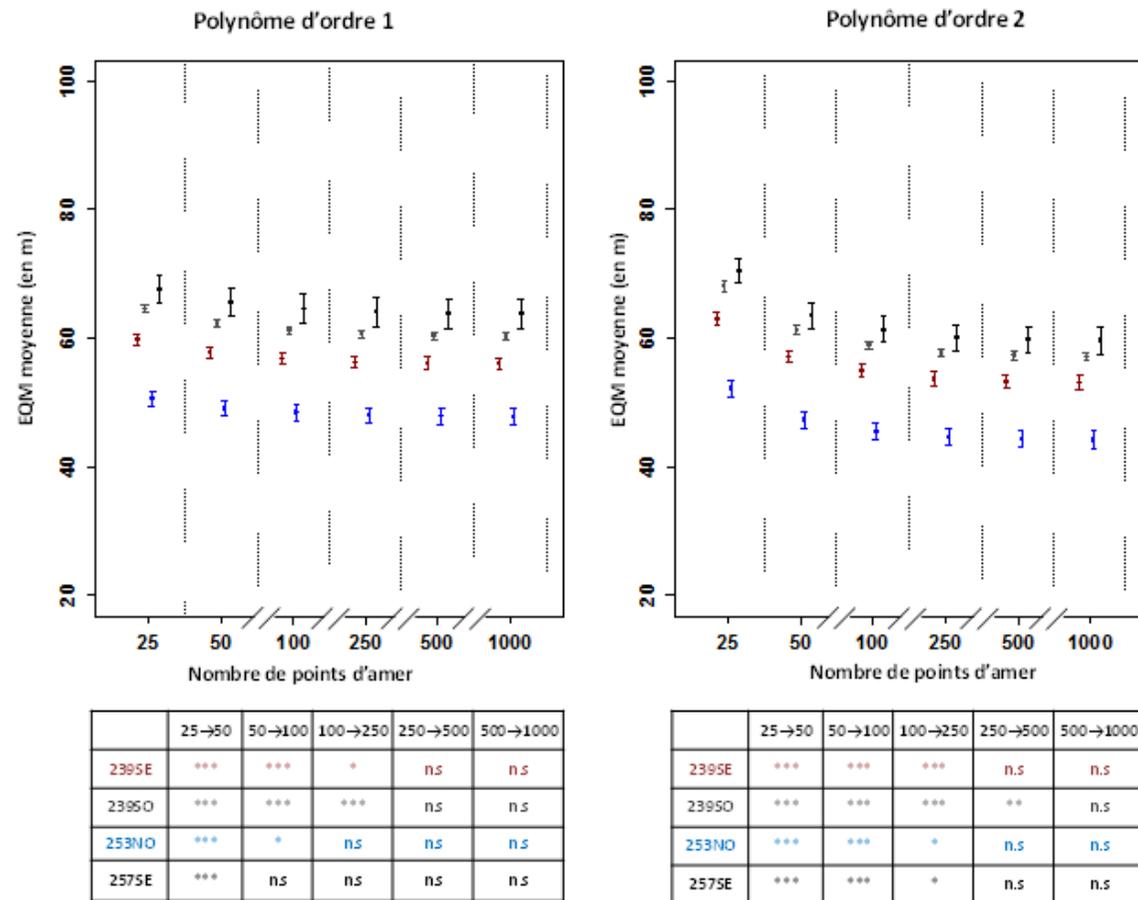


Figure 5-12. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = $p\text{-value} < 0.05$; ** = $p\text{-value} < 0.01$; *** = $p\text{-value} < 0.001$). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales

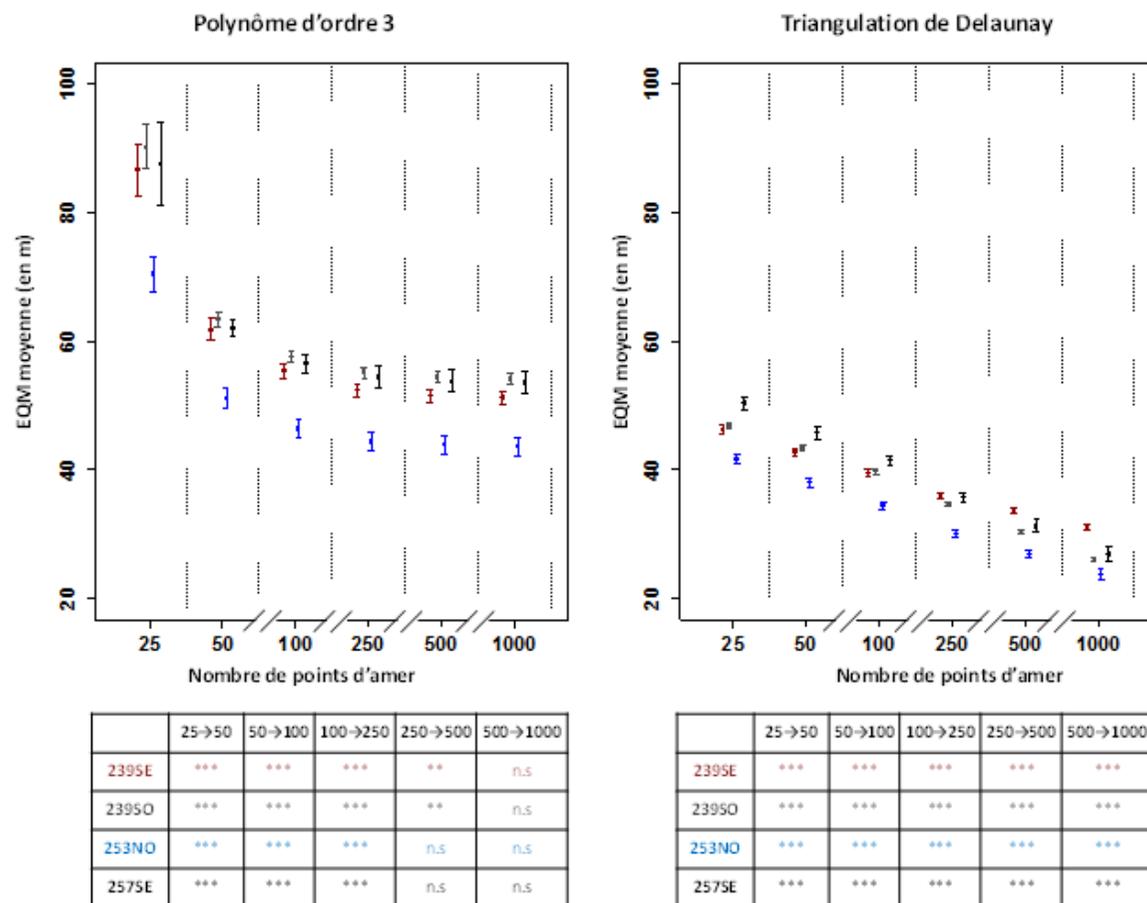


Figure 5-13. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales

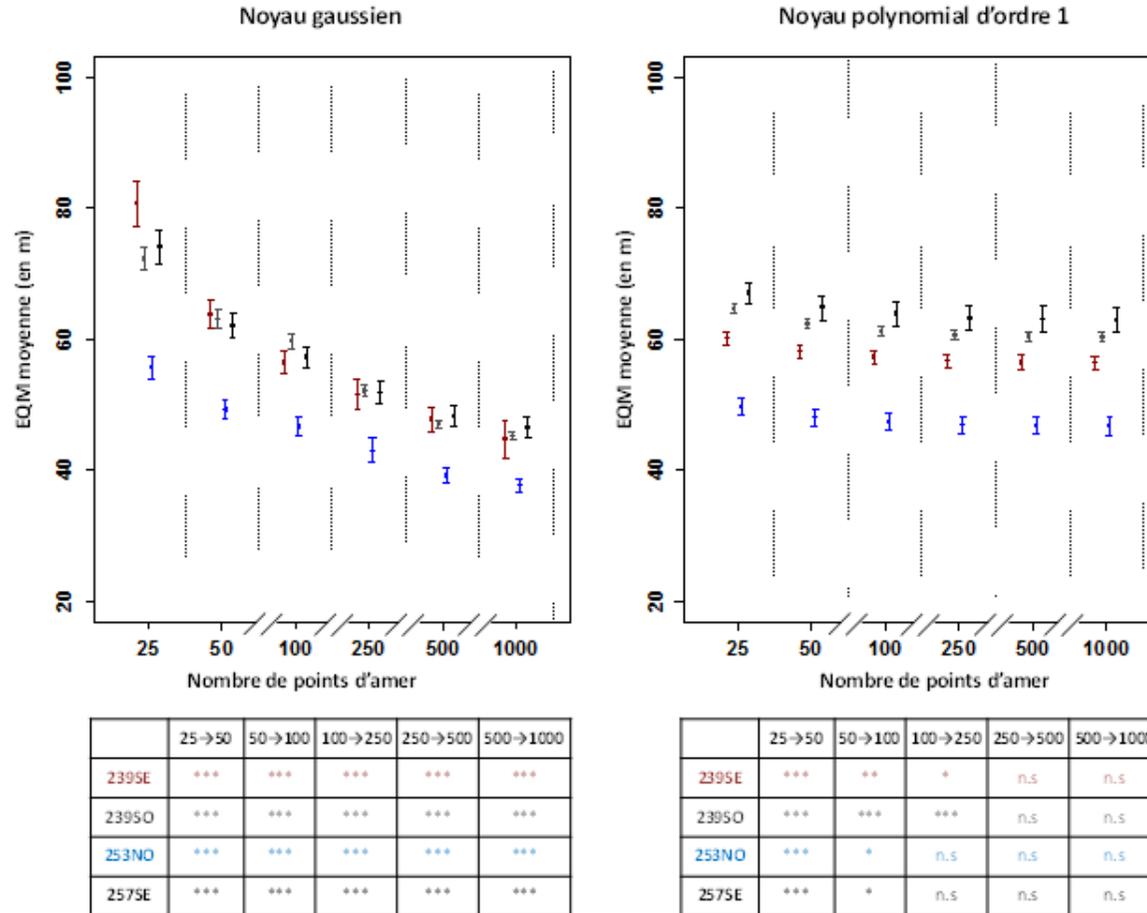


Figure 5-14. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales

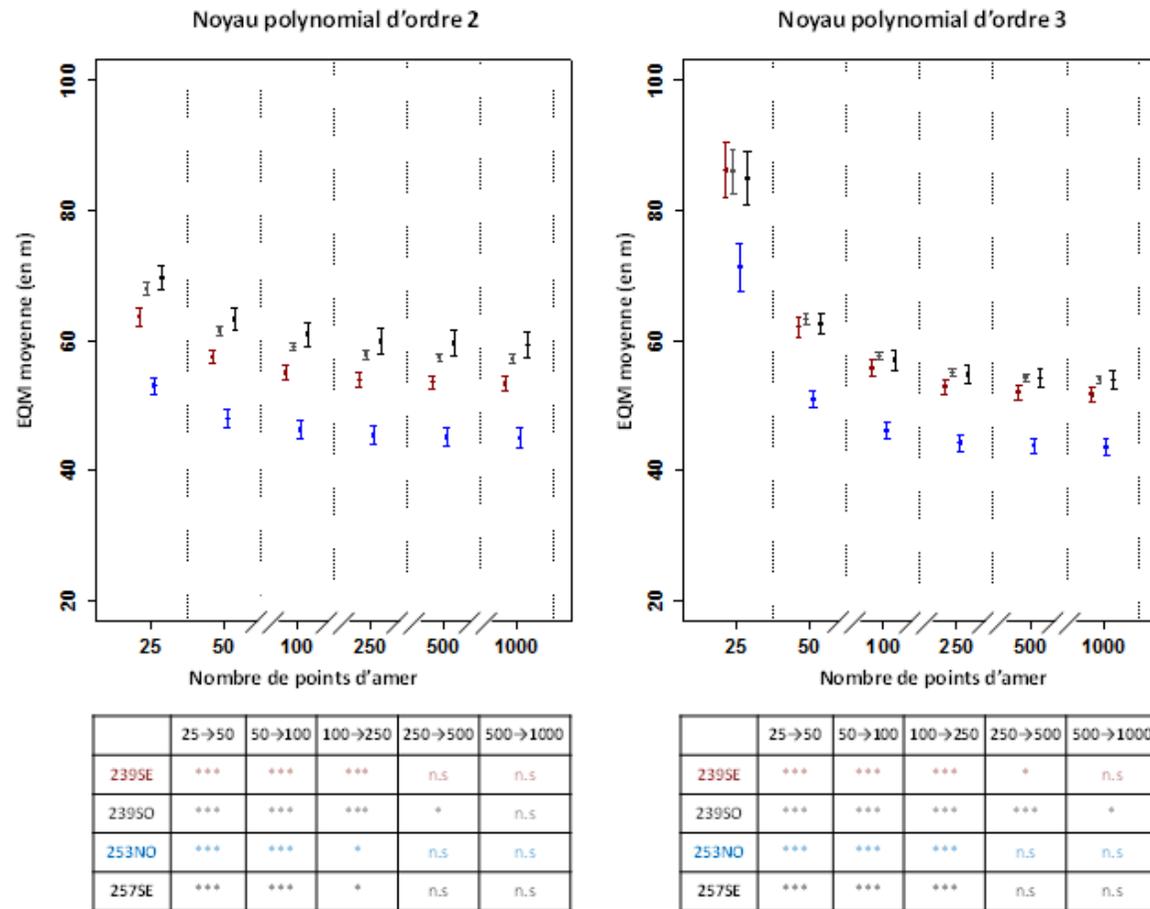


Figure 5-15. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales

5.7.4 DISCUSSION

Les résultats des tests de comparaisons effectuées à partir des quatre feuilles sélectionnées permettent d'apporter un certain nombre d'éléments quant à la mise en place d'un protocole optimal pour la correction géométrique des minutes d'Etat-major à l'échelle nationale.

- La qualité géométrique est dépendante du contexte orographique et de la disponibilité de points d'amers.

Si les tests de simulation ont clairement mis en évidence deux groupes de cartes, ils ont surtout permis de montrer que la qualité de correction géométrique était généralement moins bonne pour les deux cartes qui disposaient du relief le plus marqué (257SE et 239SO). Cela s'est notamment vérifié pour les méthodes polynomiales et les noyaux polynomiaux. Le relief est certainement responsable d'un certain nombre d'imprécisions de levés ayant des conséquences directes sur la qualité de correction géométrique. La faible quantité d'amers fiables géométriquement accentue également les difficultés qui existent pour corriger ces cartes : les possibilités offertes pour la récolte de GCPs portent sur des limites de régions souvent plus floues et imprécises que des points remarquables ou croisements de routes par exemple.

- Les méthodes sensibles localement sont les plus efficaces

Le modèle de Delaunay est la méthode la plus performante quel que soit l'effectif de points d'amers. Ceci est cohérent avec les résultats obtenus dans le précédent protocole de simulation même si nous avons observé que la méthode du noyau gaussien était plus efficace sous certaines configurations. Ici, le noyau gaussien fournit également de très bons résultats de correction à partir de 50 points. En revanche, il ne permet pas d'aboutir à une qualité aussi élevée qu'avec la méthode de Delaunay. De manière générale, ces méthodes gèrent également plus efficacement les cartes aux caractéristiques plus marquées. Les écarts de qualité géométrique entre cartes ont d'ailleurs tendance à baisser avec l'application de ces deux modèles.

- La hausse de l'effectif conduit généralement à une meilleure qualité de correction géométrique

Comme dans le précédent protocole de simulation, l'accroissement de la taille de l'effectif permet d'aboutir à de meilleurs résultats d'EMQ. Pour les méthodes polynomiales et les noyaux polynomiaux, l'amélioration des corrections géométriques en lien avec la hausse de l'effectif se vérifie pour la totalité des cartes jusqu'à 75 points, seuil déjà identifié dans le précédent protocole. La baisse est toutefois plus nette dans le cas des modèles du troisième ordre montrant que ces méthodes sont nettement plus performantes avec l'utilisation d'un jeu de points d'amer fourni. Pour les méthodes sensibles aux distorsions locales (i.e. Delaunay et le noyau gaussien), augmenter l'effectif assure systématiquement de meilleurs résultats de correction. Ainsi, plus la densité de GCPs est importante sur la zone à corriger, plus ces modèles corrigent efficacement les déformations locales et non-linéaires, ce qui était déjà le cas dans le premier protocole.

5.7.5 Bilan

Au regard des résultats obtenus dans le cadre des deux protocoles, nous sommes en mesure de fournir quelques préconisations.

Les modèles sensibles localement sont les plus efficaces et permettent de corriger plus efficacement les cartes aux déformations plus importantes. La méthode de Delaunay est celle qui fournit les résultats les plus intéressants.

Nous préconisons également de s'appuyer sur une répartition aléatoire des points d'amer. Elle « capte » mieux les déformations des cartes anciennes qui ne suivent généralement pas un patron régulier sur la zone à corriger. La répartition régulière constitue également une contrainte qui peut s'avérer contre-productive dans la récolte de GCPs.

Une taille d'échantillon à 75 points d'amer constitue un compromis intéressant EMQ/temps de traitement. Au-delà, la taille de l'effectif ne garantit pas systématiquement une amélioration de la qualité exceptée pour les modèles sensibles localement. D'un point de vue opérationnel, il paraît également difficile de sélectionner d'avantage de GCPs dans l'optique de corriger les dessins-minutes à l'échelle nationale.

Enfin, il est recommandé de s'appuyer en premier lieu sur les points remarquables pour constituer le jeu de points d'amer. Lorsqu'ils ne sont pas présents en nombre suffisant, les croisements de routes et les coins de bâtiments sont des alternatives intéressantes. Ces deux catégories aboutissent d'ailleurs à des résultats quasi-similaires.

Les travaux relatifs à ce chapitre ont fait l'objet de valorisation scientifique :

Herrault P.-A., Sheeren D., Fauvel M., Monteil M., & Paegelow M. 2012. A comparative study of geometric transformation models for the historical 'Map of France' registration , *International Symposium on Historical Maps in Environmental Geosciences (HMEG'2012)* , Cluj Napoca (Romania), 21-23 June 2012

Herrault P.-A., Sheeren D., Fauvel M., Monteil M., & Paegelow M. 2013, A comparative study of geometric transformation models for the historical 'Map of France' registration",. *Geographia Technica* , 17(1), pp. 34-46

Herrault P.-A., Sheeren D, Fauvel M, Paegelow M. 2013. Qualité des données spatiales anciennes : exemple du processus de géoréférencement, *Séminaire du GdR MAGIS (AP 'Incertitude épistémique - des données aux modèles en Géomatique')*, 19 avril 2013, St Etienne (France)

CHAPITRE 6 .VECTORISATION MANUELLE DES ESPACES FORESTIERS

SOMMAIRE

- 6.1. LA VECTORISATION MANUELLE : PRINCIPE, AVANTAGES ET LIMITES**
- 6.2 MISE EN ŒUVRE DANS LES DESSINS-MINUTES D'ETAT-MAJOR AU 1/40 000^e**
- 6.3 MISE EN ŒUVRE DANS LA REVISION TYPE 1889**
- 6.4 MISE EN ŒUVRE DANS LES PHOTOGRAPHIES AERIENNES**
- 6.5 RESULTATS DE LA VECTORISATION MANUELLE**

Ce chapitre présente la démarche employée pour extraire manuellement les espaces forestiers des différentes sources de données spatiales sélectionnées ainsi que les résultats d'extraction. Nous débuterons ce chapitre en rappelant le principe de la vectorisation manuelle ainsi que ses avantages et limites. Ensuite, pour chaque source de données spatiales retenues (cf. Chapitre 4), nous détaillerons les bases de règles de saisie (i.e. les spécifications) et les paramètres définis pour mettre en œuvre cette démarche de vectorisation. Un dernier point traitera de l'analyse succincte des résultats d'extraction dans chacune des sources retenues. Pour rappel, ces résultats seront mobilisés lors des parties C et D, c'est-à-dire pour caractériser les évolutions spatio-temporelles des espaces forestiers et modéliser les relations entre ces évolutions et la présence actuelle de biodiversité.

6.1. LA VECTORISATION MANUELLE : PRINCIPE, AVANTAGES ET LIMITES

La vectorisation manuelle (ou digitalisation) consiste à fabriquer autant d'entités (polygones, lignes, points) qu'il n'y a d'éléments séparés de même nature sur la carte (Favre et al. 2012; Robert et Autran 2012). Pour ce faire, le producteur s'appuie sur des spécifications qui définissent les critères de représentation, de contenu, en fonction des besoins de l'étude ou de l'échelle de la source de données (Sheeren 2005). Dans le cas de la vectorisation manuelle, définir les spécifications revient à déterminer une base de règles de saisie. Elle doit être définie avant de procéder à la digitalisation (Figure 6-1).

Dans ce travail, nous définirons une base de règles de saisie pour chacune des sources sélectionnées. Cela permettra ensuite de produire une base de données contenant les différents états forestiers aux cinq dates considérées dans cette étude.

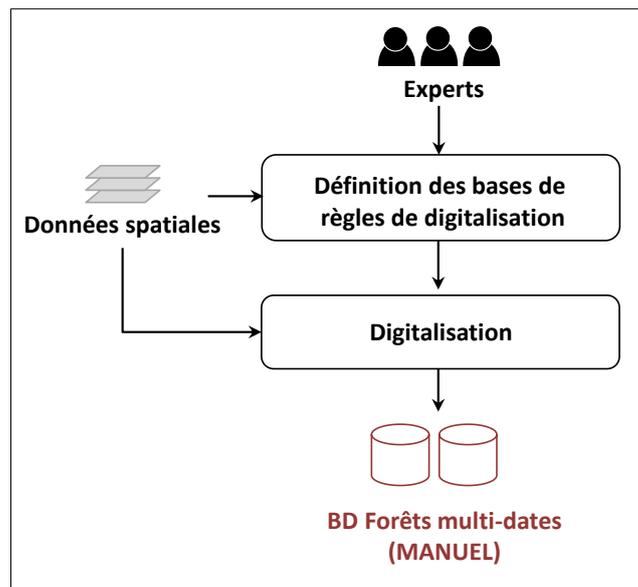


Figure 6-1. Démarche de digitalisation manuelle appliquée pour la production des BD Forêts Multi-dates (MANUEL)

L'approche de vectorisation manuelle est réputée fournir, si elle est menée consciencieusement, une représentation exhaustive et fidèle du paysage selon les spécifications définies. Elle n'est pas pour autant exempt d'erreurs puisque les spécifications peuvent être implicites ou incomplètes et qu'il est de fait impossible d'aboutir au même niveau de précision pour tous les objectifs fixés (Ruas 1999; Sheeren 2005). Ce point renforce également le caractère subjectif de ce processus. Le résultat de la digitalisation des objets peut varier selon la personne en charge de la saisie et l'interprétation qu'elle fera des règles de spécifications. Ce procédé est aussi reconnu

comme une tâche longue, fastidieuse et très couteuse en moyens humains lorsqu'il s'agit de traiter de vastes étendus.

6.2. MISE EN ŒUVRE DANS LES DESSINS-MINUTES D'ETAT-MAJOR AU 1/40 000^e

Nous détaillons ici la démarche entreprise pour vectoriser les forêts dans les minutes d'Etat-Major. Cette démarche se déroule en deux étapes : (1) la définition de la base de règles de digitalisation et (2) la phase de digitalisation manuelle (figure 6-1).

6.2.1 LA BASE DE REGLES

La base de règles de digitalisation définie pour extraire manuellement les espaces forestiers des minutes d'Etat-Major s'est appuyée sur le document technique produit par les chercheurs de l'INRA Nancy (Favre et al. 2012). Ce document détaille les différentes étapes nécessaires à la digitalisation des usages des sols dans ces cartes en tentant de fournir un cadre opérationnel explicite aux utilisateurs potentiels. Les règles fixées dans ce document sont en adéquation avec les objectifs de notre étude. Elles prévoient de numériser les éléments à une échelle fine (1/2000^e) et ont été conçues dans l'optique de s'appuyer sur les objets extraits pour mener des études d'écologie historique.

Les règles de digitalisation définies dans ce travail sont présentées dans le Tableau 6-1. Elles détaillent principalement la définition du contour des espaces forestiers ainsi que les conditions de numérisation des objets se superposant aux forêts.

La base de règles définie prévoit de numériser toutes les forêts présentes dans les minutes d'Etat-major sur la base des différentes représentations énoncées en chapitre 4. Celles-ci doivent être représentées sous la forme de polygones quel que soit leur taille. A noter qu'aucun attribut n'a été associé aux polygones digitalisées et que leur caractérisation interviendra plus tard en partie C. La délimitation du contour des bois doit être réalisée en s'appuyant sur le trait noir bordant les aplats forestiers ou bien sur la limite entre deux tâches de couleurs lorsque ce trait noir n'apparaît pas. La principale interrogation de ce travail a concerné la numérisation ou non des éléments planimétriques se superposant aux aplats forestiers. Nous avons fait le choix de n'en digitaliser aucun. Les routes représentées par un double trait auraient pu être numérisées par exemple mais leur largeur est largement surestimée par rapport à la réalité et varie suivant les cartes. Ce choix implique forcément des biais potentiels dans l'interprétation écologique puisqu'en ne tenant pas compte de ces objets, les forêts ne sont pas fragmentées alors qu'elles l'étaient dans la réalité. Toutefois, étant donné que la même règle a été appliquée à toutes les forêts et que nous chercherons à mettre en évidence les évolutions relatives de chacune d'entre elles, ce choix n'a pas d'incidence sur notre analyse.

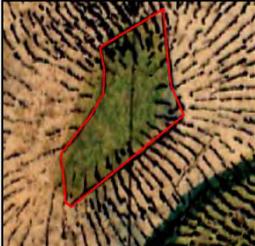
REGLES DE SELECTION	REGLES DE REPRESENTATION	
Toutes les forêts sont saisies sans exception	Une forêt est représentée par un polygone quelle que soit sa taille	
EXPLICITATION DES REGLES DE REPRESENTATION		
REGLES	SOURCE	OBJET VECTORISE
Lorsque l'entité est délimitée par un trait noir alors nous suivrons le milieu de ce trait noir		
Lorsque l'entité est délimitée par un trait noir et que l'aplatissement de couleur dépasse ce trait noir alors nous suivrons le milieu de ce trait noir		
Lorsque l'entité n'est pas délimitée par un trait noir, nous suivrons la limite entre les deux taches de couleur en ne s'éloignant jamais plus de 1 pixel (1,7mètres)		
Les éléments planimétriques se superposant aux forêts (cours d'eau, textes, voies de communication, hachures de relief) ne sont pas numérisés		
Lorsque l'entité présente une perforation, cette perforation sera délimitée		

Tableau 6-1. Base de règles définie pour la vectorisation des forêts dans les minutes de l'Etat-Major au 1/40 000^e (Source : d'après Favre et al.,2012)

6.2.2 LA DIGITALISATION

La digitalisation a été réalisée à l'aide du logiciel Arcgis 10.2 pour l'ensemble de sources traitées.

Premièrement, il est important de préciser que la digitalisation a concerné toutes les forêts échantillonnées dans le cadre du protocole écologique présentée en Chapitre 3 ainsi que celles qui étaient présentes dans leur voisinage. Ce point sera explicitée plus en détail par la suite au moment de caractériser la structure spatiale et contextuelle des forêts (Partie C, chapitre 8).

Ensuite, contrairement aux recommandations effectuées par Favre et al. (2012.), nous avons décidé de numériser les forêts après transformation géométrique de la carte. Ce choix s'explique notamment par la méthode de recalage géométrique entreprise. Favre et al. (2012) proposent de numériser les occupations du sol sur la carte non-corrigée, de transformer une première fois la carte à l'aide d'une fonction affine avant d'entreprendre un recalage élastique (*rubber-sheeting*) basée notamment sur les sommets des entités numérisées. Cette dernière étape explique le fait qu'ils aient eu besoin de digitaliser les occupations du sol avant de corriger la carte. Dans notre cas, la démarche de correction proposée dans cette thèse est indépendante du processus d'extraction, c'est la raison pour laquelle l'extraction des éléments intervient dans une seconde phase.

Tout au long de la numérisation des forêts, nous avons travaillé à l'échelle du 1/3000^e (soit 2 pixels de l'image scannée par mm d'écran) étant donné que les feuilles ont été scannées à 600 dpi. Cette échelle a l'avantage de fournir un cadre de travail opérationnel (en visualisant l'environnement de l'objet à digitaliser) tout en permettant de délimiter les forêts avec une grande précision.

Les forêts se situant à cheval entre deux ou plusieurs minutes d'Etat -Major ont fait l'objet d'un travail particulier. Nous avons adopté la méthode fournie dans Favre et al, (2012) qui propose de s'appuyer sur un recalage élastique pour assembler les composantes d'un même objet réparties sur plusieurs feuilles (les feuilles ayant été corrigées séparément). Ainsi, des points d'amer ont été placés en bordure de feuilles, directement sur les parties de forêts numérisées afin de les faire correspondre spatialement. Une dernière étape de fusion a ensuite été nécessaire pour regrouper ces différentes composantes sous une seule et même entité.

6.3. MISE EN ŒUVRE DANS LA REVISION TYPE 1889

Tout comme les minutes d'Etat-major, nous avons défini des règles de digitalisation avant de redessiner les forêts pour la révision Type 1889 de la carte d'Etat-major au 1/80 000^e.

6.3.1 LA BASE DE REGLES

Dans les feuilles de la Révision Type 1889, les forêts sont représentées par une trame de poncifs noirs non-délimitée dans la majorité des cas (Cf. Chapitre 4). Nous précisons les règles de digitalisation adoptée pour cette carte dans le tableau 6-2.

Tout comme dans les dessins-minutes d'Etat-major, la totalité des forêts présentes ont été numérisées sous la forme de polygones tandis que les éléments planimétriques se surimposant aux forêts n'ont pas été pris en compte pour des raisons identiques à celles énoncées en section 6.2.1. La principale difficulté de ce travail a porté sur la délimitation des trames de poncifs. Lorsque des pointillés noir bordait cette trame, nous nous sommes appuyés sur ce pointillés pour la délimiter mais ce cas de figure est loin d'être systématique (cf. chapitre 4). Lorsque le trait n'apparaît pas (cas le plus fréquent), la délimitation de l'objet a été réalisée sur la base de la bordure extérieure des poncifs en ne s'éloignant jamais plus de 1 pixel (3.89 mètres). Il va de soi que ce choix implique une imprécision dans le contour des objets et donc sur leur superficie puisqu'il est défini uniquement à partir de notre examen de la carte et de notre interprétation.

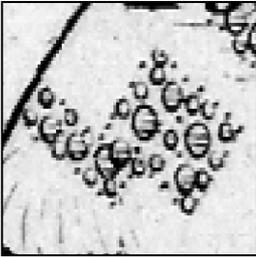
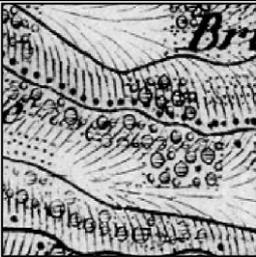
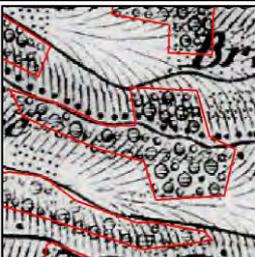
REGLES DE SELECTION	REGLES DE REPRESENTATION	
Toutes les forêts sont saisies sans exception	Une forêt est représentée par un polygone quelle que soit sa taille	
EXPLICITATION DES REGLES DE REPRESENTATION		
REGLE	SOURCE	OBJET VECTORISE
Lorsque les forêts sont délimitées par des pointillés noirs (très rares), nous suivrons ces pointillés		
Lorsque les forêts ne sont pas délimitées par des pointillés noirs, nous suivrons la bordure extérieure des poncifs en ne s'éloignant jamais plus d'un pixel		
Les éléments planimétriques se superposant aux forêts (cours d'eau, textes, voies de communication, hachures de relief) ne sont pas numérisés		
Lorsque l'entité présente une perforation, cette perforation sera délimitée		

Tableau 6-2. Base de règles définie pour la vectorisation des forêts dans les feuilles de la Révision Type 1889

6.3.2 LA DIGITALISATION

La digitalisation a été opérée après correction géométrique de la carte à l'échelle du 1/3 000^e. Bien que la résolution spatiale de ces cartes est plus grossière que les minutes au 1/40 000^e (4,04 mètres), l'absence de contour de délimitation des forêts nous a contraint à travailler à une échelle fine afin d'ajuster précisément la bordure extérieure des poncifs. L'assemblage des entités forestières se situant sur plusieurs feuilles a été réalisé par le même procédé que celui appliqué dans les minutes d'Etat-major au 1/40 000^e.

6.4. MISE EN ŒUVRE DANS LES PHOTOGRAPHIES AERIENNES

6.4.1 PRETRAITEMENTS ET ORTHORECTIFICATION DES PHOTOGRAPHIES AERIENNES ANCIENNES

L'orthorectification des photographies aériennes anciennes noir et blanc de 1954 et 1979 (cf. Chapitre 4) a été réalisée à l'aide du logiciel Erdas 5.0. Nous avons utilisé le modèle numérique de terrain BD Alti® de l'IGN de 2010 à 25 m de résolution spatiale ainsi que la BD Ortho® IRC de 2010 à 0.5 m de résolution spatiale pour définir les points d'amer. Seules les photographies prises à ces deux dates ont fait l'objet d'un tel travail puisque la BD Ortho® IRC de 2010 est une collection de clichés orthorectifiés et mosaïqués (cf. chapitre 4).

Trente à quarante points d'amer par cliché ont été sélectionnés pour procéder à l'orthorectification soit approximativement 1500 points pour les photographies de 1953-54 et 900 points pour les photographies de 1979. Le nombre de points d'amer par photographie varie suivant l'importance du relief dans la zone : davantage de points ont été récoltés dans les zones de plus fortes pentes. La fonction de transformation utilisée est une transformation polynomiale d'ordre 1. La qualité de l'ortho rectification a été évaluée au moyen de l'EMQ. En moyenne, l'erreur enregistrée est de l'ordre de 31 ± 5.7 mètres pour les photographies de 1954 et 28 ± 5.4 mètres pour les photographies de 1979.

Enfin, l'ultime étape des corrections géométriques a consisté à effectuer une mosaïque des différentes photographies aériennes ortho rectifiées à chaque date (1954 et 1979, Figure 6-2). Le mosaïquage consiste à « recoller » les différents clichés entre eux pour en faire une couverture globale de la zone traitée. Cette étape a été réalisée à l'aide du logiciel Erdas 5.0.

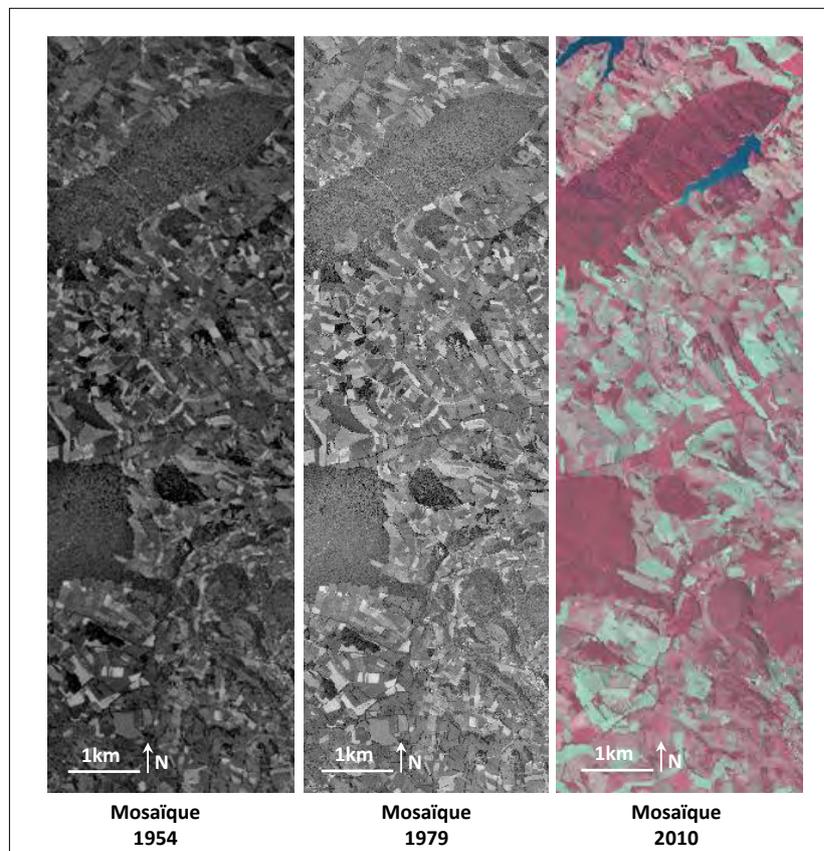


Figure 6-2 Extraits des photographies aériennes mosaïquées aux différentes dates.

6.4.2 LA BASE DE REGLES DE DIGITALISATION

Une base de règles commune a été définie pour la vectorisation des espaces forestiers dans les photographies aériennes anciennes de 1954 et de 1979 ainsi que dans la BD Ortho®IRC de 2010 (tableau 6-3).

Premièrement, la règle de sélection des forêts à digitaliser a été défini dans le but de rendre compatible la représentation des forêts dans les cartes anciennes avec celle des photographies aériennes. A noter que nous nous basons uniquement sur le critère de surface pour définir cette règle sans tenir compte de la composition des forêts, ni leur structure, leur forme ou leur densité de peuplement. La plus petite forêt observée dans les dessins-minutes d'Etat-major mesure 0.06 ha (cf. section 6.5) soit une valeur très proche du seuil fixé par l'IFN dans sa définition de la forêt (cf. chapitre 2). Nous nous sommes donc appuyés sur le seuil de l'IFN pour définir les objets à vectoriser dans les photographies aériennes et avons digitalisé tous les espaces boisés supérieurs à 0.05ha. Ce choix n'est pas sans incidence et pourrait introduire des biais dans les modèles écologiques (Partie D). Les spécifications des dessins-minutes ne permettant pas de définir correctement la forêt (chapitre 4, section 4.2.3 et 4.3.2), le seuil minimal de représentation a été défini empiriquement sur la base des objets extraits dans ce travail. Par conséquent, il pourrait être potentiellement plus ou moins élevé sur d'autres feuilles mais cet examen n'a pu être réalisé puisque nous ne disposons pas des couches de forêts extraites sur l'ensemble des feuilles à l'échelle nationale.

La délimitation du contour géométrique des forêts s'est appuyée sur la ligne séparant la lisière extérieure de la forêt et la zone homogène adjacente à cette lisière pour être conforme aux règles de saisie dans les cartes anciennes et notamment celle de la Révision Type 1889 qui prévoit de s'appuyer sur la bordure extérieure des poncifs pour délimiter les trames lorsqu'aucun trait noir ne la délimite.

Les routes et les cours d'eau traversant les forêts n'ont pas été vectorisés. Ceci s'explique en partie par le choix qui a été fait précédemment de ne pas retenir les éléments planimétriques se superposant aux entités forestières dans les cartes anciennes. L'objectif principal étant de combiner les différentes couches extraites dans le but de suivre les évolutions spatio-temporelles, il était important de limiter les écarts de définition existants entre les sources historiques et les sources plus récentes.

Dans les cas où une forêt présente une perforation, nous avons fait le choix de maintenir la perforation dans la représentation de l'objet digitalisé afin de rester au plus près de la forme originelle des objets. Bien que très peu de cas de perforations apparaissent dans les forêts des cartes historiques, les îlots d'occupation du sol au sein de ces forêts ont été maintenus et constituent par conséquent des zones de perforation également. Là aussi, une incertitude existe. Les différentes instructions associées aux cartes anciennes (section 4.2.2, chapitre 4) ne précisant pas les cas où le producteur devait restituer ces perforations, certaines ont pu être gommées selon le producteur en charge de la saisie. Des différences relatives entre les forêts aux différentes dates pourraient donc être observées sans que celles-ci soient liées à de réelles évolutions.

Enfin, les îlots forestiers reliés par « un bandeau d'arbres » supérieur à 20 mètres ont été fusionnés avec ce même bandeau (tableau 6-3). Ce seuil de largeur vise à différencier ces bandeaux des haies qui, elles, n'ont pas été vectorisées.

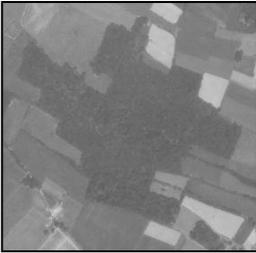
REGLES DE SELECTION	REGLES DE REPRESENTATION	
Toutes les forêts supérieures à 0.05ha sont saisies	Une forêt est représentée par un polygone quelle que soit sa taille	
EXPLICITATION DES REGLES DE REPRESENTATION		
REGLE	SOURCE	OBJET VECTORISE
Pour délimiter le contour géométrique des forêts, nous suivrons la ligne séparant la lisière extérieure de la forêt et la zone homogène adjacente à cette lisière		
Les routes et les cours d'eau traversant les surfaces forestières ne seront pas numérisés		
Lorsque la forêt présente une perforation, la zone intérieure non-recouverte par des ligneux ne sera pas digitalisée		
Lorsque deux forêts sont reliées par un bandeau d'arbres dont la largeur est supérieure à 20 mètres alors nous fusionnerons ces deux forêts avec même bandeau sous une seule et même entité		

Tableau 6-3 Base de règles définie pour la vectorisation des forêts dans les collections de photographies aériennes

6.4.3 LA DIGITALISATION

La vectorisation manuelle des forêts sur les photographies aériennes a été effectuée au 1/2 000^e, soit une échelle plus grande que celles adoptées pour la digitalisation des forêts dans les cartes historiques. La définition du contour géométrique s'appuyant sur la ligne de séparation entre deux entités naturelles (lisière/zone homogène adjacente) implique de fonctionner à une très grande échelle. A noter que les forêts ont été numérisées après orthorectification des clichés aériens et la phase de mosaïquage (pour les photographies N/B) afin de s'affranchir des problèmes de superposition entre photographies.

6.5. RESULTATS DE LA VECTORISATION MANUELLE

Un rapide bilan statistique des couches contenant les forêts digitalisées à chacune des dates montrent que 992 objets ont été vectorisés dans les minutes de 1850, 940 dans la Révision Type

1889 et un peu plus de 1 500 dans chacune des collections de photographies de 1954, 1979 et 2010 (Figure 6-3, Tableau 6-4).

On observe premièrement que la majeure partie des forêts vectorisées dans chacune des sources a une surface inférieure à 20 hectares et que la surface minimale des forêts vectorisées varie de 0,05 ha (à partir des photographies aériennes de 1954) à 0,23ha (à partir de la Révision Type 1889). Les histogrammes de surface aux différentes dates affichent des profils de Pareto qui sont assez classiques pour les objets naturels. Cela renvoie au débat actuel suggérant qu'il doit exister plusieurs moyens de penser et de ne pas seulement considérer que la plupart des objets du monde peut être modélisée sous une forme gaussienne comme c'est le cas aujourd'hui dans les analyses géospatiales (Jiang 2014). A noter que la plus grande forêt vectorisée a été retirée des histogrammes de distribution par souci de visualisation. Dans l'ensemble des sources traitées, cette forêt est nettement plus grande que les autres entités (environ 700 ha).

Dans les cartes historiques, la moyenne de surface des forêts vectorisées se situe autour de 7 ha tandis qu'elle oscille entre 4,14 et 5,49 ha dans les photographies aériennes de 1954 à 2010. La plus petite forêt vectorisée est à recenser parmi celles obtenues à partir des photographies aériennes de 1953-1954 avec une surface de 0,05 ha (seuil de sélection que nous avons fixé). On retrouve des valeurs presque similaires pour les forêts issues des autres photographies aériennes et des minutes de 1850 puisque les plus petites forêts vectorisées dans ces sources disposent d'une surface approximativement égale à 0,06-0,07 ha. Toutefois, la plus petite forêt vectorisée dans la Révision Type 1889 est nettement plus grande que dans les autres sources avec une surface de 0.23ha. Cette différence est probablement liée au fait que cette carte soit issue de la carte d'Etat-major au 1/80 000^e ayant entraîné un processus de généralisation plus important. Néanmoins, les spécifications incomplètes de ces documents ne nous permettent pas de fournir des indications précises au sujet de cette valeur seuil.

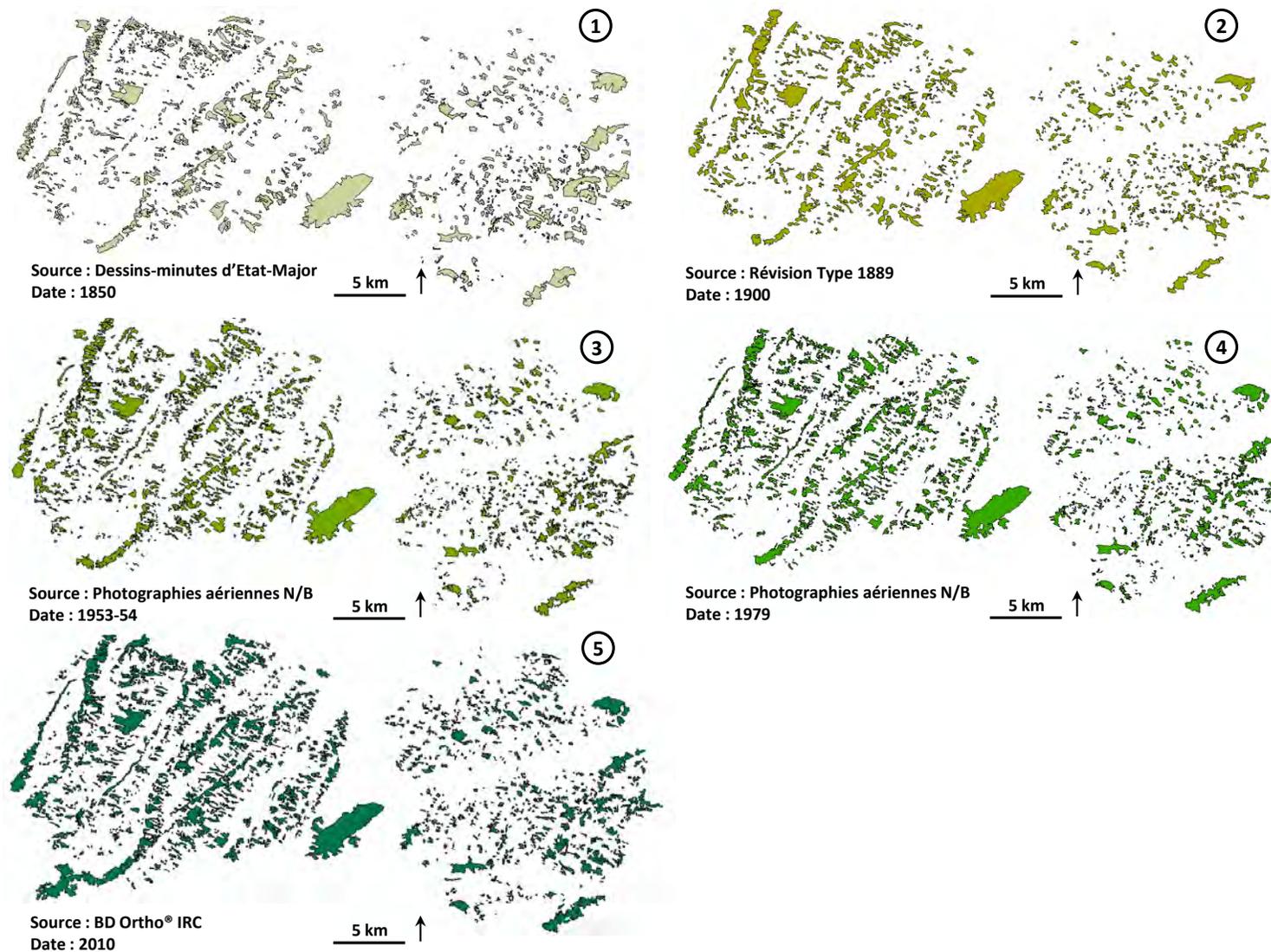


Figure 6-3 Aperçu des couvertures forestières numérisées aux différentes dates

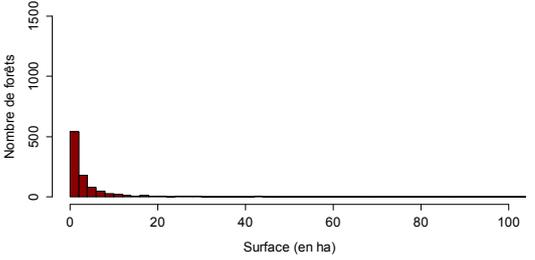
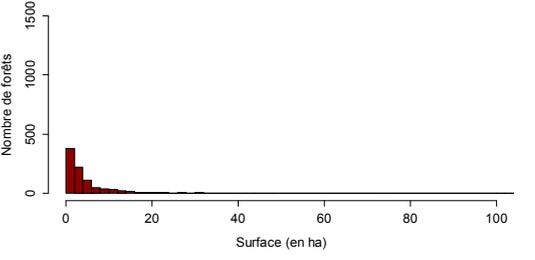
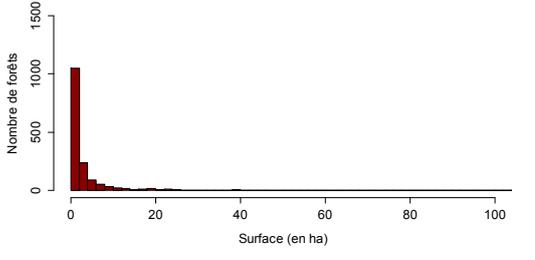
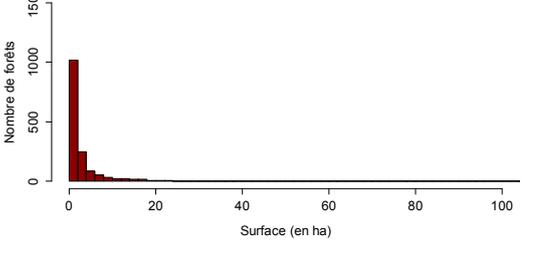
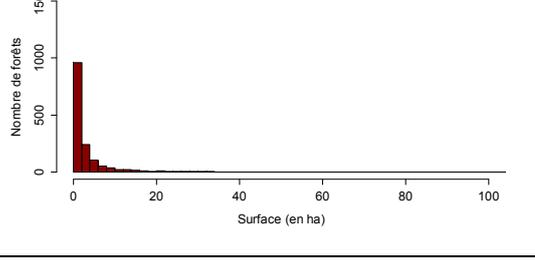
SOURCE DE DONNEES SPATIALES	HISTOGRAMME DES FREQUENCES DE DISTRIBUTION DE SURFACE	STATISTIQUES DESCRIPTIVES
<p>Dessins-minutes de la carte d'Etat-Major (1850)</p>	<p>Histogram of as.matrix(data1850)</p> 	<p>Nb objets : 992 Max (surface) : 757,48ha Min (surface) : 0,06ha Moy (surface) : 6,76ha Ecart-type (surface): 29,13ha</p>
<p>Révision Type 1889 de la carte d'Etat-Major (1900)</p>	<p>Histogram of as.matrix(data1900)</p> 	<p>Nb objets : 940 Max (surface) : 727,59ha Min (surface) : 0,23ha Moy (surface) : 7,63ha Ecart-type (surface): 28,45ha</p>
<p>Photographies aériennes N&B (1953-1954)</p>	<p>Histogram of as.matrix(data1954)</p> 	<p>Nb objets : 1578 Max (surface) : 705,71ha Min (surface) : 0,05ha Moy (surface) : 4,14ha Ecart-type (surface): 20,33ha</p>
<p>Photographies aériennes N&B (1979)</p>	<p>Histogram of as.matrix(data1979)</p> 	<p>Nb objets : 1573 Max (surface) : 676,81ha Min (surface) : 0,07ha Moy (surface) : 4,44ha Ecart-type (surface): 20,47ha</p>
<p>BD ortho IRC (2010)</p>	<p>Histogram of as.matrix(data2010)</p> 	<p>Nb objets : 1549 Max (surface) : 717,32ha Min (surface) : 0,06ha Moy (surface) : 5,49ha Ecart-type (surface): 30,75ha</p>

Tableau 6-4 Statistiques élémentaires des forêts extraites manuellement à partir des différentes sources spatiales

Synthèse de chapitre :

Les démarches de vectorisation employées ainsi que les bases de règles définies pour extraire les forêts manuellement des différentes sources ont été explicitées. Cela a permis d'effectuer une première évaluation des états forestiers aux différentes dates retenues. Nous avons pu observer que le nombre de forêts vectorisés a eu tendance à augmenter avec le temps : cet accroissement s'est principalement traduit par un nombre plus élevé de petites forêts indiquant de fait un morcellement de l'état boisé depuis un siècle et demi.

Ensuite, la définition des règles de saisies implique de faire des choix susceptibles d'introduire des biais potentiels. Par exemple, la plus petite forêt vectorisée en 1900 mesure 0.23 ha alors que le seuil minimal de représentation avoisine les 0.05 ha sur toutes les autres sources. Par conséquent des différences entre dates pourraient apparaître non pas en raison d'une évolution mais d'une différence de représentation ou de contenu. Ce point sera traité dans le chapitre 11.

Enfin, la démarche manuelle entreprise ici pour vectoriser les forêts est apparue comme une tâche longue et fastidieuse. Il devient par conséquent nécessaire d'évaluer la possibilité de mettre en œuvre des approches automatisées dans le but de rendre cette étape applicable à de zones plus vastes. Nous aborderons ce point dans le chapitre 7.

CHAPITRE 7.VECTORISATION AUTOMATIQUE DES ESPACES FORESTIERS

SOMMAIRE :

7.1. ENJEUX DE L'EXTRACTION AUTOMATIQUE

7.2. PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE AUX
MINUTES D'ETAT-MAJOR AU 1/40 000^e

7.3. PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE A LA BD
ORTHO IRC

Dans ce chapitre, nous proposons plusieurs méthodes de traitement d'images destinées à automatiser la vectorisation des forêts à partir de sources de données spatiales anciennes et actuelles. La première méthode concerne l'analyse des minutes de la carte d'Etat-Major de 1850. La seconde méthode est adaptée à l'extraction des forêts à partir de la BD ORTHO® IRC de 2010. Chacune d'entre elles est présentée, dans la suite de ce chapitre, en détaillant successivement les résultats obtenus et en discutant leurs performances pour une portée opérationnelle.

7.1. ENJEUX DE L'EXTRACTION AUTOMATIQUE

Comme mentionné dans le chapitre précédent, la vectorisation manuelle est couramment utilisée en écologie du paysage lorsqu'il s'agit de cartographier les éléments à partir de données spatiales. Celle-ci comporte plusieurs limites :

- (1) elle est très coûteuse en temps et, par conséquent, en ressources humaines
- (2) elle est difficilement reproductible sur de vastes étendues géographiques
- (3) elle est partiellement subjective car elle contraint l'utilisateur à effectuer à la fois des choix thématiques et géométriques (dans la délimitation des zones homogènes).

Ces limites ont conduit au développement d'approches automatisées. L'extraction automatique d'objets du paysage à partir d'images (issues de capteurs spatiaux ou aéroportés) est une pratique récurrente en télédétection. De nombreux auteurs ont proposé des méthodes de classification pour produire des cartes d'occupation et d'utilisation des sols à partir d'images satellitaires (Willhauck et al. 2000; Wei et al. 2004; Bacher et Mayer 2005; Boyd et Danson 2005; Lu et Weng 2007; Yu et al. 2014), de photographies aériennes (Cocquerez et Devars 1985; Shoshany et Degani 1992; Pal et al. 2003) et même de cartes historiques (Leyk et Boesch 2010; Iwanowski et Kozak 2012). Ces approches peuvent porter sur différents milieux : urbain (Puissant 2003; Sheeren et al. 2007; Skupinski et al. 2009), naturels ou semi-naturels (Collins et Woodcock 1996; Wilson et Sader 2002; Fauvel et al. 2012), agricoles (Yang et Lo 2002; Deng et al. 2008) ou encore les voies de communication (Gruen et Li 1995; Fortier et al. 1999; Péteri et al. 2003). Il ressort de tous ces travaux une large gamme de méthodologies disponibles souffrant toutefois d'un manque de généralité : elles sont souvent spécifiques à un type de donnée ou à un type d'objet en particulier.

7.2. PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE AUX FORETS DES MINUTES DE L'ETAT-MAJOR

7.2.1 L'EXTRACTION AUTOMATIQUE DANS LES CARTES SCANNEES : APPROCHES EXISTANTES

L'extraction automatique d'entités dans les cartes historiques numérisées soulève de nombreuses difficultés en géomatique pour plusieurs raisons (Leyk et al. 2005). Premièrement, de nombreux éléments planimétriques se superposent les uns aux autres, comme les routes et la toponymie sur les aplats de couleurs représentant l'occupation du sol. Cela pose des problèmes pour cartographier les aplats sous forme de polygones, en effet ces éléments superposés créent des trous au moment de leur extraction. Deuxièmement, les procédures de scannage ou de compression d'images peuvent réduire la qualité des cartes numérisées, ce qui est susceptible d'affecter les performances d'extraction (Daniil et al. 2003). Enfin, de nombreuses cartes historiques ont été produites en noir et blanc, sans autres couleurs. Pour celles-ci, il est nécessaire de développer des approches d'extraction très spécifiques et particulièrement complexe étant donné que la reconnaissance des objets ne peut pas se fonder sur leur teinte (Boesch 1996)(Figure7-1).



Figure 7-1. Illustration de la superposition d'éléments planimétriques dans les cartes anciennes. L'image (a) est un extrait d'une minute d'Etat-major de 1850 tandis que l'image (b) est un extrait de la Révision Type 1889. A noter que l'absence de couleur dans l'image (b) complexifie l'extraction automatique des éléments paysagers.

Plusieurs auteurs ont déjà proposé des méthodes automatiques pour vectoriser les objets géographiques dans les cartes thématiques scannées actuelles et anciennes (Ansoult et al. 1990; Lichtner). De nombreuses cartes ont fait l'objet d'un travail de reconnaissance d'entités telles que les cartes topographiques des Etats-Unis (Li et al, 1999), les cartes militaires de l'Institut Géographique Polonais (Iwanowski et Kozak 2012) ou encore les cartes historiques nationales Suisses (Leyk et Boesch 2010).

De manière générale, la procédure d'extraction automatique peut être divisée en trois étapes : (1) une étape de filtrage, (2) une étape de reconnaissance/extraction des objets d'intérêt et (3) une étape de post-traitements. La première étape vise à rendre l'extraction plus facile alors que la dernière étape a pour objectif de corriger et d'améliorer la qualité des entités extraites. Les étapes 2 et 3 peuvent être vues comme un seul et unique traitement selon les auteurs (Wise 1999). Nous décrivons ci-dessous les approches existantes pour le traitement des cartes historiques en combinant les approches pré et post-traitements (étapes 1 et 3) dans la même partie en raison de la similarité des techniques utilisées.

Méthodes de pré et post-traitements existantes

Les cartes thématiques ou les cartes historiques scannées sont des sources complexes à traiter automatiquement et contiennent beaucoup de « bruit » par rapport aux objets d'intérêt. En effet, les différents éléments superposés (voies de communication, lignes d'élévation, textes) peuvent être vus comme du « bruit » lorsque l'utilisateur souhaite extraire des informations spécifiques. Ce bruit rend alors difficile la mise en place de procédures automatiques visant à vectoriser les objets souhaités (forêts, prairies, bâtiments...). De plus, les procédures de scannage et la détérioration des documents anciens avec le temps (déchirure, pliure, humidité...) peuvent aussi conduire à des effets d'ombrages sur les cartes et altérer la qualité du document original.

Pour pallier ce problème, les cartes numérisées sont souvent filtrées durant une étape de prétraitement, à l'aide de techniques de traitement d'image (Samet et Soffer 1998; Iwanowski et Kozak 2012; Chiang et al. 2013). Les filtres consistent à appliquer une fenêtre mobile pour attribuer une nouvelle valeur à chaque pixel de l'image en utilisant les valeurs des pixels de son voisinage. Ils sont généralement bien adaptés à la réduction du bruit dans les images tout en préservant les éléments structuraux tels que les contours. Différents filtres existent en traitement d'image allant des filtres de convolution aux filtres morphologiques (Cocquerez et Philipp-Foliguet 1995).

L'application d'un filtre morphologique consiste à appliquer un ou plusieurs opérateurs morphologiques à l'aide d'un masque ou d'un motif géométrique prédéfini de taille et de forme variables (figure 7-2) (Sheeren et al. 2007; Soille 1999). Le motif (appelé *élément structurant - ES*) est déplacé dans l'image afin que son centre occupe toutes les positions dans l'espace (les pixels - p). Pour chaque pixel, on applique l'opérateur sélectionné (ex : une érosion, une dilatation). Il résulte de cette application une image transformée dans laquelle les éléments X de l'image peuvent avoir été fusionnés, éliminés, amincis, etc. (Figure 7-2(a)). Cette dernière figure montre par exemple qu'après application de l'élément ES sur l'image à traiter, les pixels isolés ont été éliminés.

Le filtrage de convolution quant à lui consiste à recalculer la valeur d'un pixel en se basant sur la valeur du pixel lui-même et sur les valeurs des pixels avoisinants (le voisinage étant fixé non pas par un élément structurant mais une fenêtre de convolution de taille prédéfinie). Les opérateurs de filtrage peuvent être très variés (ex. moyenne, max, médiane) et dépendent de l'objectif du filtrage (réduction du bruit, rehaussement de contours, lissage) (Figure 7-2(b)). La figure 7-2b illustre le filtrage médian. La valeur du pixel central (au centre de la fenêtre carrée 3x3) égale à 2 à l'origine a été remplacée par la médiane des valeurs de pixels dans sa fenêtre de voisinage.

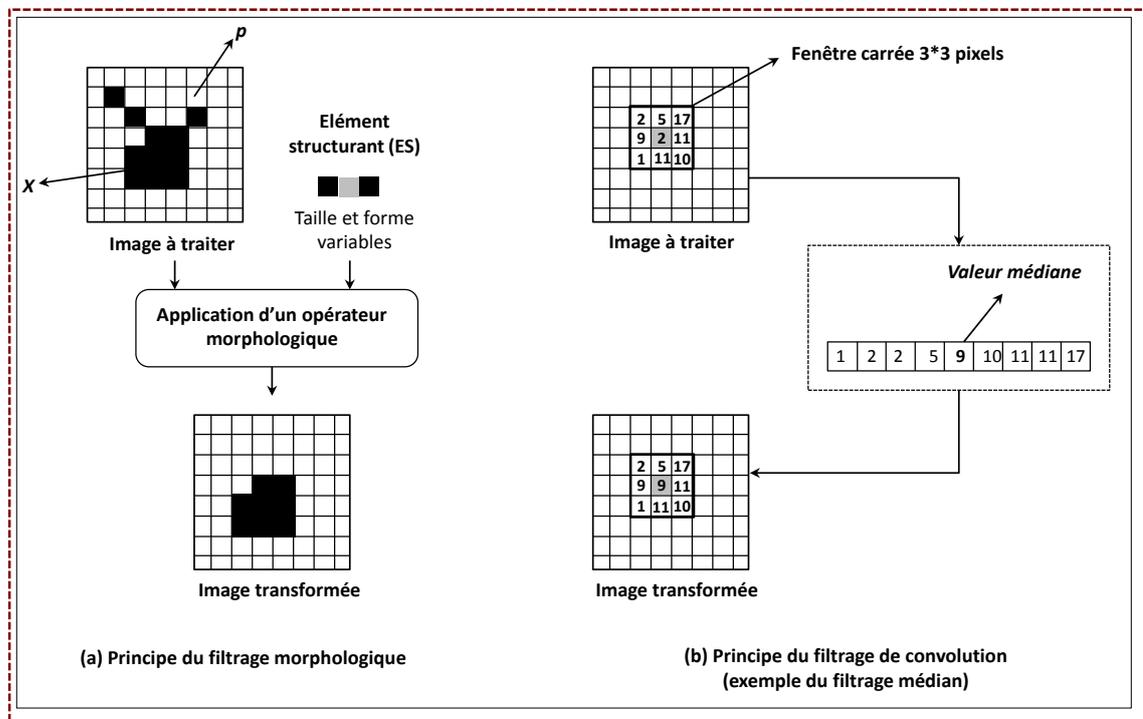


Figure 7-2. Principe du filtrage morphologique (érosion) (a) et du filtrage de convolution médian (b) (Source : Sheeren et al. 2007, adapté).

Ansoult et al. (1990) ont proposé une étape de filtrage morphologique dans une démarche de traitement automatisée d'une carte thématique. Ils ont montré comment l'opérateur de *dilatation* et le processus de *reconstruction morphologique* pouvaient être utilisés successivement pour supprimer les éléments textuels surchargeant les régions dans une carte des sols. L'approche, décrite dans cette étude, a l'avantage de fournir un cadre méthodologique général pouvant s'appliquer à d'autres cartes thématiques. Chiang et al. (2013) ont aussi adopté les filtres morphologiques dans une étape de prétraitement pour vectoriser une carte topographique scannée de l'USGS (United States Geological Survey). Ils ont appliqué plusieurs érosions successives pour supprimer les courbes de niveau avant de procéder à une extraction des voies routières. Une approche similaire a été conduite par Samet et al. (2010) mais cette fois, dans une phase de post-traitements. Un opérateur de dilatation a été appliqué pour reconstruire les lignes de contour des objets vectorisés et réduire le

bruit dans l'image après une étape de segmentation basée sur la couleur. Iwanowski et Kozak (2012) ont appliqué cette procédure de nettoyage en deux temps dans leur procédure d'extraction. Ces derniers ont utilisé un opérateur morphologique pour reconstruire les objets extraits avant d'appliquer un filtrage de convolution permettant de supprimer les pixels isolés apparus au cours de la première étape.

Cette première sous-section a permis de décrire les approches employées pour réduire le bruit dans les cartes anciennes et ainsi faciliter la mise en place de procédure d'extraction automatisée. Nous nous focalisons par la suite sur les méthodes d'extraction des éléments proprement dites.

Méthodes d'extraction existantes

Les travaux sur l'automatisation de l'extraction des objets dans les cartes historiques portent sur différents éléments tels que (a) le texte (toponymes principalement), (b) les régions (occupation du sol, bâtiments, cours d'eau...), (c) les symboles (objets ponctuels comme les églises, les moulins...) ou (d) les lignes (courbes de niveau, hachures, contours d'objets, routes, fleuves...). Etant donné la grande variété des représentations, il est souvent difficile de définir une méthode générique et bien souvent, les techniques proposées sont spécifiques à la carte et à l'objet recherché. Nous détaillons ci-dessous quelques approches existantes.

L'extraction des éléments textuels dans les cartes anciennes est particulièrement ardue à automatiser. En effet, le texte vient systématiquement se surimposer aux autres éléments de la carte, ce qui peut modifier la forme des caractères. Par ailleurs, les polices peuvent être écrites à la main rendant la reconnaissance des caractères plus délicate puisque l'écriture n'est pas homogène sur la zone d'intérêt. Cao et Tan (2002) ont proposé une méthode basée sur le constat que les chaînes de caractères ont généralement de plus courts segments que ceux des autres éléments cartographiques. Les auteurs ont alors proposé de s'appuyer sur un ratio longueur/surface des lignes présentes dans l'image pour distinguer les types de lignes et permettre une reconnaissance des motifs textuels. Le texte peut aussi être extrait grâce à la couleur, lorsque la couleur des caractères diffère de celle des autres objets. C'est l'approche adoptée par (Centeno 1998) qui a pu isoler le texte en convertissant l'espace colorimétrique initial d'une carte actuelle par la transformée de Karhunen-Loeve. Enfin, Myers et al. (1996) ont développé une approche de reconnaissance automatique du texte fondée sur une base de connaissances a priori des objets textuels dans une carte topographique scannée. Cette approche s'appuie sur des méthodes de reconnaissance de caractères et a l'avantage de s'affranchir de l'étape de segmentation. Les résultats se sont avérés très convaincants même si des erreurs de confusion peuvent apparaître avec les symboles.

Concernant les régions, les méthodes existantes de reconnaissance automatique sont principalement basées sur la segmentation d'images ou les techniques de classification. Shaw et Bajcsy (2011) ont utilisé une combinaison d'approches de segmentation par croissance de région et d'opérateurs morphologiques pour extraire le lac Ontario de plusieurs cartes anciennes produites entre le 16^e et le 19^e siècle. Les résultats montrent que cette technique est limitée lorsqu'il y a de nombreux éléments superposés. Mahmoud (2012) a proposé une approche similaire comprenant une étape de segmentation multi-résolution et une étape de classification supervisée basée à la fois sur la couleur et des attributs spatiaux (distance au plus proche voisin et critères d'adjacence) pour extraire le réseau hydrographique et la végétation d'une carte topographique du Nigéria produite entre 1968 et 1973. D'autres travaux s'affranchissent de l'étape de segmentation et tentent de reconnaître directement les objets à travers une classification des pixels. Si les cartes contiennent des aplats de couleurs (une couleur ayant une signification thématique), la reconnaissance de la couleur peut suffire à extraire les régions (Leyk et Boesch 2010). Cette approche est toutefois dépendante de l'homogénéité des teintes, et d'une manière plus générale, de la qualité des cartes (Leyk et Boesch 2010). Certains auteurs utilisent de fait des algorithmes de classification non supervisés (tel que les *K-moyennes* ou *K-means* en anglais). C'est le cas de Chiang et al. (2013) qui

s'intéressent à l'extraction de routes représentées sous la forme de régions dans une carte topographique scannée fournie par l'USGS (United State Geological Survey). Les résultats montrent que l'algorithme est efficace mais que les utilisateurs ont besoin de déterminer *a priori* un grand nombre de classes (clusters) pour séparer les différentes entités et isoler les routes. Henderson et al. (2009) ont comparé la performance des algorithmes *K-moyennes* et EM (*Expectation-Maximization* ou *Mélange gaussien*) pour cartographier les différentes classes présentes sur une carte topographique scannée de l'USGS. L'algorithme des *k-moyennes* est celui qui fournit les meilleurs résultats. Enfin, les méthodes de seuillage sont aussi utilisées par plusieurs auteurs pour séparer les éléments en niveaux de gris des éléments colorés. Cette technique est notamment employée comme traitement préliminaire avant de procéder à une segmentation de l'image (Ansoult et al. 1990; Chi et Yan 1993). Elle a l'avantage d'effectuer une première séparation des objets sur laquelle il est possible de s'appuyer pour affiner la segmentation.

L'extraction des symboles est aussi une question délicate dans le traitement automatique de cartes anciennes. En effet, ces derniers peuvent avoir des représentations très variées et non standardisées qui rendent difficile la mise en œuvre d'une procédure automatisée. Une majorité de travaux emploie les classifications à dire d'expert, basées sur un ensemble de règles décrivant les objets présents pour conduire l'extraction de symboles (Arias et al. 1993; Deseilligny et al. 1998). Ces règles décrivent la forme des symboles ou la longueur de leur contour. D'autres approches s'appuient sur une base de motifs à rechercher dans l'image (*Template matching*, (Samet et Soffer 1998)). Il est aussi possible de considérer les symboles comme de courts segments et s'appuyer sur leurs caractéristiques de longueur pour les reconnaître (Boesch 1996). Enfin, les travaux du projet Geopeuple ont eu pour objectif d'extraire automatiquement les symboles (corps de garde, chapelles..) dans la carte de Cassini. L'approche utilisée s'appuie sur une méthode de classification supervisée (Support Vector Machine) pour cartographier les symboles à la fois décrits par des descripteurs visuels (couleur, texture, forme) et des descripteurs contextuels (contexte géographique, géométrique, sémantique) (Guyomard 2013).

Enfin, les entités linéaires telles que les contours, les hachures, les voies de communication ou les cours d'eau ont aussi fait l'objet de procédures d'extraction automatisées. Kaneko (1992) s'est appuyé par exemple sur un opérateur de distance directionnelle pour mesurer la largeur et la longueur des éléments graphiques dans une carte actuelle scannée. La méthode prévoit ensuite de caractériser chaque ligne par des indicateurs géométriques avancés pour finalement les regrouper sur la base de ces indices. Toutefois, la majorité des approches conduites s'appuie sur le constat que les lignes sont généralement représentées par des teintes sombres facilement distinguables des autres éléments sur la carte. Dans ce cas, le seuillage, la segmentation multi-échelle et de manière générale les techniques employées pour l'extraction de régions sont utilisées et fournissent généralement de très bons résultats (Ansoult et al. 1990; Arrighi et Soille 1999).

Cette revue des approches existantes pour automatiser l'extraction des éléments dans les cartes scannées permet d'entrevoir la démarche à adopter. Pour rappel, les forêts dans les dessins-minutes d'Etat-major sont représentées sous la forme d'aplats colorés (du jaune au vert) le plus souvent bordées par un trait noir épais.

Premièrement, l'utilisation de filtres de lissage dans une phase de prétraitement semble être une solution efficace pour nettoyer la carte et réduire le bruit lié à la superposition des éléments planimétriques. Ils peuvent être aussi utilisés pour corriger les résultats d'extraction en phase de post-traitement. Deux types de filtre sont généralement appliqués : les filtres morphologiques et les filtres de convolution.

Ensuite, les méthodes d'extraction, à proprement parler, varient selon le type d'objet étudié. Dans notre cas, les forêts sont représentées sous la forme d'aplats colorés qu'il est *a priori* utile d'utiliser.. Les travaux sur l'extraction des régions basés sur la couleur proposent soit de s'appuyer

sur des techniques de seuillage, soit d'appliquer des algorithmes de classification supervisée ou non supervisée. Le seuillage semble inadapté car la teinte des forêts varie assez fortement, il sera par conséquent nécessaire d'adapter le seuil à chaque carte ou à chaque région d'une carte. Les algorithmes non-supervisés type *K-moyennes* pourraient être adaptés car ils offrent la possibilité de faire varier le nombre de classes et de prendre en compte la variété des représentations. L'expertise de l'utilisateur sera toutefois requise pour regrouper les classes dans un second temps si jamais il est nécessaire de les agréger. Enfin, les algorithmes supervisés impliquent de construire une base d'apprentissage et donc de tenir compte aussi des différentes représentations. L'étape de sélection des échantillons impliquera l'intervention de l'utilisateur qui devra être en mesure d'identifier la classe d'intérêt a priori pour construire cette base d'apprentissage.

7.2.2 DEMARCHE GENERALE PROPOSEE

La démarche de vectorisation automatique que nous proposons concerne les minutes d'Etat-major de 1850. Elle s'inspire des travaux existants ainsi que de l'examen visuel minutieux que nous avons pu réaliser sur ces documents pour définir une démarche la plus appropriée.

Ainsi, comme évoqué précédemment (chapitre 4, section 4.2.3), les minutes d'Etat-major sont en couleur. Cela signifie que nous pourrions nous appuyer sur cette information pour extraire automatiquement les forêts de ces documents. Cette couleur n'est toutefois pas uniforme pour un même aplats ce qui pourrait poser problème dans la phase d'extraction. Par ailleurs, de nombreux objets planimétriques se superposent et les hachures représentant l'orographie peuvent assombrir les forêts. Enfin, il existe une importante diversité de représentation des forêts dans les cartes (cf. chapitre 4). La démarche développée devra tenir compte de tous ces points.

La méthodologie d'extraction proposée ici se déroule en quatre étapes principales que nous prendrons soin de détailler dans les sous-sections qui suivent (figure 7-3):

- (1) Les prétraitements par filtrage qui visent à éliminer les hachures de relief tout en préservant le contour et la couleur.
- (2) La conversion de l'espace RGB vers l'espace CIElab qui a pour objectif d'améliorer l'uniformité de l'image et d'accroître la distinction des éléments.
- (3) La phase de classification qui vise à reconnaître les forêts dans les minutes. Plusieurs méthodes de classification sont proposées.
- (4) Les post-traitements par filtrage qui permettent d'améliorer le résultat de la classification en supprimant des artefacts.

Les résultats d'extraction automatique seront évalués par comparaison avec les objets extraits manuellement (BD Forêts MANUEL) sur la totalité de la zone d'étude.

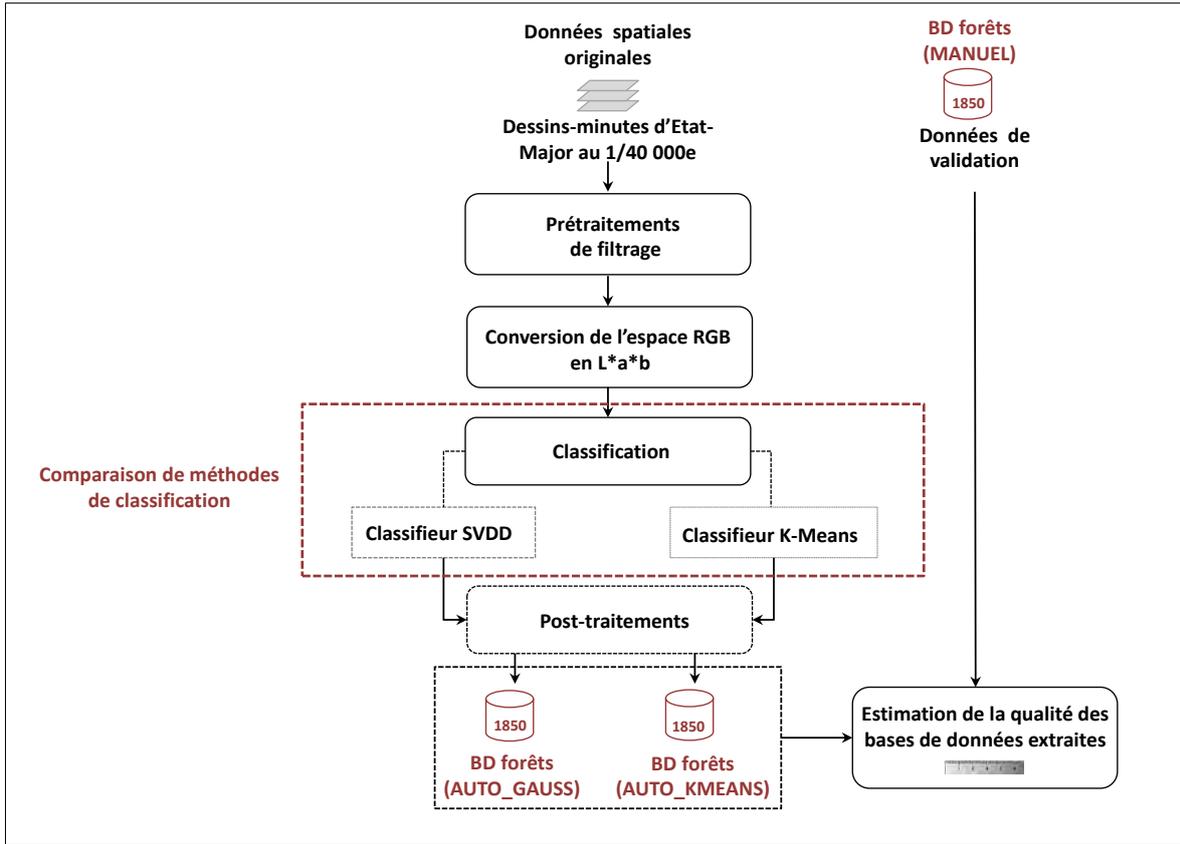


Figure 7-3. Démarche de vectorisation automatique pour extraire les forêts des dessins minutes d'Etat-major

7.2.3 PRETRAITEMENTS ET FILTRAGE DES ELEMENTS SURCHARGEANT LA FORET

Avant de procéder à une classification de l'image basée sur la couleur, une étape de nettoyage est appliquée pour éliminer les hachures de relief qui se superposent aux forêts. L'objectif est d'éliminer ces lignes tout en préservant les contours et la couleur des régions présentes sur la carte. Trois processus successifs ont été appliqués : (1) une dilatation morphologique, (2) un filtrage médian et (2) un filtrage passe-bas (Figure 7-4).

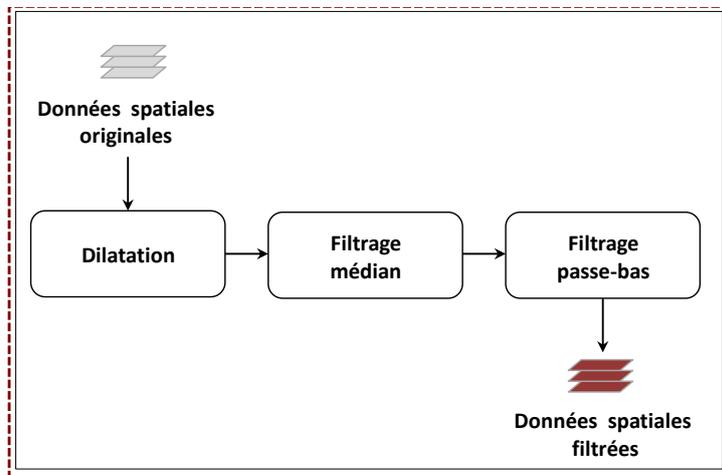


Figure 7-4. Trois étapes successives de filtrage dans la phase de prétraitements

La dilatation morphologique (Figure 7-5(1)) est dédiée au remplissage des « trous » dans les forêts créés par le texte, les symboles, les hachures de relief ou les routes. Nous avons choisi d'utiliser un opérateur de dilatation avec une fenêtre d'analyse de taille 5x5 pixels. Cet opérateur a été appliqué sur chaque bande de l'image (Rouge, Vert et Bleu) en niveaux de gris. L'histogramme associé à la figure 7-5b montre qu'une grande part des pixels situés dans les basses fréquences (niveau de gris <100) ont été supprimés.

Le filtrage médian (Figure 7-5(2)) est ensuite mis en œuvre pour éliminer les hachures de relief résiduelles tout en préservant les contours et la couleur. Contrairement à un filtre moyen, ce filtre médian présente l'intérêt de respecter les contours des objets. Nous avons donc choisi de l'utiliser dans une fenêtre de 5x5 pixels sur chaque bande dilatée (Rouge, Vert et Bleu). Le nombre d'applications du filtre varie en fonction de la densité des hachures présente sur la carte et donc du relief de la zone. Ce nombre est déterminé de manière empirique. Cette étape a permis de compléter la suppression des pixels dans les basses fréquences (Figure 7-5(c)).

Un filtre passe-bas de taille 5x5 pixels est enfin appliqué pour supprimer le bruit de fond persistant (pics apparents dans les basses fréquences de l'histogramme de la figure 7-5(c)). Il présente l'avantage de supprimer les détails d'une image en conservant les basses fréquences. Il en résulte une image corrigée illustrée par l'histogramme (d) où les pics qui apparaissaient dans les basses fréquences ont été lissés.

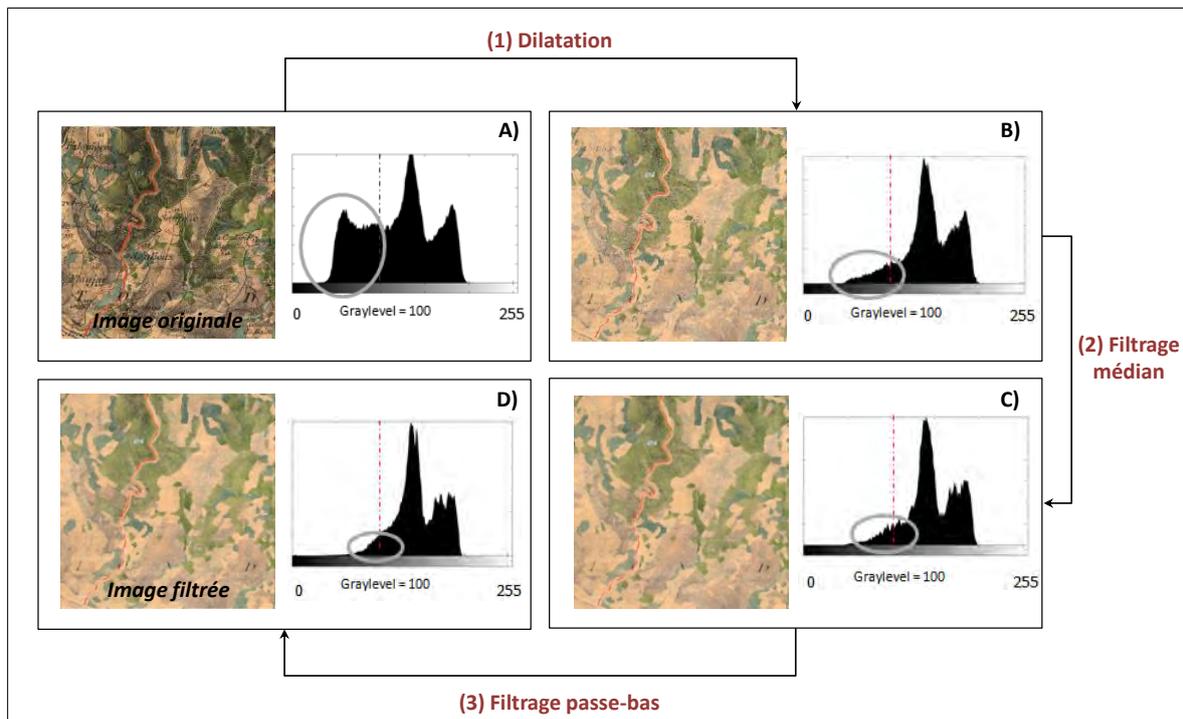


Figure 7-5. Illustration des corrections apportées par les différentes étapes de filtrage

7.2.4 TRANSFORMATION DE L'ESPACE COLORIMETRIQUE : DU RVB AU CIELAB

Les étapes précédentes nous ont permis de filtrer les objets en vue de faciliter leur classification cependant les couleurs ont été légèrement altérées durant ce processus (Figure 7-5). En effet, en fonction du relief présent dans la zone, la surcharge des hachures n'est pas la même et le filtrage peut conduire à assombrir les régions plus accidentées. D'une manière plus générale, la couleur est rarement homogène dans les cartes historiques, comme déjà évoqué, et il est donc

nécessaire de prendre en compte les différences de teinte et d'intensité dans la procédure de vectorisation.

Pour résoudre ce problème, une réflexion sur l'espace colorimétrique le plus adapté s'est imposée. Le choix d'un espace colorimétrique conditionne souvent la qualité du résultat final (Iwanowski et Kozak 2012). L'espace traditionnellement utilisé est l'espace RVB qui présente certaines limites pour la question qui nous occupe : la luminosité n'est pas uniforme en plus du fait que cet espace est plus éloigné de la perception humaine (Angulo et Serra 2003). Ainsi, les espaces colorimétriques cylindriques dans lesquels la luminosité est isolée sont mieux adaptés à la reconnaissance d'objets dont la luminosité n'est pas uniforme (Salvatore et Guitton 2004).

Dans notre approche, nous proposons de transformer l'image initiale dans l'espace colorimétrique CIELab (ou L*a*b) (Dorff et al. 1998). Ce choix s'inspire de travaux précédents montrant que le CIELab est bien approprié aux images bruitées colorées (Angulo et Serra 2003; Ganesan et al. 2010).

Dans l'espace CIELab, la composante de clarté de la couleur est reportée sur l'axe L (*Luminosity*), variant de 0 à 100, et perpendiculaire à une pile de plan « ab » (Figure 7-6). Les composantes « a » et « b » sont définies par une gamme de valeurs variant respectivement du rouge au vert en passant par le gris, et du jaune au bleu en passant par le gris. Chacun de ces plans contient ainsi toutes les couleurs possibles pour une valeur de clarté. Dans notre cas, nous pouvons avoir plusieurs valeurs de clarté (i.e. plusieurs valeurs de L) pour une même couleur (ici le vert pour les forêts). La reconnaissance de ces objets est alors fondée sur la recherche des différentes partitions de vert existant pour chaque valeur possible de L. Cette recherche est effectuée à l'aide d'un algorithme de classification que nous détaillons-ci après.

Pour passer de l'espace RVB au L*a*b, une étape première étape consiste à passer des composantes RVB aux composantes XYZ. Le système L*a*b étant défini à partir de ce dernier système (<http://www.tsi.telecomparistech.fr>).

On utilise pour cela une matrice de conversion pour résoudre le système suivant :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{bmatrix} 0.618 & 0.177 & 0.205 \\ 0.299 & 0.587 & 0.114 \\ 0 & 0.054 & 0.3944 \end{bmatrix} \cdot \begin{pmatrix} R \\ V \\ B \end{pmatrix}$$

Ensuite, la seconde étape implique de passer de l'espace XYZ à l'espace L*a*b. Les formules suivantes doivent être appliquées :

$$L = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16 \quad \text{pour } \frac{Y}{Y_n} > 0.008856$$

$$L = 903.3 \left(\frac{Y}{Y_n} \right) \quad \text{pour } \frac{Y}{Y_n} < 0.008856$$

$$a = 500 \left(f \left(\frac{X}{X_n} \right) - f \left(\frac{Y}{Y_n} \right) \right)$$

$$b = 200 \left(f \left(\frac{X}{X_n} \right) - f \left(\frac{Y}{Y_n} \right) \right)$$

$$\text{pour } t > 0.008856 \quad f(t) = \sqrt[3]{t}$$

$$\text{pour } t < 0.008856 \quad f(t) = 7.7787t + \frac{16}{116}$$

Xn, Yn et Zn correspondent aux valeurs du blanc dans l'espace XYZ. On les obtient pour RVB = (255, 255, 255).

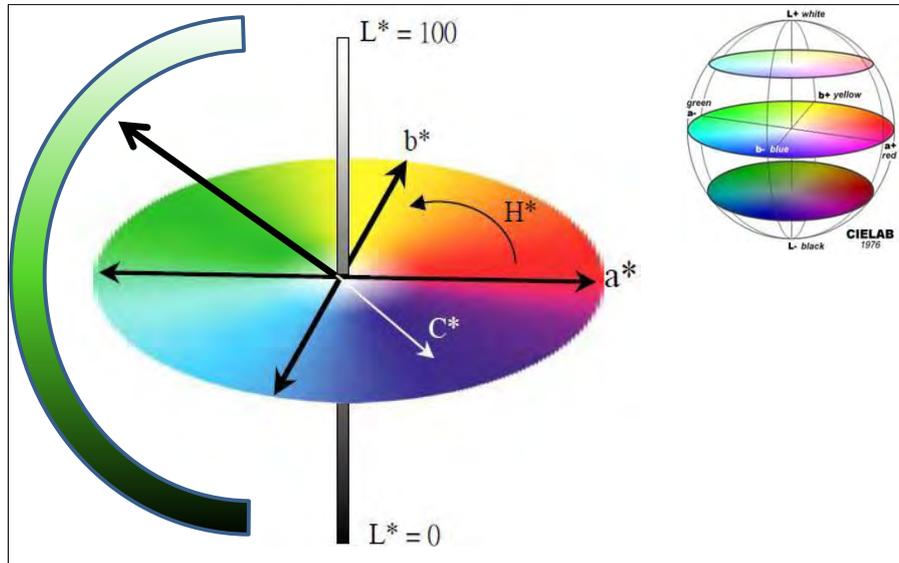


Figure 7-6. L'espace colorimétrique CIE Lab: L exprime la clarté (Luminance) qui varie de 0 (noir) à 100 (blanc) ; a et b expriment une gamme de couleurs définies par un mélange de vert (-a) à magenta (+a) et un mélange de bleu (-b) à jaune (+b) ; toutes les couleurs de même luminosité s'inscrivent dans un plan circulaire traversé par les axes a* et b* (source : http://www.bonne-mesure.com/cie_lab.php)

7.2.5 RECONNAISSANCE AUTOMATIQUE DES FORETS PAR CLASSIFICATION SUPERVISEE ET NON SUPERVISEE

La classification est un processus permettant d'organiser et regrouper des données (ici des pixels) qui se ressemblent selon un ou plusieurs critères choisis. Cette classification peut être de deux types : (1) non-supervisée et (2) supervisée (Rochon 1992).

(1) La classification non-supervisée permet d'identifier des groupes de pixels aux valeurs similaires et relativement bien séparés dans un ensemble de données (ici une image). On cherche ensuite à affecter à chaque pixel un label (ou valeur de classe) qui matérialise l'appartenance de ce pixel à un des groupes identifiés préalablement. La particularité de cette approche repose sur le fait qu'aucune connaissance *a priori* n'est introduite dans la méthode pour constituer les classes. Celles-ci sont générées automatiquement par l'algorithme et doivent être interprétées *a posteriori*.

(2) A l'inverse, la classification supervisée s'appuie des observations déjà labélisées (échantillons d'apprentissage). La méthode vise alors à affecter à une nouvelle observation une classe prédéfinie dont la distribution est apprise à partir des données d'apprentissage. La particularité de cette approche est donc la sélection des données d'apprentissage et l'obtention d'informations *a priori* sur la composition des classes dans l'image.

Dans cette logique, nous avons choisi de comparer deux méthodes de classification qui suivent ces deux approches pour le problème qui nous préoccupe : la reconnaissance automatique des forêts. Chaque approche présente des avantages et inconvénients (notamment du point de vue de la rapidité d'exécution) et nous souhaitons comparer leur performance. Ainsi, deux algorithmes

classiques ont été retenus : (1) l'algorithme des *k-moyennes*, approche non-supervisée et (2) le modèle de mélange gaussien, approche supervisée

Le Classifieur non supervisé : modèle des K-moyennes

L'algorithme des K-moyennes, aussi appelé « nuées dynamiques », est une méthode de classification automatique qui sépare les données en K classes. Il consiste dans un premier temps (phase d'initialisation) à positionner aléatoirement K noyaux dans l'espace des variables. Ces noyaux vont constituer les centres des K classes. Chaque observation (pixel) est alors affectée au noyau le plus proche sur la base d'un critère de distance, généralement euclidienne. Une fois tous les pixels associés à un noyau, la position centrale de ce noyau (centre de gravité) est recalculée. On réaffecte alors chaque pixel au nouveau noyau le plus proche. Cette procédure est répétée jusqu'à convergence (partitions stables qui ne modifie plus la position des noyaux) ou atteinte d'un critère d'arrêt. L'avantage de cette approche est qu'elle est rapide à mettre en œuvre et ne nécessite pas d'échantillons d'apprentissage. Les limites reposent sur le fait qu'il faut choisir *a priori* le nombre de classes K. Par ailleurs, l'algorithme peut converger vers un minimum local et le résultat dépend de l'initialisation aléatoire. Il faut donc l'exécuter plusieurs fois.

L'algorithme des K-moyennes classe les données à l'aide de la règle de décision suivante :

$$x \in C_i \text{ si } d(x, \mu_i)^2 < d(x, \mu_j)^2 \quad \forall j \neq i \quad \text{avec} \quad \mu_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i$$

où x est le pixel à classer, μ_i et μ_j les centres des classes i et j et n le nombre de pixels à classer. La distance utilisée pour mesurer l'écart d'un pixel à un centre de classe est la distance euclidienne notée d .

Dans une démarche de classification d'images, deux paramètres de la règle exposée ci-dessus doivent être définis par l'utilisateur : le nombre de classes et le nombre d'itérations pour faire en sorte que l'algorithme converge. Le nombre de classes est souvent surestimé par l'utilisateur dans le but d'éviter les classes mélangées (c'ad contenant des pixels aux valeurs très éloignées). Une étape de regroupement des classes peut alors être nécessaire dans un second temps pour agréger les classes et répondre aux objectifs spécifiques.

Le Classifieur supervisé : modèle de mélange gaussien

Le modèle de mélange gaussien (que l'on nomme souvent *maximum de vraisemblance* en télédétection) est une méthode d'apprentissage supervisé très répandue (Rochon 1992; Zammit 2008; Mather et Koch 2011). Contrairement aux K-moyennes où une observation est affectée à la classe dont le centre de gravité est le plus proche, le classifieur gaussien utilise des échantillons d'apprentissage pour modéliser la distribution de chaque classe selon une loi de probabilité supposée gaussienne (loi normale). Le modèle global pour l'image est donc un modèle de mélange gaussien (somme de gaussiennes relatives à chaque classe). La règle de classification sur laquelle s'appuie le mélange gaussien cherche à maximiser la probabilité qu'un pixel x appartienne à la classe C_i tel que :

$$x \in C_i \text{ si } p(C_i|x) > p(C_j|x), \quad \forall j \neq i$$

Le maximum à posteriori est défini grâce à l'emploi du théorème de Bayes :

$$p(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)}$$

où :

- $p(x|C_i)$ est la fonction de vraisemblance de C_i soit la probabilité d'observer une valeur du pixel x pour une classe C_i .

- $p(C_i)$ est la probabilité à priori de C_i soit la probabilité d'occurrence de la classe C_i dans l'image. Cette probabilité ainsi que la précédente ($p(x|C_i)$) sont définies à partir des échantillons d'apprentissage.

- $p(x)$ est la probabilité à priori de x soit la probabilité que la valeur du pixel x soit présente dans l'image. Cette probabilité est non discriminante et commune toutes les classes.

La règle de décision finale s'écrit de la manière suivante :

$$x \in C_i \text{ si } p(x|C_i)p(C_i) > p(x|C_j)p(C_j), \quad \forall j \neq i$$

Dans une démarche de classification d'image, l'utilisateur doit avoir une connaissance à priori du nombre de classes à représenter. Il doit lui-même sélectionner les échantillons d'apprentissage afin de déterminer la fonction de vraisemblance et la probabilité à priori de C_i . Plus le nombre d'échantillons est important, plus l'algorithme pourra calculer précisément ces différentes probabilités et améliorer sa robustesse.

7.2.6 TRAITEMENTS POST-CLASSIFICATION

A l'issue de la détection des forêts effectuées à l'étape précédente (image classée, figure 7-7), un post-traitement a été réalisé pour corriger certains artéfacts.

Tout d'abord, la première étape de post-traitement a consisté à boucher les « trous » persistants au sein des forêts. Pour ce faire, nous avons opté pour une labélisation automatique de chaque groupe de pixels contigus obtenus de même classe (dénommé ci-après région). Chaque région se voit attribuer un numéro différent dans l'image (ce qui reviendrait en mode vecteur, à fournir un identifiant unique à chaque objet). Ainsi, les pixels classés en 0 au sein des forêts (classés en 1) se sont vu attribuer la classe 1. Puis, l'algorithme recherche les régions totalement enclavées dans les forêts avant de les reclassifier en forêts (en classe 1). La figure 7-7 illustre les résultats de ce processus. Les régions « non-classées » (en noir sur l'image classée) au sein de la forêt principale ont été affectées à la même classe que cette même forêt sur l'image classée et corrigée.

Une seconde étape a consisté à éliminer les pixels de forêt isolés. Afin de réaliser cette tâche, nous nous sommes appuyés sur un opérateur morphologique d'ouverture basée sur un élément structurant de forme carrée de 3*3 pixels. Les pixels isolés au sud de la forêt principale sur l'image classée ont par exemple été supprimés sur l'image corrigée (Figure 7-7).

Etapes

Etape 4 : Post-traitements



Image originale

Image classée

Image classée et corrigée

Figure 7-7. Illustration des corrections apportées par les processus de prétraitements : les trous enclavés au sein des forêts ont été bouchés et les pixels isolés supprimés.

7.2.7 IMPLEMENTATION DE LA METHODE AUTOMATIQUE DANS LA PLATEFORME ENVI/IDL

Nous avons initialement développé cette démarche dans l'environnement MatLab R2009B. Toutefois, afin de diffuser plus largement la méthode et la rendre exploitable pour une communauté plus large, elle a été réimplémentée en langage IDL sous la forme d'une brique logiciel dans la plateforme de traitement d'images ENVI 4.7 (Loussert et Guerec 2014). Chaque étape de la méthode peut être exécutée à partir du menu déroulant de l'outil « *Historical Maps* » faisant appel à des fonctions utilisables de manière indépendante (figure 7-8). Cela permet à l'utilisateur, s'il le souhaite, d'adapter la démarche selon sa propre source ou ses objectifs d'extraction.

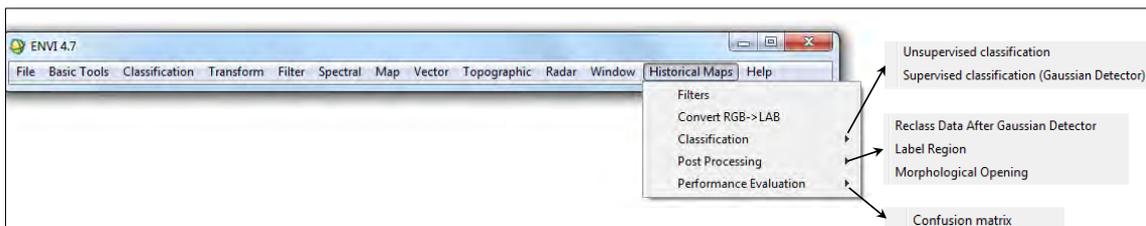


Figure 7-8. Aperçu de l'interface « *Historical Maps* » disponible sous ENVI/IDL. Cette extension regroupe la totalité des outils nécessaires à la mise en place de la démarche d'extraction proposée.

L'outil est disponible à cette adresse : <http://faudel.mathieu.free.fr/software.html>.

7.2.8 EXPERIMENTATION ET MISE EN ŒUVRE SUR LE SITE D'ETUDE

Dans un premier temps, dans le but de mettre au point la démarche et d'estimer ses performances, nous l'avons testée sur trois extraits des minutes de l'Etat-major (Figure 7-9). Ces trois extraits disposent d'une taille identique (1500 x1500 pixels) et présentent des représentations différentes des surfaces boisées (surcharge de hachures, toponymes, etc.).

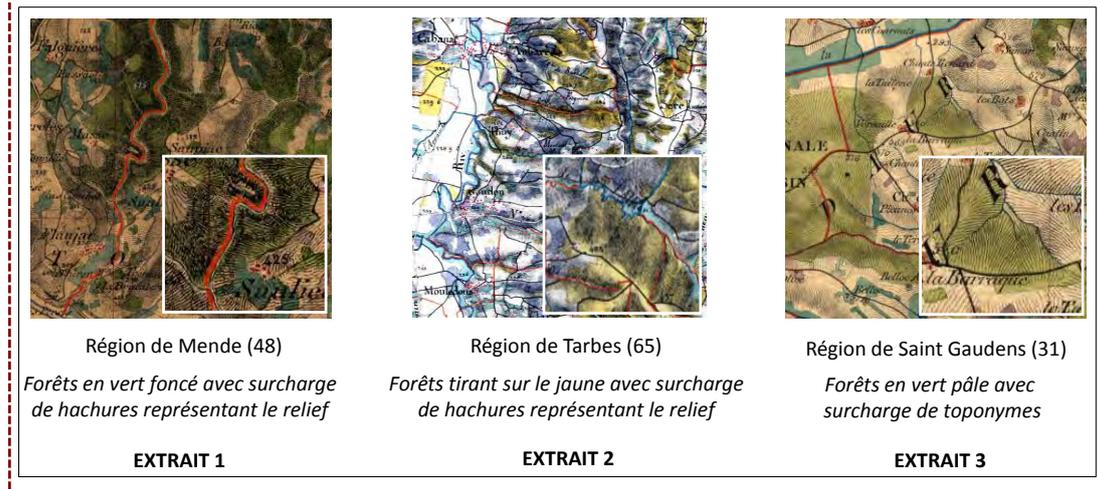


Figure 7-9. Trois extraits sélectionnés pour l'application de la méthode proposée. L'extrait 1 est localisé dans la région de Mende et se caractérise par un relief marqué ainsi que des espaces boisés représentés en vert foncé. ; l'extrait 2 est situé dans la région de Tarbes avec des forêts représentées en jaune et une présence importante de hachures (relief) ; l'extrait 3, dans la région de Saint Gaudens, est caractérisé par un relief moins marqué, des forêts colorées en vert pâle et une surcharge planimétrique causé par les toponymes.

Pour chacun des extraits, nous avons appliqué deux fois la démarche proposée pour identifier automatiquement les forêts : l'une avec la méthode du mélange gaussien, l'autre avec l'algorithme des K-moyennes. Ensuite, la méthode a été utilisée pour extraire automatiquement les forêts sur l'ensemble du site d'étude présenté en chapitre 3 et ce, à partir du classifieur supervisé.

La validation des résultats d'extraction automatique a été effectuée par comparaison avec les couches de forêts vectorisées manuellement (cf. chapitre 6). La performance a été obtenue en calculant une matrice de confusion à partir de laquelle il est possible de déduire plusieurs indicateurs de qualité tels que (Figure 7-10) (Rochon 1992).

- La précision globale P_0 : proportion de pixels bien classés par rapport au nombre de pixels évalués. Cet indicateur n'informe pas sur la distribution des erreurs par classe.
- La précision du producteur PP : probabilité qu'un pixel de classe C_i sur le terrain soit classé en C_i dans l'image. Cet indicateur informe sur les taux d'erreur d'omission.
- La précision de l'utilisateur PU : probabilité qu'un pixel de classe C_i dans l'image appartienne à la classe C_i sur le terrain. Cet indicateur informe sur les taux d'erreurs de commission
- L'indice Kappa (K) : taux d'accord (P_0) corrigé de l'effet du hasard (P_E). Ce coefficient est compris entre -1 et 1.

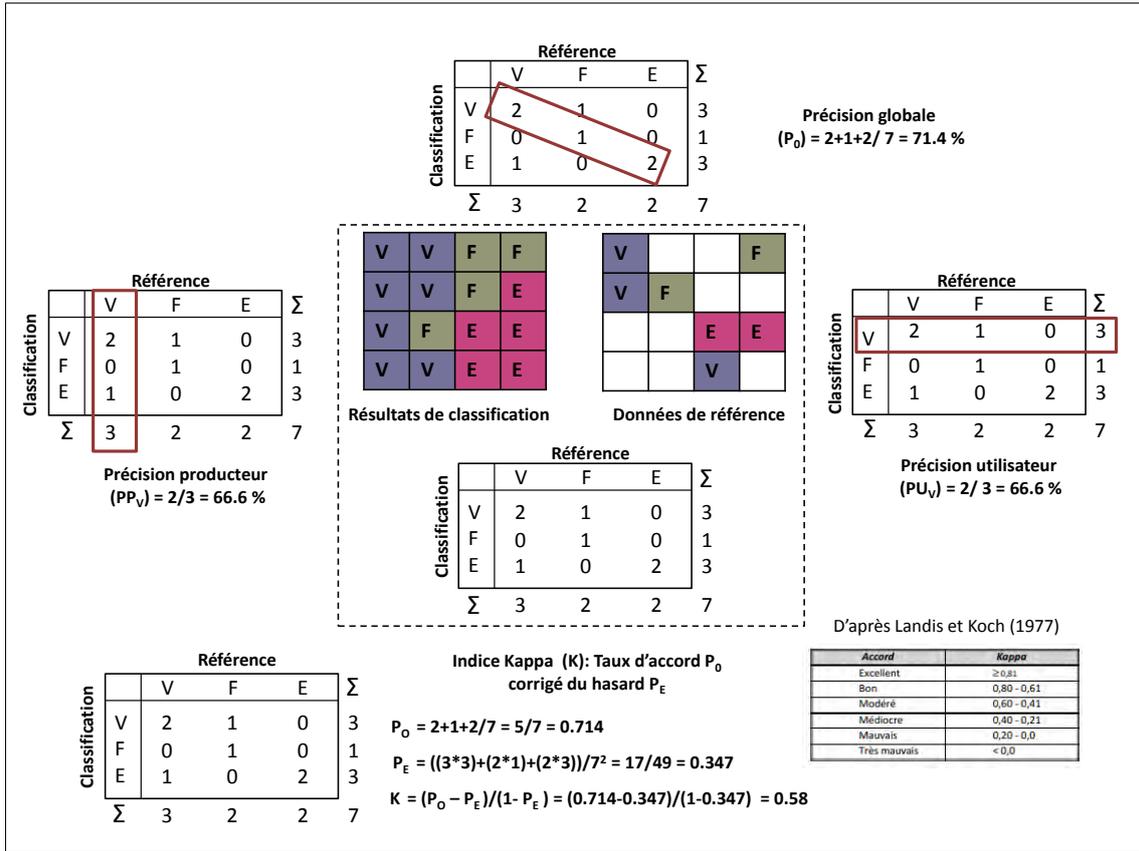


Figure 7-10. Description des indicateurs de qualité dérivés d'une matrice de confusion

7.2.9 RESULTATS ET DISCUSSION

Nous présentons ici les résultats d'application de la démarche proposée. Premièrement, nous analyserons et discuterons des résultats obtenus à partir des trois extraits sélectionnés. Deuxièmement, nous étudierons la qualité de la base de données extraite automatiquement sur l'ensemble du site d'étude.

A PARTIR DES TROIS EXTRAITS SELECTIONNES

Avant de procéder à la description des résultats obtenus, rappelons que **les forêts extraites automatiquement ont été validés par comparaison avec les forêts extraites manuellement sur la totalité de la zone étudiée.**

D'une manière générale, on observe une très bonne performance de la démarche de vectorisation proposée quel que soit le classifieur utilisé avec une précision globale de classification Forêt/Non forêt variant approximativement de 94 à 98% et un indice Kappa supérieur à 0.86 (Tableau 7-1). Si on se focalise sur la qualité de l'extraction des forêts proprement dite, on observe également des taux d'accord élevés (>83%) avec des erreurs de déficit et d'excédent variant de 0,37 à 16%. Les résultats concernant l'extrait 3 sont toutefois supérieurs à ceux obtenus à partir des deux autres extraits. Pour ce dernier, on constate une précision globale d'environ 98% et un indice Kappa supérieur à 0.94. Globalement, les taux de déficit dépassent les taux d'excédent, la tendance est alors à la sous détection.

Concernant les performances des deux classifieurs, les résultats sont très proches d'une méthode à une autre mais différent quelque peu selon le contexte orographique de la carte. Pour

l'extrait 2, les résultats sont similaires puisque les deux méthodes permettent d'obtenir une précision globale d'environ 96% et un indice Kappa de 0.90. Cependant, le classifieur gaussien offre de meilleurs résultats que la méthode des K-moyennes sur l'extrait 1 avec un gain de 2% en termes de précision globale et un meilleur indice Kappa ($0.91 > 0.86$). Il permet surtout d'améliorer considérablement le déficit de la classe « forêt » avec une différence de 9%. Enfin, la méthode des K-moyennes est plus performante que le classifieur gaussien pour l'extrait 3. L'écart de performance entre les deux méthodes est très faible du point de vue de la précision globale (1% d'écart) et légèrement plus important si l'on observe l'indice Kappa ($0.97 > 0.94$). Le taux de déficit de la classe « forêt » est en revanche deux fois inférieur avec la méthode des K-moyennes pour cet extrait. L'ensemble des matrices de confusion est disponible en annexe C.

La performance plus élevée obtenue pour l'extrait 3 s'explique assez facilement au regard de la carte. La zone géographique couverte présente moins de relief et donc de hachures surchargeant les forêts. Pour les deux autres extraits, les erreurs de déficit sont plus importantes. Ces défauts sont principalement liés à l'étape de filtrage réalisée en amont ou en aval de la détection des forêts. Ainsi, les déficits se caractérisent non pas par l'absence complète de certaines forêts mais plutôt par le manque de parties d'objets (en lisière notamment) (Figure 7-11). Toutefois, on observe dans peu de cas des erreurs qui couvrent une superficie plus importante et qui concernent la totalité des objets (Figure 7-11). Pour ces erreurs, l'objet n'a pas été détecté en raison d'une surcharge trop importante de hachures conduisant à la perte de la couleur et donc de l'objet, après filtrage (cas du déficit de grande superficie sur l'extrait 1, Figure 7-11).

Le classifieur gaussien est apparu plus efficace sur l'extrait 1 soit l'extrait le plus surchargé par les hachures de relief, problématique récurrente dans les minutes d'Etat-major de 1850. Cette méthode offre en effet l'avantage de pouvoir sélectionner des échantillons d'apprentissage dans les zones très surchargées. En effectuant cela, on élargit ainsi les possibilités du détecteur qui prend en compte ces zones complexes, bien qu'elles disposent de comptes numériques assez éloignées des autres pixels « forêt ». L'effet de cet échantillonnage dirigé est surtout retranscrit dans la valeur de déficit de la classe « forêt » : celle-ci affiche une valeur deux fois moins importante qu'avec l'autre détecteur.

Enfin, nous avons observé une tendance inverse sur l'extrait 3, extrait le moins impacté par la présence de bruit. Dans ce cas précis, l'algorithme des K-moyennes s'est montré plus efficace mais les performances du classifieur gaussien restent néanmoins très bonnes avec un taux d'accord de la classe « forêt » de 93.5 %. Au regard des cartes d'erreurs de l'extrait 3 présentées en figure 7-11, le classifieur des K-moyennes a surtout mieux identifié les marges des différents objets « forêt » présent. Il est probable que les échantillons sélectionnés pour apprendre le modèle du mélange gaussien n'aient pas été suffisamment récoltés sur les contours d'objets. De fait, les pixels de contour d'objets qui peuvent potentiellement être des pixels « mixtes » en raison de la proximité à d'autres occupations du sol n'ont pas été considérés par la procédure d'apprentissage.

INDICATEURS DE QUALITE	MELANGE GAUSSIEN			K-MOYENNES		
	EXTRAIT 1	EXTRAIT 2	EXTRAIT 3	EXTRAIT 1	EXTRAIT 2	EXTRAIT 3
PRECISION GLOBALE %	96.55	96.67	97.77	94.36	96.67	98.72
INDICE KAPPA	0.91	0.90	0.94	0.86	0.90	0.97
ACCORD % (CLASSE « FORET »)	92.05	88.46	93.50	83.02	88.82	96.22
EXCEDENT % (CLASSE « FORET »)	3.58	2.06	0.78	1.90	2.15	0.37
DEFICIT % (CLASSE « FORET »)	7.95	11.54	6.50	16.98	11.18	3.78

Tableau 7-1. Résultats des extractions opérées sur les trois extraits sélectionnés

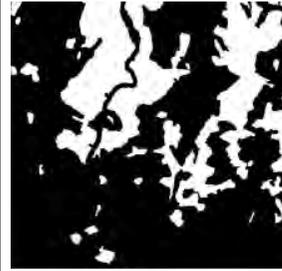
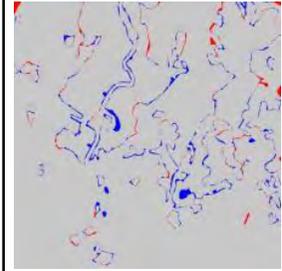
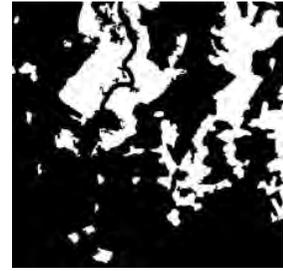
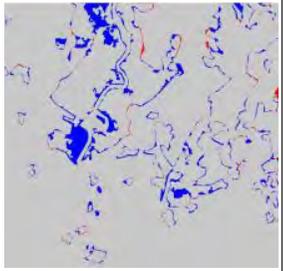
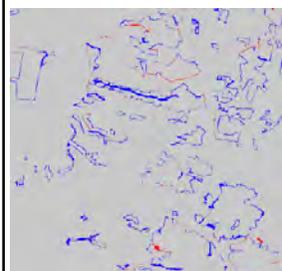
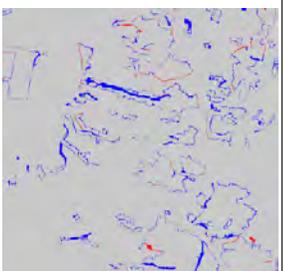
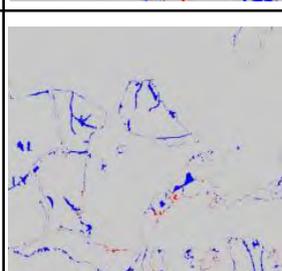
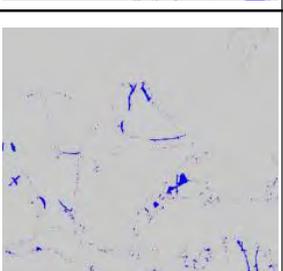
Source	Extraction manuelle	Extraction auto (mélange gaussien)	Excédents et déficits (mélange gaussien)	Extraction auto (K-moyennes)	Excédents et déficits (K-moyennes)
					
					
					

Figure 7-11. Comparaison des extractions automatiques et de la référence avec une spatialisation des erreurs d'excédents et de déficits

A PARTIR DES DONNEES COUVRANT LE SITE D'ETUDE

Nous appliquons ici notre outil d'extraction à l'ensemble du site d'étude (répartir sur 4 feuilles d'Etat-major). Il s'agit d'évaluer les performances de la démarche proposée lorsqu'elle est appliquée à une zone de plus vaste étendue.

Les résultats de validation indiquent que la méthode est très performante à une échelle large. La précision globale obtenue est de 98.64 % avec un indice Kappa de 0.89. Concernant plus spécifiquement la classe « forêt », on observe un taux d'accord de 86.71 % avec un taux d'excédent et de déficit respectivement égaux à 5.21% et 13.29 % (Tableau 7-2). La matrice de confusion est disponible en annexe C.

INDICATEURS DE QUALITE	VALEURS
PRECISION GLOBALE %	98.64
INDICE KAPPA	0.89
TAUX D'ACCORD % (CLASSE FORET)	86.71
OMISSION % (CLASSE FORET)	13.29
COMMISSION % (CLASSE FORET)	5.21

Tableau 7-2. Résultats de validation de la couche extraite à partir de la méthode de vectorisation automatique

La présence d'erreurs de classification et leur impact potentiel dépendent de l'application à laquelle sont destinées ces données extraites. Dans notre cas, les résultats d'extraction produits par l'outil développé ont vocation à être utilisés pour modéliser des relations espèces/habitats (cf. Partie D). Cela requiert de minimiser les erreurs de confusion et d'exhaustivité mais également les erreurs géométriques. En effet, les modèles écologiques envisagés s'appuient sur des variables spatiales qui décrivent d'une part, la surface des forêts et la longueur de leur lisière et d'autre part, leur connectivité aux autres forêts environnantes. La qualité de représentation des forêts extraites est donc tout aussi importante que la qualité sémantique et l'exhaustivité.

Ainsi, afin de déterminer l'adéquation potentielle de la méthode à notre cadre d'application écologique, nous avons comparé les valeurs de quatre indicateurs géométriques mesurés à l'échelle des forêts et vérifié s'il existait une différence significative (à partir du test de Wilcoxon) entre les objets vectorisés automatiquement et ceux vectorisés manuellement (Figure 7-12). Les boîtes à moustache montrent que les objets extraits automatiquement sont significativement différents de ceux extraits manuellement pour trois des quatre métriques calculées à savoir : la surface, le rapport périmètre/surface et la distance au plus proche voisin. Les objets extraits automatiquement disposent globalement d'une surface moins importante en raison des différentes étapes de filtrage appliquées qui ont eu tendance à éroder les contours d'objets. Ils apparaissent également plus compacts et affichent des valeurs de distances au plus proche voisin moins importantes.

Les exemples d'erreurs provoquées par le processus automatique (Figure 7-12c) montrent que la méthode peut parfois morceler ou perforer les objets en détectant différents éléments se superposant aux forêts non pris en compte par les étapes de pré et post-traitements. Les objets extraits sont alors plus compacts et les distances entre ces derniers ont été de fait réduites.

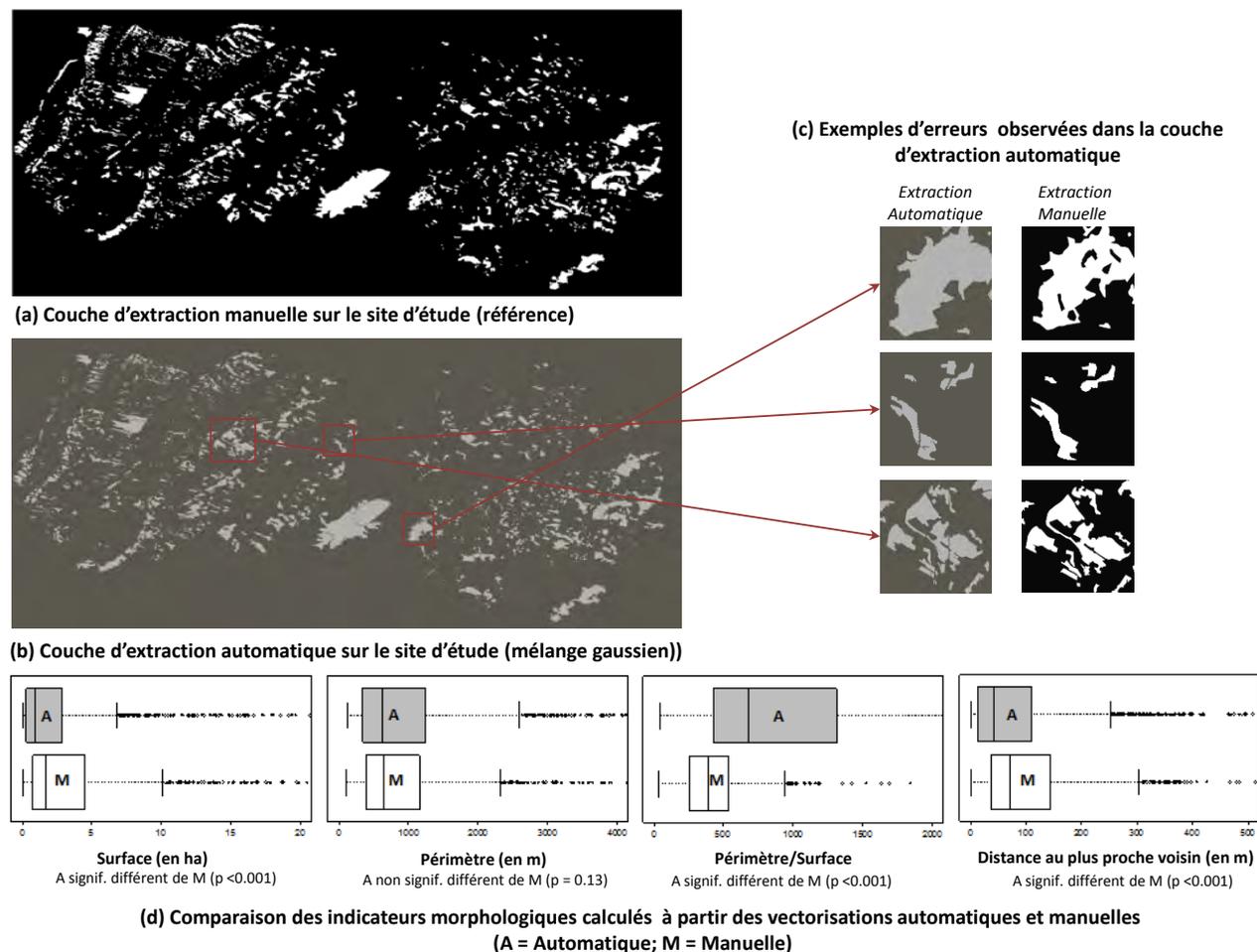


Figure 7-12. Résultats d'extraction automatique des forêts sur le site d'étude : a) et b) présentent respectivement les couches de forêts extraites manuellement et automatiquement ; c) répertorie trois exemples d'erreurs type observées sur la couche extraite automatiquement ; d) compare les distributions des valeurs de métriques calculées à partir des objets extraits automatiquement et manuellement. Le test de comparaison de moyenne utilisé est le test de Wilcoxon pour lequel une p -valeur inférieure à 0.05 permet de rejeter H_0 qui signifie que les deux moyennes sont égales.

Ces observations sont utiles pour la suite de ce travail de thèse. Bien que la méthode de vectorisation semble fournir des résultats d'extraction performants au regard des indicateurs de qualité sémantique et d'exhaustivité issus de la matrice de confusion, les indicateurs géométriques calculés à l'échelle de l'objet montrent des écarts à la référence plus importants. Cela indique que la démarche de vectorisation développée pourrait introduire un biais dans les modèles visant à relier la structure des forêts et les indicateurs de biodiversité. Celui-ci sera évalué en partie D de la thèse. Ces résultats soulignent aussi l'importance de prêter attention à la qualité géométrique des objets extraits en plus de la qualité sémantique et d'exhaustivité. En télédétection, cette composante de la qualité est rarement évaluée (Stehman et Wickham 2011; Radoux et Bogaert 2014) alors que dans le monde des bases de données spatiales vectorielles, elle est systématique (Devillers et Jeansoulin 2005). Nous y reviendrons par la suite (chapitre 11).

7.2.10 MISE EN ŒUVRE DANS LE CADRE DU PROJET OCS ANCIEN DE L'IGN

Les expérimentations ont été menées sur des feuilles proposées par les partenaires du projet (tableau 7-3, figure 7-13). Ces feuilles se distinguent par des contextes géographiques différents et pour deux d'entre elles, par des représentations particulièrement hétérogènes, y compris pour un même thème d'occupation des sols. Cela permet ainsi de confronter la méthode aux situations *a priori* les plus difficiles à traiter automatiquement.

Notre méthode de vectorisation supervisée (algorithme du mélange gaussien) a été comparée à deux autres méthodes :

(1) celle proposée par l'INRA Nancy qui n'est pas automatique. Elle consiste à vectoriser manuellement les entités en tenant compte d'un ensemble de règles prédéfinies (Favre et al. 2012). Les couches produites à partir de cette méthode feront figure de référence pour valider les couches extraites automatiquement.

(2) celle développée par le laboratoire MATIS de l'IGN qui est actuellement utilisée en production pour réaliser la couche végétation de la BDTopo®. Nous ne sommes pas en mesure de fournir des informations précises au sujet de cette approche. Toutefois, il s'agit d'une méthode structurée en trois étapes principales (Bulletin IGN n°75, p 47-51):

- rehaussement des couleurs de l'image pour améliorer le contraste
- application d'un algorithme de segmentation fondé sur l'approche hiérarchique. Il est par conséquent possible d'adapter la taille de la segmentation en fonction de la taille des objets et de faire le lien entre ces niveaux pour les agréger.
- affectation des objets segmentés par apprentissage supervisé (méthode des forêts aléatoires ou *random forest*).

Les objectifs de comparaison visent à évaluer:

- L'opérationnalité et la reproductibilité de la méthode : il s'agit d'estimer le temps de traitement d'une minute d'Etat-major, d'évaluer le besoin potentiel d'une intervention de l'utilisateur tout en mesurant les performances à proprement parler de la méthode.
- L'applicabilité de la méthode pour extraire différents thèmes d'occupation du sol sans se focaliser uniquement sur les forêts (tableau 7-3). Ces expérimentations doivent permettre de déterminer si les méthodes de vectorisation automatique peuvent être utilisées pour produire, dans un contexte opérationnel (IGN), une carte des usages anciens à l'échelle de la France.

MINUTES D'ETAT-MAJOR 1850	CONTEXTE ET LOCALISATION GEOGRAPHIQUE	Thèmes d'OCS vectorisés
191NE (Teste de Buch, 33)	<ul style="list-style-type: none"> • Zone de plaine • Faible surcharge • Représentation uniforme des couches d'occupation du sol • Bonne qualité de conservation 	<ul style="list-style-type: none"> • Forêts • Terres humides • Vignes
198NE (Privas, 07)	<ul style="list-style-type: none"> • Zone de piémont • Surcharge modérée • Diversité des représentations d'occupation du sol • Bonne qualité de conservation 	<ul style="list-style-type: none"> • Terres humides • Forêts
257NO (Prades, 66)	<ul style="list-style-type: none"> • Zone de montagne • Surcharge importante • Diversité des représentations d'occupation du sol • Mauvaise qualité de conservation 	<ul style="list-style-type: none"> • Forêts • Terres humides

Tableau 7-3. *Caractéristiques des feuilles sélectionnées pour les tests d'extraction du projet OCS ANCIEN*



Figure 7-13. Aperçu des dessins-minutes d'Etat-Major sélectionnés dans le cadre du projet OCS Ancien pour les tests d'extraction. 1) Feuille 191NE ; 2) Feuille 198NE ; 3) Feuille 257NO

Résultats de comparaison et discussion :

Les résultats de vectorisation obtenus dans ce contexte d'étude sont présentés en tableau 7-4. Un aperçu des résultats est aussi fourni en figure 7-15. Pour rappel, les matrices de confusion ont été estimées en considérant la carte produite manuellement par l'INRA Nancy comme référence. Elles sont disponibles en annexe D.

D'une manière générale, les résultats indiquent que l'efficacité des méthodes testées est très dépendante de la carte traitée. Premièrement, la carte 191NE (zone de plaine) est sans conteste celle qui pose le moins de difficultés *à priori* de par l'absence de relief (donc l'absence de hachures) et l'uniformité des aplats de couleur pour un même thème d'occupation du sol. Globalement, les performances des deux méthodes automatiques (IGN et celle que nous proposons, nommée ici méthode DYNAFOR) sont satisfaisantes avec des indices Kappa allant de 0.77 à 0.90. La méthode développée par l'IGN apporte des résultats légèrement meilleurs pour deux des thèmes extraits. Le taux d'accord de la classe « forêt » est de 86,63% contre 82,28% dans le cas de la méthode DYNAFOR tandis que celui de la classe « Terres humides » est de 74.86% contre 65.36 %. Les indicateurs d'omission et de commission montrent que les deux méthodes ont plutôt tendance à sous-détecter ces deux classes. Pour la classe « Vigne » en revanche, la méthode DYNAFOR est légèrement plus performante avec un taux d'accord de 78,21% contre 75,34% pour la méthode IGN avec de nouveau une tendance à la sous-détection. Ces résultats s'expliquent en partie par la phase de prétraitements prévue dans le cadre de la méthode DYNAFOR. Effectivement, celle-ci est nécessaire pour éliminer les éléments planimétriques superposés aux objets de la carte mais il est probable qu'elle soit légèrement mal adaptée dans les cas où il n'y pas ou peu de hachures sur la carte en lissant beaucoup trop les comptes numériques de l'image (Figure 7-14). Ainsi, les classes « Forêts » et « Terres humides » disposent d'une couleur trop pâle après cette étape pouvant s'apparenter avec le fond de carte, ce qui explique que certains objets de ces deux classes ont été mal identifiés.



Figure 7-14. Illustration des conséquences provoquées par le processus de filtrage. A gauche, l'image originale, à droite, l'image après filtrage. Les forêts (en vert) et les terres humides (en bleu) ont été lissés très fortement, le contraste entre le fond de carte (en beige) et ces deux classes est par conséquent nettement moins marqué.

Les résultats sur les deux autres feuilles, plus complexes, montrent une tendance inverse. La méthode IGN est moins performante, voire inadaptée pour certains thèmes en comparaison à la méthode DYNAFOR. Ainsi, l'IGN n'a pu extraire que le thème « Forêt » pour la feuille 198 NO. Les autres thèmes extraits par segmentation ont été jugés inexploitable par les chercheurs du laboratoire MATIS en raison d'erreurs trop importantes causés par la présence des hachures. Cette

approche ne prévoit pas de phase de filtrage en amont de la segmentation, ceci expliquant la faible opérationnalité des algorithmes de segmentation sur des cartes très chargées. La méthode DYNAFOR a, elle, fourni des résultats satisfaisants pour la feuille 198NE et moins performants sur la feuille 257NO mais acceptables au vu de la complexité de la carte (relief très accidenté). Pour la feuille 198E, la couche « forêt » extraite par la méthode DYNAFOR dispose d'un taux d'accord de 88.99% contre 61.59% pour celle fournie par la méthode IGN. Pour cette dernière, les erreurs s'expliquent notamment par des confusions importantes avec les prairies. La méthode DYNAFOR a aussi permis d'extraire les « Terres humides » avec un taux d'accord relativement bon (76.56%). Les erreurs sont principalement dues aux confusions avec le réseau hydrographique dont la teinte est très proche voire similaire (les deux occupations du sol étant représentées par des aplats bleus). Enfin, comme attendu, les résultats d'extraction sur la carte 257NO sont de moins bonne qualité mais restent acceptables au regard du niveau de difficulté de celle-ci et de sa qualité de conservation. Notre méthode a permis d'extraire les forêts avec un taux d'accord de 71.4% et les terres humides avec un taux d'accord de 64%. Les difficultés ont principalement résidé dans l'élimination des hachures de relief très denses sur cette carte (assombrissant trop fortement la teinte après filtrage) et dans la prise en compte de la diversité des représentations pour un même thème. Ainsi, les erreurs observées pour la couche « Terres humides » sont une nouvelle fois liées aux confusions avec le réseau hydrographique. Ces confusions pourraient être réduites en détectant préalablement le réseau hydrographique par exemple. Si les comptes numériques de ces deux classes semblent trop proches pour les distinguer, la forme linéaire des fleuves et des rivières pourrait constituer un critère de discrimination supplémentaire et ainsi réduire les confusions. Enfin, les erreurs apparues dans la couche « forêt » correspondent à des erreurs grossières de confusions, dues à une trop grande densité de hachures. Sélectionner d'avantage de pixels dans ces zones ne servira rien puisque leurs valeurs tendront vers le gris/noir et ne permettront pas de reconstruire l'objet d'intérêt, c'est donc en amont (prétraitements) que la possibilité de réduire ces erreurs pourrait exister

N° MINUTE ETAT-MAJOR	THEME EXTRAIT	TAUX D'ACCORD (EN %)		INDICE KAPPA		COMMISSION (EN %)		OMISSION (EN %)	
		DYN	IGN	DYN	IGN	DYN	IGN	DYN	IGN
191 NE	Forêts	82.28	86.63	0.87	0.90	0.80	1.73	17.72	13.37
	Terres humides	65.36	74.86	0.75	0.82	8.91	5.83	34.64	25.14
	Vignes	78.21	75.34	0.82	0.77	11.24	18.33	21.79	24.66
198 NO	Forêts	88.99	61.59	0.85	0.56	7.72	18.27	11.01	38.41
	Terres humides	76.56	n.s	0.73	n.s	28.18	n.s	23.44	n.s
257 NO	Forêts	71.4	n.s	0.68	n.s	29.79	n.s	31.24	n.s
	Terres humides	64.02	n.s	0.60	n.s	38.54	n.s	35.98	n.s

Tableau 7-4. Résultats des tests d'extraction effectués dans le cadre du Projet OCS ANCIEN. Les acronymes DYN et IGN indiquent respectivement la méthode présentée dans ce travail de thèse (DYNAFOR) et celle développée par l'IGN. Les n.s signifient que les résultats d'extractions ont été jugé trop peu fiables pour les présenter.

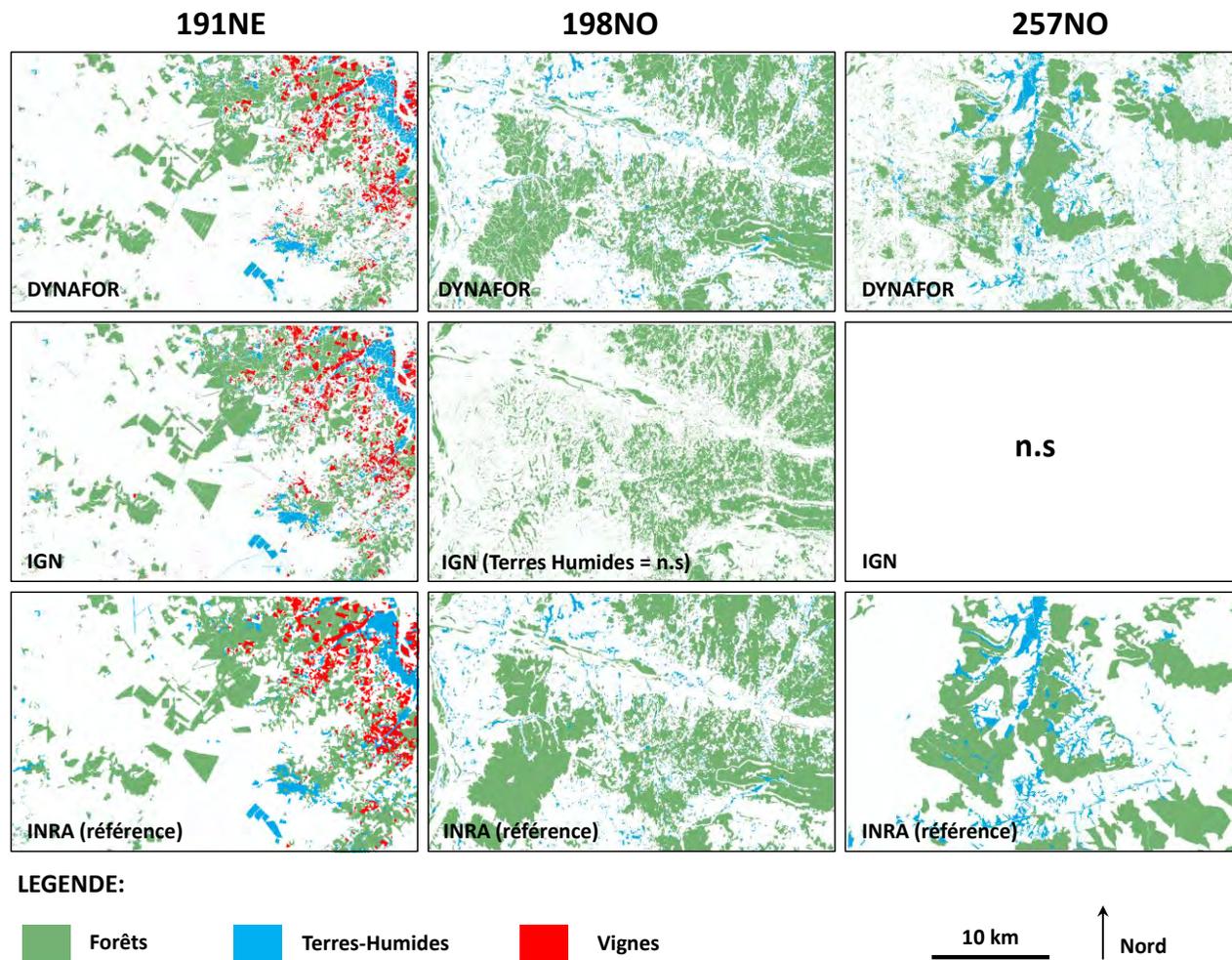


Figure 7-15. Résultats d'extraction à partir des trois dessins-minutes d'Etat-major sélectionnés dans le projet OCS Ancien. Les n.s (non-significatif) indiquent que les thèmes en question n'ont pu être extraits suffisamment précisément pour être présentés.

7.2.11 BILAN

La démarche d'extraction développée dans le cadre de cette partie a démontré un haut de niveau de performance en entraînant toutefois certaines erreurs. Ces dernières varient principalement selon le type de carte à traiter et plus précisément en fonction de leur niveau de surcharge planimétrique et de leur qualité.

La première phase de test appliqué aux trois extraits sélectionnés (section 7.2.8) a permis de mettre en avant la capacité de notre approche à tenir compte des variétés de représentation des forêts et à s'adapter au niveau de surcharge liée à la présence de relief (hachures). Les indicateurs de qualité traduisent un haut niveau de performance sur les trois extraits traités. Les résultats fournis par les deux classifieurs testés sont quant à eux très proches même si le modèle du mélange gaussien s'est avéré le plus efficace pour les zones les plus surchargées et complexes à traiter.

Ensuite, l'application de la démarche à notre site d'étude a offert la possibilité d'évaluer sa reproductibilité sur de plus vastes étendues. Les indicateurs de qualité sémantique ont affiché un très bon niveau de performance de la méthode à l'échelle du site mais l'étude des critères géométriques a mis en lumière un certain nombre de biais potentiels. Trois des quatre métriques calculées (surface, ratio périmètre/surface et distance au plus proche voisin) pour les objets extraits automatiquement sont significativement différentes de celles définies à partir des objets cartographiés manuellement. Les erreurs à l'origine de constat sont principalement dues un morcellement ou à une perforation des objets. Par conséquent, nous devons tenir compte de cette observation au moment de modéliser les relations entre la structure spatiale des bois et la richesse de syrphes (Partie D). Ces erreurs pourraient potentiellement baisser la réponse des modèles espèces-habitats et conduire à les interpréter de manière erronée.

Enfin, la dernière phase d'expérimentation a été menée en contexte opérationnel dans le cadre du projet OCS Ancien. Ce travail a eu pour objectif d'évaluer l'adéquation de notre démarche à une perspective de mise en production pour une cartographie des anciens usages des sols à partir des dessins-minutes d'Etat-major. Elle a été comparée à deux méthodes principales : l'une a été développé au laboratoire MATIS et est déjà utilisée au sein de l'IGN pour la production de la BD Topo®. L'autre est la méthode couramment employée pour numériser les éléments manuellement dans les dessins minutes d'Etat-major et constitue notre référence (Favre et al. 2012).

Globalement, nous pouvons dire que notre approche et celle de l'IGN peuvent être toutes les deux appliquées aux cartes de niveau « facile » c'est-à-dire lorsque la densité de hachure est faible et la représentation des forêts uniforme (191NE). Notre approche est légèrement moins performante sur ce type de carte mais pourrait être améliorée en amont de la phase de classification. Un travail sur l'adaptation des paramètres de filtrage pourrait s'avérer intéressant dans l'optique d'altérer moins fortement la teinte de certains objets et faciliter leur extraction.

Lorsque le niveau de difficulté des cartes augmente (198NE et 257NO), la performance globale des méthodes proposées baisse mais reste néanmoins très satisfaisante en ce qui concerne notre approche. Tout d'abord, la méthode IGN n'a été en mesure de fournir qu'un seul thème d'occupation du sol sur deux (les forêts) dans la carte 198NE et aucun sur la feuille 257NO. L'absence d'une phase de nettoyage en amont de la segmentation prévue par cette approche explique en partie cette faiblesse. Concernant notre approche, elle a permis d'extraire efficacement les thèmes demandés sur la première feuille (198NE) alors que les résultats à partir de la carte 257NO sont très encourageants. La baisse de qualité des résultats s'explique majoritairement par des erreurs grossières provoquées par une densité de hachures très importante masquant partiellement les aplats de forêts. Une nouvelle fois, il serait nécessaire d'adapter la phase de filtrage pour tenir compte plus efficacement de ces hachures. Une possibilité serait par exemple de s'appuyer sur un Modèle Numérique de Terrain (MNT) pour adapter la taille de fenêtre de convolution ou des opérateurs morphologiques en

fonction des plus fortes pentes. Il semble néanmoins indispensable d'avoir recours à l'intervention humaine en appui d'une approche automatique pour gérer efficacement ce type de difficultés.

7.3. PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE A LA BD ORTHO® IRC

Nous décrivons ci-après la démarche employée pour extraire automatiquement les forêts de la BD Ortho® IRC de 2010. A noter que l'approche développée n'est pas spécifique aux objets étudiés (les forêts) ni à la BD Ortho® IRC. Nous souhaitons développer une approche reproductible susceptible d'être appliquée à différents types d'objets et à toutes les photographies aériennes.

7.3.1 ENJEUX DE L'EXTRACTION AUTOMATIQUE DES FORETS DANS LES PHOTOGRAPHIES AERIENNES

Les photographies aériennes sont des sources de données spatiales très utiles pour cartographier les écosystèmes et les entités paysagères à grande échelle (Morgan et al. 2010). Elles disposent d'une résolution radiométrique basique (3 bandes) et fournissent des détails spatiaux importants. Elles permettent aussi d'obtenir des informations de manière quasi continue des années 1940 jusqu'à aujourd'hui. Dans le cas des forêts, elles sont largement utilisées pour les inventaires forestiers (Tuominen et Pekkarinen 2005; Boureau 2008), l'estimation de paramètres de structure (St-Onge et Cavayas 1997) ou de composition (Kayitakire et al. 2002), la réalisation de cartes de dégâts (à la suite d'une tempête ou d'une maladie par exemple), l'estimation de leur biomasse (Lefsky et al. 1999) ou encore, la gestion des habitats (Coops et Catling 1997).

De nombreuses opérations de suivi des écosystèmes s'appuient sur des cartes dérivées des photographies aériennes (Pasqualini et al. 1998; Zharikov et al. 2005; Morgan et al. 2010). Néanmoins, la disponibilité accrue des images satellites ces dernières décennies a quelque peu modifié ces pratiques (Boyd et Danson 2005). Elles offrent souvent une couverture spatiale plus importante ainsi qu'une plus grande fréquence temporelle d'acquisition. Les images satellites ne sont cependant pas toujours adaptées à tous les objectifs d'étude. Elles manquent de recul temporel (acquisition depuis les années 1970), leur utilisation est ainsi réduite aux quarante-cinq dernières années. Leur résolution spatiale est aussi souvent plus faible que celle des clichés aériens même si aujourd'hui, cet écart tend à diminuer avec les nouveaux capteurs satellites à très haute résolution spatiale (<5m).

De nombreux travaux existent sur l'automatisation de l'extraction à partir des photographies aériennes (Boyd et Danson 2005). Les travaux portant sur les écosystèmes forestiers sont d'ailleurs très anciens et un colloque consacré exclusivement à ce sujet existe depuis de nombreuses années (ForestSAT). Néanmoins, en écologie, la majeure partie des études utilisent traditionnellement la photo-interprétation pour être en mesure de cartographier précisément les éléments et s'affranchir des potentielles erreurs liées à aux classifications automatiques. Cependant, les limites de la digitalisation sont connues et ont été décrites dans le chapitre 6.

Deux approches sont généralement employées pour extraire automatiquement les éléments des photographies ou des images satellitaires. L'approche « pixel » consiste à effectuer une classification à partir des comptes numériques des pixels. Il s'agit de l'approche classique en télédétection que nous avons d'ailleurs adoptée pour reconnaître les forêts dans les cartes d'Etat-major. Puis, l'émergence des données à Très Haute Résolution Spatiale (THRS) a conduit au développement d'une nouvelle approche orientée « objet » (GEOBIA, Hay et al. 2003). Elle est généralement plus adaptée que l'approche pixel lorsque l'image présente un haut niveau de détail (ou hétérogénéité colorimétrique ou spectrale), comme c'est le cas dans les photographies aériennes, car un pixel n'est plus un mélange d'objets mais une partie seulement d'un objet

(Blaschke 2010). Par ailleurs, la combinaison de différents critères pour reconnaître les objets s'apparente davantage au processus cognitif de photo-interprétation.

Cette approche orientée-objet consiste à segmenter l'image pour construire des objets colorimétriquement homogènes. Les objets peuvent ensuite être classés à l'aide de critères spectraux (valeurs moyennes des canaux par exemple) mais aussi, de critères spatiaux (forme, taille), texturaux¹ ou encore, de critères relationnels (proximité ou adjacence à d'autres objets notamment). Concernant la végétation, une mention particulière est à attribuer à la texture, indicateur largement repris dans plusieurs études (Woodcock et al. 1988; Lévesque et King 2003; Lilles et al. 2004; Morgan et al. 2010; Kollár et al. 2013; Thierion et al. 2014). La plupart des études exploitant la texture portent sur la caractérisation de l'hétérogénéité des habitats, les paramètres biophysiques ou les caractéristiques structurales des forêts (Kollar et al, 2013 d'après Morgan, 2010). Shang et Waite (1991) ont par exemple proposé une combinaison de la texture et de la teinte pour identifier les forêts boréales au Canada. Les résultats ont montré que la texture permettait d'améliorer de 30% la précision globale de classification par rapport à une situation où seule la teinte était utilisée. Kollar et al (2013) ont également démontré l'apport de ces indicateurs pour cartographier les habitats végétaux dans une zone humide en Hongrie à partir de photographies aériennes. La combinaison des indicateurs texturaux et spectraux s'est avérée la plus performante parmi plusieurs combinaisons testées. A noter qu'il existe un grand nombre de métriques texturales, la plupart fondées sur les matrices de co-occurrence (Haralick 1979). Il est souvent préconisé de faire un tri de ces métriques au préalable pour ne pas trop accroître les temps de calcul et s'affranchir des problèmes de corrélations entre métriques (Bindel et al. 2011; Mahmoud 2012).

Enfin, la végétation dispose de propriétés optiques spécifiques. La chlorophylle absorbe moins les longueurs d'ondes dans les longueurs d'onde du proche infrarouge. De fait, dans ce domaine spectral, la végétation a toujours une réflectance plus importante que les autres objets (sols, eau, minéraux). De nombreux auteurs utilisent par conséquent des indices de végétation construits à partir des canaux infrarouges de l'image pour cartographier des espaces forestiers (Jackson et Gaston 1994; King 1995; Holopainen et Wang 1998, Hyponen 1999). Il en existe plusieurs et le plus couramment utilisé est le NDVI (*Normalized Difference Vegetation Index*).

7.3.2 DEMARCHE GENERALE PROPOSEE POUR EXTRAIRE LES FORETS DE LA BD ORTHO® IRC

Dans cette section, nous nous attachons à décrire l'approche proposée pour cartographier automatiquement les forêts à partir de la BD Ortho® IRC de 2010. Seule cette source a fait l'objet d'une telle application puisque les photographies aériennes anciennes en niveau de gris nécessitent une démarche spécifique pour traiter les problèmes de vignettages (Lefebvre 2011) ou l'homogénéisation des comptes numériques. Par faute de temps, nous n'avons pu nous focaliser précisément sur ces points et proposer une démarche automatisée pour ces photographies.

La démarche proposée ci-après se déroule en six étapes qui seront détaillées dans la suite du manuscrit (figure 7-16):

(1) Etape de segmentation: cette étape permet d'agréger des pixels pour former des segments spectralement homogènes (appelés aussi « objets »). L'algorithme utilisé est fondé sur une approche par croissance de régions (ou segments) permettant une segmentation multi-résolution (Baatz et Schäpe 2000).

(2) Etape de caractérisation des segments : elle permet de calculer un vecteur d'attributs de différente nature (spectraux, spatiaux, etc.) qui est utilisé par la suite dans la classification.

¹ texture : organisation spatiale des niveaux de gris dans un voisinage

(3) Etape de construction d'une base d'apprentissage : cette troisième étape a consisté à sélectionner des segments qualifiés (ou labélisés) afin de construire un jeu d'exemples d'apprentissage à deux classes (exemple positif : forêt ; exemple négatif : non forêt). Cette base d'exemples sera mobilisée dans la phase suivante relative à l'apprentissage proprement dit d'un modèle de classification (approche supervisée).

(4) Etape d'apprentissage d'un modèle : cette étape permet de construire les règles de classification qui serviront ensuite à reconnaître les forêts dans toute l'image. Elles ont été apprises automatiquement, par la construction d'un arbre de décision (Quinlan 1986). Il s'agit d'une méthode supervisée non paramétrique.

(5) Etape de classification : les règles apprises précédemment sont exécutées pour classer l'ensemble des segments en « forêt » ou « non forêt ».

(6) Etape de post-traitement : elle a servi à améliorer les résultats de classification (élimination d'artefacts).

(7) Etape de validation : la classification obtenue a été comparée à la classification de référence correspondant à la vectorisation manuelle détaillée en chapitre 6. Les indicateurs classiques de qualité issus d'une matrice de confusion ont ensuite été calculés.

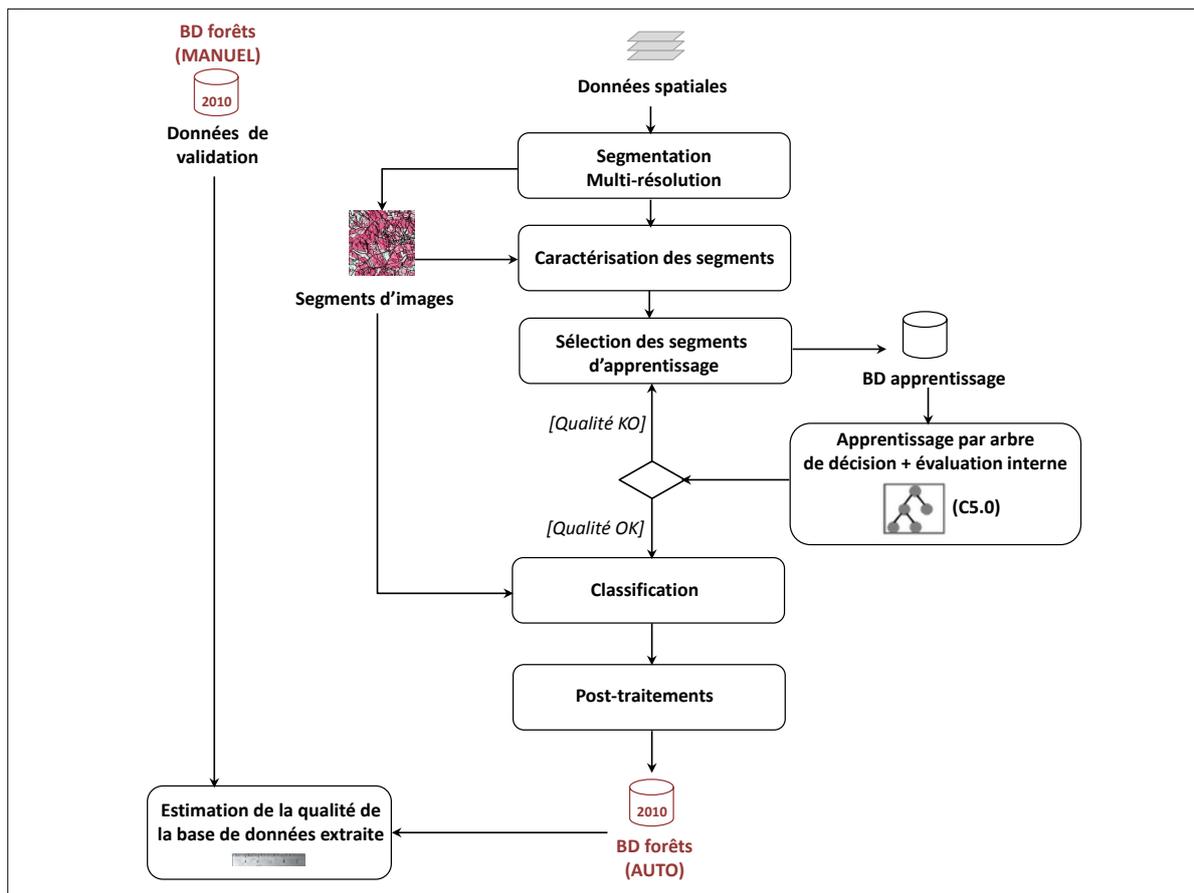


Figure 7-16 Démarche générale de la méthode proposée pour extraire automatiquement les forêts de la BD ortho IRC

SEGMENTATION MULTI-RESOLUTION

La procédure de segmentation est une procédure itérative pouvant être divisée en cinq blocs fonctionnels : les mesures, le critère, le contrôle, la modification et l'arrêt (Benoit-Cattin et al. 2000, Lefebvre 2011) (Figure 7-17).

(1) La segmentation vise à partitionner une image en segments (ou objets) homogènes. Par conséquent, les *mesures* servent à estimer des caractéristiques d'un segment par rapport à ses voisins. Elle peut reposer sur des propriétés spectrales (telles que le niveau de gris moyen des pixels ou la texture) ou des propriétés relatives à leur contour ou leur géométrie.

(2) Le *critère* sert à définir l'homogénéité ou la similarité entre deux segments. Le bloc *critère* reçoit l'ensemble des mesures du bloc *mesure* précédent et fournit un critère par segment. Il tend à diminuer lorsque la segmentation s'approche du résultat attendu.

(3) Le *contrôle* intervient pour prendre la décision de faire évoluer la taille des segments. Soit les segments doivent diminuer (approche descendante) et cela résulte en un découpage des segments ; soit les segments doivent s'agrandir (approche ascendante) et cela résulte en une agrégation de pixels ou une fusion de plusieurs segments.

(4) La *modification* est l'étape phare du processus de segmentation (Rose 2008). Elle fait varier la segmentation en respectant une base d'instructions indiquées par le bloc *contrôle*. Trois types peuvent être envisagés :

- modification avec un nombre constant de segments,
- modification avec création de nouveaux segments,
- modification avec agrégation de segments.

(5) Enfin, l'*arrêt* permet de stopper la chaîne de traitements au moment où la carte de segmentation n'est plus modifiée ou si le nombre d'itérations fixé est atteint.

Différentes méthodes de segmentation par régions peuvent être utilisées. Elles diffèrent notamment par leur étape de *modification* des segments qui conditionne directement la structure des segments en sortie.

Les méthodes basées sur une modification avec un nombre constant de segments comprennent trois grands types de méthodes. On recense les méthodes s'appuyant sur le **seuillage des histogrammes** qui utilisent uniquement les valeurs spectrales des pixels pour les regrouper (Fukunaga et Hostetler 1975; Carreira-Perpindn 2007). Ces méthodes souffrent d'un problème important : elles ne tiennent pas compte du lien spatial entre les pixels. D'autres approches consistent à employer le **croissance des segments** à partir d'un pixel de départ appelé germe (Batz et Schäpe 2000). Elles sont mises à disposition par le logiciel que nous utilisons ici (eCognition) qui facilite d'ailleurs leur utilisation. Enfin, la segmentation par **ligne de partage des eaux** constitue une troisième approche possible de ce groupe. Dans cette approche fondée sur la morphologie mathématique, l'image est considérée comme une surface topographique constituée de bassins versants et de crêtes où chaque pixel correspond à une mesure d'élévation. A partir de cette représentation, les régions sont extraites par un processus d'immersion de la surface (Vincent et Soille 1991; Beucher 1992). Un des principaux inconvénients de cette méthode est la segmentation trop fine de l'image. L'algorithme est très sensible au bruit et génère souvent des mini bassins-versants qui ne correspondent pas à des objets dans leur ensemble dans l'image (Benoit-Cattin et al. 2000).

Les méthodes fondées sur l'agrégation de régions emploient majoritairement les graphes d'adjacence à partir de sur-segmentations. A partir du graphe, les couples de nœuds adjacents sont identifiés comme des candidats. Les régions correspondant à chaque couple de nœud sont séparées. Si elles sont jugées homogènes sur la base d'un ensemble de critères, elles sont agrégées et la liste des couples est mise à jour. Cette méthode est très dépendante de l'ordre de fusion (départ au coin supérieur droit ou gauche par exemple) car une région agrégée dans un sens pourrait ne pas l'être dans l'autre mais agrégée à une autre région (Sonka et al. 2014).

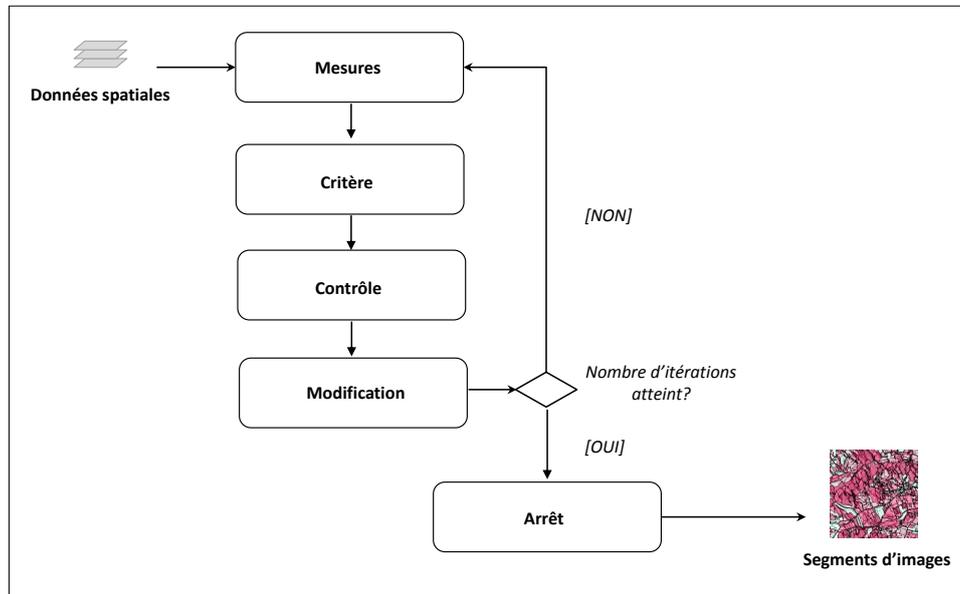


Figure 7-17 Principe général de la segmentation (source : Benoit-Cattin et al. 2000)

Les méthodes fondées sur la division des régions reposent sur une approche multi-échelles de type descendante (top-down). Elle commence par l'image entière sans vérifier le critère choisi puis l'image est divisée successivement en blocs rectangulaires. Si le bloc est homogène, il n'est plus divisé sinon il est à nouveau divisé en n blocs lors de l'itération suivante. L'algorithme le plus répandu est le *quad-tree* et a déjà été utilisé à de nombreuses reprises en télédétection (Arellano et al. 1998; Wuest et Zhang 2009). Celui-ci s'apparente à une division successive des régions basée sur une organisation hiérarchique des blocs de différentes tailles. Il est toutefois réputé pour manquer de précision dans la délimitation des contours qui correspondent uniquement aux blocs d'analyse sans reposer sur une étude des contours de l'image.

- **Principe de la segmentation multi-résolution**

L'algorithme multi-résolution tente de regrouper les pixels sur la base des critères d'homogénéité relative. Le processus de fusion s'arrête lorsque l'homogénéité des objets candidats à la fusion dépasse l'hétérogénéité maximale acceptée (contrôlée par le paramètre d'échelle qui renvoie à une échelle spatiale des objets identifiés – voir ci-dessous). Le critère d'homogénéité est calculé comme une combinaison de couleurs (critères *spectraux*) et de propriétés de la *forme*. L'homogénéité de *couleur* est estimée sur la base de l'écart-type du spectre de couleurs pour chaque segment généré et l'homogénéité de la *forme* sur la base de l'écart à une forme de référence définie par un degré de compacité. Ces deux propriétés, *couleur* et *forme* définissant le critère d'homogénéité peuvent être pondérés, ce qui permet de privilégier si nécessaire des segments compacts (poids de la forme à 0.9 et couleur à 0.1).

L'originalité de cet algorithme repose sur sa capacité à segmenter de façon hiérarchique (plusieurs niveaux possibles). Les objets obtenus sont représentés dans un arbre hiérarchique dans

lequel chaque objet connaît son contexte, ses voisins ainsi que ses objets « parents » et « enfants » caractérisés par de nombreux attributs (Figure 7-18). (Burnett et Blaschke 2003).

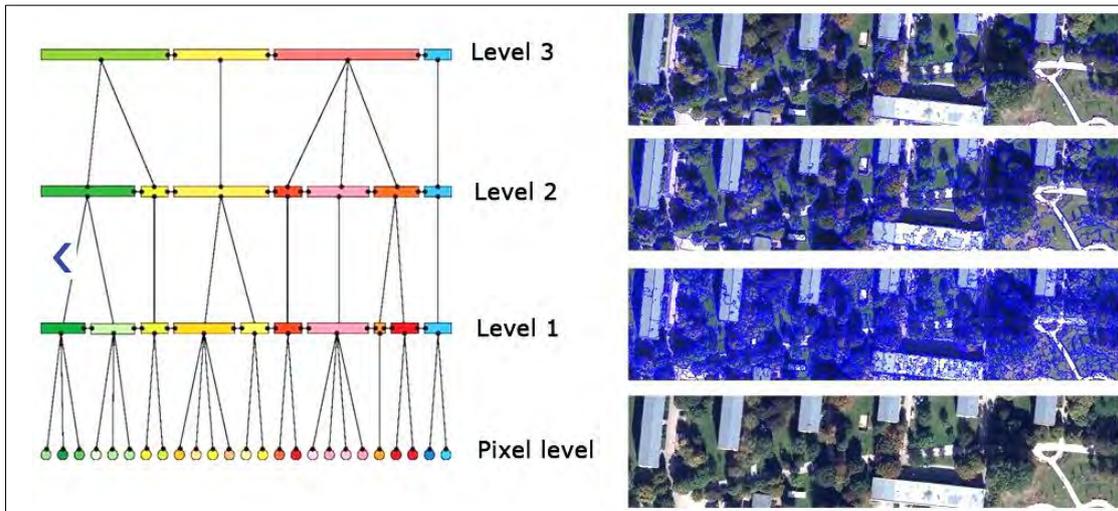


Figure 7-18 Principe de la segmentation multi-résolution. Différents niveaux de segmentation peuvent être produits et reliés par une structure hiérarchique (Willhauck et al. 2000)

- **L'échelle de segmentation**

Un des paramètres principaux de la segmentation multi-résolution concerne le paramètre d'échelle qui conditionne la taille des objets segmentés. L'extension ESP (*Estimation Scale Parameter*) développée par Drăguț et al. (2010) (Figure 7-19) permet d'identifier le paramètre d'échelle qui minimise la variance locale de la réflectance à l'échelle des objets générés. Les auteurs proposent aussi de s'appuyer sur le taux de changement de la variance locale pour identifier le seuil à partir duquel l'image peut être segmentée. Les « pics » apparents sur cette dernière courbe (figure 7-20) montrent qu'à ce seuil (la valeur d'échelle à laquelle apparaît le pic), les objets segmentés seront caractérisés par une homogénéité relative équivalente.

Cette procédure a été appliquée à 10 dalles de la BD Ortho® IRC localisée sur notre site d'étude (cf. chapitre 4) dans le but d'identifier un seuil de segmentation adapté à nos objets d'intérêt. Premièrement, afin d'accorder le maximum d'importance aux critères spectraux, nous avons fixé les paramètres de forme et de compacité à 0.1. Nous avons ensuite fait varier le paramètre d'échelle de 150 à 250 avec un pas de 10 (Figure 7-19). Cet intervalle de valeur a été choisi empiriquement et a été jugé pertinent pour les objectifs fixés.

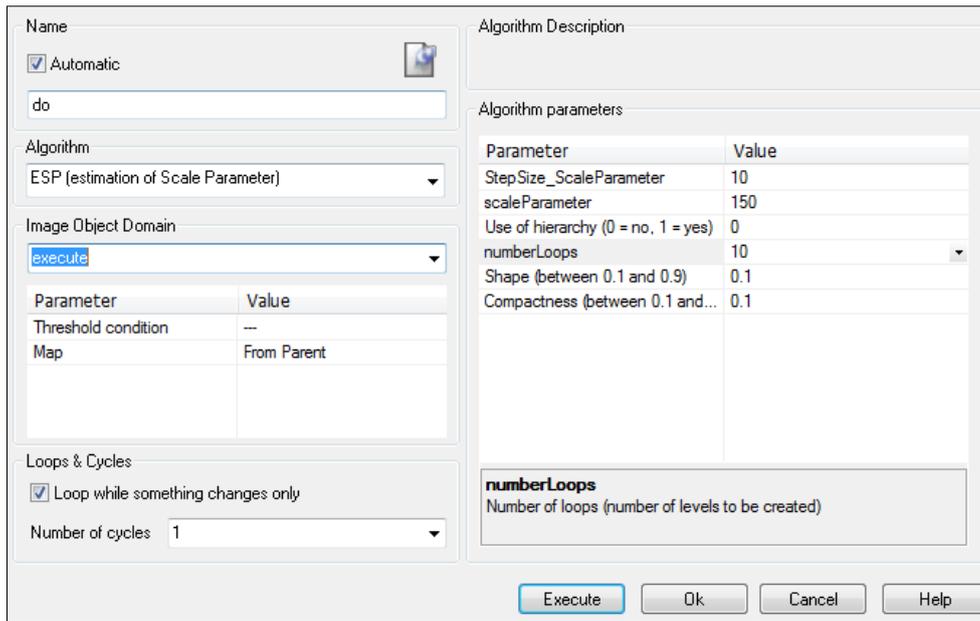


Figure 7-19 Interface de l'outil ESP Tool

Comme il est possible de voir sur la Figure 7-20, aucun aplatissement de la courbe représentant la variance locale (Locale Variance) n'apparaît en fonction des valeurs d'échelle. Nous nous sommes donc concentrés sur la courbe représentant son taux de changement (Rate of change). Globalement (7 cas sur 10), un premier pic est apparu à une valeur d'échelle de 180 comme dans les deux exemples ci-dessous. Nous avons par conséquent retenu ce paramètre d'échelle pour segmenter nos images.

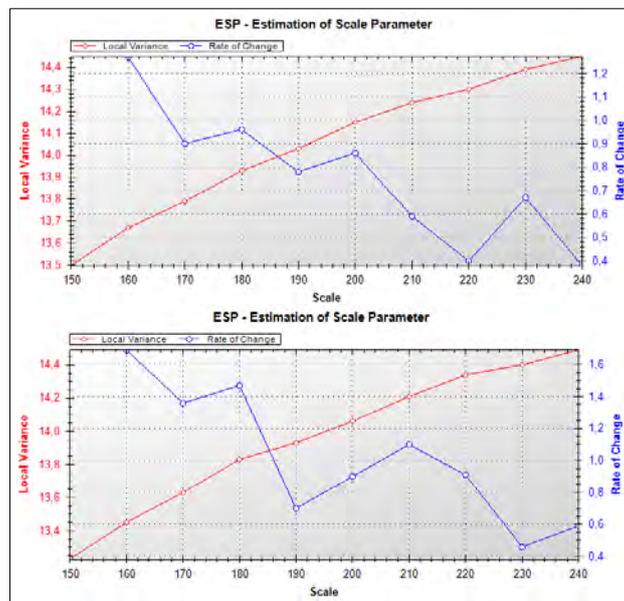


Figure 7-20 Exemples de sorties de l'outil ESP Tool pour deux dalles de la BD Ortho IRC. La courbe de la variance locale n'affiche aucun points d'inflexion tandis que le premier pic apparent sur la courbe du taux de changement est situé à une valeur d'échelle de 180.

La figure 7-21 donne un aperçu d'une dalle segmentée. Nous constatons que les régions créées reflètent correctement la structure spatiale des parcelles agricoles. Cela est important car la plupart des confusions qui pourraient intervenir avec les forêts concerne les cultures ou les prairies qui

peuvent aussi avoir une forte réflectance dans le canal du proche-infrarouge selon les saisons ou les stades de rotation.

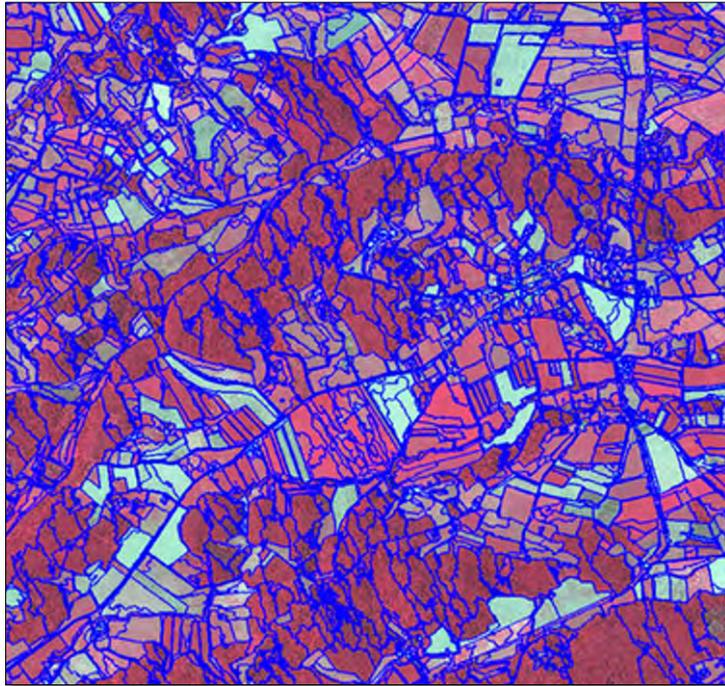


Figure 7-21 Exemple d'une dalle de la BD Ortho IRC segmentée avec les paramètres retenus

CARACTERISATION DES SEGMENTS D'IMAGES

Une fois la segmentation réalisée, l'étape suivante vise à caractériser par classification les segments par un ensemble d'attributs (ou descripteurs). Différents types d'attributs existent.

- (1) Les attributs spectraux qui sont directement dérivés des comptes numériques des pixels composant l'objet. Chaque bande spectrale de l'image peut être exploitée pour dériver ces attributs.
- (2) Les attributs spatiaux qui visent à caractériser la structure de l'objet. Ils fournissent notamment des informations sur leur taille (surface, périmètre, longueur) et leur forme (complexité, compacité, élongation).
- (3) Les attributs texturaux qui informent sur l'organisation spatiale locale des comptes numériques au sein d'un objet. Les attributs les plus simples décrivent la dispersion ou variabilité des valeurs des pixels au niveau de l'objet (texture de 1^{er} ordre). D'autres rendent compte d'une organisation spatiale plus complexe par une mesure d'entropie, de contraste ou d'un moment angulaire (texture de 2^{ème} ordre).
- (4) Les attributs contextuels qui informent sur les relations de proximité ou les relations topologiques qu'entretiennent les objets avec leurs voisins (adjacence, inclusion, etc.).

Au total, 12 attributs ont été calculés (Tableau 7-5). Nous avons principalement axé notre choix sur des variables spectrales et texturales au regard des travaux existants concernant l'extraction des forêts dans les photographies aériennes. Un indice de forme a toutefois été ajouté (la rectangularité) sous l'hypothèse qu'il pourrait être une mesure pertinente pour différencier les forêts des parcelles agricoles.

ATTRIBUTS	ACRONYMES	DEFINITIONS
Moyenne (canal Vert)	Mean_B1	Moyenne des valeurs des pixels d'un segment dans le canal Vert
Moyenne (canal Rouge)	Mean_B2	Moyenne des valeurs des pixels d'un segment dans le canal Rouge
Moyenne (canal IR)	Mean_B3	Moyenne des valeurs des pixels d'un segment dans le canal Infrarouge
Ecart-type (canal Vert)	Standard_d_B1	Ecart-type des valeurs des pixels d'un segment dans le canal Vert
Ecart-type (canal Rouge)	Standard_d_B2	Ecart-type des valeurs des pixels d'un segment dans le canal Rouge
Ecart-type (canal IR)	Standard_d_B3	Ecart-type des valeurs des pixels d'un segment dans le canal Infrarouge
Brillance	Brighthness	Moyenne des niveaux de gris d'un segment
Texture (homogénéité)	GLCM_Homog	Indice d'homogénéité de l'organisation spatiale de la texture d'un segment
Texture (entropie)	GLCM_Entro	Indice d'entropie de l'organisation spatiale de la texture d'un segment
Texture (moyenne)	GLCM_Mean	Moyenne des valeurs de texture
Texture (écart-type)	GLCM_StdDe	Ecart-Type des valeurs de texture
Rectangularité	Rectangular fit	Rapport entre l'aire du segment et le plus petit rectangle inscrit dans le segment.

Tableau 7-5. Description des attributs sélectionnés pour caractériser les segments d'image
(eCognition, Reference Guide)

CREATION DE LA BASE D'APPRENTISSAGE

Pour être en mesure d'obtenir une base d'exemples d'apprentissage représentative de notre zone à classer, nous avons sélectionné des exemples à trois endroits différents de notre zone d'étude (Figure 7-22).

Le choix et la labellisation des exemples ont été effectués par photo-interprétation. Deux labels de classe ont été attribués aux échantillons : forêt (exemple positif) ou non-forêt (négatif). Seules ces deux classes sont nécessaires sachant que nous projetons d'utiliser un algorithme de classification non-paramétrique (les arbres de décision). Nous n'avons pas besoin de connaissances *a priori* sur les différentes classes de l'image et donc de sélectionner des échantillons pour chacune d'entre elles. Au total, 250 échantillons ont été sélectionnés dont 50% labélisés « forêt » et 50% « non-forêt ».



Figure 7-22. Localisation des trois zones d'échantillonnage. Les polygones gris représentent la couche de forêts extraite manuellement et vise à faciliter la localisation des zones d'apprentissage. Les polygones verts sont les segments d'exemples labélisés « forêt » et les polygones jaunes sont les segments d'exemples labélisés « non-forêt ».

RECONNAISSANCE AUTOMATIQUE DES FORETS PAR APPRENTISSAGE

La reconnaissance automatique des forêts a été effectuée à l'aide d'une méthode de classification non-paramétrique : les arbres de décision.

Etant donné que nous souhaitons uniquement faire la différenciation entre les forêts et le reste du paysage, l'approche non-paramétrique est bien adaptée puisqu'elle ne requiert pas d'a priori sur la distribution des classes présentes dans l'image. Ensuite, l'approche à base de règles binaires et floues (logique floue) est la plus courante pour classer les éléments à la suite d'une segmentation. La plupart du temps, ces règles sont généralement définies empiriquement alors que nous utiliserons l'apprentissage pour les définir. Enfin, les règles de décision ont l'avantage d'être intelligibles, ce qui n'est pas le cas d'un modèle de mélange gaussien par exemple. Elles sont aussi facilement reproductibles pour traiter de vastes étendues.

Pour établir un arbre, les algorithmes fonctionnent généralement de manière descendante. Ils tentent de diviser les exemples d'apprentissage de manière itérative en s'appuyant sur les descripteurs (attributs) jusqu'à obtenir des sous-ensembles d'exemples appartenant presque tous à la même classe (nœuds purs). Au fur et à mesure de la construction de cet arbre, des questions doivent être résolues : le nœud est-il terminal ou pas ? si oui, l'algorithme doit affecter une classe à la feuille ; si non, un nouveau test sur les descripteurs doit être formulé pour connaître la structure des sous-arbres (Sheeren 2005)

Pour choisir cet attribut de manière automatique, les algorithmes s'appuient sur des fonctions permettant de mesurer le degré de mélange des exemples entre les différentes classes. L'algorithme C4.5 développé par (Quinlan 1986; Quinlan 2014) exploite la fonction d'entropie tandis qu'un autre algorithme comme CART utilise l'indice de Gini (Breiman 1996).

L'objectif global d'un modèle d'apprentissage vise à construire un modèle de classification qui minimise les erreurs sur les exemples d'apprentissage mais également sur les exemples à classer. Deux méthodes peuvent être utilisées pour estimer cette erreur. La première consiste à diviser le jeu

d'exemples d'apprentissage en deux sous-groupes, l'un pour apprendre le modèle et l'autre pour mesurer l'erreur (le jeu « test »). Etant donné que ce jeu test n'est pas utilisé dans la phase d'apprentissage du modèle, on estime qu'il permet de faire une approximation satisfaisante de l'erreur réelle. Malheureusement, cette méthode n'est pas toujours possible en raison d'un trop faible nombre d'échantillons. Dans ce cas de figure, une seconde approche consiste à utiliser la validation croisée (cf. chapitre 5).

- **Apprentissage du modèle de classification**

L'étape d'apprentissage du modèle de classification a été réalisée au sein du logiciel WEKA. Celui-ci a été développé à l'université de Waikato en Nouvelle Zélande (Witten et al. 1999). Il a l'avantage de regrouper plusieurs algorithmes d'apprentissage supervisés sous la forme d'arbre de décision. L'algorithme de classification choisi est l'algorithme C5.0 (Quinlan 2014) et l'erreur d'apprentissage du modèle a été évaluée en interne par validation croisée.

La précision globale du modèle retenu (évaluée en interne par validation croisée) affiche un haut niveau de performance (Précision globale = 99.15% ; Kappa = 0.98). Ces résultats ne sont toutefois qu'une approximation de l'erreur réelle et les tests de validation réalisés à partir d'exemples indépendants sont moins optimistes (cf. section 7.3.3).

L'algorithme a permis d'établir un arbre de décision relativement simple (quatre feuilles) en ne retenant que deux attributs, l'écart-type des valeurs de teinte dans la bande 1 (le vert) et la brillance (Figure 7-23). On remarque qu'il y a un léger sur-ajustement pour les cinq derniers exemples : l'arbre aurait pu être coupé plus haut, c'est-à-dire avant la quatrième feuille. En effet, ce sont surtout les deux premières branches qui permettent de classer un grand nombre d'exemples (respectivement 144 et 89) et sont par conséquent plus génériques.

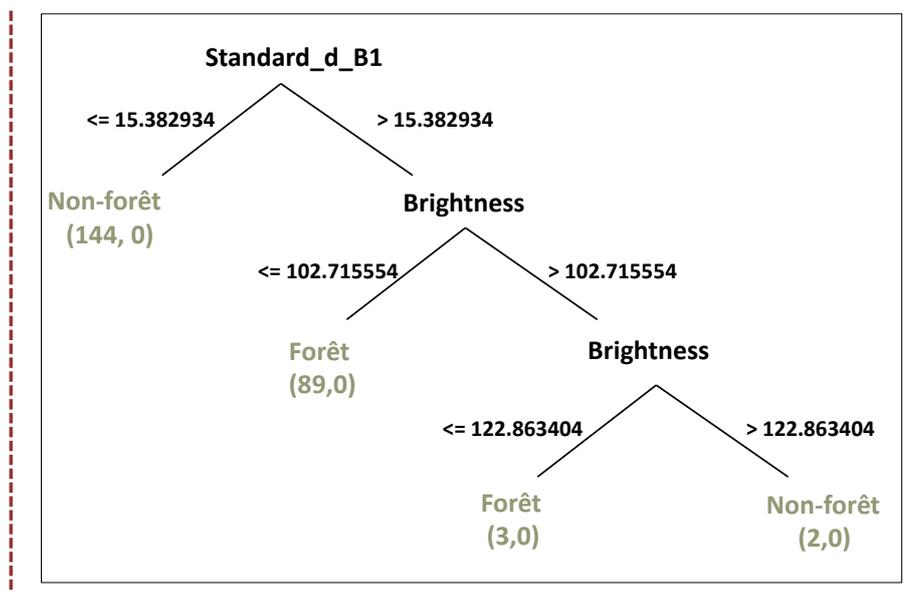


Figure 7-23 Arbre de décision obtenu après application de l'algorithme C5.0

- **L'application des règles de décision pour classer les objets**

La formulation des règles de décision et l'application du modèle de classification ont été effectuées avec le logiciel commercial eCognition. Cet outil permet de formuler différentes règles de décision pour les appliquer successivement au sein d'une structure hiérarchique appelée « rule set ».

Les règles de classification ont été appliquées à chacune des dalles retenues de la BDOrtho IRC préalablement segmentées (Figure 7-24).

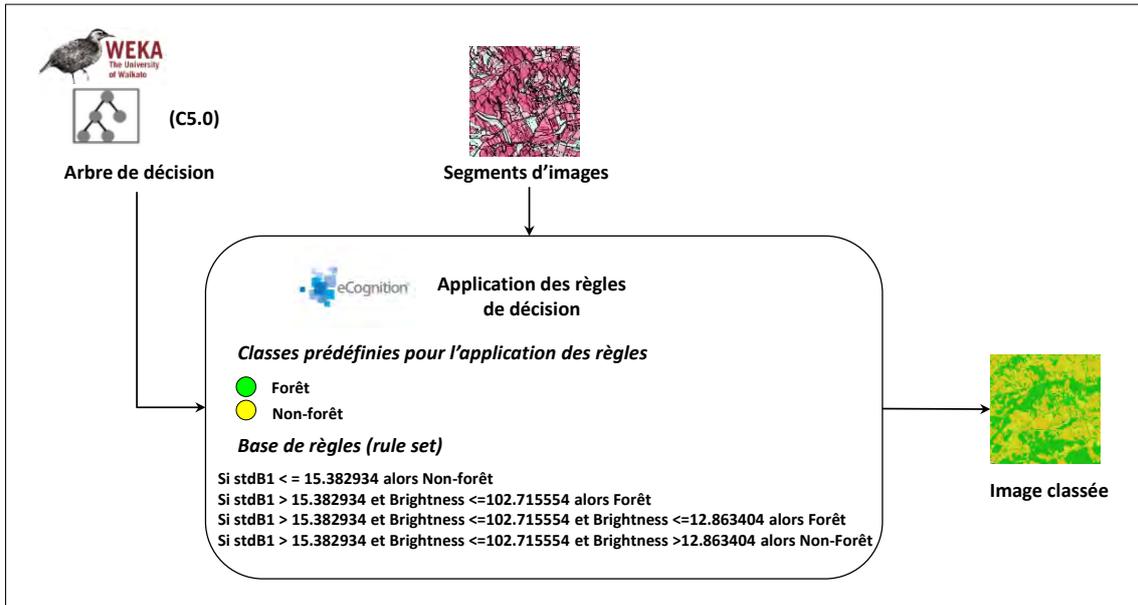


Figure 7-24. Application des règles de décision dans la plateforme eCognition à partir d'un arbre de décision construit avec le logiciel WEKA

TRAITEMENTS POST-CLASSIFICATION

Les traitements post-classification appliqués dans cette thèse ont consisté principalement à éliminer les arbres ou bosquets détectés alors qu'ils ne sont pas en réalité des éléments forestiers. Ces ligneux sont majoritairement des haies qui disposent de valeurs de comptes numériques identiques aux forêts dans les différents canaux.

Les traitements ont été appliqués sur le masque vectorisé de la couche forêt extraite. Premièrement, nous avons créé une zone tampon positive d'une valeur de 5 mètres autour de chaque objet dans le but d'élargir leur surface. Cette étape vise à préserver les objets d'intérêts (les forêts) de la seconde étape qui suit. Nous avons ensuite appliqué une zone tampon négative sur les objets « élargis » dans le but de supprimer les haies tout en préservant nos objets d'intérêt (les forêts) (Figure 7-25).

On peut observer sur l'image 1 les haies à éliminer. Le premier opérateur (zone tampon positive) a eu pour effet de « gonfler » leur superficie (image 2). L'image 3 montre que la seconde étape (zone tampon négative) a permis d'éliminer les haies en tronquant toutefois légèrement les bords des forêts.

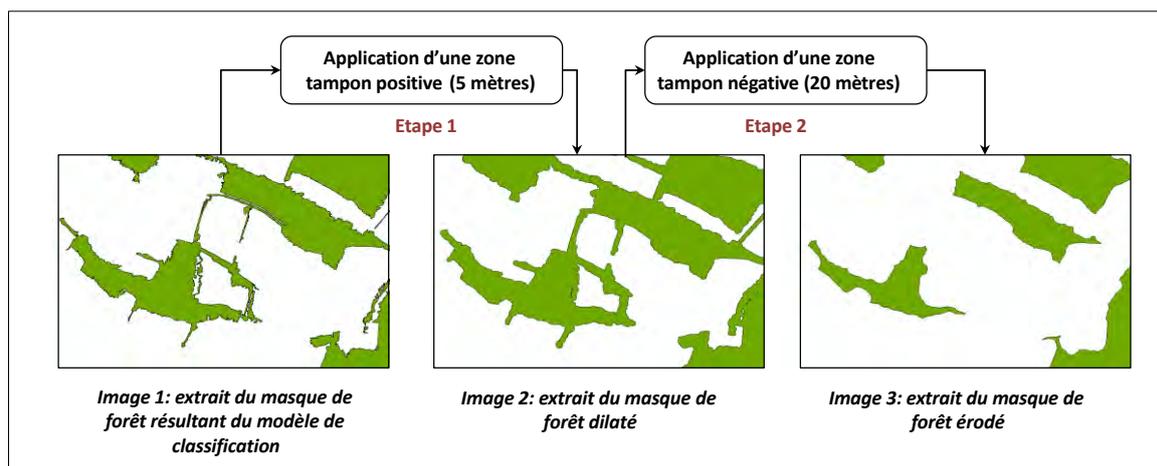


Figure 7-25 Exemple de résultats observés après application des post-traitements : les haies ont été supprimées et les objets « forêt » ont été isolés les uns des autres

7.3.3 RESULTATS ET DISCUSSION

Nous présentons ici les résultats d'extraction obtenus après application de la démarche proposée à l'échelle du site d'étude. Nous nous sommes appuyés sur les indicateurs dérivés de la matrice de confusion et avons considéré la BD Forêts 2010 (MANUEL) comme la couche de référence (cf. chapitre 6). Une spatialisation des erreurs sera toutefois effectuée dans le but d'alimenter la discussion des résultats.

Les résultats d'extraction à l'échelle du site d'étude montrent un niveau de performance encourageant de la démarche proposée (Figure 7-26 et tableau 7-6). Si l'indicateur de précision globale affiche un haut niveau de performance (95.45 %), il semble néanmoins que cette valeur soit très fortement tirée par le taux d'accord de la classe non-forêt. Effectivement, l'indice Kappa est nettement moins élevé (0.72) tandis que le taux d'accord de la classe forêt n'est que de 69.36%. Les indicateurs montrent également que la méthode a plutôt tendance à sous-détecter les forêts (taux de déficit = 30,64%) qu'à les sur-détecter (taux d'excédent = 17,26%).

INDICATEURS DE QUALITE	VALEURS
PRECISION GLOBALE %	95.45
INDICE KAPPA	0.72
TAUX D'ACCORD % (CLASSE FORET)	69.36
OMISSION %	30.64
COMMISSION %	17.26

Tableau 7-6. Résultats de validation de la BD Forêts 2010 (AUTO)

La spatialisation des erreurs de classification permet de mieux comprendre les limites de la méthode proposée (Figure 7-26(3)). Les erreurs d'omission sont liées à deux raisons principales :

(1) une segmentation peu adaptée aux forêts de petite taille (paramètre d'échelle trop élevé). Les objets non détectés disposent globalement d'une surface beaucoup moins importante que les objets détectés (Figure 7-26). La plupart de ces petits objets n'ont pas été individualisés lors de la phase de segmentation. Ils ont été intégrés dans des segments de plus grande taille, plus hétérogènes. Ces segments disposent, par conséquent, de caractéristiques spectrales différentes des autres forêts utilisés dans la phase d'apprentissage et n'ont pu être affectées à cette classe. Le paramètre d'échelle identifié n'est donc pas entièrement adapté puisqu'il génère une sous-segmentation. L'utilisation de deux, voire plusieurs niveaux aurait permis de couvrir une plus large gamme de possibilités et d'éviter ainsi les problèmes de segments hétérogènes. A noter que le

paramétrage des algorithmes de segmentation est un problème récurrent (Blaschke et al. 2014). La détermination d'une segmentation optimale dépassait l'objectif de ce travail. Nous n'avons pas cherché à corriger ces défauts mais nous étudierons plus loin les effets que ces erreurs peuvent engendrer dans les modèles écologiques (chapitre 11).

(2) une imprécision dans la délimitation des objets. Les forêts détectées ont des contours qui ne sont pas toujours rigoureusement exacts du point de vue de leur position. Ces erreurs ont été générées durant la phase de post-traitements (élimination des haies) qui a eu tendance à rogner les objets. Bien que la zone tampon positive (dilatation) ait eu pour conséquence de préserver les objets non visés par la seconde étape (application d'une zone tampon négative - érosion), les paramètres appliqués n'ont pu les tenir à l'écart de certaines suppressions. Une nouvelle fois, l'application de paramètres uniques à l'échelle de zones aussi vastes peine à traiter efficacement les différents cas possibles. Il serait par conséquent nécessaire de les adapter. Une nouvelle fois, nous n'avons pas cherché à optimiser la définition de ces paramètres mais nous en tiendrons compte dans la phase de modélisation écologique.

Concernant les erreurs d'excédents, elles sont beaucoup plus localisées et portent sur des régions de taille plus élevée (Figure 7-26(3)). La plupart du temps, ce sont des régions mixtes composées de forêts et d'autres éléments accolés aux forêts de plus grande étendue. Ce problème est lié une nouvelle fois à la segmentation qui a conduit à créer des objets hétérogènes. Ces derniers ont été affectés à la classe « forêt » par le modèle de classification alors qu'une partie de leur surface appartenait à une autre classe d'occupation du sol.

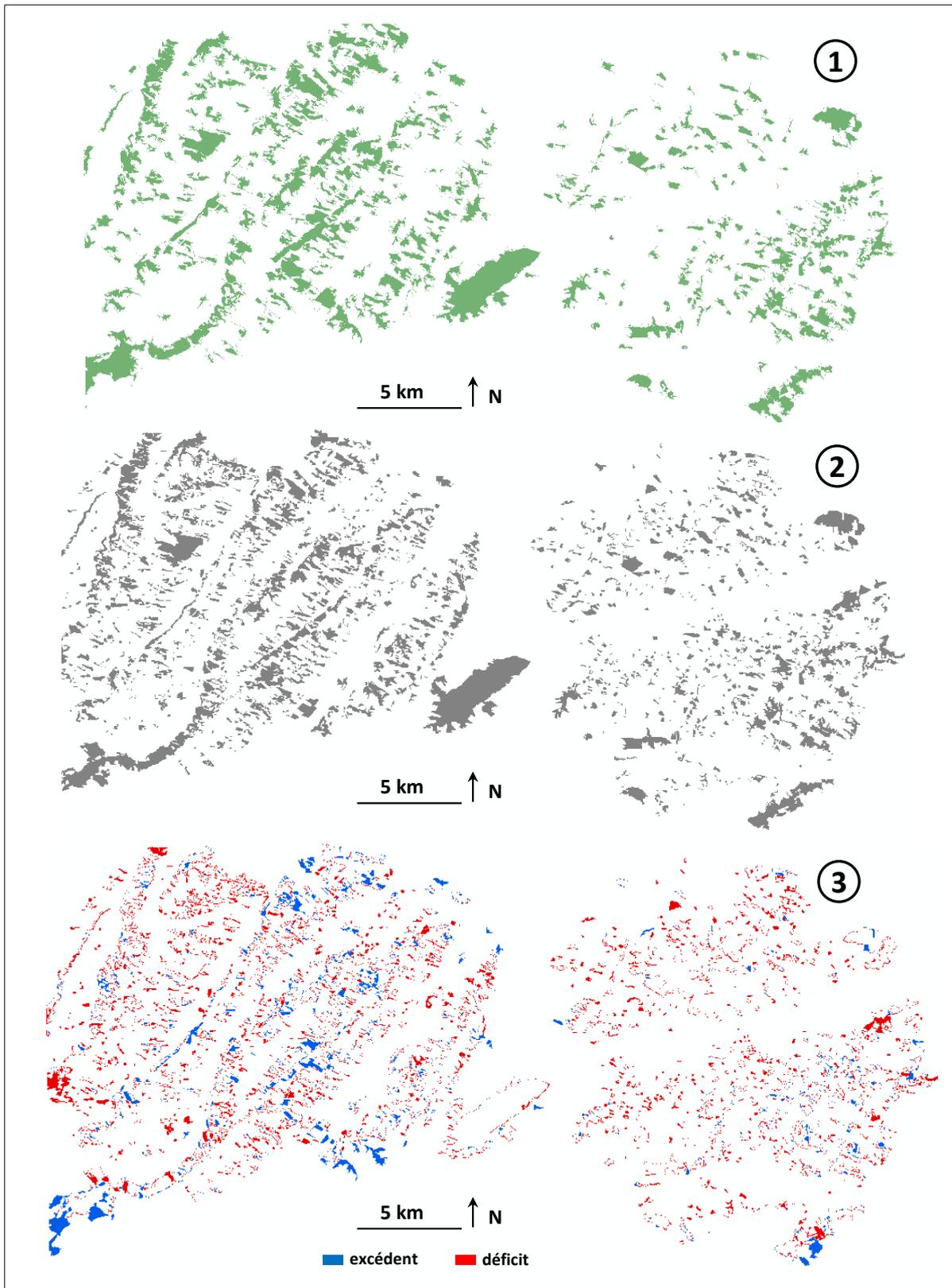


Figure 7-26. Aperçu des résultats d'extraction à partir de la BD Ortho IRC sur le site d'étude. (1) la couche de forêts extraites automatiquement; (2) la couche de forêts extraite manuellement (référence); (3) spatialisation des erreurs de déficit et d'excédent.

Synthèse de chapitre :

Le développement d'approches automatisées pour extraire les éléments des données spatiales devient aujourd'hui nécessaire au regard de la faible opérationnalité de la digitalisation manuelle, en particulier sur des études concernées par de vastes étendues géographiques.

Le premier volet de ce chapitre a permis d'évaluer les performances d'une méthode d'extraction automatique appliquée aux minutes d'Etat-major de 1850. L'outil proposé a montré un haut niveau de performance lorsqu'il est appliqué à des feuilles de la carte d'Etat-major dont le niveau de surcharge permet de visualiser correctement les occupations du sol présentes. Lorsque ce niveau de surcharge est accentué et permet difficilement d'entrevoir les aplats de couleur inscrits sur les feuilles, les résultats sont encourageants mais une intervention humaine paraît nécessaire pour corriger certaines erreurs grossières.

La seconde partie a consisté à développer une méthode qui permet d'identifier automatiquement les forêts sur la BD Ortho IRC de 2010. Nous nous sommes appuyés sur l'approche orientée-objet ainsi qu'une procédure d'apprentissage par arbre de décision. Les résultats indiquent une performance assez bonne de la méthode à l'échelle du site d'étude mais un certain nombre d'erreurs est à déplorer. Elles sont en partie liées à l'échelle de segmentation choisie qui n'a pu prendre en compte la variabilité des forêts observées à l'échelle du site.

D'une manière plus générale, l'apport des procédures automatiques dans les travaux de reconstitution paysagère historique est prometteur. Certes, les méthodes entraînent différentes erreurs qu'il conviendra de gérer à un moment ou un autre de la chaîne de traitement. Elles ont aussi montré leur efficacité à large échelle et permettent de cartographier avec précision les éléments inscrits sur ces documents, notamment la méthode appliquée aux dessins-minutes d'Etat-major. Enfin, s'il apparaît difficile de se passer de toute intervention humaine, ces approches pourraient permettre d'effectuer un travail en amont important permettant de réduire certains biais introduits par la digitalisation manuelle (subjectivité de l'interprétation et de l'échelle d'analyse).

Les travaux relatifs à ce chapitre ont fait l'objet de valorisation scientifique :

HERRAULT P.-A., SHEEREN D., FAUVEL M., PAEGELOW M, Vectorisation automatique des forêts dans les minutes de la carte d'état-major du 19e siècle. *Revue Internationale de Géomatique* 25(1): 35-51 (2015)

HERRAULT P.-A., SHEEREN D., FAUVEL M. et PAEGELOW M. 2013. Vectorisation automatique des forêts dans les minutes de la carte d'Etat-Major du XIXème siècle, *International Conference on Spatial Analysis et GEOmatics (SAGEO'13)*, Brest (France), 24-26 septembre 2013.

HERRAULT P.-A., SHEEREN D., FAUVEL M., PAEGELOW M, Automatic Extraction of Forests from Historical Maps Based on Unsupervised Classification in the CIElab Color Space. [AGILE Conference. 2013](#): 95-112

CONCLUSION DE LA PARTIE B

La première étape, celle de la sélection des données spatiales, a conduit à retenir cinq sources s'échelonnant de 1850 à 2010. Cela nous permet de disposer d'informations spatialisées concernant les surfaces forestières avec un recul suffisant pour évaluer le rôle de l'histoire sur la biodiversité. Ces sources de données présentent toutefois des différences géométriques, sémantiques ou de complétude dont nous devons tenir compte au moment de mettre en évidence les évolutions spatio-temporelles des forêts échantillonnées (Partie C).

Le second chapitre s'est intéressé au problème de la correction géométrique des cartes anciennes et plus précisément à celui des cartes d'Etat-major. Nous avons proposé une nouvelle méthode de correction géométrique fondée sur la régression ridge à noyau gaussien et avons évalué les différents paramètres de la procédure de correction dans le but de définir une stratégie optimale: le nombre de points d'amer, la répartition spatiale des points d'amer, la nature des points d'amer et le choix de la méthode de correction. Deux protocoles de simulation ont alors été appliqués : le premier s'est appuyé sur un des dessins-minutes retenus dans le cadre de notre étude (la feuille 241NE) tandis que le second a été développé dans le cadre du projet OCS ANCIEN et appliqué à quatre dessins minutes aux caractéristiques différentes sélectionnés pour ce projet. D'une manière générale, notre méthode (la régression ridge à noyau gaussien) a fourni de meilleurs résultats que les modèles globaux conventionnels (i.e. les modèles polynomiaux) mais la méthode la plus performante reste la triangulation de Delaunay. Une répartition aléatoire des points d'amers est apparue également plus efficace tandis que la nature des points n'a pas montré d'effets significatifs sur la qualité de correction. Le déterminant le plus important concerne l'effectif de points d'amer : si la hausse de l'effectif améliore systématiquement la qualité de correction pour les méthodes de Delaunay et du noyau gaussien, nous préconisons toutefois de sélectionner au maximum 75 points d'amer sur la zone d'intérêt. Ce seuil est apparu comme le compromis le plus intéressant au regard de l'erreur de correction enregistrée par rapport au temps de traitement nécessaire pour la récolte des amers.

Le troisième chapitre a consisté à détailler les bases de règles de saisie pour vectoriser manuellement les objets dans les différentes sources. Les choix effectués ont permis de limiter les écarts de représentation et de contenu entre sources mais certains d'entre eux seraient susceptibles d'entraîner certains biais dans les modèles écologiques. En effet, certaines différences entre dates pourraient être observées non pas en raison d'une évolution de l'objet mais en raison d'une différence de contenu ou de représentation entre sources. Un rapide bilan statistique des objets vectorisés aux différentes dates a ensuite permis de mettre en évidence un accroissement du nombre d'objets numérisés au cours du temps. Ces observations fournissent des premières indications sur le morcellement du paysage forestier depuis un siècle (fragmentation) et demi dans les coteaux de Gascogne.

Enfin, le quatrième et dernier volet de cette partie a abordé la possibilité de proposer des approches automatisées pour extraire les éléments des sources de données spatiales dans l'optique de faciliter les travaux d'écologie historique. Nous en avons proposé deux :

(1) L'une a permis d'extraire automatiquement les forêts des dessins-minutes d'Etat-major avec un haut niveau de performance en gérant à la fois les problèmes de « bruit » très présents dans ces documents et la variété des représentations pour un même thème. Elle a d'ailleurs été mobilisée dans le cadre du projet OCS ANCIEN et a permis de fournir des résultats très satisfaisants pour différentes feuilles d'Etat-major aux caractéristiques diverses. Ces tests de comparaison ont mis en lumière la possibilité d'appliquer notre approche à différents thèmes d'occupation du sol et la capacité de la méthode à s'adapter à différents niveaux de difficulté. Une intervention humaine

semble toutefois s'imposer pour les cas les plus complexes, c'est-à-dire quand le niveau de surcharge de la carte est tel qu'il masque partiellement voire complètement les aplats d'occupation du sol.

(2) La seconde approche développée dans ce chapitre s'est focalisée sur la détection des forêts de la BD Ortho® IRC. Nous avons, pour cela, mis au point une approche fondée sur la segmentation multi-résolution et un algorithme de classification par apprentissage (les arbres de décision). Les résultats ont démontré un degré de performance encourageant de notre approche mais celle-ci a toutefois entraîné des erreurs grossières. Ces dernières sont principalement dues à une mauvaise prise en compte des objets de petite taille par l'étape de segmentation. La méthode pourrait par conséquent être améliorée en définissant deux voire plusieurs niveaux d'échelle de segmentation pour prendre en compte la variabilité de taille des espaces forestiers.

A noter que ces erreurs de classifications, tout comme celles entraînées par la méthode développée à partir des dessins-minutes, feront l'objet d'un travail spécifique au chapitre 11 : nous évaluerons leur impact sur la réponse des modèles espèces-habitat.

PARTIE C

Apparier les espaces forestiers et caractériser les évolutions spatio-temporelles

INTRODUCTION DE LA PARTIE C

L'objectif global de cette troisième partie vise à appairer les espaces forestiers et caractériser les évolutions spatio-temporelles. Elle est constituée de deux chapitres (chapitres 8 et 9) dont l'un vise à expliciter la démarche proposée pour répondre aux objectifs fixés tandis que l'autre détaillera la mise en œuvre de la démarche proposée.

Le chapitre 8 décrit ainsi la démarche proposée et se déroule en quatre étapes majeures

Nous présenterons dans un premier temps une méthodologie d'appariement des bases de données afin de déterminer la filiation de chaque forêt suivie. Cette phase doit permettre d'explicitier les liens entre objets homologues de chaque base à chacune des dates et d'informer sur la généalogie de chaque objet. La cardinalité des liens d'appariement offrira en plus l'opportunité d'identifier un certain nombre de processus spatio-temporels dont on devra tenir compte dans la démarche de caractérisation à proprement parler.

La caractérisation des états constitue la seconde phase de notre démarche. Il s'agit de qualifier la structure spatiale de l'état des forêts chaque date à l'aide de métriques quantitatives afin de pouvoir étudier l'évolution de la morphologie et de l'isolement de chaque bois. Cette caractérisation doit être mise en œuvre à l'échelle de l'objet puisque nous souhaitons expliquer la richesse spécifique actuelle de syrphes à l'échelle d'une forêt échantillonnée.

Dans le but de caractériser les évolutions spatio-temporelles, deux approches ont été mobilisées. Une première approche dite « individu » servira à dériver un lot de variables explicatives de changement pour chaque bois que nous remobiliserons en Partie 4 au moment de mettre en lien l'histoire du paysage et la richesse spécifique de syrphes forestiers (**BD Changements**). Les métriques quantitatives retenues lors de l'étape précédente seront utilisées pour construire ces différentes variables. La seconde approche dite « population » aura pour objectif de reconstruire les trajectoires d'évolution des espaces boisés. L'idée est de fournir une représentation adaptée des évolutions sur toute la période de temps étudiée dans le but de faire émerger des grandes tendances à l'échelle du groupe de forêts.

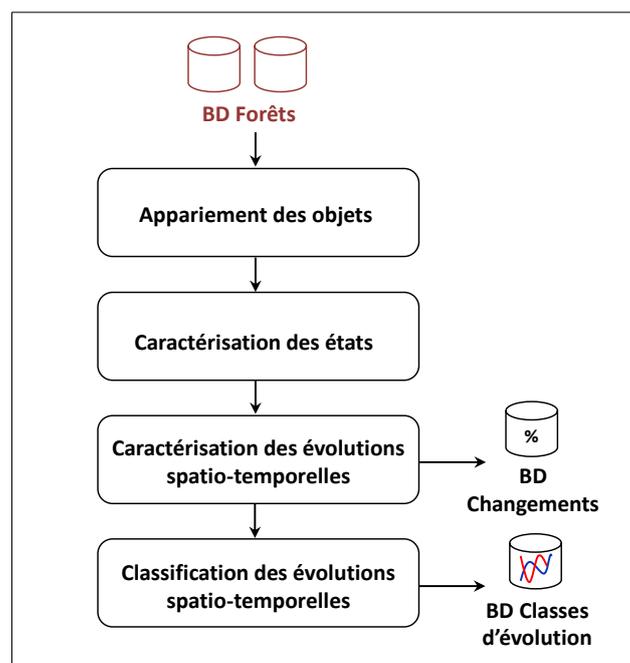


Figure C-1. Démarche générale suivie dans la partie C

La classification des évolutions spatio-temporelle est un prolongement de l'approche « population ». Les trajectoires définies feront l'objet d'une procédure de regroupement pour justement identifier des grandes catégories d'évolution. Nous effectuerons ce regroupement après examen visuel des trajectoires obtenues pour en déduire des catégories d'évolutions.

Le chapitre 9 détaillera la mise en œuvre de la démarche proposée appliquée aux 48 forêts suivies dans le cadre de notre protocole écologique. Les résultats obtenus à l'issue de chaque étape seront par conséquent décrits et discutés dans l'ordre défini par la démarche méthodologique.

CHAPITRE 8.DEMARCHE METHODOLOGIQUE PROPOSEE

SOMMAIRE

- 8.1. L'APPARIEMENT DES FORETS
- 8.2. LA CARACTERISATION DES ETATS
- 8.3. LA CARACTERISATION DES EVOLUTIONS SPATIO-TEMPORELLES
- 8.4. LA CLASSIFICATION DES EVOLUTIONS SPATIO-TEMPORELLES

8.1. L'APPARIEMENT DES FORETS

L'appariement est une étape centrale de notre démarche. Il permet d'associer les données de différents jeux créés précédemment et de définir des liens de filiation entre les homologues. L'appariement conduit à la création d'un ensemble de couples d'objets appariés sur lesquels va être basée la caractérisation des évolutions spatio-temporelles.

8.1.1 L'APPARIEMENT : DEFINITION ET METHODES ACTUELLES

En géomatique, l'appariement est utilisé chaque fois que l'on souhaite mettre en correspondance des objets géographiques se rapportant aux mêmes phénomènes du monde réel mais provenant de jeu de données différents (Sheeren 2002). De fait, la combinaison de couches d'entités spatiales se rapportant à différents phénomènes ne relève pas de l'appariement mais simplement d'une combinaison d'informations permettant de relier ces différents phénomènes.

L'appariement peut être appliqué dans plusieurs contextes :

- pour évaluer la qualité d'un jeu de données géographiques. Dans ce cas, l'appariement sert à mettre en relation la base de données à évaluer et un jeu de données de référence (Bel Hadj Ali, 2001)
- pour extraire et intégrer des mises à jour. L'appariement permet ainsi de détecter les différences entre les versions et faciliter la déduction des évolutions subies (Badard 2000).
- pour intégrer et créer des bases de données multi-résolution. Cela offre alors la possibilité de manipuler simultanément des objets géographiques à différents niveaux d'échelles et selon différents points de vue (Sheeren 2005; Mustière et Devogele 2008).
- pour fusionner des données de plusieurs bases en associant l'information sémantique d'une base des éléments géométriques d'une autre base.

Deux grandes stratégies d'appariement pourraient être employées. Soit on apparie les objets qui forment des couples conformes aux spécifications des deux bases, soit on apparie tous les objets pourvu que ces objets sont jugés représenter la même entité du monde réel.

La première n'entraîne pas d'erreur puisqu'elle est réalisée sur la base d'un ensemble de connaissances, en plus de la taille et la forme des objets. En pratique cela est rarement possible car la base de connaissances est souvent incomplète, il est par conséquent difficile de prévoir toutes les correspondances possibles entre les données même après une analyse approfondie des spécifications (Gesbert 2005).

La seconde ne se soucie pas de la conformité des représentations au moment de l'appariement et l'évaluation est effectuée après appariement. On peut étudier comment les objets se correspondent et apprendre les correspondances incohérentes ou équivalentes dans une seconde phase (Sheeren 2005).

La limite entre les deux approches est parfois confuse. Il apparait en effet difficile de se passer d'une base de connaissances pour apparie même si celle-ci semble incomplète ou trop généraliste. Dans notre cas, la seconde stratégie paraît la plus appropriée à notre problème. Nous souhaitons mettre en relation chaque forêt échantillonnée avec ses homologues dans les sources antérieures. L'emploi de la première approche est inadapté en raison des différences importantes entre sources et l'absence de règles de spécifications explicites pour une partie d'entre elles. De plus, la seconde approche offre la possibilité de relier toutes les forêts du moment qu'elles représentent la même entité dans le monde réel bien que les règles de représentation soient parfois mal connues.

La phase d'appariement a fait l'objet de nombreux travaux ces dernières années pour faciliter son automatisation (Devogele 1997; Bel Hadj Ali 2001; Olteanu 2008; Safra et al. 2013; Huh et al.

2014). La plupart des algorithmes sont basés sur la ressemblance des formes et la localisation géographique des objets de chaque base mais peuvent aussi s'appuyer sur des valeurs d'attributs identiques. Différentes mesures (distances, longueur, taille) peuvent être employées et les différences de représentation permettent de mettre en évidence des liens de cardinalité 0-1, 1-0, 1-1, 1-m, n-1, n-m.

Du point de vue la proximité entre objets polygonaux, la distance surfacique est la mesure la plus couramment utilisée. Elle a été définie par (Vauglin 1997) (Bel Hadj Ali 2001) et utilisée à de nombreuses reprises avec succès pour l'appariement d'entités polygonales (Bel Hadj Ali 2001; Olteanu 2008). Sa définition est donnée en figure 8-1. Il s'agit d'une distance mathématique renvoyant des valeurs comprises entre 0 (objets superposés) et 1 (objets disjoints). Une des difficultés concernant son utilisation, et ceci est vrai pour l'ensemble des mesures d'appariement, porte sur son paramétrage. L'objectif est de fixer une valeur de seuil permettant de rechercher efficacement les candidats à l'appariement. De plus le critère de distance est souvent insuffisant. Deux objets peuvent avoir une distance surfacique faible et présenter des formes très différentes indiquant qu'il ne s'agit pas de la même entité dans le monde réel. C'est la raison pour laquelle, il faut bien souvent compléter le critère de proximité à des critères morphologiques, topologiques et sémantiques (Olteanu 2008).

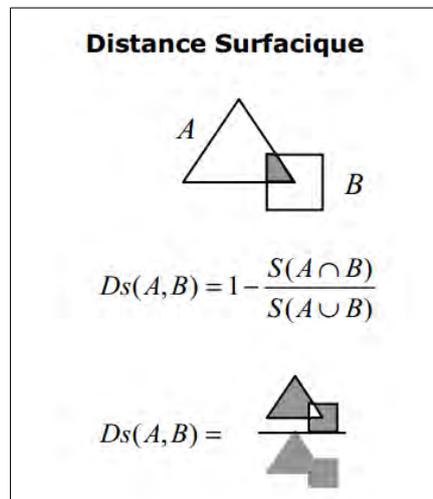


Figure 8-1 Définition de la distance surfaccique (source : Sheeren 2005)

Ainsi, malgré les nombreux travaux existants (on pourra se reporter à la littérature citée pour davantage de détails sur les méthodes), l'automatisation de l'appariement reste un problème difficile et peu d'outils opérationnels existent aujourd'hui.

8.1.2 LE PROBLEME D'APPARIEMENT DES FORETS

L'appariement des forêts est un problème qui apparait naturellement dès lors que l'on souhaite mettre en évidence les évolutions spatio-temporelles des espaces forestiers à travers le temps. Sa mise en œuvre n'en demeure pas moins complexe et tient notamment au fait que la forêt est un objet évolutif. Sa taille, sa forme, ses relations avec ses voisins ou sa simple présence sont susceptibles de varier très fortement d'une date à une autre. Pour ces raisons, la distinction entre équivalences, incohérences ou évolutions peut devenir très difficile à cerner. Ce caractère évolutif de la forêt entraîne également des fusions, des divisions ou des agrégations d'objets à travers le temps. Par conséquent, un patch actuel composé d'une seule entité peut très bien partager une histoire commune avec plusieurs entités dans une des sources antérieures.

Enfin, le fait d'étudier l'évolution des forêts sur le temps long implique d'employer des données spatiales très différentes de plus ou moins bonne qualité qui accentuent les difficultés de mise en relation des objets extraits aux différentes dates. Les homologues peuvent être anormalement éloignés sans que ce décalage soit uniforme sur l'ensemble de la zone d'intérêt. La recherche des candidats devient alors quasi-impossible à automatiser.

Pour toutes ces raisons, nous avons fait le choix d'appliquer une démarche d'appariement basée sur une identification visuelle des homologues. Cette approche est envisageable en raison du faible nombre de forêts à apparier ($N = 48$) et permet d'obtenir des couples d'appariement fiables qui pourront être réutilisés pour quantifier les évolutions spatio-temporelles. La démarche complète de notre appariement visuel est décrite dans la section qui suit.

8.1.3 DEMARCHE D'APPARIEMENT VISUEL

Nous détaillons ici la démarche entreprise pour relier les objets « forêts » vectorisés précédemment sur les différentes sources et mettre en évidence les liens de filiation (Figure 8-2). A noter que cette démarche s'applique uniquement aux forêts contenant des relevés et non à l'ensemble des forêts extraites sur notre site d'étude. Seule la généalogie de ces patches particuliers est nécessaire pour établir un lien entre leur histoire et la biodiversité qu'ils abritent.

Nous avons fait le choix d'appliquer une démarche manuelle basée sur une identification visuelle des homologues. Cette identification a été conduite selon plusieurs critères que nous décrivons en section 8.1.3.1.

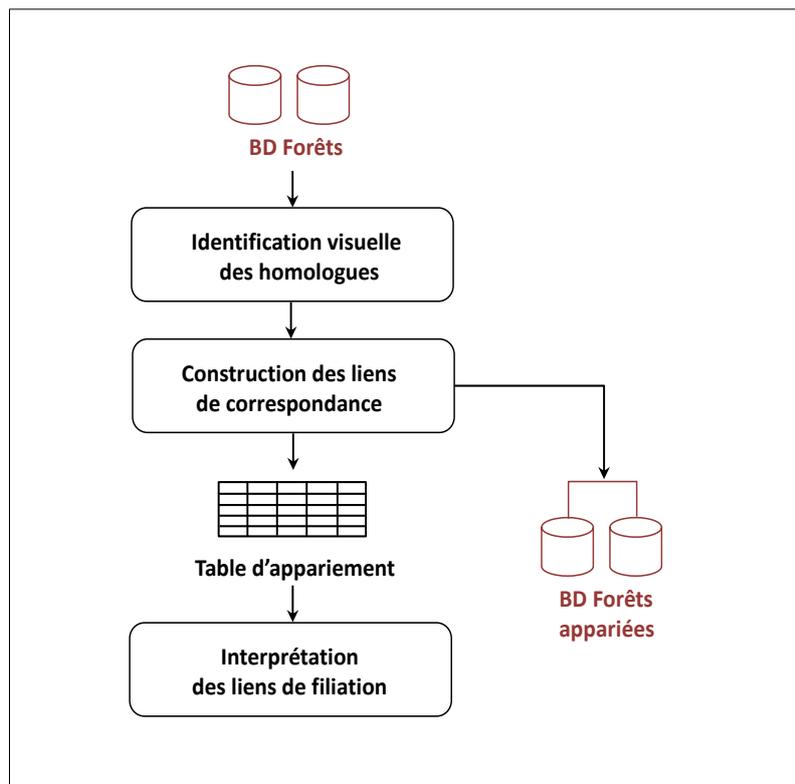


Figure 8-2. Démarche méthodologique pour apparier les espaces forestiers

Cette procédure conduira à la création des liens de filiation entre homologues qui, via une table d'appariement, pourront être exploités et interprétés. Ceci permettra d'obtenir un premier

aperçu des évolutions qualitatives des objets au cours du temps telles que la disparition, l'apparition, la fusion, la division ou la réallocation (Badard 2000; Bordin 2006).

8.1.3.1. LES CRITERES D'APPARIEMENT

L'appariement des forêts extraites sur les différentes sources a été conduit en considérant plusieurs déterminants. Bien que l'approche n'ait pas été automatisée, différents critères ont permis visuellement de rechercher les candidats à l'appariement puis d'établir les liens de filiation.

Nous nous sommes premièrement appuyés sur la ressemblance des formes et des tailles d'objets. Lorsque les forêts ont n'ont pas connu ou peu d'évolutions, ces critères se sont avérées efficaces pour relier les forêts de deux bases de données différentes. Ces critères se sont toutefois avérés insuffisants dès lors que les évolutions entre deux dates ou les différences de représentation étaient plus importantes. Les caractéristiques géométriques des objets extraits peuvent différer très fortement (Figure 8-3). On observe sur la figure que deux objets présents dans la BD 1900 et la BD 1953-54 qui représentent la même entité dans le monde réel peuvent souffrir d'une différence de surface d'environ 7ha, d'une différence de périmètre d'environ 1800m et d'un écart de compacité d'environ 34 points. Ces critères sont par conséquent inadaptés pour relier ces deux objets.

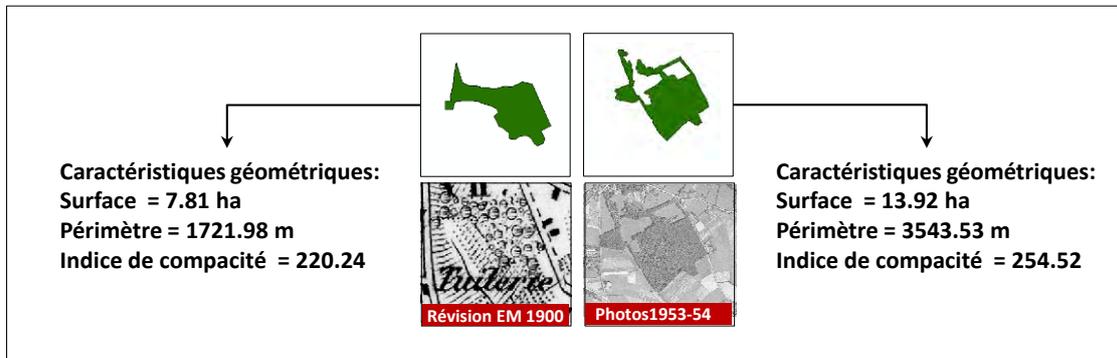


Figure 8-3. Illustration de l'écart géométrique susceptible d'apparaître entre deux homologues

Lorsque ces deux premiers critères ont été jugés inefficaces, nous nous sommes basés uniquement sur la localisation des objets et leur contexte géographique. La superposition de deux objets extraits dans deux bases de données différentes peut informer sur une potentielle histoire partagée par ces deux objets. Mais il est arrivé également que les objets ne se superposent pas ou se superposent par erreur en raison des décalages géométriques entre sources. Dans ces cas précis, le contexte géographique de l'objet a été étudié et a permis de statuer sur ses possibles relations avec les objets antérieurs ou postérieurs (Figure 8-4). Généralement, la distance aux éléments supposés plus stables dans le temps tels que les axes routiers ou le réseau hydrographique s'est avérée efficace. Il y a donc un raisonnement fondé sur une multitude d'informations différentes (forme, taille, position, contexte, relation avec les objets voisins...) qui aident à reconnaître les homologues de façon fiable.

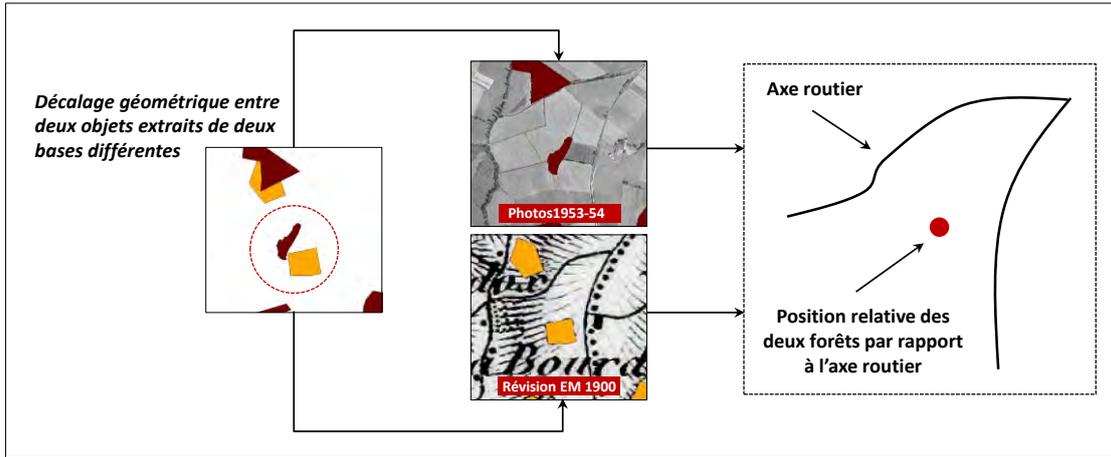


Figure 8-4. Illustration de l'écart de position entre deux homologues. Le contexte géographique des objets offre une alternative efficace pour les relier.

8.1.3.2. LE SENS D'APPARIEMENT

A ce stade de la démarche, il est important de rappeler l'objectif majeur de notre étude. Nous cherchons à mettre en relation la richesse spécifique en syrphes forestiers contenue dans un patch actuel et l'histoire de ce patch échantillonné. Par conséquent, seule nous intéresse l'histoire de ce patch qui contient précisément les relevés biologiques et non d'autres forêts ou parties de forêt qui pourraient se situer dans son voisinage. Ce point est crucial car il conditionne le sens dans lequel doit être effectué notre appariement et ne pas respecter cette logique conduirait à effectuer un appariement erroné. Nous avons donc mis en correspondance nos forêts en partant du présent vers le passé (analyse rétrospective).

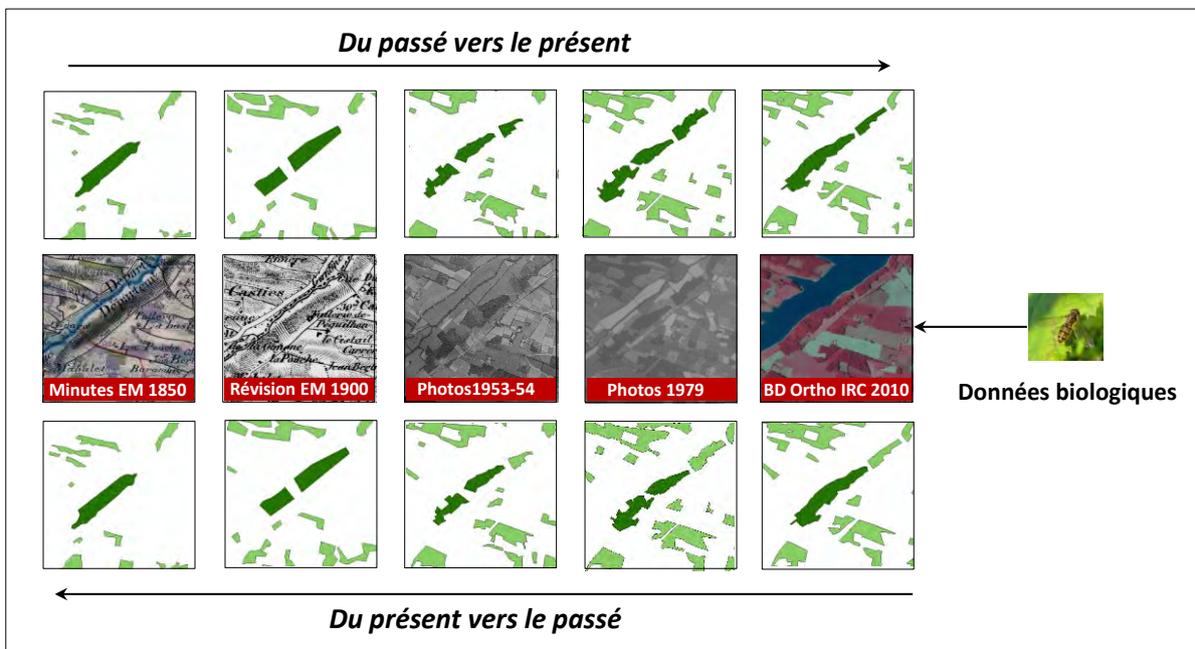


Figure 8-5. Illustration des effets provoqués par le sens d'appariement sur la somme des objets appariés à chaque date

La figure 8-5 illustre deux exemples d'appariement de la même forêt échantillonnée en 2010 réalisée dans deux sens contraires. Lorsqu'il est effectué du présent vers le passé, le point de départ est fixé par l'unique patch dans lequel les relevés ont été effectués. On observe alors que deux patches ont été appariés en 1979 tout comme en 1953-54 et en 1900. Seul un patch a été apparié en 1850. Lorsqu'il est réalisé du passé vers le présent, la somme des objets appariés à chaque date varie. Deux patches ont été appariés en 1900, trois en 1953-1954 et en 1979 puis deux en 2010. Non seulement le nombre d'objets appariés diffère en 1953-54 et en 1979 mais l'appariement a surtout conduit à considérer deux objets en 2010 dont un des deux ne contient pas de relevés biologiques. L'histoire de ce dernier patch n'a donc pas sa place dans notre analyse et même s'il entretient une possible histoire partagée avec le patch échantillonné mise en lumière par ce sens d'appariement, seule l'évolution de celui-contenant les relevés nous intéresse.

8.1.3.3. LA PROCEDURE D'APPARIEMENT

Premièrement, nous avons procédé à une étape d'enrichissement des données. Par défaut, dans une base de données, chaque objet possède un identifiant unique. Etant donné que nous disposons au total de cinq bases indépendantes mais que le principe d'identification des objets est le même pour chacune d'entre elles, des confusions peuvent apparaître dans l'interprétation des liens de filiation. Deux objets de deux bases distinctes peuvent en effet partager le même identifiant sans qu'aucune relation ne soit établie entre ces deux objets. Par conséquent, chaque base de données s'est vue attribuer une lettre allant de A à E tel que A = 2010, B = 1979, C = 1954, D = 1900 et E = 1850 permettant à chaque objet d'une base de disposer d'un identifiant chiffré qui le distingue des autres objets dans une même base et d'une lettre qui le distingue de ses potentiels homologues dans les autres bases de données.

La seconde étape concerne le processus de mise en correspondance des objets à proprement parler. Comme mentionné précédemment, une forêt est un objet évolutif susceptible d'être divisé ou fusionné à travers le temps. Compte tenu de ces possibles évolutions, plusieurs objets à une date peuvent composer un seul objet à une autre date. Cela conduit à la nécessité de prendre en compte des entités multi-parties (ou objet composé). Selon ESRI, "une entité multi-parties est composée de plusieurs parties physiques mais qui ne référence qu'un seul jeu d'attributs dans la base de données. Par exemple, dans une couche d'états, l'état d'Hawaï peut être considéré comme une entité multi-parties. Bien que composé de nombreuses îles, il est enregistré dans la base de données comme une seule entité (ESRI, help.arcgis.com). Nous avons décidé de tenir compte de ce principe sans pour autant affecter un identifiant unique à toutes les parties d'un objet composé. Pour interpréter les liens de filiation, nous avons besoin de connaître les processus de chacune des parties d'une entité, ce qui implique de distinguer chacune d'entre elles par un identifiant particulier. Ce cas précis du méta-objet sera ainsi géré au sein de la table d'appariement que nous décrivons ci-après.

Dans le but d'interpréter la cardinalité des liens d'appariement, une table a été construite pour relier les objets de chaque base et faciliter la lecture des liens de filiation (Annexe F). C'est en réalité une table des correspondances contenant cinq colonnes (une pour chaque date) dont la lecture s'effectue ligne par ligne et permet d'observer les relations qu'entretiennent les objets provenant de base de données différente. Lorsque la généalogie d'un objet conduit à la création d'entités multi-parties, le nombre de lignes croît en fonction du nombre de parties à considérer. Prenons l'exemple illustré en figure 8-6. Nous avons cherché à retracer l'évolution de la forêt n°50A présente en 2010 et l'appariement a été effectué du présent vers le passé (de 2010 vers 1850). Cette forêt est composée de deux parties en 1979 (48B et 50B) alors que l'une des deux parties (48B) était composée elle-même de deux parties en 1953-54 (15C et 16C). Le bois 50B était en revanche présent sous la même forme en 1953-54 et a pris l'identifiant n° 1579C. En 1900, les objets n°1579C et 16C n'ont pas été retrouvés (présence de 0 dans la colonne BD 1900) alors que l'objet 15C était composé de deux parties (115D et 113D). Enfin, ces deux objets ont été retrouvés dans la BD1850 sous la forme d'une seule et unique partie caractérisée par l'identifiant 180E.

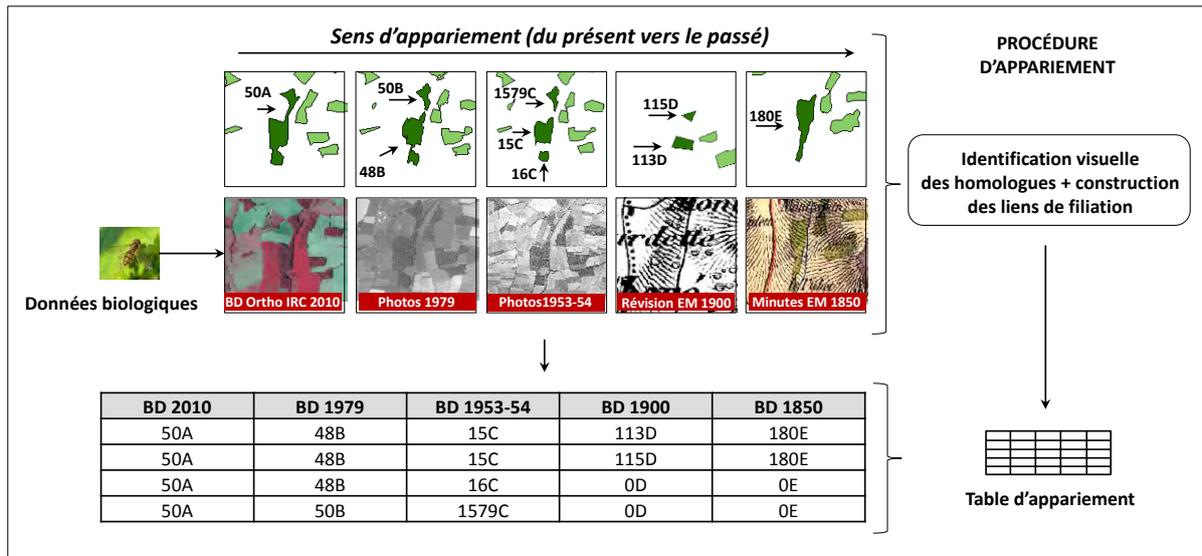


Figure 8-6. Procédure d'établissement de la table d'appariement après l'identification visuelle des homologues

Après avoir établi l'ensemble des liens de correspondance, nous avons cherché à les interpréter. (Figure 8-7). La formalisation de ces liens informe sur la nature du processus d'évolution et leur complexité (Badard 2000; Bordin 2006). Les liens simples correspondent uniquement aux processus entraînant un seul objet. Ils comprennent le lien « 1-1 », synonyme de stabilité d'un point de vue de la présence de l'objet ainsi que les liens « 1-0 » et « 0-1 » correspondant respectivement à une disparition ou une apparition de l'objet. Les liens complexes se distinguent par la présence d'entités multi-parties (ou d'objets composés). Le lien « 1-n » signifie qu'un objet dans une base partage la même histoire que plusieurs objets dans une autre base et inversement pour le lien « n-1 ». Enfin, l'évolution la plus complexe est caractérisée par un lien de type « n-m » ou plusieurs objets dans une base peuvent être mis en relation avec plusieurs objets d'une autre base.

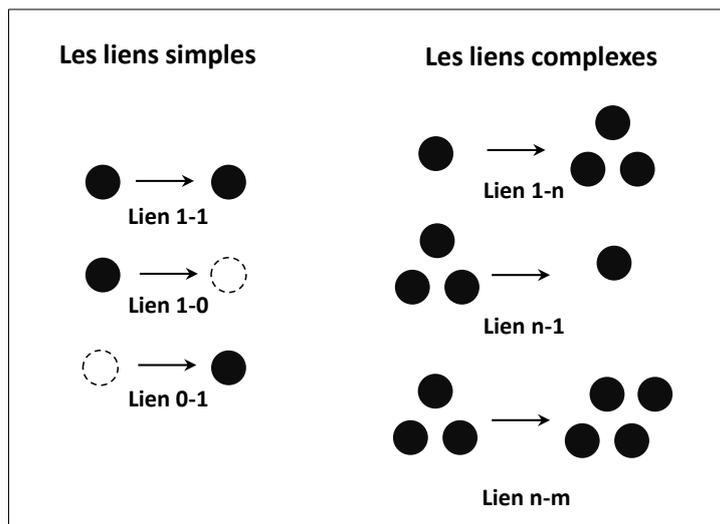


Figure 8-7. Présentation des différents liens de filiation pouvant apparaître à l'issue de la phase d'appariement

8.2. LA CARACTERISATION DES ETATS

La phase d'appariement permet de dériver des informations sur l'aspect qualitatif de l'évolution des forêts (stabilité, disparition/apparition, fusion/fragmentation). Un travail est par conséquent nécessaire pour caractériser les objets de chaque base et ainsi pouvoir quantifier les changements survenus du point de vue de leur géométrie et de leur contexte. Etant donné que nous ne disposons pas d'informations sur les pratiques sylvicoles ou encore la structure verticale des peuplements, notre approche sera essentiellement spatiale. Les données dont nous disposons permettent uniquement de constater la taille et la forme de l'objet ainsi que les relations qu'il entretient avec ses voisins, c'est-à-dire son contexte et son isolement. A noter que seuls les habitats forestiers ont été pris en compte dans le voisinage des patchs échantillonnés étant donné la charge de travail supplémentaire que représentait l'extraction des autres thèmes d'occupation du sol.

8.2.1 DECRIRE LA GEOMETRIE ET LE CONTEXTE DES OBJETS PAR DES METRIQUES

La forêt est un objet géographique. Ce dernier peut être défini comme un objet situé dans un espace ayant des limites bien définies et disposant de propriétés telle que le nom et le type. Il n'est pas uniquement situé dans l'espace mais fait partie intégrante de cet espace vu qu'il hérite forcément de sa structure et de ses propriétés géométriques, topologiques ou météorologiques par exemple.

Ainsi, l'étude des phénomènes spatiaux qui est l'essence même des sciences géographiques s'appuie sur la description de ces objets pour analyser et mesurer leurs évolutions. Cette description porte à la fois sur la géométrie de l'objet qui fait référence à sa position, sa taille et sa forme selon un système de coordonnées de référence ; son contexte qui inclut notamment ses relations avec les objets de son voisinage qu'ils soient de même nature ou non ; et sa thématique qui décrit les attributs d'une entité géographique excepté sa taille, sa forme et son contexte.

Compte tenu de la diversité et de la complexité des objets géographiques, il est difficile de définir une méthode standard pour les caractériser. Dans notre travail, nous nous intéressons aux objets géographiques de types « forêt » caractérisés par des polygones « simples » (composé d'une seule partie) ou complexe (composé de plusieurs parties) à distance inégale les uns des autres. La thématique n'a pas d'importance ici puisque nous ne traitons que des objets de même nature et que nous ne disposons pas d'informations qui permettraient de les différencier du point de vue de leur composition, leur mode de gestion, leur densité ou encore, selon la structure de leur peuplement.

L'essor de l'écologie du paysage ces dernières années a conduit à la création de métriques paysagères permettant de caractériser les entités situés au sein de l'espace qu'est le paysage (Magle et al. 2009). Le nombre de mesure visant à décrire ce type d'entités s'est fortement accru ces dernières années et offre la possibilité de qualifier le paysage à différentes échelles qu'il convient d'adapter au phénomène étudié. Compte tenu des possibilités offertes par ces métriques qui semblent particulièrement bien adaptées à notre problème, notre choix s'est automatiquement tourné vers leur utilisation.

8.2.2 LES METRIQUES PAYSAGERES EN ECOLOGIE DU PAYSAGE

Il existe aujourd'hui un grand nombre de métriques permettant de caractériser la structure du paysage (O'Neill 1986; Turner 1990; Turner et Gardner 1991; Baker et Cai 1992; Gustafson et Parker 1992; McGarigal et Marks 1995; Jaeger 2000; McGarigal 2002). Ces métriques se sont développées parallèlement à l'essor de l'écologie du paysage qui a conduit à mesurer progressivement un certain nombre d'effets du paysage sur la présence de biodiversité.

L'utilisation accrue de ces métriques dans les travaux d'écologie du paysage a également été encouragée par la mise à disposition d'outils intégrant la plupart de ces métriques. Des logiciels tels

que Fragstat, Guidos ou des extensions comme celle de *Patch Analyst* du logiciel ArcGIS facilitent leur calcul et permet de les mobiliser directement pour conduire une analyse spatiale paysagère. Le logiciel Fragstat est l'outil le plus complet en intégrant au total 75 métriques et le plus utilisé par la communauté.

Selon la documentation fournie par le logiciel Fragstat (McGarigal et al. 2002), trois niveaux de métriques peuvent être mis en évidence : le niveau « objet », le niveau « classe » et le niveau « paysage »

- les métriques au niveau « objet » s'attachent à décrire la structure spatiale d'un habitat dans sa classe ou dans un paysage (ex : la compacité d'une parcelle, sa surface, etc.)
- les métriques au niveau « classe » cherchent à caractériser un groupe d'objets appartenant à une même classe (ex : la surface moyennes des parcelles de prairies)
- les métriques au niveau « paysage » permettent de qualifier l'ensemble des entités présentes dans la zone d'étude considérée. (ex : la proportion de chaque occupation du sol dans le paysage étudié, sa diversité, etc.)

Le foisonnement de métriques paysagères et les outils qui les intègrent comportent toutefois certaines limites. Premièrement, il existe une redondance de l'information apportée par ces variables. Plusieurs d'entre elles mesurent la même chose (densité de patch ou moyenne de taille des patches par exemple) et souffrent de fortes corrélations lorsqu'elles sont intégrées simultanément dans un modèle explicatif (Riitters et al. 1995). Ces corrélations peuvent également être provoquées par la corrélation de différents aspects de la structure spatiale elle-même (ex : les patches les plus larges se situent dans les zones les moins connectés). Deuxièmement, un logiciel tel que Fragstat propose de calculer ces métriques à partir de données raster uniquement. Les biais introduits par la rasterisation peuvent alors conduire à obtenir des informations erronées quant à la structure spatiale d'un patch d'habitat en particulier (Bregt et al. 1991; Bettinger et al. 1996; Shortridge 2004). Toutes ces raisons montrent qu'il est nécessaire de faire un choix à la fois sur le type d'outil à utiliser pour calculer ces indicateurs paysagers (et donc la méthode) mais aussi sur les métriques elles-mêmes en ne sélectionnant que celles qui sont adaptées à notre objectif d'étude.

8.2.3 CHOIX DES METRIQUES PAYSAGERES POUR CARACTERISER LES FORETS

Les métriques paysagères sélectionnées doivent permettre d'analyser les effets de la structure spatiale du paysage sur la richesse spécifique actuelle de syrphes forestiers mesurée à l'échelle d'une forêt.

Comme déjà évoqué en Partie A (chapitre 2), les effets impactant la variation de la biodiversité forestière peuvent être classés en trois -groupes : les effets de la morphologie du patch forestier (cf. relation aire/espèce, chapitre 1), les effets de son isolement (cf. théorie de la biogéographie insulaire, chapitre 1) et les effets de la structure interne de son peuplement (cf. notion de niches écologiques, chapitre 1) (Deconchat et Balent 1996). Nous ne disposons pas de données permettant de mesurer les effets du troisième groupe. Nous nous concentrons donc sur les métriques capables de caractériser la taille et la forme des habitats (effets de la morphologie) et celles qui permettent de caractériser le contexte paysager dans lequel ils s'inscrivent, en particulier la présence d'autres habitats de même classe (ici les forêts) dans le voisinage de l'objet (effets de l'isolement). De plus, les métriques choisies doivent être calculées à l'échelle de l'objet afin de pouvoir mesurer les effets à l'échelle d'un patch contenant des relevés biologiques. Un des critères essentiels dans le choix des métriques concerne ainsi la capacité de celles-ci à pouvoir caractériser un objet simple (composée d'une seule partie) et un objet composé (entité multi-partie) pour faire en sorte que ces deux objets soient comparables sur la base de la même métrique. Enfin, nous avons fait le choix de calculer ces

métriques et de caractériser la structure spatiale des forêts au format vectoriel en utilisant le logiciel Arcgis 10.1 dans le but d'éviter les biais potentiels liés à la rasterisation.

8.2.3.1. LES METRIQUES POUR MESURER LES EFFETS DE LA MORPHOLOGIE DES FORETS

Les effets de la morphologie des forêts concernent à la fois les effets liés à leur taille, leur périmètre (longueur de lisière) et leur forme.

La taille d'un patch d'habitat est considérée comme un des facteurs majeurs de la biodiversité présente. (Connor et McCoy 1979; Andren 1994; Bender et al. 1998; Fahrig 2003; Ouin et al. 2006; Ewers et al. 2007; Prugh et al. 2008). La plupart des études ayant pour but de tester ces effets s'appuient sur une simple métrique de surface. Elle permet de caractériser à la fois un objet simple et un objet composé en effectuant la somme des différentes parties selon ce qu'on souhaite étudier. Son calcul peut alors porter sur l'objet dans sa totalité ou bien simplement une partie. Il est par exemple possible de déduire la surface du cœur¹ de l'objet de la surface totale pour étudier les effets du milieu intérieur du patch d'habitat. Cette métrique est toutefois très corrélée à la surface totale.

Le périmètre d'un objet « forêt » informe sur la structure de sa lisière. La lisière d'un patch est une zone tampon, une limite entre deux milieux dont l'un des deux est généralement forestier, par exemple entre une forêt et une prairie ou une forêt et une plage. Elle dispose d'une certaine largeur et se caractérise par des conditions microclimatiques et écologiques particulières, des micro-habitats spécifiques favorables ou défavorables aux espèces des milieux adjacents (Rapport BILISSE-GIP Ecofor). En écologie du paysage, et ce pour décrire le fonctionnement écologique des lisières, on parle souvent d'écotone. Cette notion décrit une zone de variation subtile entre deux milieux dont le principal déterminant est une synergie de facteurs écologiques et non un facteur physique majeur. Les effets des lisières sur la biodiversité forestière ont été largement démontrés et sont réputés s'estomper au fur et à mesure que l'on s'en éloigne (Malcolm 1994; Rand et al. 2006; Alignier 2010; Alignier et Deconchat 2011). En pratique, sa délimitation est souvent très difficile, elle ne dispose pas de limites claires, sa représentation peut alors varier très fortement. Pour mesurer ses effets, les auteurs s'appuient le plus souvent sur sa longueur (le périmètre du patch d'habitat) et sa surface (Ranta et al. 1998; Gascon et al. 1999; Fischer et Church 2003). Pour déterminer sa surface, l'intervention humaine est souvent requise pour fixer la largeur de la zone tampon et permettre son calcul.

La forme d'une forêt fait directement référence aux variations des bords ou des marges d'une surface et est considéré comme une des propriétés majeures pour décrire un patch forestier (Forman et Godron 1981). Elle décrit les caractéristiques des limites d'un objet qui peuvent ainsi traduire un certain degré de perturbation du patch considéré. Dans ce sens, les forêts impactées par les feux peuvent par exemple présenter des limites très complexes et découpées (Haydon et al. 2000; Lindemann et Baker 2001). Les forêts fluviales dont la forme est le plus souvent allongée sont très sensibles à certains processus tels que les inondations ou la sédimentation. Une forêt allongée est plus vulnérable car le cœur est restreint voire inexistant (si forme trop allongée). Il a également été montré que les vieilles forêts matures du Wisconsin disposaient d'une forme beaucoup plus complexe que les forêts gérées par l'Homme (Mladenoff et al. 1993) ou que les forêts natives étaient à la fois plus complexes et plus allongés que les forêts exotiques dans le nord-ouest de l'Espagne (Saura et Carballal 2004). Il existe un grand nombre de métriques de forme. Ces dernières peuvent être divisées en deux sous-classes. L'une est destinée à caractériser les caractéristiques basiques de l'objet telles que la rectangularité, la rondeur ou le degré d'élongation de cet objet. L'autre vise plutôt à décrire la complexité de l'objet comme la dimension fractale d'un polygone. Toutes ces métriques ont aussi la particularité d'être fortement corrélés et il n'est pas utile d'en calculer plusieurs lorsqu'il s'agit de mesurer la forme. Cela évite en effet d'introduire des variables colinéaires

¹ Surface de cœur : surface totale de l'objet – surface de lisière

dans les modèles (cf. partie D). (Pour une revue complète des métriques de forme, voir (Jiao et al. 2012).

Au regard des différents aspects dont il faut tenir compte pour évaluer les effets de la morphologie des bois sur la variation de la biodiversité forestière et des problèmes de corrélations entre variables, nous avons fait le choix d'utiliser une métrique simple par type d'effets. Une métrique de surface sera mobilisée pour analyser les effets de la taille du bois. Dans le cas d'entités multi-parties, nous calculerons la somme totale des surfaces des différentes parties. Le périmètre permettra de qualifier la longueur de lisière des forêts. Là aussi, nous effectuerons la somme totale des périmètres des différentes parties pour caractériser un objet composé. Enfin, nous emploierons une métrique simple de compacité (périmètre/surface) pour mesurer l'effet de la complexité de la forêt sur la biodiversité. Cette mesure permet d'obtenir une information à la fois sur la structure du milieu intérieur et de l'objet et la complexité de ses limites. Pour affecter une valeur de compacité à un objet composé, nous avons cette fois calculé la moyenne des valeurs d'indices obtenues pour les différentes parties.

8.2.3.2. LES METRIQUES POUR MESURER LES EFFETS D'ISOLEMENT DES FORETS

Lorsqu'un paysage devient discontinu, la capacité de déplacement effectué par les organismes entre les patchs d'habitats est un critère essentiel pour maintenir la stabilité génétique et démographique des populations (Hanski 1994; Hanski et Gaggiotti 2004). De plus, le degré d'isolement du patch est couramment considéré comme un facteur déterminant de la distribution et de l'abondance des espèces (Partie A, chapitre A) (Crooks 2002; Crooks et Sanjayan 2006). De nombreuses métriques existent pour quantifier la connectivité (qui peut être vue comme l'isolement) d'un patch allant de la plus simple (simple distance entre les patchs) à la plus complexe (modèles intégrant les surfaces de coûts ou la théorie des graphes) (Magle et al. 2009). Ces métriques peuvent être classées en trois catégories distinctes.

Les métriques structurelles visent à quantifier la simple configuration physique des patchs d'habitat. Parmi elles, la distance euclidienne entre patchs ou encore la distance euclidienne pondérée par la surface des patchs voisins sont régulièrement utilisées (Magle et al. 2009).

La connectivité potentielle a pour objectif de mesurer une quantité d'habitat dans un rayon donné autour du patch d'intérêt et peut être complexifiée en intégrant des paramètres traduisant les capacités de mouvements de l'organisme considéré. Il a récemment été montré que les effets de la taille et de la connectivité d'un patch pouvaient être mesurés conjointement par ce type de métriques (Fahrig 2013). En effet, il est possible que l'évolution de la surface d'un patch tende à faire croître significativement la quantité d'habitat dans un rayon donné si celui-ci est pris en compte dans ce calcul. L'effet de l'isolement d'un habitat sur la richesse spécifique peut également être visible uniquement à travers une quantité d'habitat plutôt qu'à travers une distance inter-patchs. Ces métriques sont également facilement adaptables à un groupe d'objets. A contrario, la métrique « Effective Mesh Size » particulièrement utilisée dans les études de fragmentation du paysage offre la possibilité de mesurer la connectivité d'un groupe d'objets en décrivant la surface prise par les différentes parties de cet objet composé (Jaeger 2000) Cette métrique calcule en réalité la surface moyenne des différentes parties d'un objet composé pondérée par ces différentes valeurs de surface. La pondération vise à obtenir une moyenne plus « réaliste » de cette taille moyenne lorsque deux ou plusieurs objets aux surfaces très éloignées composent une entité multi-partie.

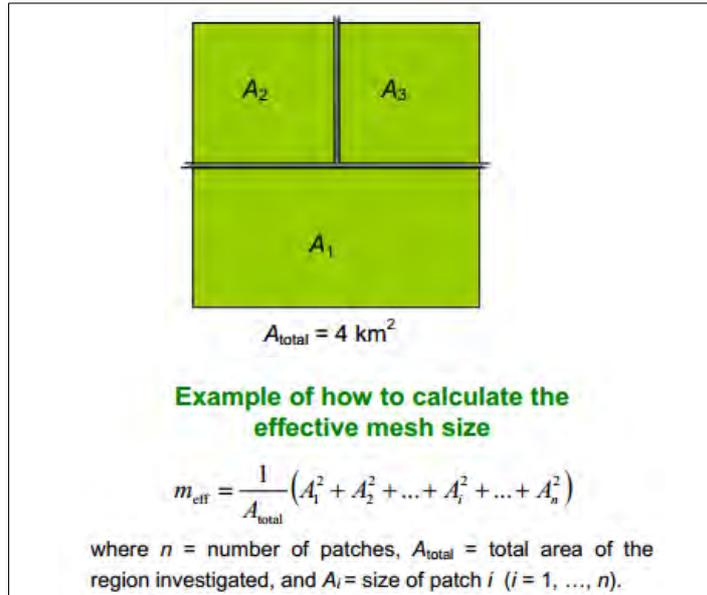


Figure 8-8. Définition de la métrique « effective mesh-size » (source : Jaeger 2000)

Enfin, les métriques de connectivité fonctionnelle retranscrivent la capacité d'un groupe d'espèces à se déplacer dans un paysage en tenant compte de la composition et de la configuration de la matrice entre les patches d'habitats d'intérêt. Par exemple, une espèce forestière pourrait avoir une capacité de déplacement d'une forêt à une autre plus aisée dans un paysage contenant des forêts entourées majoritairement de prairies et friches que dans un paysage entouré de cultures. Parmi elles, les métriques s'appuyant sur la théorie des graphes et les surfaces de coûts ont fourni des résultats très performants bien qu'un grand nombre de paramètres soit nécessaire pour les calibrer (Theobald et al. 2006; Foltête et al. 2012).

Nous avons privilégié une seule métrique pour mesurer les effets de l'isolement d'un habitat sur la richesse spécifique qu'il abrite. Nous calculerons un pourcentage d'habitats forestier dans un rayon donné autour du patch d'intérêt. Cette métrique est reconnue comme une métrique simple mais suffisante pour estimer le degré de connectivité d'un habitat (Magle et al. 2009). Comme nous ne souhaitons pas confondre les effets de la taille et de l'isolement, une enveloppe convexe a été dessinée autour du patch d'intérêt ou du groupe de patches d'intérêt (dans le cas d'une entité multi-partie) pour prendre uniquement en considération les patches voisins (Figure 8-9). L'aire d'étude a été finalement définie sur la base de l'enveloppe convexe par une zone tampon (rayon fixé à 4800 mètres). Cette valeur a été retenue sur la base des travaux de (Ouin et al. 2006) qui ont identifié cette échelle comme la plus adaptée pour évaluer les effets de l'isolement sur la richesse de syrphes dans les forêts des coteaux de Gascogne. A noter que toutes les surfaces boisées intersectant cette aire d'étude ont été prises en compte dans le calcul du rapport entre la surface totale des patches voisins et l'aire d'étude à l'exception de l'objet lui-même. Ceci a pour but de favoriser la continuité des habitats en considérant qu'un syrpe capable de coloniser un bois profitera de la connectivité de ce bois pour le coloniser dans sa totalité.

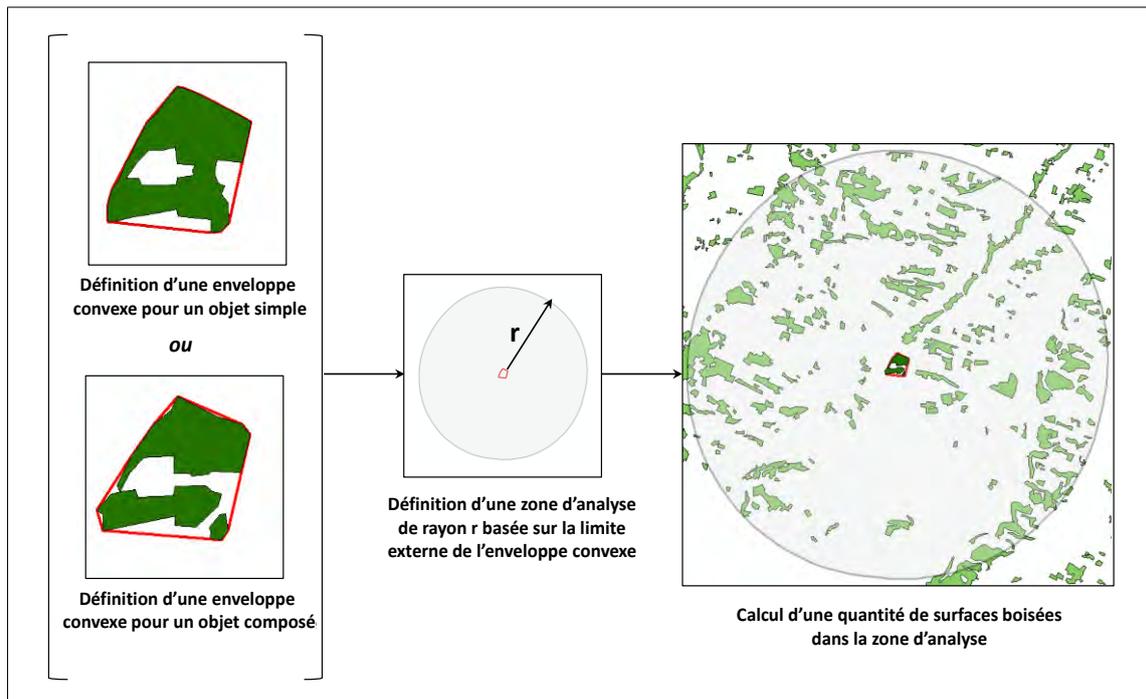


Figure 8-9. Procédure de traitement pour déterminer la connectivité potentielle d'un patch d'habitat vue comme une densité d'habitat dans son voisinage

8.3. LA CARACTERISATION DES ENTITES SPATIO-TEMPORELLES

Nous présentons maintenant la démarche employée pour étudier l'évolution spatio-temporelle des espaces boisés sur une période de temps approximativement égale à un siècle et demi. Pour commencer, nous effectuerons un rappel des principaux concepts en lien avec la prise en compte du temps dans le suivi des phénomènes géographiques puis décrirons les approches existantes pouvant s'appliquer ou non à notre problématique. L'analyse des évolutions possibles des surfaces boisées permettra ensuite d'identifier toutes les possibilités dont on devra tenir compte dans notre démarche avant de décrire à proprement parler la procédure mise en œuvre.

8.3.1 ETATS, EVENEMENTS ET PROCESSUS : DEFINITIONS

L'étude de l'évolution des phénomènes est une thématique récurrente en géographie. Selon (Grenon et Smith 2004), l'évolution des objets du monde réel peut être appréhendée de deux manières différentes : (1) soit on s'appuie sur les objets géographiques se situant dans l'espace et sur une certaine durée bien que leur nature soit susceptible de changer ; (2) soit on tient compte des processus (ou événements) qu'il est possible de situer dans le temps mais pas dans l'espace. Les évolutions sont par conséquent structurées à la fois par une logique spatiale (les objets géographiques) et/ou une logique temporelle (les événements et les processus) (Buard 2013).

La première approche implique d'observer les états d'un phénomène à plusieurs dates données. L'état d'un phénomène géographique peut être défini comme une photographie d'un objet géographique prise à un instant de son évolution. De fait, son évolution n'est pas directement observable, c'est donc à l'utilisateur de comparer ces différents états pour évaluer les changements des objets géographiques. Ces changements peuvent porter sur plusieurs propriétés : (1) les propriétés thématiques dans le cas où les attributs d'objets varient : nature d'occupation du sol, nombre d'habitants dans une ville ; (2) les propriétés spatiales si la surface de l'objet, sa forme ou sa localisation évoluent ; (3) les propriétés de « généalogie » si l'objet se scinde en plusieurs objets, se

fusionne, s'il disparaît ou apparaît. Grenon et Smith (1994) (d'après Mathian et Sanders, 2014) proposent de définir les entités mobilisées dans ce type d'approche comme des entités SNAP (ou endurantes). Elles ont la capacité de persister dans le temps et sont observables à tout moment de leur existence. Elles ne disposent pas de « parties temporelles » (Mathian et Sanders, 2014) ; si elles évoluent, elles maintiennent leur identité.

La seconde approche suggère de s'appuyer directement sur les processus (ou évènements). Les processus sont des situations dynamiques, relativement homogènes et continues (sans fin) tandis que les évènements sont des modèles de situation non homogènes et limitées dans le temps (Galton et Mizoguchi, 2009). Ils impliquent tous les deux des transformations géométriques ou sémantiques entre des états successifs d'un objet géographique. Par opposition aux entités SNAP, Grenon et Smith (1994) désignent ici les entités comme des entités SPAN (ou perdurantes). Il s'agit d'entités de type trajectoire: elles ont un début et une fin et sont caractérisées par une succession d'intervalles de temps (Livet, 2009), chacun correspondant à un changement (Mathian et Sanders, 2014).

Ces deux approches ne sont pas pour autant antinomiques mais peuvent être complémentaires. Dans le cas du suivi du phénomène d'évolution forestier, une forêt est une entité SNAP alors que sa trajectoire de vie est une entité SPAN composée de différentes phases de jeunesse et de maturité. Les deux entités sont liées : une coupe de l'entité SPAN correspond à un état (en terme d'existence et de propriétés) de l'entité SNAP associée (Bittner et al. 2004). Le tableau 8-1 reprend les couples d'entités SNAP et SPAN qui entretiennent d'étroites relations entre elles. On observe d'un côté les objets avec leurs propriétés, de l'autre les trajectoires qui structurent ces objets dans le temps sous l'influence de processus et d'évènements. Par exemple, les entités SNAP de type forêts ou lac sont reliés respectivement aux processus d'embroussaillage et de salinisation (entités perdurantes) qui donnent lieu à une fermeture du milieu forestier ou à l'augmentation de la salinité de l'eau.

TERMINOLOGIE TYPES D'ENTITES	ENTITES SNAP (OU ENDURANTES)	ENTITES SPAN (OU PERDURANTES)
POINT DE VUE ONTOLOGIQUE	<ul style="list-style-type: none"> Objets (fixes ou mobiles) Propriétés 	<ul style="list-style-type: none"> Trajectoires Processus et évènements
POINT DE VUE THEMATIQUE	<ul style="list-style-type: none"> individus (êtres humains, navires, troupeaux) villes pays forêts, lacs routes 	<ul style="list-style-type: none"> parcours, traversées, déplacements urbanisation, congestion, croissance création, recomposition embroussaillage, salinisation trafic

Tableau 8-1 Différents rapports au temps des entités : points de vue ontologique et thématique (source : Mathian et Sanders, 2014)

8.3.2 DECRIRE L'EVOLUTION DES OBJETS GEOGRAPHIQUES : PLUSIEURS MODELES

Après avoir décrit les deux grandes approches couramment employées pour suivre l'évolution d'un phénomène géographique, nous présentons ici les différents modèles d'application selon ces deux courants. On distingue les modèles basés sur les états et les changements des modèles basés sur les évènements et les processus.

Avant de procéder à une revue des différents modèles existants, il est important de rappeler que nous nous plaçons dans un cadre de modélisation descriptive. Nous souhaitons décrire des états, des changements ou des rythmes de changements sans vouloir toutefois en expliquer les causes (qui

sort du cadre de ce travail). Seules des informations sur l'évolution de la structure spatiale des forêts sont analysables à partir de nos données spatialisées et seront finalement reliées aux données biologiques actuelles dont nous disposons.

Les modèles basés sur les états ou les changements

Les modèles de ce type les plus couramment utilisés sont les modèles **Snapshot** (Plumejeaud 2011, Harbelot et al.2015). Ils s'appuient sur des états spatiaux ordonnés à différents temps où à chaque temps est disponible une couche d'entités indépendante sans relation directe avec les autres couches. L'intervalle entre deux états peut varier et n'a pas d'incidence sur le modèle (Del Mondo 2011). Dans la lignée des modèles snapshot, le modèle **Space Time Composite (STC)** intègre les changements dans les attributs des entités (Langran et Chrisman 1988). Le phénomène est représenté par un espace homogène dans lequel les attributs temporels changent et sont capturés à des temps discrets (figure 8-10). Sa principale limite repose sur le fait qu'il n'est pas possible de repérer un changement dans l'espace.

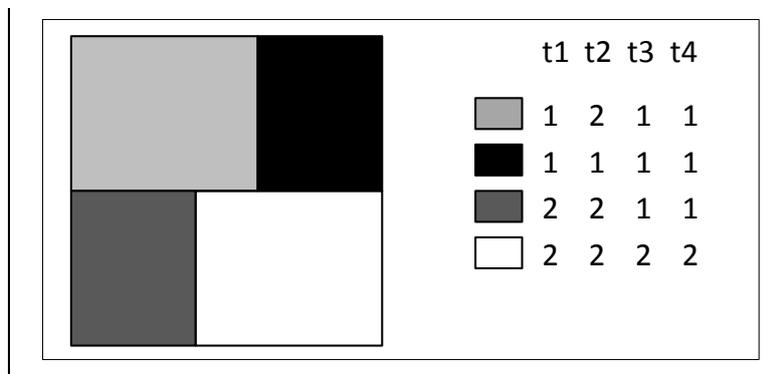


Figure 8-10. Exemple d'un modèle STC pour la rotation des cultures (1 = Blé, 2 = Maïs)

Le modèle **Spatio Temporal Object (STO)** représente les changements sur une structure de temps discrète et linéaire. Il s'appuie sur les empreintes spatiales de l'entité aux temps auxquels elle existe mais ne permet pas de mettre en évidence les changements de généalogie de l'objet. D'autres modèles ont ainsi vu le jour pour pallier ce problème. Le **langage descriptif (Change Description Language)** développé par (Hornsby et Egenhofer 2000) est basé sur les concepts d'existence ou de non existence de l'entité. Il a pour objectif de décrire qualitativement les changements survenus via des opérateurs qui agissent sur l'entité tels que la destruction, la création ou la séparation (Figure 8-11). Il est toutefois difficile de représenter l'évolution globale de l'objet puisque les opérateurs logiques proposés permettent de modéliser les changements uniquement entre deux instants.

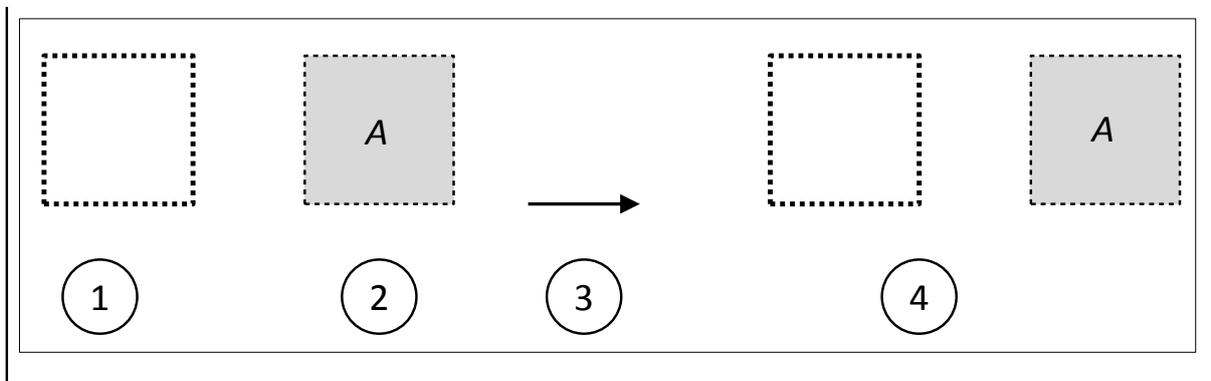


Figure 8-11 Les primitives de base du langage CDL: (1) entité qui n'existe pas; (2) entité qui existe; (3) transition entre deux instants (4) création d'une entité (source : Del Mondo 2011 d'après Hornsby et Egenhofer 2000)

Enfin, des modèles de trajectoires offrent la possibilité de suivre l'évolution spatiale des entités en décrivant les déplacements d'une entité en lien avec une activité (Hagerstrand 1968). Cette forme de représentation sera notamment reprise par (Hornsby et Egenhofer 2002) travers des **géolifelines**. Ce dernier modèle s'appuie sur des périodes de temps pendant lesquelles une entité occupe un espace, l'évolution de l'entité est ainsi représentée par une *geospatial lifeline* de différentes formes selon le type de mouvement adopté (figure 8-12). La représentation spatiale y est explicite tandis que l'analyse des changements d'états n'est pas possible.

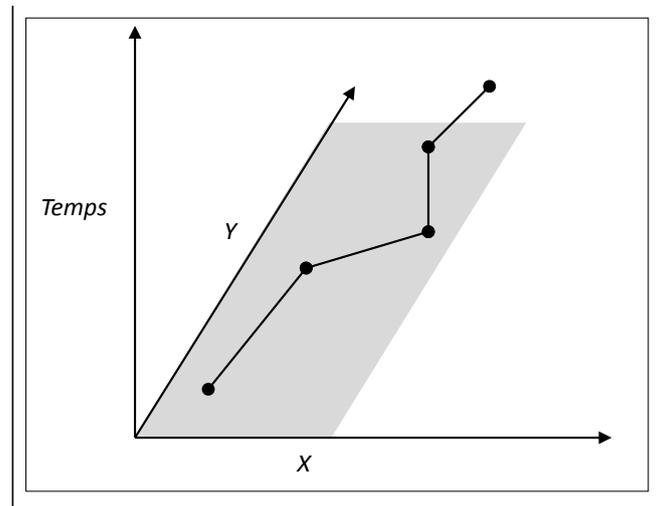


Figure 8-12 Représentation d'une géolifeline

Les modèles basés sur les événements et les processus

Peuquet et Duan (1995) ont proposé le modèle **Event Based Spatio Temporal Data Model (ESTDM)**. Les événements sont placés le long d'une droite temporelle sous la forme d'une liste et un nouvel événement est créé si un changement est détecté. Si l'information est bien inscrite temporellement dans ce modèle, elle demeure néanmoins statique et fragmentée. **La typologie des processus spatio-temporels de** (Claramunt et Thériault 1995) propose de formaliser les liens connus entre les événements et leurs conséquences (Figure 8-13). Au premier groupe traitant de l'évolution d'un objet simple, ils ajoutent deux autres catégories portant sur les relations entre objets : (1) un groupe où les relations concernent la généalogie et la diffusion spatiale ; (2) un groupe où les changements affectent la structure spatiale d'un ensemble d'objets. Cet enrichissement est fondamental puisqu'il donne la possibilité de tenir compte de toutes les relations entre objets qu'elles soient spatiales, d'appartenance ou fonctionnelles (Mathian et Sanders, 2014). (Worboys 2005) apportera lui aussi un formalisme et un langage basés sur les événements pour représenter les changements mais cette fois-ci localisés dans l'espace. Repris par (Jiang et Worboys 2009), les événements sont différenciés et qualifiés de **topological event**. Pour chaque snapshot, la structure de l'événement est modélisée sous forme d'un arbre et il devient possible d'évaluer les changements en analysant l'évolution de la forme de l'arbre.

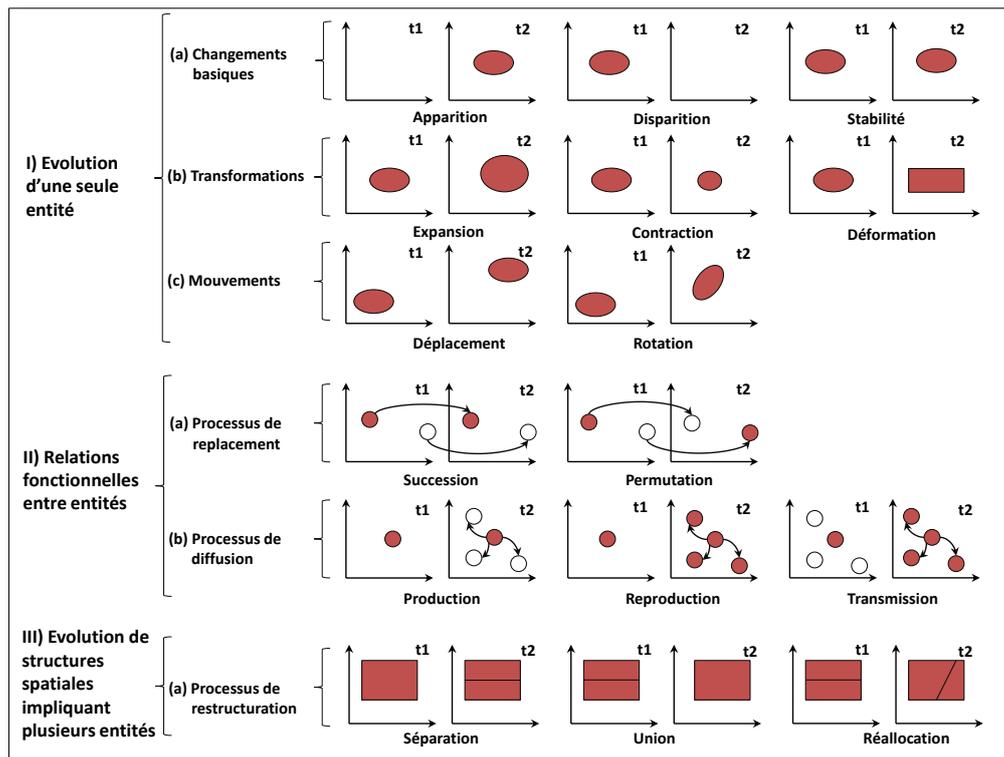


Figure 8-13. Les types de processus spatio-temporels (source: Claramunt et Thériault 1995)

8.3.3 LES EVOLUTIONS POSSIBLES DES FORETS

Nous cherchons maintenant à identifier toutes les évolutions possibles des forêts dont on devra tenir compte dans la mise en œuvre de la modélisation.

La typologie proposée par le Joint European Research Center (JRC) distingue quatre types d'évolutions pouvant avoir potentiellement des effets néfastes sur la biodiversité forestière : l'attrition et le rétrécissement qui modifient la taille de l'habitat (totalement disparu ou taille réduite), la perforation qui provoque une baisse de surface et une modification de la longueur de lisière et la fragmentation qui a tendance à dégrader la lisière forestière et modifier l'isolement (Figure 8-14). Ce sont ensuite ces perturbations qui impactent plus ou moins fortement le microclimat des habitats forestiers, les différentes niches des espèces et par conséquent, leur population. Il semble néanmoins difficile d'obtenir une vision claire de toutes les évolutions possibles à l'aide de cette typologie uniquement. Seuls les évolutions « négatives » (au sens de la perte de surface) sont considérées alors que les évolutions « positives » (cad qui entraînent un gain de surface ou de connectivité) font également varier la biodiversité forestière et sont susceptibles d'apparaître en détectant l'évolution de nos forêts. Nous décidons par conséquent de l'enrichir à l'aide de la typologie de (Claramunt et Thériault 1995).

La Figure 8-15 répertorie au total neuf processus spatio-temporels pouvant potentiellement intervenir dans l'évolution de chaque forêt suivie. Six d'entre eux impliquent une seule entité tels que : la disparition, l'apparition, la stabilité, la contraction, l'expansion et la perforation. La stabilité, la disparition et l'apparition provoquent une modification de la généalogie (généalogie = existence) de la forêt suivie (continuité ou interruption ou démarrage d'une nouvelle filiation) et ont des effets directs sur l'ancienneté de l'habitat (continuité temporelle). La contraction, l'expansion et la perforation entraînent une modification de la structure spatiale de la forêt qui peut ensuite impacter la morphologie du bois mais également sa connectivité. Les processus de restructuration comme la division, l'union (ou fusion) et la réallocation impliquent eux plusieurs entités. Ils bouleversent à la

fois la généalogie de la forêt suivie mais aussi sa structure spatiale. Par exemple, une union entre deux forêts conduira à une expansion de sa structure tandis qu'une division mènera plutôt à une contraction. Ces perturbations provoquent ainsi des effets sur la morphologie de l'habitat forestier et son isolement. A noter enfin que la totalité des processus listés provoquent également des effets indirects sur les habitats voisins notamment en termes d'isolement. Dès lors qu'un objet apparaît, disparaît, persiste, se contracte ou s'étend, les relations qu'entretiennent les habitats voisins avec l'habitat d'intérêt varie puisque la quantité de surfaces boisées dans un rayon donné change et la longueur de lisière évolue.

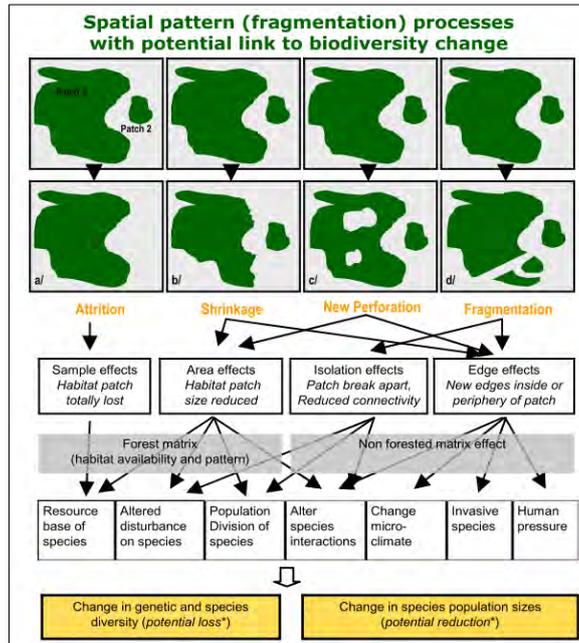


Figure 8-14. Typologie des processus spatio-temporels forestiers ayant potentiellement un impact sur la biodiversité forestière (source : Joint Research Center)

PROCESSUS SPATIO-TEMPOREL	REPRÉSENTATION	PROPRIÉTÉS IMPACTÉES	EFFETS SUR L'HABITAT D'INTERET	EFFETS SUR LES HABITATS VOISINS
Disparition	→	•Généalogie	• Δ ancienneté	•Δ isolement
Apparition	→	•Généalogie	• Δ ancienneté	•Δ isolement
Stabilité	→	•Généalogie	• Δ ancienneté	• Δ isolement
Contraction	→	•Spatiales	•Δ morphologie • Δ isolement	•Δ isolement
Expansion	→	•Spatiales	•Δ morphologie • Δ isolement	•Δ isolement
Perforation	→	•Spatiales	•Δ morphologie • Δ isolement	•Δ isolement
Division	→	•Généalogie •Spatiales	•Δ morphologie • Δ isolement	•Δ isolement
Union	→	•Généalogie • Spatiales	•Δ morphologie • Δ isolement	•Δ isolement
Réallocation	→	•Généalogie • Spatiales	•Δ morphologie • Δ isolement	•Δ isolement

Figure 8-15. Liste des processus spatio-temporels susceptibles d'apparaître dans l'analyse d'évolution des surfaces boisées. Les Delta (Δ) indiquent une modification des caractéristiques d'une forêt. Pour rappel, la généalogie fait référence à l'existence de l'objet.

8.3.4 CHOIX D'UN MODELE DESCRIPTIF POUR CARACTERISER LES EVOLUTIONS DES FORETS

Il s'agit maintenant d'adopter un modèle descriptif qui permette de tenir compte de l'ensemble des processus spatio-temporels listés et qui soit adaptés aux données dont nous disposons.

Nous avons décidé d'utiliser un modèle **snapshot** pour plusieurs raisons. Premièrement, la primitive temporelle utilisée dans ce modèle est l'instant et la structure de temps discrète et linéaire, ce qui est en adéquation avec les données dont nous disposons. Nous possédons cinq bases de données spatiales représentant l'état du paysage forestier à cinq instants particuliers d'une ligne de temps allant de 1850 à 2010 (Figure 8-16). Deuxièmement, nous devons être en mesure d'identifier des processus affectant à la fois la structure spatiale et la généalogie de l'objet. Le fait que le temps soit implicitement représenté dans ce modèle et qu'il faille comparer les états pour identifier les changements représente pour nous une certaine flexibilité. Nous pouvons gérer les évolutions basiques des objets comme une séquence d'attributs tout en s'appuyant sur leurs identifiants pour pouvoir suivre leur filiation. Enfin, une des limites du modèle snapshot est qu'il ne permet pas de tenir compte de l'intervalle temporel entre deux dates. Cela n'a pas d'incidence sur notre analyse puisque nous ne souhaitons pas décrire l'intensité du processus ni sa vitesse mais uniquement décrire les changements survenus entre deux dates.

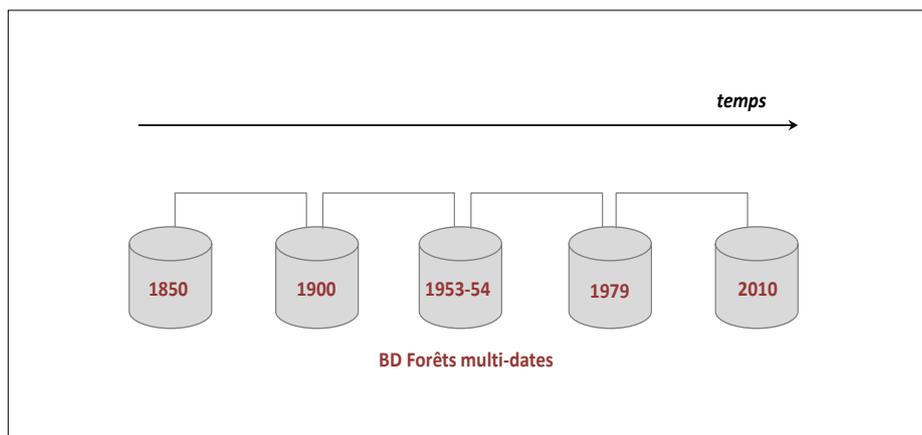


Figure 8-16. Description du modèle « snapshot » pour la mise en évidence des évolutions

8.3.5 MESURER LES EVOLUTIONS A TRAVERS LES CHANGEMENTS D'ETATS : UNE APPROCHE « INDIVIDU »

Comme mentionné précédemment, le modèle **snapshot** intègre le temps de manière implicite. C'est à l'utilisateur de comparer les états pour mettre en évidence les changements survenus.

Dans cette logique, cette comparaison d'états constituera notre première approche pour mesurer les évolutions. Nous l'avons baptisé « **approche individu** » car chaque forêt suivie sera comparée individuellement avec ses homologues dans la base de données qui la précède ou qui la succède. **Ces taux de changements permettront par la suite de mesurer les effets de l'histoire sur la biodiversité (Partie 4) en les intégrant comme des variables explicatives dans les modèles paysage-biodiversité.**

Concernant les comparaisons à proprement parler, nous nous appuyerons sur les bases de données spatiales forestières caractérisées et appariées. Les changements seront ainsi exprimés pour chacune des métriques paysagères sélectionnées par l'intermédiaire de taux d'évolution (Figure 8-

17). Cela conduira à la création d'une base de données « changements » où chaque période entre deux dates successives sera caractérisée par les taux d'évolution de chaque métrique sélectionnée : Δ surface (%), Δ périmètre (%), Δ compacité (%), Δ connectivité potentielle (%).

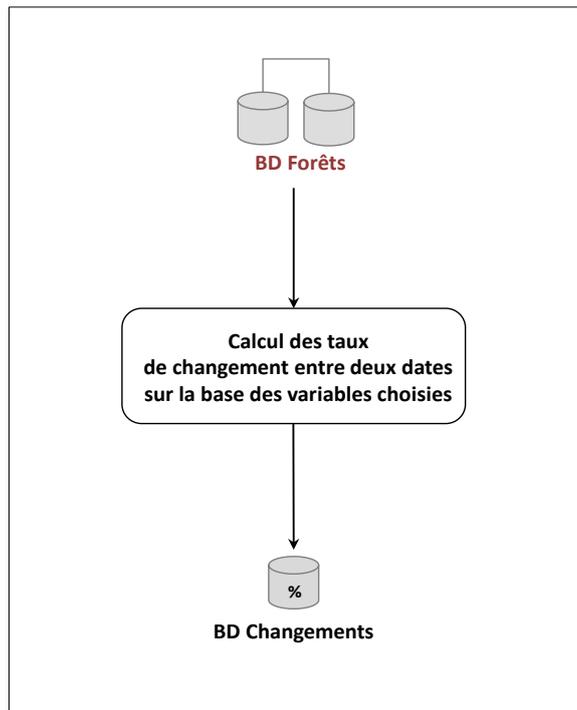


Figure 8-17. Description de l'approche « individu » pour quantifier les évolutions de la structure spatiale des bois à travers les changements d'états. Ce modèle s'applique à chaque couple de dates.

8.4. CLASSER LES EVOLUTIONS SPATIO-TEMPORELLES : UNE APPROCHE « POPULATION »

La classification des évolutions spatio-temporelle constitue la dernière étape de notre démarche. Elle a pour objectif d'identifier et de construire des types d'évolution des forêts suivies. On va ainsi chercher à caractériser la diversité des trajectoires à partir d'une procédure de classification.

8.4.1 CONSTRUIRE, DECRIRE ET CATEGORISER DES TYPES D'EVOLUTIONS

L'enjeu consiste ici à identifier à posteriori des types de changements. Pour ce faire, nous nous appuyons sur une méthode d'analyse multidimensionnelle qui est appropriée pour rendre compte de la diversité des évolutions d'un ensemble d'entités, à savoir la classification et l'analyse en composantes principales (ACP). Les individus statistiques sont clairement identifiés (les forêts actuelles échantillonnées et leurs homologues), décrits par les mêmes métriques à plusieurs dates. Il est par conséquent possible de les représenter sous la forme d'un tableau (cas multi varié). L'ACP a pour objectif de synthétiser les ressemblances respectives entre les lignes et les colonnes de ce tableau.

8.4.2 DEMARCHE ADOPTEE POUR IDENTIFIER DES TYPES D'EVOLUTIONS : UNE APPROCHE « POPULATION »

Par opposition à l'approche « individu » développée plus haut, la méthode employée ici se veut plus globale en essayant de reconstituer l'ensemble de la trajectoire d'une forêt (sur toute la ligne de temps) et en tenant compte de l'ensemble des individus de la population pour définir des types d'évolution. Ceci doit permettre d'analyser les évolutions des forêts à l'échelle de la population et d'identifier des tendances qui permettent de caractériser l'ensemble de notre jeu de données d'où le terme d'approche « population ».

La démarche appliquée se déroule en trois étapes majeures.

Nous avons tout d'abord conduit une ACP sur les individus de la BD Forêts 2010 (MANUEL). Le plan factoriel obtenu à l'issue de cette analyse a été considéré comme notre plan factoriel de référence et a servi de support à l'analyse des évolutions.

Deuxièmement, les individus des autres dates ont été projetés dans le plan factoriel de référence dans l'objectif de suivre l'évolution de chaque forêt depuis 2010. La trajectoire d'un objet est donc caractérisée par une suite de points correspondant respectivement aux positions prises par chaque forêt à chaque date dans le plan factoriel.

Enfin, une fois les trajectoires établies, nous procéderons à un regroupement des trajectoires pour identifier des tendances d'évolution. Cette dernière opération mènera à la création de la base de données « Classes d'évolution ».

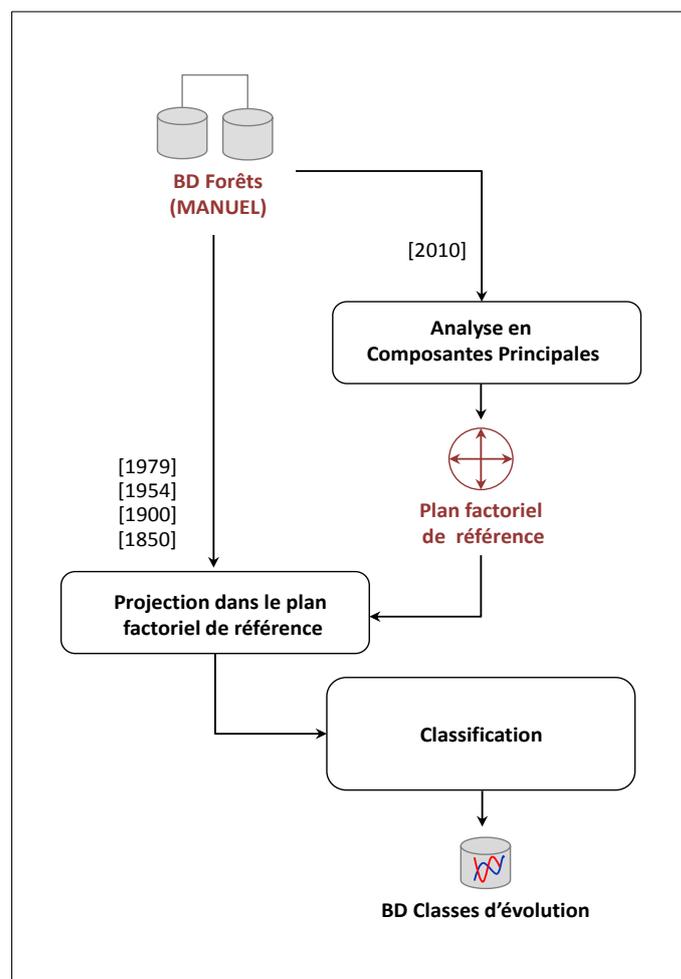


Figure 8-18. Description de l'approche « population » pour mettre en évidence des tendances d'évolution à travers une typologie de trajectoires

8.4.3 REALISATION D'UNE ACP SUR L'ANNEE DE REFERENCE

L'Analyse en Composantes Principales (ACP) est un grand classique de l'analyse des données pour l'étude exploratoire ou la compression d'un grand tableau $n \times p$ de données quantitatives. Son objectif est de revenir à un espace de dimension réduite en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales (Jolliffe 2002). L'idée est d'étudier la dispersion des données en s'appuyant sur la matrice de variances-covariances. De cette matrice, on va extraire, par un procédé mathématique adéquat, les facteurs explicatifs, en petit nombre. Ces facteurs permettent par la suite d'effectuer les analyses désirées dans un espace à n dimensions où n correspond au nombre de facteurs retenus, en déformant le moins possible la configuration globale des individus selon l'ensemble des variables initiales (remplacées par des facteurs).

Pour appliquer une analyse multi-variée classique, il est nécessaire de disposer d'un tableau statistique à deux dimensions. Il faut alors choisir de projeter le temps sur les variables ou les individus. Ce choix a des implications. Dans le premier cas, le profil d'un individu, ici la forêt, sera sa structure spatiale comptée à chacune des cinq dates selon les quatre métriques paysagères choisies (5*4 colonnes). Dans le cas où le choix est de projeter le temps sur les individus, chaque forêt apparaîtra autant de fois qu'il y a de dates, soit cinq fois dans notre cas. Si l'on cherche à « suivre » les variables paysagères dans le temps dans la manière dont elles différencient les forêts alors la projection du temps sur les variables est la plus intéressante. Notre questionnement porte en revanche sur la différenciation des trajectoires des forêts, c'est donc la projection du temps sur les forêts (les individus) qui est la plus adaptée. Une fois cette logique adoptée, deux solutions apparaissent. Soit on décide d'effectuer une ACP sur un tableau à 5*n lignes (5 étant le nombre de dates et n le nombre d'individus) et 4 colonnes (le nombre de variables). Le plan référentiel de l'analyse sera ainsi un plan moyen de la structure spatiale des forêts toutes dates confondues. Soit on effectue l'analyse sur une date référence pour analyser les évolutions par rapport à cette même date. La première solution ne nous satisfait pas car l'établissement d'un plan moyen aura tendance à lisser les dynamiques forestières. Nous choisissons donc de considérer l'année de 2010 comme l'année de référence (année correspondant aux relevés biologiques) afin de mettre en évidence le changement de la structure spatiale des bois relativement à cette date en particulier. Celle-ci s'appuie sur l'ensemble des forêts présentes (48) caractérisées par quatre variables paysagères (4) soit un tableau de contingence de 48 lignes*4 colonnes.

8.4.4 PROJECTION DES AUTRES DATES DANS LE PLAN FACTORIEL DE REFERENCE

Après avoir choisi 2010 comme année de référence pour la réalisation de l'ACP, il est nécessaire maintenant de projeter les individus des autres dates dans le plan factoriel de référence. Ces individus « supplémentaires » n'interviennent pas dans la structure de différenciation mais seront uniquement reprojétés dans ce plan factoriel. Cette opération a été effectuée date par date (de la plus récente vers la plus ancienne) et a permis de calculer les coordonnées de chaque homologue dans ce nouveau plan. L'analyse de la trajectoire consiste à évaluer le déplacement de la position des homologues vis-à-vis de la position prise par la forêt en 2010.

8.4.5 LA CLASSIFICATION DES TRAJECTOIRES

La classification des trajectoires constitue l'ultime étape de notre démarche. Elle doit permettre d'identifier des tendances communes parmi les trajectoires établies pour les regrouper. Ce regroupement conduira à la création d'une nouvelle base de données d'évolution (BD « classe d'évolution ») où les forêts seront classées selon leur dynamique historique.

Notons tout d'abord que la procédure de classification détaillée ci-dessous a été appliquée uniquement aux forêts présentes dans l'ensemble des sources spatiales ($N = 28$) c'est-à-dire les

forêts dont il est possible de suivre l'évolution depuis 1850 jusqu' à 2010 sans discontinuité. Les forêts apparues à partir de 1900 ou ensuite (N=20) constituent un groupe particulier que nous détaillerons en plusieurs classes selon la date d'apparition de chaque bois (cf. section 10.2.1).

Plusieurs méthodes auraient pu être employées pour classer automatiquement nos trajectoires (Classification ascendante hiérarchique, K-moyennes,...). Néanmoins, nous verrons au chapitre 9 que les trajectoires obtenues sont très hétérogènes. Nous avons par conséquent procédé à un regroupement manuel basé sur un examen visuel des différentes trajectoires.

CHAPITRE 9.MISE EN ŒUVRE DE LA DEMARCHE METHODOLOGIQUE PROPOSEE

SOMMAIRE

- 9.1. REPRESENTATION ET INTERPRETATION DES LIENS D'APPARIEMENT
- 9.2. CALCUL ET ANALYSE DES METRIQUES PAYSAGERES PAR DATE
- 9.3. ANALYSE DES CHANGEMENTS D'ETATS
- 9.4. TYPOLOGIE DES EVOLUTIONS SPATIO-TEMPORELLES

Ce chapitre aborde la mise en application de la démarche méthodologique proposée pour appairer les surfaces boisées et caractériser les évolutions spatio-temporelles. Nous analysons et discutons des résultats obtenus pour les forêts suivies dans le cadre de notre protocole écologique. La présentation des résultats est effectuée dans l'ordre de la démarche proposée à savoir de la phase d'appariement vers la typologie des évolutions spatio-temporelles.

9.1. REPRÉSENTATION ET INTERPRÉTATION DES LIENS D'APPARIEMENT

La phase d'appariement des forêts a permis de mettre en évidence la généalogie de chaque forêt ayant fait l'objet de relevés biologiques dans le cadre du protocole « Syrphes ». Une identification visuelle a été conduite pour construire les liens de filiation entre homologues et ces liens ont été répertoriés dans une table d'appariement. A partir de cette table, il devient possible d'étudier les processus d'évolution affectant la généalogie des objets suivis sur l'ensemble de la ligne de temps.

Pour disposer d'une vision globale de la généalogie de chaque forêt, nous décidons d'adopter un mode de représentation qui permette d'analyser plus « facilement » ces liens. Chacun des processus affectant la généalogie identifiés en chapitre 8 est associé à une primitive géométrique (Figure 9-1). L'assemblage de ces primitives permet de construire des arbres d'évolutions illustrant l'histoire identitaire de chaque bois. La figure 9-1 présente six primitives correspondant à six processus spatio-temporels susceptibles d'intervenir entre deux dates. La lecture de celles-ci s'effectue du bas (état au temps t) vers le haut (état au temps $t+1$). Un nœud correspond à la présence d'un fragment forestier tandis qu'un segment correspond à une relation de filiation directe entre deux ou plusieurs fragments ou bien entre un fragment et un « vide » qui signifie que le fragment n'est pas présent à la date considérée (disparition ou apparition).

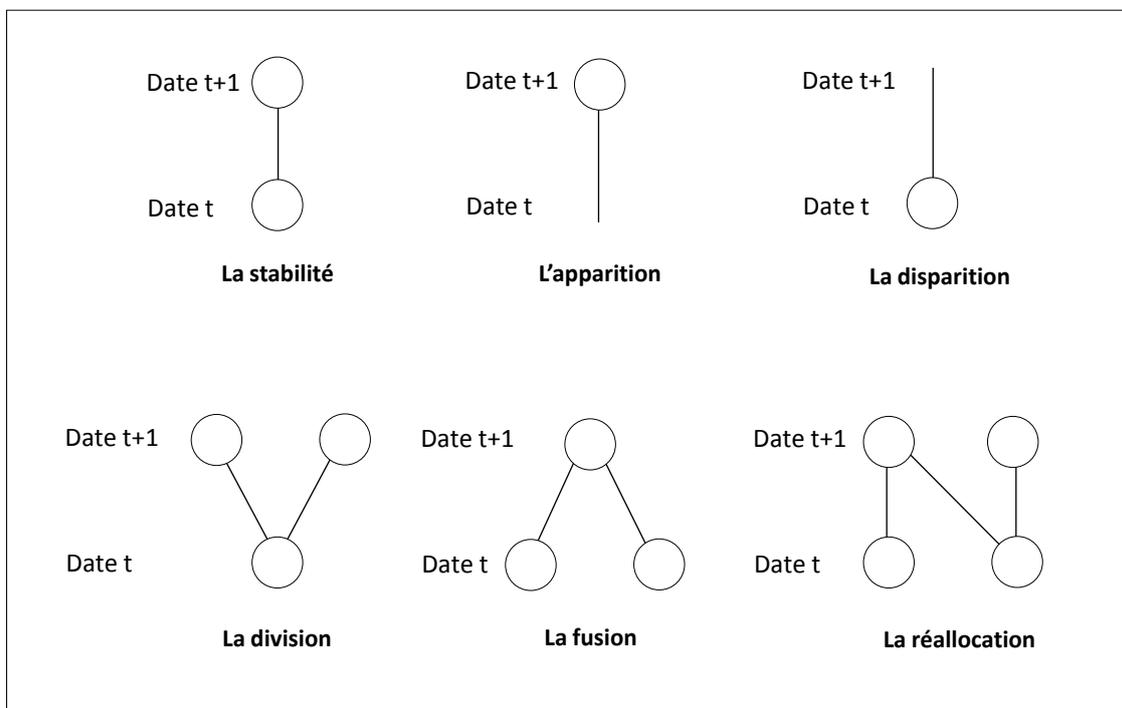


Figure 9-1. Représentation des processus spatio-temporels (entre deux dates) pouvant affecter la généalogie des forêts par des primitives géométriques.

9.1.1 DESCRIPTION DES ARBRES D'ÉVOLUTION

Les arbres d'évolutions mis en évidence pour les 48 forêts suivies permettent tout d'abord d'identifier deux grands groupes caractérisés par une durée d'existence différente (Figure 9-2. La première catégorie regroupe les forêts apparues après 1850 et englobe les arbres dont le premier nœud est observable à partir de 1900 ou sur une date postérieure (N=20, ex : ID9, ID37, ID45, etc..). Ces forêts « récentes » dispose alors d'une date d'apparition différente : trois sont apparues en 1900, 13 en 1954 et seulement deux en 1979. A contrario, le second groupe est constitué des forêts dont il a été possible de suivre l'évolution sur toute la période de temps puisqu'elles ont été retrouvées sur chacune des sources spatiales utilisées (N =28, ex : ID2, ID 14, ID40, etc). Pour ces dernières, les arbres sont caractérisés par une durée de présence maximale ainsi que par la présence d'au moins un nœud à chaque niveau de la ligne de temps (à chaque date). Elles sont perçues comme les forêts « anciennes » (au sens de forêts ayant une continuité dans le temps). Les autres seront nommées "récentes" par la suite.

Si l'on s'attache maintenant à décrire la structure horizontale des arbres, deux niveaux de complexité apparaissent. Les arbres disposant d'un seul nœud à chaque niveau de leur évolution désignent les forêts n'ayant pas subis de fragmentation, de fusion ou de réallocation durant la période d'étude (N = 35). Les autres arbres se distinguent par un nombre de nœud supérieur à 1 sur au moins un des niveaux de leur structure d'évolution (N = 13). La présence de plusieurs nœuds aux différents niveaux indiquent que le fragment étudié a subi un évènement ayant conduit à le diviser ou à le faire fusionner avec un autre fragment : la somme des nœuds renseignant de fait sur l'intensité de cet épisode. La plupart de ces arbres sont des forêts anciennes (10 sur 13 cas détectés, ex : ID2, ID27, ID40, etc..) indiquant que les forêts apparues plus récemment ont été majoritairement moins impactés que les anciennes par ce type de processus.

Parmi ces forêts dont l'évolution a été jugée plus complexe (N= 13), huit d'entre elles ont subi une division avant 1954 (ex : ID2, ID8, ID27) et aucune n'en a connu ensuite : quatre forêts ont connu une division entre 1850 et 1900 (ID2, 21, 27 et 33) et quatre entre 1900 et 1954 (ID8, 11, 18 et 33). La période 1954-1979 est marquée majoritairement par des processus de fusion (8 cas sur 13 détectés) et rompt avec les processus observés pour ces forêts avant 1954. Cette période est ainsi caractérisée par une simplification de la structure globale des forêts. Enfin, trois forêts appartenant à ce groupe ont été alimentés historiquement par l'apparition d'une ou plusieurs forêts durant la période d'étude (ID27, 30 et 40). Cela indique que ces forêts sont composées de fragments aux âges différents.

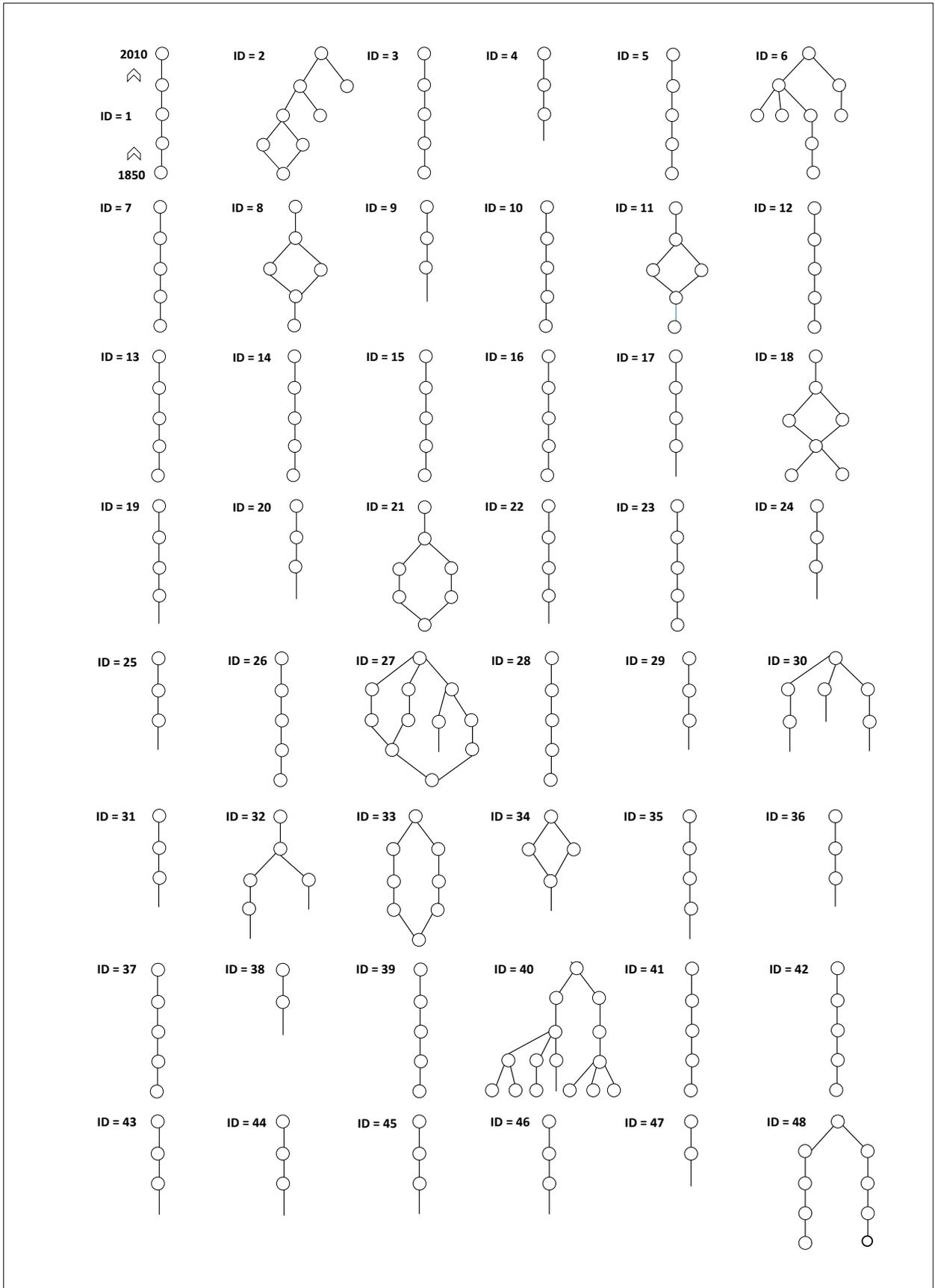


Figure 9-2. Représentation de la cardinalité des liens d'appariement pour les 48 forêts suivies

9.1.2 DISCUSSION

Les forêts suivies dans ce protocole affichent des évolutions très variées.

Une partie d'entre elles a connu une évolution relativement simple sans connaître de fragmentation, de fusion ou de réallocation et ont été observées sur chacune des sources spatiales utilisées. C'est le cas pour la plupart des forêts récentes (18 sur 20) et d'une partie des forêts anciennes (18 sur 28). Ce type d'évolution pourrait s'avérer par conséquent plutôt favorable au maintien et au développement de la biodiversité en considérant que ces bois n'aient pas connu d'événements majeurs de perturbation. Ces propos sont toutefois à nuancer puisque nous ne possédons pas, à ce stade, d'informations quantitatives sur leur évolution de leur taille, leur périmètre, leur isolement, etc.

D'autres ont également été observées depuis 1850 mais ont subi une évolution plus complexe entraînant la fusion, la réallocation ou la division d'un de leurs fragments. Celles qui ont connu une division l'ont subie pour la plupart avant 1954 et pourraient être des bois concernés par un potentiel effet de l'histoire sur leur biodiversité. En effet, certaines espèces pourraient avoir réagi plus tardivement à ces perturbations et se montrer moins nombreuses dans ces patchs aujourd'hui. Toutefois, la description des arbres d'évolution a montré que les forêts aux évolutions complexes ont eu tendance à fusionner après 1954 et à se simplifier d'un point de vue de leur structure après cette date. Cette simplification provoquée par la réunion de plusieurs patchs pourrait aussi avoir des conséquences non négligeables sur la biodiversité. L'assemblage de plusieurs fragments aux milieux internes divers pourrait conduire à une forêt potentiellement plus riche en termes de ressources. Enfin, certains fragments sont apparus à différents niveaux de la généalogie des forêts et ont fusionné avec des fragments présents depuis une plus longue période de temps. Ce regroupement de bois aux âges dissemblables peut aussi jouer un rôle important. Il est possible par exemple qu'un bois aux critères de naturalité très développés ait été mis en relation avec un fragment abritant des stades de développement moins avancés, ceci ayant pour possible conséquence de réunir des communautés d'espèces inféodées à différents types de milieux.

9.2. CALCUL ET ANALYSE DES METRIQUES PAYSAGERES PAR DATE

Les métriques paysagères ont été calculées pour toutes les forêts appariées dans les différentes BD Forêts. Nous présentons ici les résultats pour les BD Forêts MANUEL uniquement et procéderons à une description des résultats métriques par métriques (Tableau 9-1, Figure 9-3).

En 1954, 1979 et 2010, les forêts présentes disposaient d'une surface moyenne située aux alentours des 10 ha avec des valeurs d'écart-types elles aussi très proches. Cette surface moyenne est toutefois plus élevée en 1900 et 1850 avec respectivement des valeurs de 16,08 et 22,73 ha. La surface minimale des bois est également plus élevée dans ces deux dernières bases de données (0,39ha en 1900 contre 0,08 ha en 1954 par exemple) que dans les trois autres. Ce différentiel de surface entre les deux bases les plus anciennes et les plus récentes s'expliquent par l'apparition de forêts au cours du temps. Treize d'entre elles sont survenues entre 1900 et 1954, ce qui veut dire que leur taille est relativement limitée tout en provoquant la baisse de la surface moyenne calculée.

BD FORETS		2010(N = 48)	1979(N = 48)	1954(N = 46)	1900(N = 33)	1850(N = 28)
METRIQUES						
SURFACE (en ha)	MIN	0.13	0.23	0.08	0.39	0.50
	MAX	186.62	182.44	196.92	206.96	180.59
	MOY	10.81	10.15	10.20	16.08	22.73
	SD	28.31	27.37	30.01	36.72	40.03
PERIMETRE (en m)	MIN	138.66	234.1	122.61	264.41	297.65
	MAX	13692.28	17907.85	16254.12	11051.78	15956.08
	MOY	1665.59	1739.84	1764.67	1871.48	2837.70
	SD	2296.44	2806.72	2640.72	2097.26	3824.82
COMPACITE	MIN	0.01	0.01	0.01	0.01	0.01
	MAX	0.11	0.11	0.15	0.07	0.12
	MOY	0.04	0.04	0.05	0.03	0.03
	SD	0.02	0.03	0.03	0.02	0.02
CONNECTIVITE (en %)	MIN	7.73	6.56	6.48	2.77	3.24
	MAX	31.44	25.73	25.12	41.93	39.9
	MOY	18.56	14.70	13.14	12.85	11.16
	SD	7.02	5.04	4.65	7.98	7.09

Tableau 9-1. Résultats des calculs des métriques pour toutes les forêts identifiées sur chaque source. Il s'agit ici des résultats moyennés à l'échelle du groupe de forêt mais chacune de ces métriques a été calculée pour chaque individu à chaque date.

Les forêts présentes en 2010 et appariées en 1979, 1954 et 1900 sont caractérisées par un périmètre moyen oscillant entre 1665 et 1872 ha tandis que celui des forêts de 1850 est plus élevé avec une valeur de 2837.70 ha. L'écart-type enregistré pour cette dernière base de données est le plus élevé de tous (3824.82 m) indiquant de fait une grande hétérogénéité de structure parmi les forêts présentes à cette date. On remarque également que les forêts des BD 1979 et 1954 affichent des valeurs d'écart-types plus importantes qu'en 2010 ou 1900. Il semble en effet qu'un certain nombre de forêts disposaient d'une longueur de lisière plus importantes en 1979 et 1954 puisque les périmètres maximaux enregistrés pour ces deux dates sont respectivement de 17907.85 et 16254.12m soit environ 3000 m de plus qu'en 2010 et 5000 m de plus qu'en 1900.

La compacité moyenne la plus élevée a été recensée en 1954 avec une valeur de 0.05 ± 0.03 . C'est également à cette date que la compacité maximale est la plus haute avec un indice moyen de 0.15 soit 0.04 points de plus qu'en 2010 et 1979 et respectivement 0.08 et 0.03 points de mieux qu'en 1900 et 1850. Là aussi, l'apparition d'un grand nombre de forêts entre 1900 et 1954 a laissé place à de petits objets aux contours géométriques complexes qui contribuent à faire augmenter la compacité moyenne en 1954. Globalement, on observe assez peu de disparités parmi les autres bases de données. La valeur minimale enregistrée est la même à chaque date (compacité min = 0.01) et estis que les valeurs obtenues pour 2010 sont très proches de celles calculées à partir des forêts de 1850.

Les forêts de 2010 étaient globalement les mieux connectées avec une connectivité moyenne de $18.56 \pm 7.02\%$ soit 4 points de plus qu'en 1979 ou 7 points de plus qu'en 1850. Globalement, cette connectivité moyenne n'a cessé de croître au cours du temps. Toutefois, les valeurs de connectivité maximales sont plus importantes en 1850 et 1900 que dans les trois bases les plus récentes avec notamment un écart de 10 points avec celle enregistrée pour 2010. On constate en revanche que la connectivité minimale est plus élevée en 2010, 1979 et 1954 que dans les bases les plus anciennes. Ceci indique que la connectivité potentielle s'est légèrement dégradée pour les forêts historiquement mieux connectées alors qu'elle a eu tendance à s'améliorer pour les bois faiblement connectées en 1850 et en 1900.

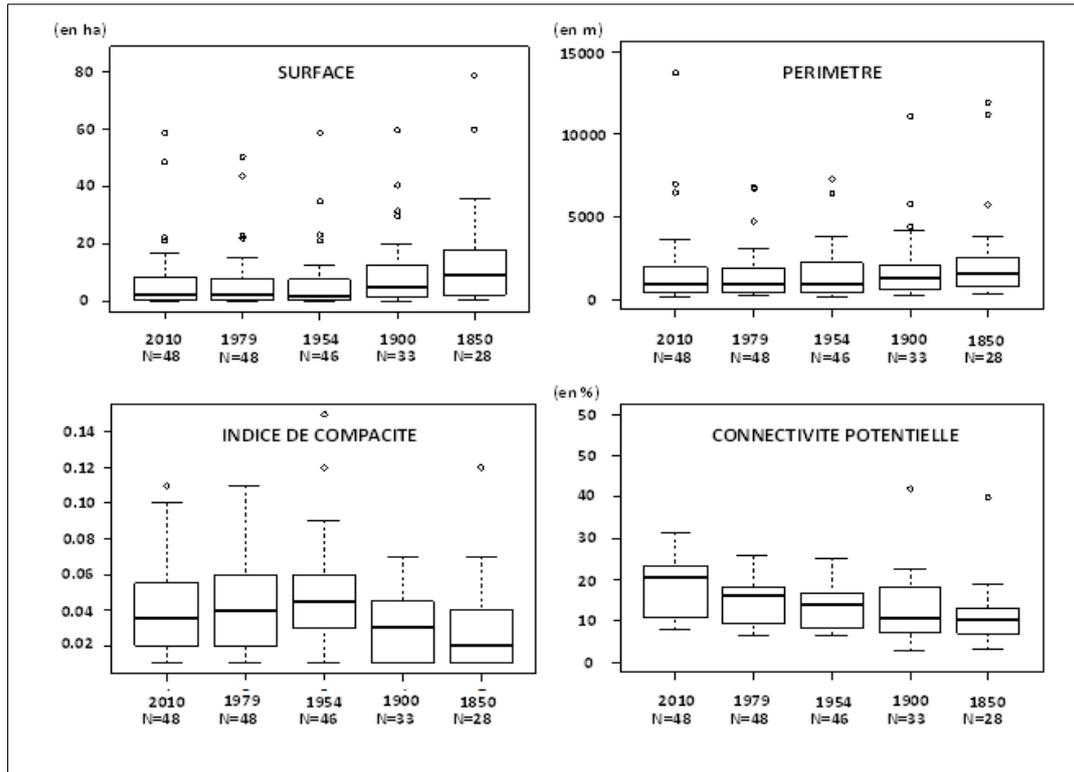


Figure 9-3. Distribution des valeurs de métriques pour toutes les forêts identifiées sur chaque source

9.3. ANALYSE DES CHANGEMENTS D'ETATS

Le tableau 9-2 répertorie les taux de changements par métriques calculés entre deux dates. Le nombre d'individus croît à chaque transition en raison de l'apparition de nouvelles forêts au cours du temps. Nous commenterons et discuterons par conséquent ces résultats période par période.

La période 1850-1900

Au total, 28 forêts sont concernées par cette période de temps.

Le taux de changement moyen de la surface indique une baisse de -5.56% accompagnée d'un fort écart-type ($\sigma = 57.33$). La plupart des forêts ont en effet connu une contraction (N= 20) tandis que 8 d'entre elles ont été concernées par une expansion. Le périmètre a lui aussi baissé en moyenne de 7.04%. Néanmoins, seulement 16 forêts ont connu une baisse alors que 12 d'entre elles ont évolué positivement du point de vue de cette métrique. On note que la compacité a en moyenne augmenté de 38.48 ± 112.13 %. Ceci s'avère être la tendance pour 17 forêts alors les onze autres bois ont, eux, connu une baisse de compacité. Enfin, la connectivité s'est en moyenne améliorée de 22.48 ± 55.13 %. Seulement 18 forêts ont en réalité profité de cette amélioration.

La période 1900-1954

Puisque cinq forêts sont apparues entre 1850 et 1900, les taux de changements entre 1900 et 1954 concernent cette fois 33 patches.

On remarque que la surface affiche un taux d'évolution moyen positif de 0.27% marquée par un écart-type très important ($\sigma = 76.89\%$). Plus de la moitié des bois ont en effet perdu de la surface tandis que seulement 10 d'entre eux ont profité d'une hausse. Le périmètre a évolué en moyenne positivement ($23.95\% \pm 60.44\%$). Contrairement aux observations effectuées à propos de la surface,

cette hausse concerne la majorité des bois (N = 21). Les changements les plus importants portent sur la compacité. Nous avons enregistré une hausse moyenne de $175.60 \pm 388.95\%$ soit le taux d'évolution le plus important recensé parmi toutes les comparaisons opérées. Cette hausse ne concerne toutefois que deux tiers des patchs puisque neuf d'entre eux ont connu une réduction de leur compacité. Enfin, si l'évolution moyenne de la connectivité affiche une valeur positive de $15.02 \pm 59.46\%$, cela ne semble pas être la tendance générale. Vingt-quatre forêts sur trente-trois ont été concernées par une baisse de connectivité.

La période 1954-1979

Treize nouvelles forêts ont vu le jour entre 1900 et 1954, nous décrivons par conséquent l'évolution de la structure de 46 bois entre 1954 et 1979.

La surface a évolué en moyenne positivement ($31.32\% \pm 80.99\%$) durant cette période et cela concerne la majorité des forêts (N = 33). Le périmètre a lui aussi augmenté en moyenne avec un taux de changement de $23.95\% \pm 60.44\%$. Contrairement à la surface, cette hausse ne concerne que la moitié des bois puisque 23 d'entre eux ont connu une baisse de la longueur de leur contour. La tendance d'évolution moyenne de la compacité est à la baisse ($-13.07 \pm 24.00\%$) et se vérifie pour environ trois quarts des bois (N = 32). Finalement, la connectivité moyenne des forêts marque une progression avec un taux de changement positif de $10.69\% \pm 7.35\%$. La quasi-totalité des bois sont concernées par cette amélioration (N = 44).

La période 1979-2010

Toutes les forêts suivies étant présentes en 1979, l'analyse des changements d'états concerne cette fois la totalité des patchs échantillonnés.

Cette dernière période de temps est marquée par une évolution moyenne positive de la surface avec un taux de changement moyen de $9.08 \pm 40.08\%$. Les évolutions au sein même de ce groupe sont toutefois à nuancer car seules 25 forêts sur 48 ont connu une augmentation de leur surface. Les changements enregistrés concernant le périmètre suivent une logique similaire. Le taux de changement moyen est de $2.64 \pm 24.10\%$ mais cette évolution positive n'est vraie que pour 22 bois sur 48. La compacité moyenne est en baisse (-3.38 ± 24.10). Cela est la tendance générale puisque seulement 17 forêts ont connu une amélioration du point de vue de cet aspect de leur structure. Enfin, cette ultime transition est caractérisée par une évolution positive de la connectivité égale à $24.69 \pm 11.98\%$ soit l'évolution la plus marquante enregistrée concernant cette métrique. A noter que tous les bois suivis durant cette période ont vu leur connectivité augmenter.

PERIODES	METRIQUES	SURFACE	PERIMETRE	COMPACITE	CONNECTIVITE
1850-1900 (N = 28)	TAUX DE CHANGEMENT MOYEN (EN %)	-5.56	-7.04	38.48	22.48
	DEVIATION STANDARD	57.33	30.91	112.13	55.13
	N(+)	8	12	17	18
	N(-)	20	16	11	10
1900-1954 (N = 33)	TAUX DE CHANGEMENT MOYEN (EN %)	0.27	23.95	175.60	15.02
	DEVIATION STANDARD	76.89	60.44	388.95	59.46
	N(+)	10	21	24	9
	N(-)	23	12	9	24
1954-1979 (N = 46)	TAUX DE CHANGEMENT MOYEN (EN %)	31.32	7.24	-13.07	10.69
	DEVIATION STANDARD	80.99	38.20	24.00	7.35
	N(+)	33	23	12	44
	N(-)	13	23	32	2
1979-2010 (N = 48)	TAUX DE CHANGEMENT MOYEN (EN %)	9.08	2.64	-3.38	24.69
	DEVIATION STANDARD	40.86	29.14	24.10	11.98
	N(+)	25	22	17	48
	N(-)	23	26	31	0

Tableau 9-2. Résultats des calculs des taux de changement entre deux dates. N(+) indiquent le nombre d'individus ayant connu une évolution positive tandis que les N(-) indiquent le nombre d'individus ayant connu une évolution négative.

Discussion

Cette analyse a permis de démontrer une grande hétérogénéité dans les évolutions des forêts suivies.

Nous avons premièrement observé de fortes valeurs d'écart types pour la quasi-totalité des taux de changement moyens calculés. Cette grande variabilité indique que les forêts ont connu des changements d'intensité différente et que les événements qui ont conduit à les faire évoluer étaient d'une ampleur variable.

Ensuite, la tendance des changements moyens n'est pas en accord avec la tendance d'évolution de la totalité des bois. Certaines évoluent négativement, d'autres positivement, il est par conséquent difficile de tirer des conclusions génériques sur l'évolution du groupe de forêts au cours de la période considérée. Là aussi, les événements ayant agi pour modifier la structure des forêts sont disparates et ont probablement eu des effets contraires sur la variation de la biodiversité au sein de ces bois.

Compte tenu de ces observations, il semble plus approprié de s'appuyer sur la tendance majoritaire pour discuter des changements survenus à l'échelle du groupe excepté les cas où il paraît difficile de dégager une tendance claire.

La première période de temps (1850-1900) est caractérisée majoritairement par une perte de surface et une baisse de périmètre. Bien que le minimum forestier soit situé aux alentours de 1850, certaines forêts ont continué à se dégrader du point de vue de leur structure jusqu'en 1900. La compacité a majoritairement augmenté, ce qui indique que les contours des bois ont eu tendance à se complexifier. Il est probable que des événements de coupes destinés au bois de chauffage aient eu lieu en bordure de forêts, ceci participant d'une part à réduire leur surface et d'autre part à morceler le contour des bois (De Warnaffe et al. 2006; Deconchat et al. 2007). Quant à la

connectivité, la majeure partie des bois a vu sa situation s'améliorer. La tendance majoritaire étant plutôt à la baisse de surface, il est probable que cette amélioration soit liée à l'apparition de nouveaux patchs dans le paysage. Ceci ayant eu pour conséquence d'augmenter la quantité de surfaces boisées dans le voisinage des patchs suivis.

La seconde période est marquée par une tendance d'évolution surfacique identique à la première période sans pour autant que le périmètre des patchs en soit impacté. Il semble y avoir eu une continuité du processus de contraction des patchs mais cela semble être plutôt dû à une modification de leur forme. L'évolution des patchs disposant de forme plus contractée vers une forme plus allongée peut ainsi faire baisser la surface tout en accroissant la longueur de contour des objets. La compacité des patchs a augmenté et va de pair avec les propos tenus précédemment. La dynamique d'évolution des patchs a conduit à créer des objets plus complexes au cours du temps. Enfin, contrairement à ce qu'affiche le taux de changement moyen, la connectivité a diminué pour la majorité des patchs durant cette période. Cela pourrait être lié à la baisse de surface d'une grande quantité des patchs dans le voisinage des patchs étudiés.

La période 1954-1979 contraste avec les dynamiques jusque-là observées. La majorité des bois étudiés ont gagné de la surface et ceci constitue la première évolution positive depuis 1850 en ce qui concerne cet aspect de leur structure. Il est toutefois difficile de se prononcer sur l'évolution du périmètre puisque la moitié des forêts a connu une baisse alors que la seconde moitié s'est améliorée de ce point de vue. En revanche, le niveau de complexité des patchs a baissé pour la majorité des bois indiquant que l'expansion observée durant cette période a eu tendance à simplifier les forêts suivies. On observe également une tendance claire de la connectivité : 44 forêts sur les 46 suivies ont profité d'une hausse de la quantité de surfaces boisées dans leur voisinage.

L'ultime période de temps étudiée (1979-2010) est caractérisée par une croissance positive de la connectivité pour la totalité des forêts suivies. Le phénomène d'afforestation qui s'est élargi à l'ensemble du paysage dès 1954 s'est donc poursuivi pour favoriser à priori les flux de colonisation des espèces à travers les bois. Les dynamiques de surface et de périmètre sont relativement partagées et n'affichent pas de tendance claire bien que les taux moyen d'évolution sont positifs pour ces deux métriques. En revanche, la compacité a une nouvelle fois baissé pour la grande majorité des bois (31 cas 48 détectés) laissant place à des forêts aux formes relativement simplifiées indiquant peut être qu'elles sont de plus en plus gérées par l'Homme.

9.4. TYPOLOGIE DES EVOLUTIONS SPATIO-TEMPORELLES

La typologie des évolutions doit nous permettre d'identifier des tendances d'évolutions parmi les forêts suivies et de les regrouper. Ce regroupement aura pour objectif de savoir si la biodiversité contenue dans les patchs d'une même classe est similaire ou non. Cette vérification apportera ainsi des informations précieuses au sujet de l'impact que pourrait avoir l'histoire sur la variation de syrphes forestiers au sein des patchs actuels.

9.4.1 ACP SUR L'ANNEE DE REFERENCE

Afin d'avoir une idée de la structuration des variables, il est nécessaire d'examiner les parts relatives de la variance pour chaque composante. Le graphique 1 de la Figure 9-4 montre que 57,43 % de l'information contenue dans le tableau initial est résumée par le premier facteur. Le pouvoir discriminant des autres facteurs est relativement faible bien que 26,45 % de l'information du tableau initial soit résumée par le second facteur. Au total, environ 83% de la variance est expliquée par les deux premières composantes : nous choisissons donc d'utiliser uniquement ce premier plan factoriel.

Il s'agit maintenant d'interpréter la relation qu'entretiennent les variables avec les composantes retenues (Figure 9-4(2)). Pour cela, nous analysons les coefficients de corrélation entre

les variables et les axes. Le cercle des corrélations montre clairement une structure orthogonale. Nous observons que la surface et le périmètre montrent une forte corrélation linéaire avec l'axe 1 en affichant des coefficients de corrélation proches de 1. L'axe 2 rend compte d'une très forte corrélation avec la variable de connectivité illustrée par un coefficient de corrélation égal à -0,94. Enfin, la variable de compacité affiche un coefficient de corrélation modéré avec l'axe 1 (-0,68). A noter que le signe négatif de ce dernier coefficient indique que la compacité des forêts est négativement corrélée aux variables de surface et de périmètre. Le plan factoriel retenu permet ainsi de dégager très clairement deux axes d'analyse, à savoir la morphologie des forêts sur l'axe 1 et leur connectivité sur l'axe 2.

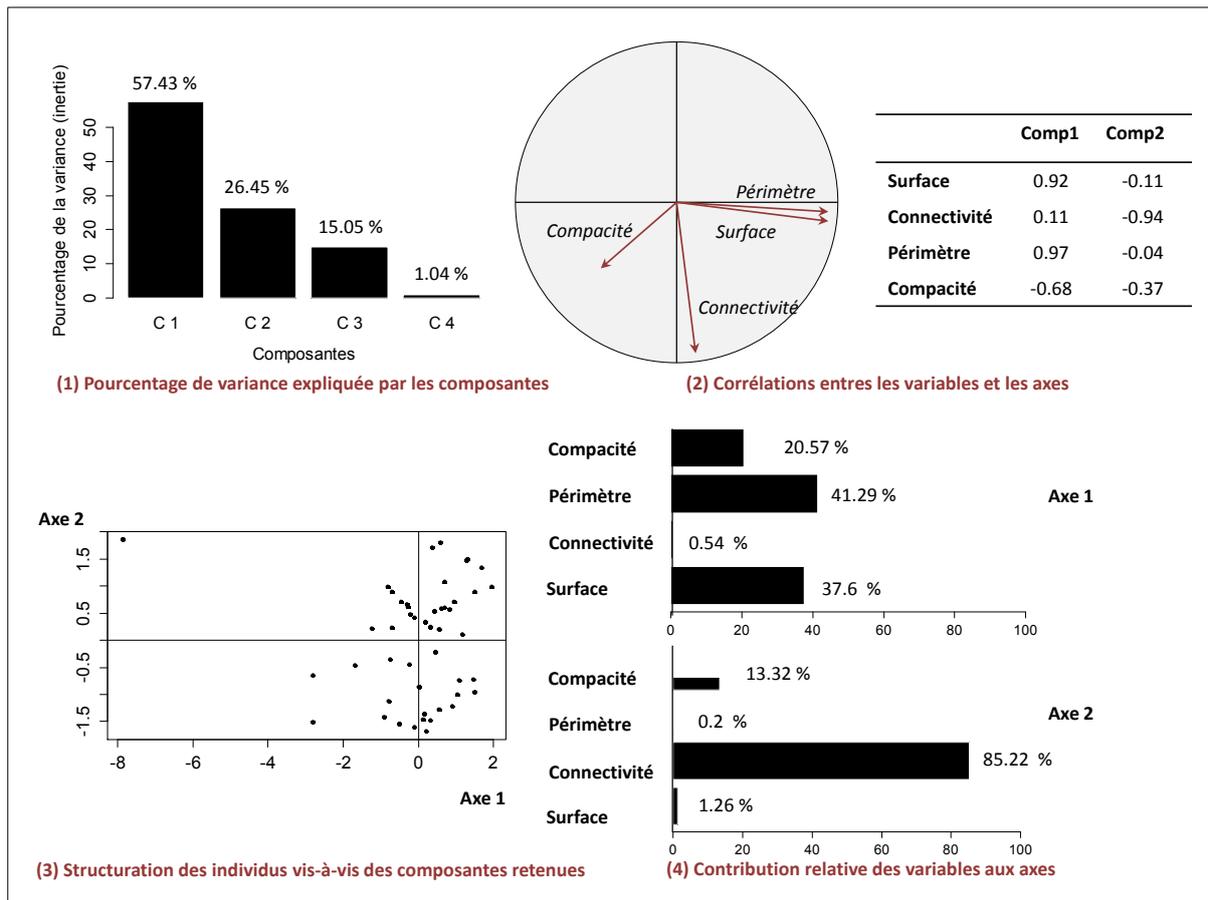


Figure 9-4. Résultats de l'Analyse en Composantes Principale basée sur les forêts de 2010

Ensuite, nous avons analysé la contribution de chaque variable à la construction des axes dans le but de savoir lesquelles avaient participé le plus à construire les composantes retenues dans notre plan factoriel (Figure 9-4(4)). Ainsi, l'axe 1 a été construit majoritairement et équitablement par la surface et le périmètre. Ces deux métriques affichent des valeurs de contribution relative à l'axe 1 respectivement égales à 37,6 et 41,29%. La compacité n'a apporté qu'une faible contribution (20,57%) alors que l'apport de la connectivité pour cet axe a été quasi nul (0,54%). L'axe 2 a lui été majoritairement construit à partir de la variable de connectivité (85,22%). La compacité a joué un rôle mineur en contribuant à hauteur de 13,32% alors que les variables de surface et de périmètre ont très faiblement apporté elles aussi (respectivement 1,26 et 0,2%).

Enfin, la projection des individus sur le plan factoriel offre la possibilité d'observer comment ils se répartissent vis-à-vis des deux composantes retenues (Figure 9-4(3)). Le graphique montre que les individus sont répartis équitablement sur les axes 1 et 2 avec des valeurs d'amplitude quasi-équivalente. L'axe 1 oppose ainsi les patchs les plus larges situés à droite (axe1= 2) et les patchs les

plus petits situés tout à gauche de l'axe (-3). L'axe 2 renseigne sur le niveau de connectivité des patches où les forêts les mieux connectés se trouvent tout en haut (axe 2 =2) alors que les forêts moins bien pourvues du point de vue de cet aspect sont situés en bas de l'axe (axe 2 = -2).

9.4.2 LES TRAJECTOIRES D'EVOLUTION

L'ACP réalisée sur les forêts de 2010 a permis de faire ressortir une structuration claire des individus selon deux aspects principaux : leur taille (surface et périmètre) et leur isolement (ou connectivité). L'analyse des trajectoires d'évolution consistera par conséquent à analyser le déplacement des forêts en fonction du temps sur la base de ces deux aspects de leur structure.

Dans cette logique, chacune des forêts suivies a fait l'objet d'une trajectoire d'évolution. Quatre types de trajectoires émergent immédiatement sur la base de leur date d'apparition (Figure 9-5). Un premier groupe reprend ainsi toutes les trajectoires des forêts présentes depuis 1850 jusqu'à aujourd'hui sans discontinuité (i.e. les forêts anciennes, N = 28). Ces trajectoires sont caractérisées par la présence de cinq points illustrant la position du bois à cinq instants de notre ligne de temps (1850, 1900, 1953, 1979 et 2010). Elles affichent des évolutions très variées et semblent se structurer principalement autour de l'axe 2 (i.e. la connectivité). Viennent ensuite les trajectoires des forêts observées à partir de 1900. Elles sont au nombre de 5 et sont constituées de trois segments reliant les quatre positions historiques de la forêt suivie. Les deux dernières classes de trajectoires contiennent respectivement les forêts observées à partir de 1954 (N = 13) et observées à partir de 1979 (N = 2). Ces trois dernières classes (i.e. les forêts récentes) affichent des évolutions moins complexes et semblent également se structurer autour de l'axe 2.

Les forêts récentes (i.e. apparues en 1900 ou ensuite) feront l'objet d'un regroupement particulier à savoir des classes d'âge selon leur date d'apparition. Elles ne sont donc pas concernées par la phase de classification décrite ci-dessous pour plusieurs raisons. Premièrement, il n'est pas possible de comparer toutes ces forêts sur la même durée d'existence en raison du décalage lié à leur date d'apparition. Deuxièmement, la création d'un premier niveau de regroupement sur la base de l'ancienneté est pertinente d'un point de vue écologique. **Cette distinction offrira la possibilité de vérifier si la durée d'existence joue un rôle sur la présence actuelle de syrphes forestiers.**

Les forêts anciennes affichent elles des évolutions plus hétérogènes et doivent faire l'objet d'une attention particulière. La classification des trajectoires menée dans la section ci-dessous porte donc principalement sur ce groupe et doit faire émerger des grandes tendances d'évolution. Cela permettra de vérifier si les dynamiques d'évolution qu'ont connues ces bois durant un siècle et demi ont conduit à faire varier différemment la biodiversité qu'ils contiennent aujourd'hui.

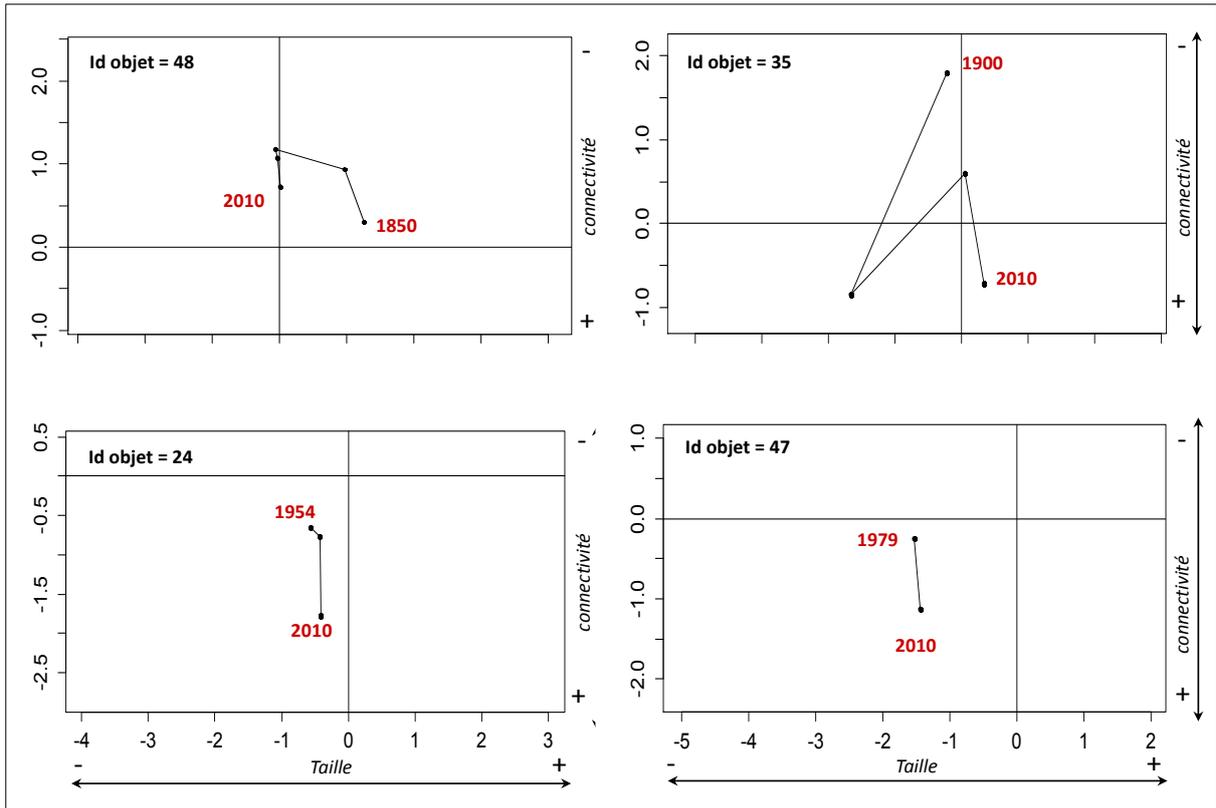


Figure 9-5. Exemples de quatre trajectoires forestières correspondant à quatre durées d'existence différente

9.4.3 LES GROUPES DE TRAJECTOIRES

Les trajectoires obtenues sont globalement très hétérogènes tant du point de vue de la direction des changements (i.e. une hausse ou une baisse de la connectivité ou de surface) que de leur intensité (cf. Annexe h pour visualiser les trajectoires). Comme démontré précédemment (section 9.3), il est d'autant plus difficile de faire ressortir des tendances claires étant donné que ce sont rarement les mêmes forêts qui sont concernées par la même évolution d'une période à une autre. Pour ces différentes raisons, nous avons fait le choix d'analyser principalement le déplacement global de la position des forêts (i.e. entre 1850 et 2010) sur les deux axes de notre plan factoriel pour les regrouper.

Deux grandes groupes ont été observés : (1) les forêts ayant très peu évolué en surface et en connectivité ; (2) les forêts ayant connu une nette hausse de connectivité entre 1850 et 2010 sans évolution majeure de leur surface (Tableau 9-3).

Premièrement, au sein du premier groupe (N = 12), la plupart des forêts sont de petite taille (8 cas sur 12, surface 2010 <7ha, ex : id1, id22, id7) caractérisées par une trajectoire repliée sur elle-même et des coordonnées globalement agglomérés au même endroit dans le plan factoriel. Les autres (4 cas sur 12, ex : id10, id14, id16) disposent d'une trajectoire similaire mais sont des forêts de plus grande superficie (cadre droit du plan factoriel, surface2010>7ha).

Le second groupe (N = 12) est composé de cinq forêts de petites tailles ayant connu une forte hausse de connectivité entre 1850 et 2010 (ex : id2, id37, id39). Les trajectoires de ces forêts sont caractérisées par un déplacement important sur l'axe 2, du haut vers le bas (i.e. de 1850 vers 2010). Les autres forêts de ce groupe, au nombre de 7, ont aussi subi une amélioration de leur connectivité

sur cette période mais disposent d'une plus grande superficie (surface 2010 > 7,2ha, cadre droit du plan factoriel, ex : id11, id21, id28).

Enfin, les quatre forêts restantes (ex : id3, id6 et id40, id48) sont de grande taille et caractérisées par un déplacement vertical sur l'axe 2 du bas vers le haut. Elles ont par conséquent subies une perte globale de connectivité entre 1850 et 2010.

Malgré la difficulté de créer une typologie des forêts à partir des trajectoires obtenues, ce travail a permis de formuler plusieurs hypothèses sur la variation de biodiversité potentielle au sein des groupes observés (cf. Tableau 9-3, Figure 9-6).

Les forêts du premier groupe ont peu évolué et disposent d'une faible superficie. On pourrait donc s'attendre à une faible richesse dans ces bois aujourd'hui pour deux raisons. Leurs petites surfaces impliquent une faible diversité des ressources. Leur morphologie et leur isolement ne se sont pas améliorées et n'ont par conséquent pas pu compenser ces mauvaises conditions initiales. Les forêts du second groupe ne se sont pas améliorées non plus du point de vue de ces deux aspects (morphologie et isolement). Elles disposent en revanche d'une plus grande surface donc potentiellement d'une plus grande diversité de ressources qui leur permet d'accueillir plus d'espèces différentes. Ainsi, le fait qu'elles aient été peu concernées par des modifications est potentiellement un signe positif, ceci ayant permis aux bonnes conditions initiales de persister sur le long terme.

Pour les forêts du groupe 3, la hausse de connectivité entre 1850 et 2010 a pu compenser leur faible surface en favorisant les flux d'espèces entre les patchs forestiers. Elles pourraient donc abriter plus d'espèces aujourd'hui qu'en 1850. Les forêts du groupe 4 ont à la fois connu cette hausse mais disposent d'une plus grande superficie peu impactée au cours du temps. La richesse de ces bois devrait par conséquent être plus élevée aujourd'hui et potentiellement plus élevée que dans les autres groupes. En effet, le maintien d'une grande diversité de ressources et l'amélioration des conditions d'isolement constituent des conditions favorables à l'installation d'une grande diversité. Enfin, les forêts du groupe 5 disposent d'une grande superficie mais leur connectivité s'est dégradée au cours du temps. Leur richesse devrait par conséquent être moins riche aujourd'hui qu'en 1850 et les espèces abritées dans ces bois pourraient avoir été concernées par des mécanismes d'extinction (cf. chapitre 10).

GROUPE	EFFECTIF	CARACTERISTIQUES	ID
Groupe 1	N = 8	Petites forêts stables	1, 5, 8, 7, 12, 13, 15,23
Groupe 2	N = 4	Grandes forêts stables	10, 14, 16, 18
Groupe 3	N = 5	Petites forêts ayant connu une hausse de connectivité entre 1850 et 2010	2, 37, 39, 41,42
Groupe 4	N = 7	Grandes forêts ayant connu une hausse de connectivité entre 1850 et 2010	11, 21, 26, 27, 28, 33,37
Groupe 5	N = 4	Grandes forêts ayant connu une perte de connectivité entre 1850 et 2010	3, 6, 48,40

Tableau 9-3. Caractéristiques des groupes de forêts identifiés à la suite de la procédure de regroupement

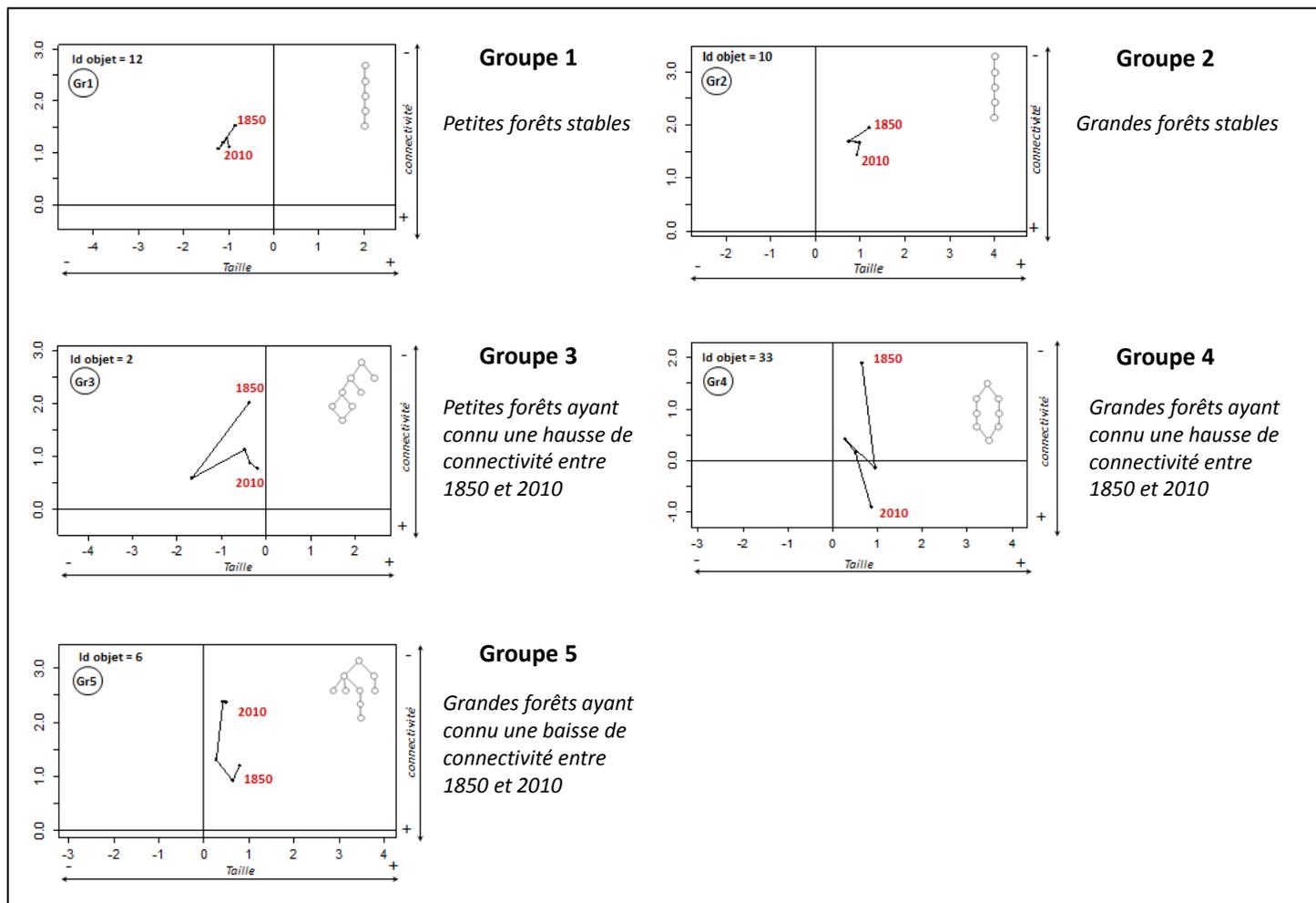


Figure 9-6. Aperçu des trajectoires caractéristiques des cinq groupes identifiés.

CONCLUSION DE LA PARTIE C

L'intérêt de cette partie était de proposer une démarche qui permette d'apparier les espaces forestiers et de caractériser les évolutions spatio-temporelles de 1850 à 2010. La démarche proposée a été appliquée ensuite aux espaces forestiers que nous suivons dans le cadre de notre protocole écologique.

Par le biais d'une approche manuelle, nous avons cherché à apparier les espaces forestiers. Cette étape centrale de notre démarche a permis d'identifier les différents homologues des forêts échantillonnées sur les sources précédentes et de reconstruire la généalogie de chacune d'entre elles. L'analyse de la cardinalité des liens a ainsi permis de donner un aperçu global des processus spatio-temporels ayant affecté la généalogie des forêts suivies.

La caractérisation des trajectoires a eu pour objectif de décrire la structure spatiale des forêts. Le choix des métriques a été guidé par deux principaux déterminants. Le premier concerne leur capacité à décrire la morphologie et l'isolement des bois. Le second porte sur leur aptitude à pouvoir caractériser à la fois un objet simple mais aussi un objet composé afin qu'ils soient comparable sur la base de la même métrique. Quatre variables paysagères ont finalement été retenues. La surface, le périmètre et un indice de compacité ont permis de caractériser la morphologie des bois pour évaluer respectivement les effets de taille, de lisière et de la forme des habitats forestiers sur la richesse de syrphes. L'isolement des bois a été décrit à travers une métrique de connectivité potentielle.

La caractérisation des évolutions spatio-temporelles a été effectuée à travers deux approches. L'approche « individu » a consisté à calculer les taux de changement entre deux dates pour chaque forêt suivie à l'aide des quatre métriques choisies. Nous avons pour cela adopté un modèle « snapshot » qui a permis de produire la BD « Changement ». Cette approche permettra ainsi de mettre en relation les variables explicatives de changement associées à chaque bois avec la richesse spécifique de syrphes qu'ils contiennent en Partie 4. L'approche « population » s'est appuyée sur l'Analyse en Composantes Principales pour reconstruire les trajectoires d'évolution des forêts. Pour cela, un plan factoriel de référence a premièrement été défini en appliquant une ACP aux forêts présentes en 2010 avant que les individus des autres dates soient reprojétés dans ce même plan. Ces trajectoires ont alors été mobilisées dans la section suivante pour faire émerger des grands types d'évolution parmi les forêts étudiées.

Enfin, la classification des évolutions spatio-temporelles a eu pour objectif de regrouper les trajectoires des forêts produites dans la section précédente pour construire une typologie des évolutions. Les forêts anciennes et récentes ont dans un premier temps fait l'objet d'une première séparation. Puis, nous avons procédé à un examen visuel des trajectoires obtenues pour tenter de déceler des grands types d'évolution parmi les forêts anciennes. Cinq groupes de forêts ont été identifiés et ont permis de formuler plusieurs hypothèses sur la potentielle biodiversité qu'ils pourraient contenir.

PARTIE D

Mesurer l'effet des évolutions spatio-temporelles des forêts et de leurs incertitudes sur la biodiversité

INTRODUCTION DE LA PARTIE D

Cette dernière partie porte sur la modélisation des effets des évolutions des forêts et de leurs incertitudes sur la biodiversité. L'objectif est de mettre en relation les variables paysagères dynamiques dérivées précédemment avec la richesse actuelle de syrphes forestiers mesurées dans les forêts échantillonnées.

Le chapitre 10 présente la démarche de modélisation statistique proposée pour étudier le potentiel lien entre l'histoire des forêts et la biodiversité qu'elles abritent actuellement. Les résultats de cette étude seront présentés sous la forme d'un article publié dans la revue *Landscape Ecology*.

Le chapitre 11 a pour objectif d'évaluer les effets de l'incertitude spatiale sur la réponse des modèles écologiques définis. Après un bref rappel des grandes notions structurant les travaux de l'incertitude spatiale, nous montrerons quels types d'erreurs sont susceptibles de biaiser les modèles espèces-habitat. Nous nous focaliserons sur l'impact du processus d'extraction des forêts et évaluerons ses effets en comparant plusieurs modèles écologiques basés sur différentes variables d'entrée.

CHAPITRE 10. MODELISATION ESPECE-HABITAT : EFFET DE L'EVOLUTION DES FORETS SUR LA BIODIVERSITE

SOMMAIRE

- 10.1. LA MODELISATION ESPECE-HABITAT EN ECOLOGIE HISTORIQUE :
PRINCIPES ET SPECIFICITES
- 10.2. LA DEMARCHE DE MODELISATION PROPOSEE
- 10.3. UNE ETUDE PUBLIEE DANS L'ETSCAPE ECOLOGY

Nous nous attachons dans ce chapitre à décrire la démarche de modélisation employée pour estimer les effets potentiels de l'évolution des forêts sur la biodiversité. Un premier point sera consacré au fonctionnement général de la modélisation espèce/habitat tout en précisant les spécificités des modèles utilisés en écologie historique. L'intérêt sera ensuite d'identifier une approche de modélisation permettant d'intégrer l'histoire des paysages pour être en mesure de la relier avec la richesse de syrphes. La dernière partie de ce chapitre présentera les résultats de cette étude sous la forme d'un article soumis à *Landscape Ecology* (version révisée).

10.1. LA MODELISATION ESPECE-HABITAT EN ECOLOGIE HISTORIQUE : PRINCIPES ET SPECIFICITES

La modélisation espèce-habitat vise à prédire ou à expliquer la distribution des espèces en fonction de variables décrivant les caractéristiques de leurs habitats (milieu interne, morphologie ou isolement). Elle est utilisée pour répondre à divers objectifs de gestion et de conservation. Les résultats de ces modèles permettent de guider l'échantillonnage pour une espèce connue ou pour accélérer la détection de nouvelles espèces, définir des zones de protection ou encore guider la réintroduction d'espèces menacées. Elle est largement utilisée dans plusieurs disciplines telles que la biogéographie, la phylogéographie, l'écologie du paysage ou plus récemment en écologie historique. Elle offre alors la possibilité d'évaluer divers mécanismes comme celui de la spéciation (phylogéographie) ou encore les risques biologiques et les problématiques liées aux espèces invasives

Les espèces de syrphes présentent des schémas de distribution et d'agrégation particuliers en association avec des habitats ou des étapes particulières de leur cycle de vie (Raymond 2013; Raymond et al. 2014). La connaissance des distributions spatiales des espèces et des mécanismes qui les structurent devrait permettre une meilleure utilisation, gestion et conservation des habitats notamment lorsqu'ils sont des ressources exploitées ou affectées par les activités humaines.

Dans le cadre de la modélisation statistique de distribution des espèces et dans le but d'identifier les facteurs contrôlant cette distribution, les modèles sont utilisés pour relier la réponse d'une population à des hypothèses de contrôle externe (taille de l'habitat, température, faune associée) ou internes à la population (taille, structure en âge...) (Le Rest 2013). Ces modèles sont en mesure de fournir une estimation numérique de la réponse de ces espèces en fonction de la variation de ces facteurs. Il existe un grand nombre de modélisation statistique permettant de prédire la distribution des espèces et plusieurs méthodes peuvent être utilisées pour répondre à ces objectifs. En écologie du paysage, le schéma classique de modélisation consiste à mettre en relation un patron actuel de biodiversité avec des variables mesurant la configuration et la composition actuelle du paysage. Néanmoins, comme il est mentionné en Chapitre 2, ces facteurs peuvent avoir un effet sur la biodiversité mais ils ne l'expliquent à eux seuls. Si la plupart des espèces répondent immédiatement aux modifications de leur habitat, d'autres peuvent exprimer un délai de réponse à ces perturbations : il n'est alors plus possible de se focaliser uniquement sur des données actuelles pour expliquer ces patrons de distribution.

10.1.1 LES NOTIONS DE CREDIT DE COLONISATION ET DE DETTE D'EXTINCTION

En écologie historique, la modélisation espèce-habitat a pour objectif d'évaluer une potentielle relation entre l'histoire des habitats (leurs états passés, leurs évolutions) et la variation de la biodiversité observée actuellement dans un paysage. On cherche en réalité à savoir s'il existe un écart temporel entre la modification d'un habitat et la présence d'une espèce ou d'une communauté en lien avec cette évolution de l'habitat. Cet écart temporel (ou temps de relaxation) structure aujourd'hui une double notion clé de l'écologie historique : le crédit de colonisation ou la dette d'extinction (Hanski 2000; Vellend et al. 2006; Kuussaari et al. 2009; Cristofoli et al. 2010; Piqueray et al. 2011)

Selon Kuusaari (2009), la dette d'extinction traduit la différence entre la réponse de l'espèce observée au début de la modification de son habitat et la réponse observée à l'équilibre théorique (Figure 10-1(a)). Au point d'équilibre initial, la richesse en espèces est élevée. Après une perturbation, l'habitat est dégradé mais ne provoque pas nécessairement l'extinction immédiate de toutes les espèces avant de parvenir au nouveau point d'équilibre (new equilibrium). De fait la différence entre le nombre d'espèces restantes à la fin de la perte d'habitat et le nombre d'espèces au point d'équilibre théorique informe sur une potentielle d'extinction. La période écoulée entre le début de la perte d'habitat et le moment où le nouvel équilibre est atteint est appelée temps de relaxation (relaxation time). Ce temps de relaxation diffère selon le cycle générationnel des espèces. Les espèces disposant d'un cycle générationnel court (points orange) sont susceptibles d'atteindre plus rapidement un nouvel équilibre après la perturbation d'un habitat que les espèces au cycle générationnel plus long (points bleus). L'image de gauche (encadré bleu) présente un exemple d'une plante herbacée (cycle générationnel long) et l'image de droite est l'exemple d'une espèce de papillon soit une espèce au cycle générationnel plus court.

Ce même auteur propose une seconde approche pour évaluer la présence d'une dette d'extinction (Figure 10-1(b)). Dans un habitat continu (c'est-à-dire qui n'a pas connu de perturbation), les sous parties de cet habitat (points bleus) disposeront d'une richesse plus importante que les habitats de même surface ayant connu une perturbation (points oranges). De fait, la différence en nombre d'espèces entre ces habitats fragmentés et non fragmentés (à tailles égales) donne une estimation de la perte d'espèces au début de la perte d'habitat. Dans les habitats fragmentés, les espèces survivantes ne pourront persister à long terme, il y aura alors probablement une dette d'extinction. De plus, immédiatement après la perturbation, la dette d'extinction est proportionnellement plus importante dans les petits patchs. Ainsi, si les dettes d'extinction interviennent plus rapidement dans les habitats de plus petite taille, la dette restante à payer est généralement plus grande dans les patchs les plus larges. La photo à droite illustre plusieurs sous-parties d'un habitat continu (patch bleu) et des habitats fragmentés plus ou moins larges (patchs orange).

- Ainsi, **une dette d'extinction** peut être constatée lorsqu'une perte d'habitat a été observée dans le passé et que la réponse de l'espèce à cette perte n'a pas été vérifiée sur le paysage actuel. L'espèce n'a pas encore payé sa dette et risque de s'éteindre dans le futur. Par conséquent, les caractéristiques du paysage passé et de l'habitat considéré expliquent mieux la diversité biologique d'aujourd'hui. Dit autrement, la diversité en espèces est plus importante que celle attendue au regard des caractéristiques de l'habitat (taille et isolement notamment) (Kuusaari et al. 2009a)
- Le schéma est identique lorsqu'on parle de « **crédit** » de colonisation mais la dynamique d'évolution de l'habitat observée est positive et s'apparente à une restauration. L'espèce dispose ainsi d'un crédit de colonisation, c'est-à-dire qu'elle n'a pas encore bénéficié de cette restauration en colonisant le patch nouvellement créé ou modifié.

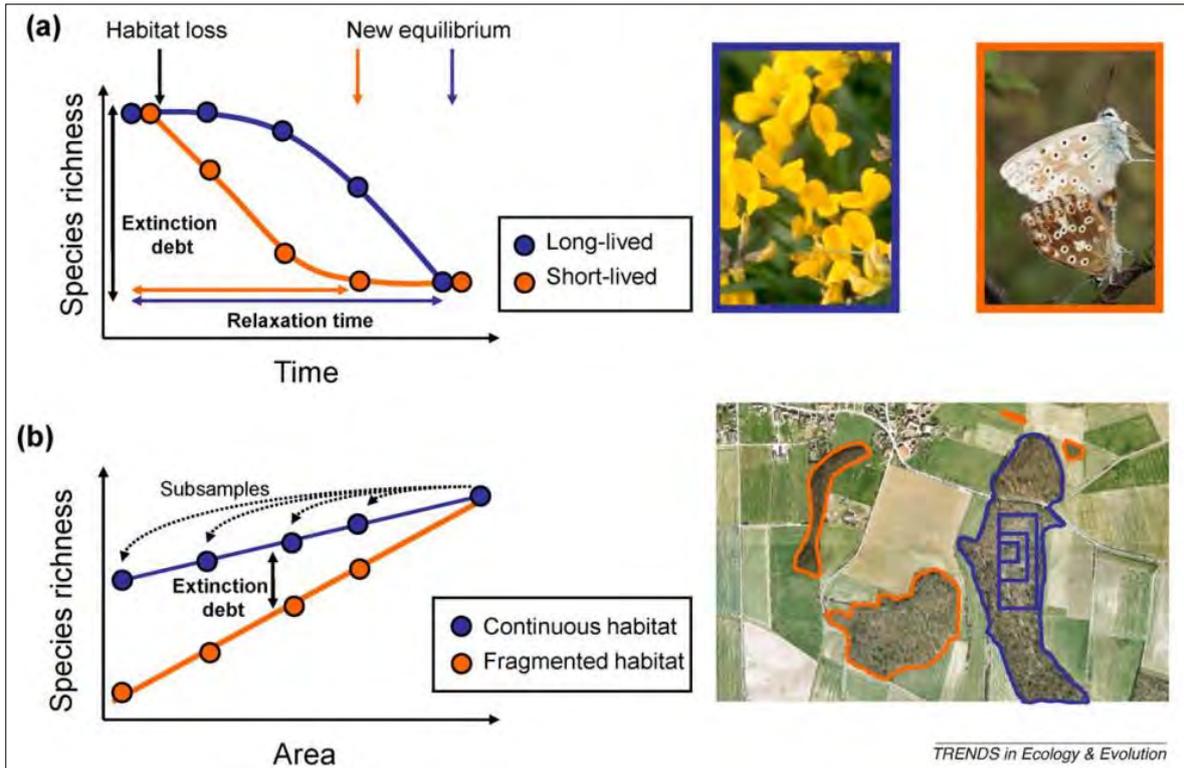


Figure 10-1. Le principe de la dette d'extinction (source : Kuussaari et al.,2009). (a) Une dette d'extinction peut être mise en évidence en observant une différence entre le nombre d'espèces présentes au début de la perturbation et le nombre d'espèces au nouveau point d'équilibre, (b) Une dette d'extinction peut être mise en évidence en observant une différence entre le nombre d'espèce présentes dans un habitat non-perturbé et le nombre d'espèces présentes dans un habitat perturbé de même taille.

10.1.2 LES DETERMINANTS DE LA DETTE ET DU CREDIT

La probabilité d'une dette d'extinction et d'un crédit de colonisation dépend principalement de trois facteurs :

(1) Premièrement, elle peut être liée aux traits de vie de l'espèce concernée (Jackson et Sax 2010). Les espèces disposant d'un taux de *turnover* important (taux de renouvellement des espèces) vont par exemple adopter un temps de relaxation plus court en réponse à une modification de leur habitat que celles dont le renouvellement s'effectue moins rapidement. Elles vont par conséquent s'adapter assez vite aux évolutions du milieu dans lequel elles se trouvent et seront de fait moins sujet à ce type de phénomène. La capacité de dispersion de l'espèce joue également un rôle crucial. Les espèces peuvent mettre plus ou moins de temps à coloniser des zones nouvellement restaurées ou créées (cas du crédit) ou bien à réinvestir des zones plus propices à leur pérennité lorsque leur habitat initial a été dégradé (cas de la dette). Les espèces disposant d'une faible capacité de dispersion (i.e. les plantes) pourront difficilement émigrer ou immigrer, ce qui n'est pas le cas des espèces plus mobiles (i.e. les oiseaux). Enfin, les espèces spécialistes sont également réputées plus vulnérables à ce type de phénomène. Elles sont plus sensibles à la disparition ou à la restauration d'un type d'habitat en particulier.

(2) La configuration spatio-temporelle des habitats peut également conditionner la détection d'une dette d'extinction (ou d'un crédit de colonisation) (Sang et al. 2010; Cousins et Vanhoenacker 2011; Piqueray et al. 2011; Dullinger et al. 2012). On observe par exemple en figure 10-1(b) que pour une même surface, la richesse spécifique est plus importante au sein d'une sous-partie d'un habitat

continu (cercles bleus)) que dans un habitat fragmenté de même taille (cercles oranges). Les différences de richesse observées entre les grandes et petites sous-parties de l'habitat continu (en bleu) donnent ainsi une approximation de la perte de richesse dans la phase initiale de la perte d'habitat. Dans les habitats fragmentés, ces mêmes espèces n'auraient pas la possibilité de persister et aurait probablement subi une dette d'extinction. On constate en effet en décalant notre attention vers la gauche du graphique que l'écart entre les deux droites s'accroît au fur et à mesure que la taille des patchs baisse. Cela indique qu'immédiatement après la perturbation, la dette d'extinction est proportionnellement plus haute dans les petits patchs isolés que dans les grands patchs connectés. Il est également possible de conclure qu'une dette interviendra moins rapidement dans les patchs de grande taille mais que cette dette sera probablement d'une plus grande importance que dans les patchs de plus petite taille.

(3) Enfin, le temps écoulé entre l'évènement dit de perturbation et la date de la réponse observée est fondamental puisqu'une dette ou un crédit ont pu survenir dans cet intervalle sans que cela ait pu être observé. Il a par exemple été montré que des plantes forestières en Belgique avaient payé leur dette 200 ans après la perturbation alors qu'au Royaume Uni, d'autres travaux ont mesuré plus de 1000 ans entre un évènement de perturbation et une réaction de ce type de plantes en réponse à cette dégradation (Peterken et Game 1984). De plus, il existe un lien étroit entre les échelles spatiales et temporelles. Nous avons vu dans le point précédent qu'une dette ou un crédit étaient susceptibles d'intervenir plus rapidement dans les petits patchs isolés (Figure 10-1(b). Cousins (2009), en reprenant différentes études mesurant l'effet de la fragmentation des prairies sur les plantes qu'elles accueillent, a montré que la plupart des dettes mises en évidence concernaient les prairies disposant d'au moins 10 % de leur surface originale. Cette relation a en revanche été très peu observée dans les prairies en contenant moins de 10%. Une possible explication est que la plupart des dettes ont eu lieu dans les paysages fortement fragmentés où le morçèlement du paysage a débuté il y a 70 à 200 ans. La magnitude de la perturbation peut aussi affecter le temps de relaxation des espèces.

10.1.3 COMMENT MESURER UNE DETTE D'EXTINCTION OU UN CREDIT DE COLONISATION ?

En raison d'un nombre important de travaux dans ce domaine depuis quelques années, il existe aujourd'hui différentes d'approches possibles pour mesurer une dette d'extinction ou un crédit de colonisation. Elles peuvent être classées en cinq groupes construits sur la base du type de données disponibles (Figure 10-2). A noter que les approches présentées et explicitées par Kuusaari et al (2009) s'appliquent à la dette d'extinction mais peuvent être toutefois remobilisées pour mesurer un crédit de colonisation (Cristofoli et al. 2010; Piqueray et al. 2011)

(A) La première approche utilise des données spatiales passés et présentes pour disposer d'informations sur la configuration spatiale des habitats. C'est aujourd'hui l'approche la plus répandue (Helm et al. 2006; Krauss et al. 2010). Elle demeure facile à mettre en œuvre mais n'informe pas sur la magnitude de la dette (i.e. la quantité d'individus ou d'espèce impactées). De fait, **si la richesse actuelle est mieux corrélée à la structure spatiale des habitats passés alors il est probable qu'il y a dette d'extinction (Figure 10-2(a)).**

(B) Une seconde approche utilise les mêmes données et vise à comparer la richesse des habitats fragmentés (i.e. non-stables) à celle des habitats continus et non perturbés (i.e. stables). **La richesse des habitats stables est considérée comme la richesse à l'état d'équilibre. Par conséquent la différence de richesse entre un habitat stable et instable, à même surface, informe sur la magnitude de la dette d'extinction (Figure 10-2(b)).**

(C) Une troisième possibilité est de s'appuyer sur des données spatiales et biologiques passés pour estimer la relation Aire/espèce de l'époque et la comparer avec la relation Aire/espèce d'aujourd'hui.

Les données biologiques passées sont toutefois difficiles à acquérir pour ce type d'approche. **L'écart entre la richesse passé et la richesse actuelle, à même surface, informe sur la magnitude de la dette d'extinction (Figure 10-2(c))**

(D) La quatrième approche mobilise des données spatiales et biologiques multi-temporelles. En effet, lorsque des données de changements spatiaux et des occurrences d'espèces avant et après perturbation sont disponibles, il est possible d'estimer directement le temps de relaxation mais de telles données sont rarement disponibles. **L'intervalle de temps qui sépare une richesse mesurée dans un patch au début de la phase de perturbation et une autre richesse mesurée dans ce même patch après perturbation informe directement sur le temps de relaxation de l'espèce (Figure 10.4(d)).**

e) Une dernière approche nécessite la présence de données empiriques concernant la croissance des populations ou leur déclin informant de fait sur le comportement adopté en réponse à une perturbation. Combiner ces données avec des approches de modélisation peut être potentiellement une méthode efficace pour évaluer une dette d'extinction. Bulman et al (2007) se sont par exemple appuyés sur ce type de données pour paramétrer un modèle de métapopulation et simuler l'évolution de six métapopulations de papillons occupant un réseau de patchs disponibles au Royaume Uni (Bulman et al. 2007).

Si les approches diffèrent principalement par le type de données qu'elles utilisent, elles sont plus ou moins faciles à mettre en œuvre du fait que certains types de données sont le plus souvent absents. De plus, ces méthodes ne permettent pas toute d'évaluer la magnitude de la dette d'extinction ou du crédit de colonisation, les interprétations possibles en sortie de modèle sont de fait différentes.

Les approches évoquées permettant de révéler un crédit de colonisation ou une dette d'extinction sont celles habituellement mises en œuvre pour mesurer les effets de l'histoire sur la biodiversité. De récents travaux s'appuient aussi sur la dynamique des habitats en considérant les changements comme un processus cumulatif (Metzger et al. 2009a; Ewers et al. 2013a). Metzger et al (2009) ont par exemple montré qu'intégrer les taux de changement de surface d'un habitat (i.e. la différence de surface relative entre deux dates) ainsi que les taux de changement de connectivité, en plus des caractéristiques spatiales des habitats actuels, permettaient d'améliorer significativement l'explication de la richesse spécifique et de l'abondance de certaines espèces (Metzger et al. 2009a). Le taux de changement des habitats pourrait modifier à travers le temps la balance extinction-colonisation d'une population et ainsi conditionner la réponse observée sur un paysage actuel (Münzbergová et al. 2005a). Une récente étude portant sur la richesse de plantes dans les landes belges a également adopté ce principe (Cristofoli et al. 2010). L'étude des dynamiques du paysages a notamment permis d'observer des évolutions complexes parmi les patchs suivis et de faire une distinction entre ceux qui ont subi une évolution positive ou négative. Ces informations sont cruciales car l'hypothèse formulée sur la présence d'une dette d'extinction ou d'un crédit de colonisation dépend directement du type de dynamique survenue dans le paysage.

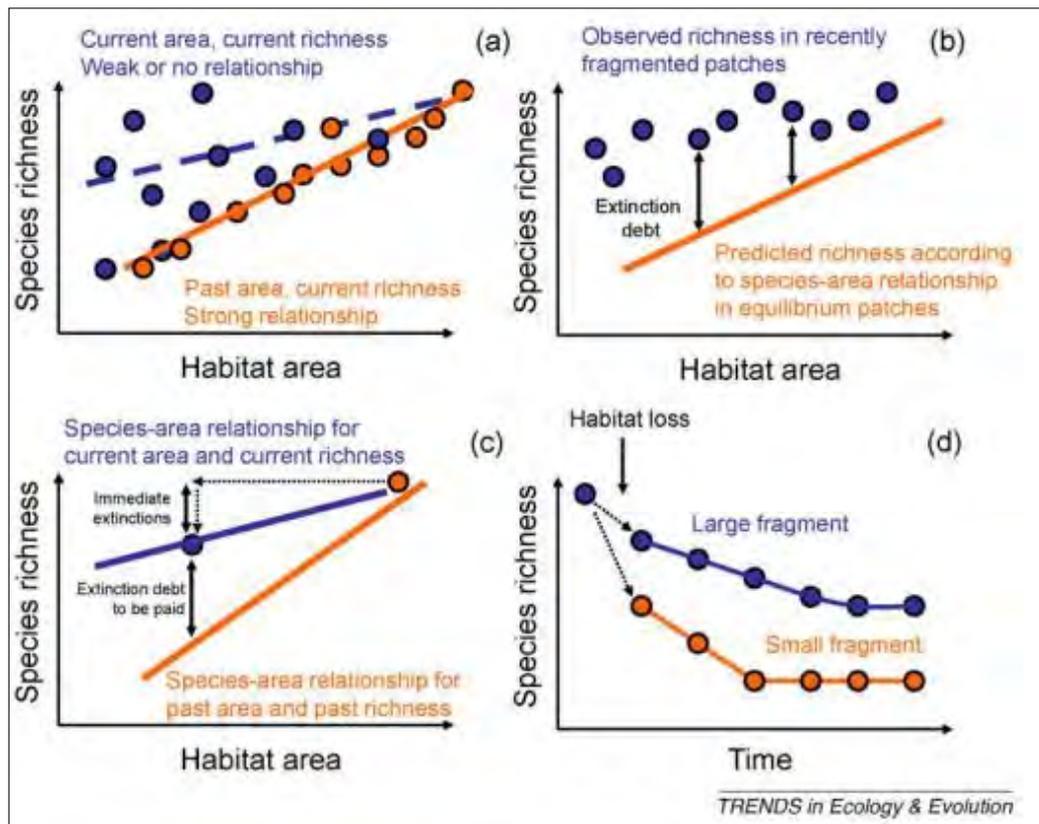


Figure 10-2. Les approches pour détecter une dette d'extinction. Ces dernières peuvent être remobilisées également pour détecter un crédit de colonisation. (source : Kuussaari et al. 2009)

10.1.4 BILAN

La majorité des travaux dans ce domaine portent principalement sur la mise en évidence de dettes d'extinction et peu s'attachent à mettre en lumière un crédit de colonisation. Ceci s'explique par les dynamiques globales d'évolution du paysage à travers le monde qui ont tendance à morceler les habitats disponibles ou à les supprimer totalement (Fahrig et Merriam 1985; Saunders et al. 1991; Young et Clarke 2000). Les dynamiques conduisant à la création de nouveaux patchs restent plus marginales. De plus, les recherches concernant les dynamiques de colonisation ont très longtemps été menées à une échelle très locale (Bossuyt et Milinkovitch 2001) en mesurant les effets des pratiques de gestion à l'échelle d'un site. Il était de fait difficile de tirer des conclusions à l'échelle d'une population par ce type de procédé. Des travaux plus récents démontrent aujourd'hui les effets qualitatifs de la structure spatiale du paysage sur les processus de colonisation des espèces (Jacquelyn et al. 2003; Verheyen et al. 2003; Vellend et al. 2006b; Cristofoli et Mahy 2010).

Dans le but de mesurer une dette d'extinction ou un crédit de colonisation, quelques préconisations doivent être prises.

Un premier point à respecter pour conduire une étude visant à vérifier la présence d'une dette d'extinction ou d'un crédit de colonisation est de s'appuyer sur des espèces spécifiquement inféodées au type d'habitat que l'on étudie. Introduire les espèces généralistes pourrait masquer la présence d'une dette ou d'un crédit. Par exemple, des espèces spécialistes des milieux prairiaux ont payé une dette d'extinction dans une zone prairiale estonienne mais pas les généralistes (Helm et al. 2006a). Cela n'empêche pas les espèces spécialistes d'adopter parfois un comportement différent en réponse à une perturbation de leur habitat. Elles peuvent utiliser une portion différente de l'habitat

et répondre à une disponibilité d'habitat à différentes échelles (Gutiérrez et al. 2001; Cozzi et al. 2008).

Deuxièmement, le choix des critères à utiliser pour caractériser les habitats est primordial. Des variables couramment utilisées comme la surface d'habitat peuvent n'entretenir aucune relation avec la présence d'une espèce alors que la connectivité ou la qualité des patchs étudiés en seront les principaux déterminants. La richesse spécifique actuelle dans les prairies suédoises a d'ailleurs affiché une relation avec la connectivité historique des habitats mais pas avec leur surface (Lindborg et Eriksson 2004). Un travail à priori sur les effets potentiels des variables est par conséquent requis pour améliorer le pouvoir explicatif des modèles construits (comme réalisé en chapitre 1 et 2). De plus, le risque d'extinction étant dépendant de plusieurs échelles spatiales, il est préconisé de sélectionner une zone d'analyse adaptée ou bien d'en tester plusieurs (Lindborg et Eriksson 2004a).

Troisièmement, la mauvaise qualité des données spatiales peut être un facteur limitant. Le suivi et la quantification des changements de la structure spatiale des habitats doit le plus souvent s'appuyer sur des données précises pour être en mesure de démontrer une dette d'extinction ou un crédit de colonisation. Par conséquent, la détection d'une dette d'extinction (ou d'un crédit de colonisation) ou la non-détection pourrait résulter des incertitudes inhérentes aux données spatiales, d'autant plus lorsque celles-ci sont issues de sources anciennes.

10.2. DEMARCHE DE MODELISATION PROPOSEE

Nous explicitons ici la démarche de modélisation statistique que nous proposons pour mesurer les effets de l'histoire des habitats forestiers sur la richesse spécifique actuelle de syrphes.

10.2.1 LES VARIABLES EXPLICATIVES ET LA VARIABLE REPONSE

A ce stade de la démarche globale de cette thèse, nous sommes en possession de deux types de variables explicatives : des variables d'état et des variables d'évolution. Les variables d'état sont issues des bases de données spatiales forestières caractérisées (et appariées) aux différentes dates retenues dans notre analyse (cf. section 8.2). De fait, chaque forêt ayant été identifiée comme partageant la généalogie d'un patch forestier échantillonné a été caractérisée par deux variables paysagères : la surface (en hectares) et le pourcentage d'habitat forestier dans un rayon de 4800 mètres (variable de connectivité). Cette base de données multi-dates a permis aussi de créer une variable d'ancienneté nommée « Age ». Chaque forêt échantillonnée s'est vu attribuer un numéro correspondant à sa date d'apparition dans notre jeu de données (1 = à partir de 1850, 2 = à partir de 1900, 3 = à partir de 1954, 4 = à partir de 1979). Enfin, les variables d'évolution proviennent de la BD « changements » produite en section 8.3.5. Elles correspondent aux taux de changement des métriques calculées pour chacune des forêts suivies (i.e. la surface et la connectivité), durant les différentes périodes de temps.

Ces variables explicatives vont être mobilisées pour expliquer la variation d'une richesse spécifique actuelle de syrphes construite à l'échelle de notre site (la variable réponse). Elle exprime le nombre total d'espèces forestières différentes récoltées à l'échelle d'un patch échantillonné.

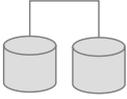
BASES DE DONNEES	VARIABLES DERIVEES	DESCRIPTION
 <p>BD Forêts Multi-dates</p>	<ul style="list-style-type: none"> • <i>Surface2010</i> Surface actuelle d'un patch forestier (BD ortho IRC 2010) • <i>Connectivité2010</i> Connectivité actuelle d'un patch forestier (BD ortho IRC 2010) • <i>Surface1979</i> Surface passée d'un patch forestier (Photographies 1979) • <i>Connectivité1979</i> Connectivité passée d'un patch forestier (Photographies1979) • <i>Surface1954</i> Surface passée d'un patch forestier (Photographies 1953-54) • <i>Connectivité1954</i> Connectivité passée d'un patch forestier (Photographies 1953-54) • <i>Surface1900</i> Surface passée d'un patch forestier (Cartes EM Type 1889) • <i>Connectivité1900</i> Connectivité passée d'un patch forestier (Cartes EM Type 1889) • <i>Surface1850</i> Surface passée d'un patch forestier (Minutes EM 1850) • <i>Connectivité1850</i> Connectivité passée d'un patch forestier (Minutes EM 1850) • <i>Age</i> Continuité historique d'un patch forestier 	
 <p>BD Changements</p>	<ul style="list-style-type: none"> • $\Delta Surf_{2010-1979}$ %Changement de surface d'un patch forestier entre 2010 et 1979 • $\Delta Con_{2010-1979}$ %Changement de connectivité d'un patch forestier entre 2010 et 1979 • $\Delta Surf_{1979-1954}$ %Changement de surface d'un patch forestier entre 1979 et 1954 • $\Delta Con_{1979-1954}$ %Changement de connectivité d'un patch forestier entre 1979 et 1954 • $\Delta Surf_{1954-1900}$ %Changement de surface d'un patch forestier entre 1954 et 1900 • $\Delta Con_{1954-1900}$ %Changement de connectivité d'un patch forestier entre 1954 et 1900 • $\Delta Surf_{1900-1850}$ %Changement de surface d'un patch forestier entre 1900 et 1850 • $\Delta Con_{1900-1850}$ %Changement de connectivité d'un patch forestier entre 1900 et 1850 • $\Delta Surf_{2010-1850}$ %Changement de surface d'un patch forestier entre 2010 et 1850 • $\Delta Con_{2010-1850}$ %Changement de connectivité d'un patch forestier entre 2010 et 1850 	

Tableau 10-1. Récapitulatif des variables explicatives générées à partir des différentes bases de données produites

10.2.2 LE CHOIX DU MODELE STATISTIQUE

L'utilisation de modèles en écologie permet d'obtenir une relation mathématique simplifiée d'un processus. En d'autres termes, on cherche à voir si les variables mesurées ont des effets ou non sur la diversité des espèces échantillonnées. Deux grandes familles de modèles sont généralement utilisées pour étudier la distribution d'une espèce : les modèles mécanistes et les modèles corrélatifs (ou empiriques ou statistiques).

Les modèles mécanistes cherchent à expliquer un phénomène à partir de connaissances variées sur divers fonctionnement. Ces données peuvent porter sur la physiologie des espèces, la phénologie ou encore les mécanismes de reproduction des espèces, etc. La combinaison de ces différents paramètres d'entrée permet ainsi de projeter de nouveaux patrons de distribution de l'espèce étudiée. Plus la base de connaissance en entrée est importante, plus le patron de distribution prédit est fiable. Ils ont l'avantage de prendre en compte les processus biologiques sous-jacents et sont réputées plus précises que les modèles corrélatifs. Toutefois, la base de connaissances est souvent partielle et manque de consistance : la fiabilité du patron de distribution dépend souvent du nombre de facteurs.

Les modèles corrélatifs s'appuient sur un ensemble de variables structurelles et environnementales explicatives pour expliquer la présence (ou présence/absence ou abondance) d'une espèce à l'échelle d'un site d'échantillonnage. La fonction de relation dérivée par le modèle peut ensuite être potentiellement utilisée pour prédire un patron de distribution de l'espèce à une

échelle qui va au-delà du site d'échantillonnage. Les modèles corrélatifs sont relativement faciles à paramétrer car ils nécessitent moins de données mais sont jugées moins précises que les mécanistes.

Au regard des données dont nous disposons, notre choix s'est automatiquement tourné vers le second type de modèle. Nous souhaitons évaluer les effets de variables paysagères actuelles et historiques (variables explicatives) sur une richesse actuelle de syrphes forestiers ayant été observée (ou mesurée) au sein d'un site d'échantillonnage (variable réponse).

A noter que nous n'avons pas pour objectif premier de prédire une distribution de l'espèce à une échelle plus large (modèle prédictif) mais simplement d'expliquer les processus qui influent sur la distribution de l'espèce (modèles explicatifs). Ainsi, un bon modèle prédictif n'est pas forcément un bon modèle explicatif et vice versa (Guisan et Zimmermann 2000). Dans le premier cas, on cherche à maximiser le pouvoir prédictif du modèle au dépend de la compréhension des processus alors que dans le second cas, c'est la compréhension du système qui prime.

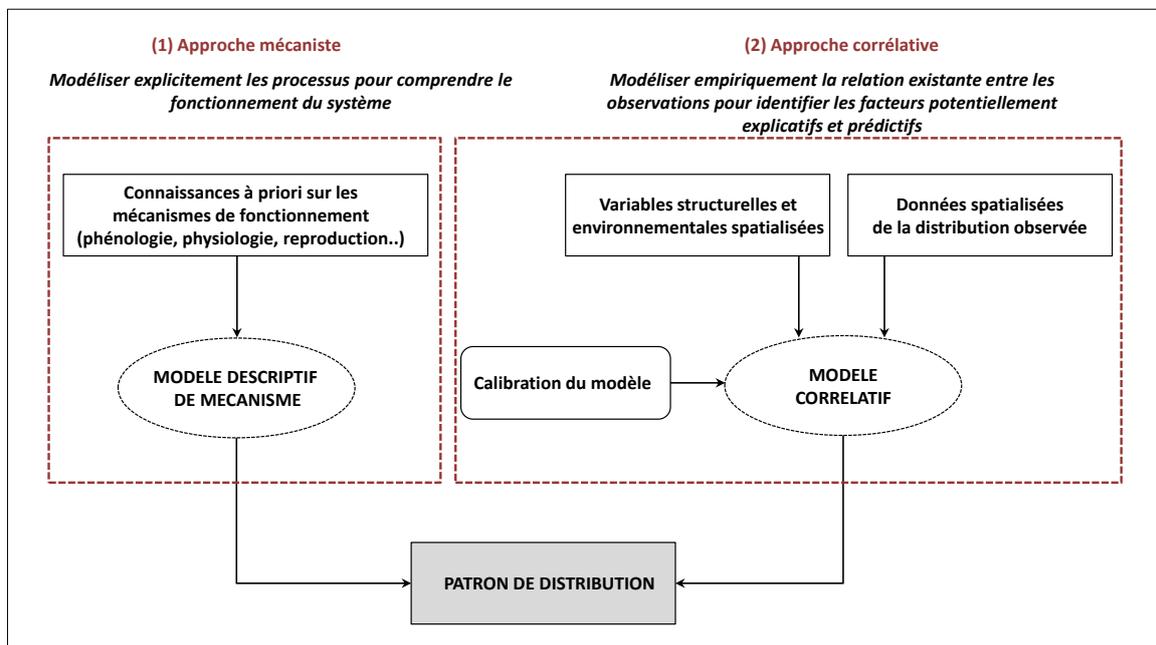


Figure 10-3. Illustration des deux approches de modélisation permettant d'étudier la distribution des espèces (source : https://www.eccorev.fr/IMG/pdf/leriche_mod2010.pdf, adapté)

L'approche la plus courante en écologie est de s'appuyer sur les modèles corrélatifs pour déterminer les liens entre les relevés de présence/absence (ou abondance, ou richesse) et des variables explicatives qui caractérisent les sites échantillonnés (Segurado et Araujo 2004; Elith et Leathwick 2009; Franklin et al. 2009). Les modèles de régression sont les plus répandus et les plus simples à interpréter (Elith et Leathwick 2009). Ils consistent à expliquer la présence ou la diversité d'une espèce étudiée par une somme pondérée des variables explicatives (cas d'un modèle multi-variés). Les modèles de régression englobent à la fois les modèles linéaires (Seal 1967), les modèles linéaires généralisés (Wedderburn 1974), les modèles additifs généralisés (Hastie et Tibshirani 1986) ou leur dérivés comme les modèles incluant des effets aléatoires ou non (Laird et Ware 1982; Breslow et Clayton 1993).

Parmi tous ces modèles, les GLM sont bien adaptés à notre objectif. Ils sont basés sur une variable réponse (l'occurrence des espèces), des variables explicatives (paysagères, climatiques, topographiques, etc.) et une fonction de lien qui décrit la relation entre la valeur attendue de la réponse et les variables explicatives. Ces modèles ont l'avantage de modéliser des relations paramétriques relativement complexes entre les deux types de variables (réponse et explicatives) et de tester les interactions qui existent entre elles (Leprieur et Rubin 2011).

10.2.3 LA PROCEDURE DE MODELISATION

La mise en œuvre d'un modèle linéaire généralisé (GLM) requiert d'adopter une démarche en plusieurs étapes visant à vérifier un certain nombre de postulats avant et après ajustement du modèle (Figure 10-4) (Wedderburn 1974)

Les vérifications pré-ajustement concernent principalement les effets de corrélation entre variables susceptibles d'introduire des biais dans n'importe quel modèle statistique (Gelman et al. 2001). Les vérifications post-ajustement portent sur deux points fondamentaux de la régression linéaire : l'homoscédasticité (i.e. l'homogénéité de la variance des résidus) et la qualité de l'ajustement (i.e. la sur-dispersion ou la sous-dispersion). Si ces conditions ne sont pas vérifiées, on doit alors procéder à un certain nombre de modifications pour corriger ces aspects.

Nous présentons ci-après les différentes étapes d'une procédure de modélisation basée sur l'utilisation des modèles linéaires généralisés (GLM) d'une manière générale. La mise en œuvre pour notre cas d'étude sera décrite plus loin dans ce travail.

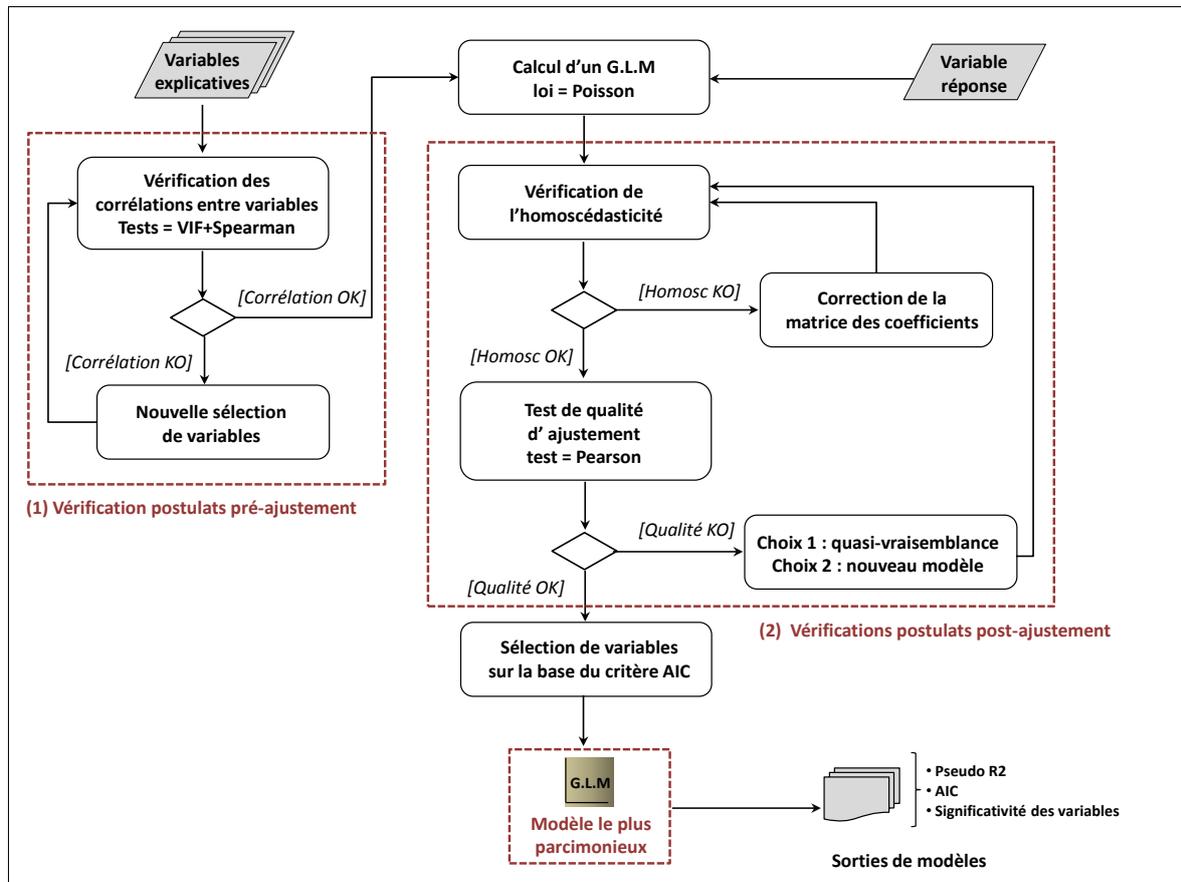


Figure 10-4. Schéma méthodologique d'une démarche modélisation basée sur l'emploi d'un modèle linéaire généralisé (GLM)

10.2.3.1. VERIFICATION DES CORRELATIONS ENTRE VARIABLES EXPLICATIVES (PRE-AJUSTEMENT)

En écologie, il arrive régulièrement que les variables explicatives générées pour expliquer la distribution d'une espèce aient été mesurées sur les mêmes unités géographiques ou les mêmes individus. Ces variables peuvent alors souffrir de corrélation entre elles. Les conséquences de la corrélation entre les variables explicatives dans un modèle linéaire sont les suivantes (Le Rest 2013) :

- les coefficients de régression estimés peuvent être surévalués
- le signe des coefficients de régression peut être contraire à l'intuition
- la variance des estimateurs peut être élevée.

La corrélation est un problème bien connu dans l'application des modèles corrélatifs (F Dormann et al. 2007). Par exemple, le fait qu'un signe de régression soit modifié par la corrélation peut être particulièrement gênant pour étudier l'effet propre d'une variable X_i sur Y .

De nombreux tests existent pour vérifier la corrélation entre variables explicatives comme ceux de Pearson lorsqu'au moins une des deux variables suit une loi normale (test paramétrique) ou de Spearman lorsque les variables ne suivent pas une loi normale (test non-paramétrique). Ces tests s'effectuent deux à deux et sont très faciles à mettre en œuvre. Il est admis que des variables corrélées à moins de 70 % peuvent être intégrées dans un même modèle sans que cela ne puisse avoir d'incidences majeures sur les coefficients d'estimation du modèle (F Dormann et al. 2007).

Une autre approche consiste à effectuer un test de multi collinéarité des variables en allant plus loin que de simples corrélations deux à deux. Le VIF (Variance Inflation Factor) mesure à quel point une variable explicative peut être expliquée par les autres variables explicatives (Zuur et al. 2010). Il est écrit tel que :

$$VIF = \frac{1}{1 - R^2}$$

Des valeurs de VIF inférieures à 2 suggèrent que les variables peuvent être intégrées conjointement dans un modèle sans altérer son fonctionnement (Zuur et al. 2010a).

Dans nos expérimentations, nous verrons que nous avons utilisé les deux approches. Premièrement, nous avons réalisé un test de Spearman entre les variables explicatives (deux à deux) puis avons calculé un VIF. Lorsque les conditions n'étaient pas respectées, nous avons retiré des variables explicatives du modèle afin qu'il remplisse toutes les conditions.

10.2.3.2. CALCUL D'UN GLM DE TYPE « POISSON »

Nous avons choisi d'utiliser un modèle linéaire généralisé (GLM). Étant donné que l'hypothèse de normalité n'est pas respectée pour les données de présence d'espèces, l'utilisation d'un GLM permet de palier ce problème. Celui-ci généralise la régression linéaire en permettant aux variables explicatives d'être reliées à la variable réponse via une fonction de lien. Cette fonction de lien donne la forme de la relation entre l'espérance de la variable réponse et les variables explicatives. Il conserve toutefois la forme de base du modèle linéaire simple, ce qui facilite son interprétation (cf. Wedderburn, 1974).

Il peut être écrit sous la forme :

$$g(E(y)) = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

où g est la fonction de lien et E l'espérance de la variable réponse y . Les b_i sont les coefficients de régression des variables explicatives X_i (i variant de 1 à n).

L'estimation des paramètres du GLM est effectuée non pas par la méthode des moindres carrés comme pour les modèles linéaires simples mais par la méthode du maximum de vraisemblance. Cette méthode d'estimation mesure la probabilité d'obtenir les données observées avec le modèle considéré. Son calcul dépend par conséquent de la famille de distribution choisie. Dans notre cas, notre variable réponse est une variable de richesse spécifique dérivée d'un comptage. Nous avons choisi d'utiliser une fonction de lien de type "log" de façon à suivre une loi de Poisson pour la distribution des erreurs, cette loi étant préconisée lorsque la variable réponse correspond à des données de comptage (ce qui est le cas de la richesse spécifique).

Enfin, si les variables explicatives ont un effet indépendant sur la variable réponse, il existe aussi la possibilité que ces effets ne soient pas constants mais varient en fonction de la valeur prise par une des autres variables indépendantes introduites dans le modèle. Ainsi, l'effet d'une variable X_1 peut dépendre de la valeur prise par X_2 : on dit alors qu'il y a une interaction entre X_1 et X_2 . Cette interaction peut être introduite comme un terme unique dans le modèle établi généralement sous la forme $X_1 : X_2$.

10.2.3.3. VERIFICATION DE L'HOMOSCEDASTICITE (POST-AJUSTEMENT)

A l'issue de la modélisation GLM, d'autres postulats doivent être vérifiées. Un problème récurrent est l'existence d'une relation entre la variance des résidus et la valeur des prédictions : la variance des résidus a tendance à croître avec les prédictions. Pour vérifier ce principe, il est possible d'effectuer une observation graphique des résidus en fonctions des valeurs prédites. Les écarts ne doivent pas croître en fonction des valeurs prédites mais doivent toujours rester du même ordre. On cherche en réalité à vérifier l'homogénéité de la variance des résidus.

Si une relation apparente se présente entre les résidus et les valeurs prédites par le modèle alors les estimations peuvent être biaisées et il ne devient plus possible d'utiliser les tests hypothèses usuels post-estimation (Student, Fisher...). Plusieurs méthodes ont été développées pour corriger la matrice variance-covariance en effectuant des régressions plus robustes. La plus répandue étant la correction par la matrice de White qui consiste à ajuster les écarts types des coefficients de régression (White 1980).

10.2.3.4. ESTIMATION DE QUALITE DE L'AJUSTEMENT (POST-AJUSTEMENT)

Une fois les paramètres estimés par le modèle, il s'agit d'évaluer la qualité de l'ajustement du modèle sur la base des différences entre les observations et les estimations. En d'autres termes, on cherche à vérifier la sur-dispersion du modèle. Nous n'aborderons pas le problème de la sous-dispersion car elle n'apparaît que très rarement en écologie sous certaines conditions particulières. La déviance expliquée et le X^2 généralisé de Pearson sont les deux outils principaux pour évaluer et comparer les modèles logistiques (Wedderburn 1974).

Ces deux mesures sont réputées fournir des résultats peu différents, nous avons par conséquent choisi d'utiliser le second test.

La statistique du test de Pearson est défini tel que :

$$X^2 = \sum_{i=1}^I \frac{(y_i - \hat{\mu}_i)^2}{\widehat{Var}(\hat{\mu}_i)}$$

où y_i désigne les observations et $\hat{\mu}_i$ les estimations.

Etant donné que l'espérance d'une loi du X^2 est son nombre de degrés de libertés (ddl), il est souvent préconisé de comparer les statistiques avec le nombre de degrés de libertés. Le modèle est jugé satisfaisant pour un rapport X^2/ddl plus petit que 1 (Lancelot et Lesnoff 2005). Si le test renvoie une valeur supérieure à 1 alors il faut tenir compte de la sur dispersion. Soit on modifie la relation entre la moyenne et la variance sans pour autant spécifier une autre distribution, ce qui conduit à appliquer une quasi-vraisemblance. Dans notre cas, cela pourrait être une distribution quasi-poisson. Soit on utilise un autre modèle de régression et une nouvelle forme de distribution comme la régression binomiale négative ou béta-binomiale par exemple (Chatfield et Goodhardt 1976).

10.2.3.5. SELECTION DE VARIABLES

Dans un modèle de régression, on cherche bien souvent à trouver un compromis entre le nombre de variables à utiliser et la qualité de l'ajustement du modèle aux données. Plus précisément, on cherche un équilibre entre le biais et la variance en contrôlant l'ajustement et la complexité du modèle.

Un modèle complexe est un modèle qui intègre un gr et nombre de variables explicatives. En général, ces modèles présentent une erreur d'ajustement faible. L'écart aux données est réduit. On dit que le biais est faible. En revanche, ce sont des modèles qui prédisent généralement moins bien. Ils présentent un sur-ajustement aux observations ce qui augmente l'erreur de prédiction. On dit que la variance de ces modèles est élevée car ils sont sensibles aux observations (ou base d'apprentissage).

On peut dès lors chercher à diminuer le nombre de variables explicatives pour augmenter la capacité de généralisation du modèle (i.e. réduire sa variance). Néanmoins, en simplifiant le modèle, on augmente le biais ce qui diminue la qualité d'ajustement (ou dit autrement, augmente l'erreur d'apprentissage).

Par conséquent, lorsqu'on construit un modèle, on va chercher à définir un modèle le plus parcimonieux, qui présente un biais et une variance aussi faibles que possibles, résultant d'un compromis.

Pour trouver ce modèle le plus parcimonieux, plusieurs critères de parcimonie ont été développés (pour une revue de ces critères, voir (George 2000)). En écologie, le critère le plus utilisé est l'*Aikake Information criteria* (Akaike 1981; Akaike 1987). L'AIC cherche à définir un modèle de dimension (complexité) raisonnable, parcimonieux, permettant de répondre à un certain nombre de questions à partir de l'ensemble des données disponibles (Shibata 1981; Le Rest 2013). Ce dernier est composé de deux parties. L'une vise à évaluer la qualité de l'ajustement (par exemple la déviance = $-2 \cdot \log$ -vraisemblance) et l'autre est une pénalité proportionnelle à la complexité du modèle.

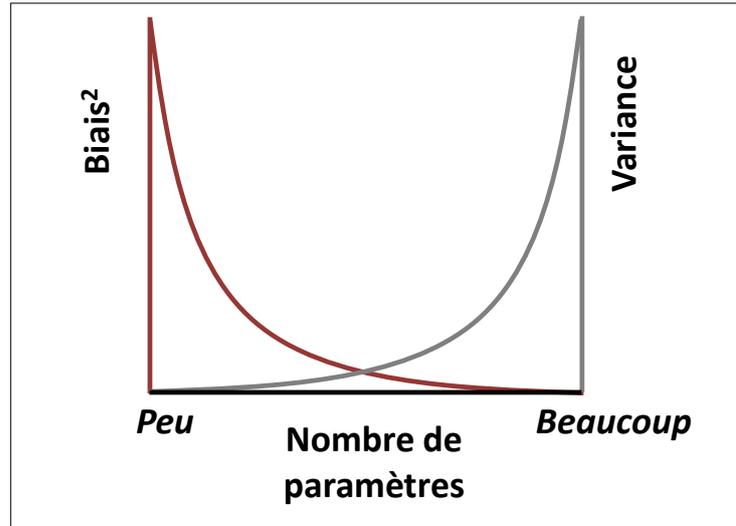


Figure 10-5 Illustration du principe de parcimonie (Burnham et Anderson 2002)

D'un point de vue pratique, la sélection de variables est effectuée étapes par étapes en testant toutes les combinaisons de variables explicatives possibles. Soit du haut vers le bas (*forward*) qui consiste à évaluer le modèle plus complexe vers le moins complexe ou bien du bas vers le haut (*backward*), c'est-à-dire du modèle le moins complexe vers le plus complexe.

Nous avons choisi d'appliquer la seconde approche (*backward*) et avons sélectionné le modèle qui minimisait le critère AIC.

10.2.3.6. ANALYSE DES SORTIES DE MODELE

Dans le but d'estimer la pertinence des modèles construits et la consistance des relations mises en évidence, nous nous sommes appuyés sur trois types d'indicateurs.

Premièrement, nous souhaitons évaluer la qualité de l'ajustement du modèle construit, c'est à dire son pouvoir explicatif. Etant donné que l'estimation des paramètres dans un modèle GLM choisi n'est pas effectuée par la méthode des moindres-carrés, il n'est pas possible de calculer un R² classique (appelé aussi D²). Plusieurs alternatives existent cependant pour dériver une mesure de pseudo R² : soit en comparant la vraisemblance du modèle étudié à celle du modèle nul² (McFadden 1973; Nagelkerke 1991), soit en comparant la déviance du modèle étudié à celle du modèle nul (Guisan et Zimmerman, 2000).

Nous avons opté pour le Pseudo R² basé sur la déviance (Guisan et Zimmermann 2000). Il est très proche conceptuellement du coefficient de détermination calculé dans une régression linéaire multiple : si Pseudo R² = 0, les variables explicatives n'expliquent pas la variable réponse (ou le modèle n'ajuste pas les observations); si Pseudo R² = 1, les variables explicatives expliquent totalement la variable réponse. Il est aussi possible d'obtenir une version ajustée de cet indicateur pour procéder à des comparaisons de modèles.

$$\text{Pseudo R}^2 \text{ (Guisan et Zimmerman, 2000)} = \frac{(\text{Déviance nulle} - \text{Déviance résiduelle})}{\text{Déviance nulle}}$$

¹ Coefficient de détermination d'une régression simple ou multiple. Il mesure la proportion de variance expliquée par la ou les variables explicatives.

² Modèle sans effet des variables. Seul l'intercept est considéré.

$$\text{Pseudo } R^2 \text{ ajusté (Guisan et Zimmerman, 2000)} = 1 - [(n - 1)/(n - p)] * (1 - R^2)$$

Où n est le nombre d'observations et p le nombre de paramètres pour le modèle à estimer.

Enfin, la démarche de modélisation que nous décrivons ci-après nous a conduit à établir des comparaisons entre différents modèles construits. Si le pseudo R^2 ajusté peut être utilisé, aucun seuil de différence n'est établi dans le but de se prononcer sur une différence significative entre deux modèles. En revanche, une différence supérieure à 2 points entre deux AIC permet de conclure que deux modèles sont différents (Burnham et Anderson 2002). Nous nous sommes par conséquent appuyé sur ce dernier critère pour évaluer les différences entre les modèles établis.

10.2.4 DEMARCHE DE MODELISATION

Nous décrivons ici la démarche de modélisation visant à évaluer les effets de l'histoire des patchs forestiers sur la richesse actuelle de syrphes. Comme évoqué précédemment (section XX), nous disposons de variables explicatives d'état et d'évolution. Celles-ci interviennent à l'échelle des objets et de leur contexte. De plus, le site LTER des coteaux de Gascogne fait l'objet depuis de nombreuses années d'un certain nombre de relevés de différente nature. Grâce à eux, il nous est donc possible de compléter notre lot de variables explicatives par des variables d'hétérogénéité interne de peuplement récoltées à l'échelle des pièges posés au sein de nos bois échantillonnées. Ces mesures ont été conduites par Laurent Larrieu dans le cadre de la campagne effectuée pour la définition de l'Indice Potentielle de Biodiversité (IBP) des forêts des coteaux de Gascogne.

L'intérêt d'alimenter notre jeu de données avec ce type de variable repose sur le fait que nous pourrions disposer d'une vision globale des effets des trois facteurs majeurs susceptibles de faire varier la biodiversité forestière : l'hétérogénéité interne du bois, sa morphologie et son isolement (chapitre 2, section 2.2). De plus, il est possible que ces trois échelles d'analyse soient complémentaires et qu'analyser les effets de l'une aidera à l'interprétation des autres effets sur la richesse actuelle de syrphes forestiers. L'objectif de ce travail vise par conséquent à évaluer les effets combinés de l'histoire, de l'hétérogénéité, de la morphologie et de l'isolement des patchs forestiers sur la richesse spécifique actuelle de syrphes forestiers dans les coteaux de Gascogne.

10.3. UNE ETUDE PUBLIEE DANS LANDSCAPE ECOLOGY JOURNAL

La démarche de modélisation détaillée ci-dessus a fait l'objet d'un travail collaboratif impliquant plusieurs membres du laboratoire Dynafor et du Swiss Federal Research Institute (WSL) où j'ai effectué un séjour de trois mois durant ma seconde année de doctorat. Nous présentons donc les résultats de cette étude sous la forme d'un article soumis à Landscape Ecology en version révisée.

Combined effects of area, connectivity, history and structural heterogeneity of woodlots on the species richness of hoverflies (Diptera: Syrphidae)

Herrault P.-A., Larrieu L., Cordier S., Gimmi U., Lachat T., Ouin A., Sarthou J.-P., Sheeren D.

Herrault P.-A. (corresponding author)

-University of Toulouse, INP-ENSAT, UMR 1201 DYNAFOR, Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, Castanet Tolosan cedex, Toulouse, 31326, France

-University of Toulouse, UTM, UMR 5602 GEODE, 5, allée A. Machado, Toulouse, 31058, France

Email: pierrealexis.herrault@ensat.fr

Tel: 05 34 32 39 26

Larrieu L.

-INRA, UMR1201 DYNAFOR, Chemin de Borde Rouge, Auzeville Tolosane, CS 52627, 31326 Castanet Tolosan Cedex, France

-CNPF/IDF, Antenne de Toulouse, 7 chemin de la Lacade, 31320 Auzeville Tolosane, France

Cordier S.

-Université of Liège-Gembloux Agro-Bio Tech, B-5030, Gembloux, Belgique.

Gimmi U., Lachat T.

-Swiss Federal Research Institute WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

Ouin A., Sheeren D.

-University of Toulouse, INP-ENSAT, UMR 1201 DYNAFOR, Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, Castanet Tolosan cedex, Toulouse, 31326, France

Sarthou J.-P.

-University of Toulouse, INP-ENSAT, UMR 1248 AGIR, Av. de l'Agrobiopôle, BP 32607, Auzeville Tolosane, Castanet Tolosan cedex, Toulouse, 31326, France

Abstract

Hoverflies (Diptera Syrphidae) play a crucial role in ecosystem functioning and are often chosen as bioindicators in evaluations of ecosystem conservation. However, little work has been undertaken to explore the relative importance of local and landscape-scale factors as well as historical changes in habitat to explain species patterns. In this study, the joint effects of woodlot area, structural heterogeneity, connectivity and history on the species richness of forest hoverflies were investigated. It was also evaluated whether hoverflies exhibit a delayed response to habitat changes, and tests were undertaken on the effect of considering spatiotemporal changes in explaining species assemblages. The current species richness of forest specialists was sampled in 48 woodlots using 99 Malaise traps. The structural explanatory variables were derived from a rapid habitat assessment protocol. Old maps and aerial photographs were used to extract past and present spatial patterns of the woodlots. Generalised linear models showed that the current habitat area, connectivity, historical continuity and average density of microhabitats played a significant part in explaining species richness ($D^2=35.57\%$). It was also observed that hoverfly diversity was affected differently by change in patch area between 1979 and 2010, depending on woodlot connectivity. For isolated woodlots, coexistence of extinction debt and colonization credit was highlighted revealing that 30 years are not enough for assemblages of forest to reach new equilibrium. Our results also showed the interest of adopting a change-oriented approach to explain the current distribution pattern of species.

Keywords: hoverflies, woodlands, heterogeneity, connectivity, history, extinction debt, colonization credit.

1. Introduction

Forests are among the most biologically diverse ecosystems on the planet (Gosselin and Laroussinie 2004). Species richness is influenced by broad spatial-scale processes, forest area, habitat heterogeneity, history of disturbances and resource continuity, stochastic events which can lead to local extinction and connectivity which allows recolonisation, inter-specific competition, etc. (Ricklefs 1987; Huston, 1994). Despite the multitude of these factors, most of the studies on biodiversity in woodlands focus on just one of them or on just on the local scale. Therefore, quantifying the relative importance of each factor and of the landscape on local species richness still presents a challenge.

Since forests have been exploited in Western Europe for thousands of years, most of disturbances are anthropogenic, and human activities such as deforestation or timber harvesting greatly affect species diversity (Larsson 2001). Habitat loss and fragmentation are currently the primary causes of biodiversity decline (Fahrig 2003; Foley et al. 2005; Hanski 2005). Local extinction of populations might not be reversed by colonisation because an increase in isolation and reduction of habitat size could lead to species extinction. On the opposite, the creation of new habitat patches or an increase in connectivity among habitats could conducive to species immigration. Some species react immediately to landscape changes while others express a time-lagged response (Hanski and Ovaskainen 2002; Lindborg and Eriksson 2004; Metzger et al. 2009; Krauss et al. 2010). This time delay in the response of certain species may lead to what is termed an extinction debt (Tilman et al. 1994; Cousins 2009; Kuussaari et al. 2009) or a colonization credit (Cristofoli et al. 2010; Piqueray et al. 2011).

Extinction debts have been observed for vascular plants in temperate regions (Helm et al. 2006b; Krauss et al. 2010; Guardiola et al. 2013) but also for other species with higher dispersal ability, such as pollinating insects (butterflies, hoverflies) or birds (Brooks et al. 1999; Ford et al. 2009). Colonization credits were also detected for wet heathland butterfly communities (Cristofoli and Mahy 2010) and plant communities in calcareous grasslands (Piqueray et al. 2011). Even though the mechanisms involved in the delayed responses of species are still unclear, specialist species are expected to be more affected by habitat changes than generalist species (Cousins and Vanhoenacker 2011b).

Empirical evidence of extinction debt is often based on the comparison of past and present habitat characteristics (Kuussaari et al. 2009b). When past habitat variables explain current species richness better than the present ones, an extinction debt is assumed (Krauss et al. 2010b). Although this comparison approach enables to detect a potential extinction debt, "past" and "current" habitat characteristics are disconnected because the state of habitats is described at one time without integrating patch history. Consequently, the temporal trajectory of habitats is mostly ignored. For colonization credit, the estimation is often based from species-area relationships in equilibrium habitat patches (Kuussaari et al. 2009b). A colonization credit is assumed when observed species richness in new habitats is lower than expected according to the habitat spatial characteristics, some species yet to colonise it. This could be evaluated by defining groups of old and new patches and by comparing their species-habitat relationships, assuming that the old ones are closest to the equilibrium (Cristofoli and Mahy 2010).

Recently, some authors adopted a dynamic approach in the detection of extinction debt or the analysis of historical factors controlling the pattern of species richness (Metzger et al. 2009b), considering habitat change as a cumulative process (Ewers et al. 2013b). Metzger et al. (2009) showed that including the rate of change of habitat area (i.e. the rate of relative change between two dates) and connectivity for several taxonomic groups (tree, bird and frog species), in addition to the present habitat characteristics, enabled to strongly improve the explanation of species richness and abundance. The rate of landscape change could influence the balance between colonization and extinction (Münzbergová et al. 2005).

Ouin et al. (2006) previously investigated the effects of landscape variables on the presence of hoverflies (Syrphidae) in a fragmented landscape of southwest France and highlighted that the current woodlot area explained roughly 30% of the variability in the species richness. Hoverflies are known to be quite mobile and are therefore sensitive to conditions on large scales (Smith et al. 2008). This family encompasses a species diverse group in terms of trophic and habitat requirements, including saproxylic species, and has been used as an indicator of disturbance or habitat quality (Sommaggio 1999). They are also used as bioindicators to identify forests for potential conservation (Good and Speight 1996). Larvae of saproxylic species participate in deadwood recycling (Speight, 1989), and adults of all the species are flower-visiting insects and probably provide pollination services for several trees and shrubs (Groot & Bevk, 2012).

Despite the crucial role of hoverflies in ecosystem functioning, little is known about the various factors that influence their contemporary patterns (Keil and Konvicka 2005). It might therefore be useful to know more about this to reinforce their role as bioindicators (Sommaggio 1999). In this paper, investigations were undertaken in the same fragmented woodlots as those in Ouin et al. (2006) on the combined effects of current landscape variables, their past dynamics and local structural heterogeneity variables on the species richness of hoverflies. Firstly, an assessment was undertaken of the relative importance of area, connectivity, structural heterogeneity and history of woodlots in controlling the species richness of forest specialists. A species-area relationship was expected with a possible lower richness in less connected woodlots. However, due to high dispersal abilities of hoverflies, a weak effect of connectivity was presumed (Ouin et al. 2006). It was also assumed that the presence of tree microhabitats (hereafter called microhabitats) and deadwood could offer favourable habitats for certain species, namely saproxylic species, depending on the historical continuity. Thus, younger woodlots could have less species than older woodlots especially for isolated ones. Evaluations were then undertaken as to whether hoverfly communities are subject to extinction debt or colonisation credit in this context. Little attention has been given to this taxonomic group even though a recent study showed that hoverflies can be affected by extinction debt: richness of specialist and generalist species was best predicted by past area (Bommarco et al. 2014). Here, we considered historical changes in area and connectivity to reveal indications of extinction debt and colonisation credit. If hoverflies react immediately to landscape changes, a positive relationship between species richness and changes in spatial properties (from negative to positive) was expected. In case of time-lagged response, extinction debt in patches with decreased area would lead to more species being present than expected while colonisation credit would lead to less species than expected in patches with increased area. Thus, a negative effect of changes on species richness could be expected. This relationship might be modulated by changes in connectivity, the number of species resulting from a colonisation-extinction dynamics (MacArthur, 1967).

2. Materials and methods

Study area

The study site is located in southwest France (43°16'28"N, 0°51'11" E, WGS-84) and covers roughly 30,000 ha, including the Long-Term Ecological Research site "Vallées et Coteaux de Gascogne" (LTER_EU_FR_003) (Fig. 1). This is a hilly region (altitude 215-400 m), dissected by north-south valleys. The climate is temperate with Pyrenean Mountain and slight Mediterranean influences. Wood cover is fragmented and covers some 15 % of the area. The dominant tree species are sessile oak (*Quercus petraea* Mattus), pedunculate oak (*Quercus robur* L.) and pubescent oak (*Quercus pubescens* Willd.) for more than 80 % of the woodland area (IFN, 2000).

In this region, small private forests are the most frequent type of woodlands. Owing to the absence of a forest management plan for approximately 90 % of the area, there is high spatial heterogeneity between woodlands. Owners generally only have empirical knowledge of forest management, which

may strongly influence the spatial patterns of forest biodiversity (De Warnaffe et al. 2006; etrieu et al. 2008).



Fig.1 Location of the 48 woodlots sampled in the study site in southwest France.

Data collection

Biological sampling

Hoverfly (Diptera: Syrphidae) sampling was conducted by Ouin et al. (2006) using Malaise traps during two one-month periods: 10 May – 10 June (spring) and 13 September – 13 October 2000 (autumn). Collection bottles were replaced every fortnight in spring and every month in autumn. 48 woodlands were selected from Ouin et al. (2006) according to a gradient of surface area, isolation and type of management, resulting in different forest structures and compositions. The trap number per woodlot was adjusted to the patch area in order to obtain an equivalent trapping intensity for each woodlot (Table 1). The large majority of small woodlots (< 5ha) contained only one trap. For larger woodlots, more traps were added. A total of 99 Malaise traps were installed in the surveyed woodlots.

Class of surface area (A)	No. of traps	No. of woodlots in the class	Average forest species richness per wood
A < 5ha	1	29	2.48 ± 1.68
5 < A < 10ha	2	9	3.77 ± 1.74
10 < A < 15ha	3	3	2.45 ± 1.37
15 < A < 25ha	4	4	2.48 ± 2.10
45 < A < 60ha	7	2	3.6 ± 0.98
150 < A < 200ha	14	1	5.67

Table 1. For each woodlot area class: number of traps, number of woodlots in the class, average forest species richness per wood.

Structural heterogeneity variables of woodlands

Structural variables were surveyed during summer 2011 using a rapid stand description protocol (Larrieu and Gonin 2008) in a 1-ha circular plot centred around each trap. 99 hectares were therefore sampled. Eight attributes of forest structure and composition were observed that were assumed to be relevant to hoverfly assemblages: (1) the number of indigenous tree species; (2) the number of large standing deadwood (diameter at breast height (dbh) > 40cm) including whole dead trees, snags or stumps taller than 1.5 m; (3) the number of large lying deadwood (diameter > 40cm and length > 1m); (4) the number of very large living trees (dbh > 70cm); (5) the number of microhabitat-bearing trees (only living trees), a tree being counted once for each type of microhabitat carried, namely empty cavities, cavities with mould, sporophores of saproxylic fungi, dendrothelms, missing bark, cracks, broken crowns, large amount (> 20 %) of deadwood in the crown; (6) a diversity index equal to the sum of microhabitat-bearing trees when considering a maximum of two trees bearing the same microhabitat type per plot; (7) the proportion (%) of open areas (clearings, edges, areas with a well-developed herb layer composed of flowering plants); and (8) the number of aquatic habitats (in reference to a list). Finally the average value of each of these variables per trap in one wood was calculated in order to test the effect of structural heterogeneity on the species richness of hoverflies. Because Malaise traps catch adults looking for flower resources (mainly in clearings, edges or woods with low tree-cover), structural variables directly described adult habitats but indirectly larvae habitats, with the hypothesis that adults do not move too away from larvae habitat, i.e. at least stay in the same woodlot.

Area, connectivity and historical variables of woodlands

Aerial colour infrared orthophotos from 2010 (BD Ortho®, French mapping agency IGN) were used to estimate the current area and connectivity of each woodlot. Connectivity was captured in terms of isolation, i.e. as the forest density in the surrounding landscape of one focal patch (Tischendorf et al. 2003; Magle et al. 2009). Forest density was measured as percentage forest cover excluding the focal patch. The buffer was created from the focal patch boundaries with a 4.8 km radius (Fig. 2), which was assumed to be the relevant scale for analysing the effect of isolation on the hoverfly community (Quin et al. 2006). Such a large scale is explained by the high dispersal capacity of hoverflies (Sarthou and Speight 2005). The forest density index showed a strong correlation between area-weighted distance-based metrics of connectivity (e.g. Spearman correlation $Rho = 0.75$, p -value < 0.001 with Hanski's index for the 2010 dataset (Hanski and Thomas 1994; Magle et al. 2009).

Historical spatial data were also used to quantify the effect of changes in woodland area and isolation on current species richness as forest continuity over time. Two historical maps were used in addition to old orthophotos from 1953-54 and 1979 (Fig. 3). The first historical map is the French military 'Etat Major' colour map from the middle of the 19th century at a scale of 1:40,000 (IGN 2011). Sheets we used in this map series are: 241 NE, 241 NO, 229 SE, 229 SO. The second map is a revised version of the first one from 1900 and produced in black and white at a scale of 1:80,000 (Nadal 2011). These historical maps were geo-referenced with a 3rd order polynomial transformation (70 ground control points, root mean square error ~ 55 metres). Woodlands were digitised at 1.7-metre resolution according to a rule base defined by cartographic experts ensuring a homogeneous representation of the objects (Favre et al. 2012). Old aerial black and white photographs from IGN (flight missions of 24/07/1953 and 12/06/1954 at the scale of 1:25,000 and mission of 18/06/1979 at the scale of 1:29,000) were ortho-rectified using a digital terrain model of 25-metre resolution and applying a linear transformation (40 ground control points, RMSE ~ 30 meters). Finally, datasets of woodlots produced for the fifth time period were matched manually in order to build the corresponding relations for related spatial entities and estimate the evolutions.

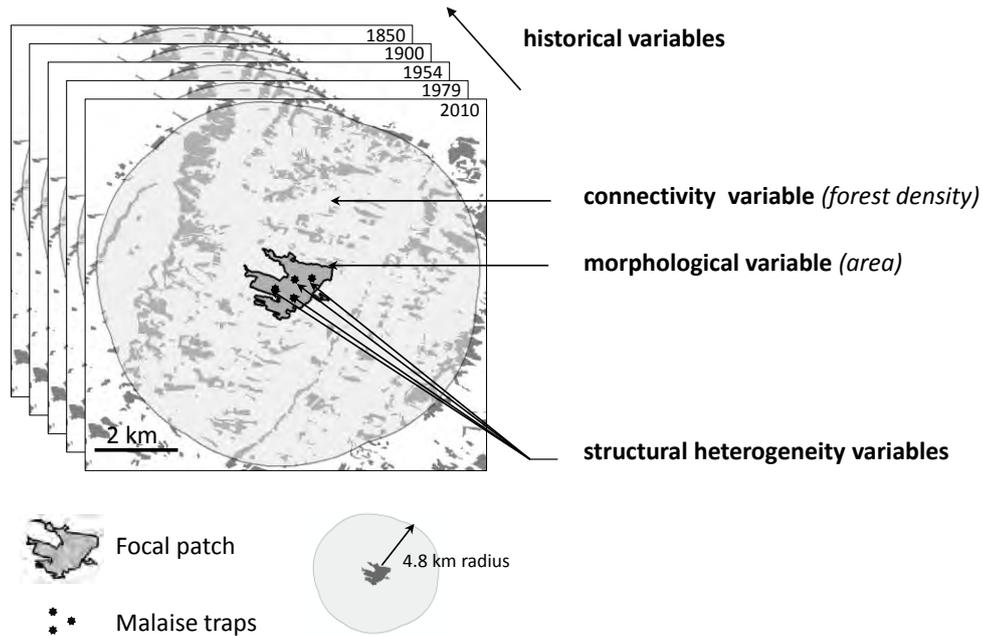


Fig.2 Structural heterogeneity, area, connectivity and historical continuity of woodlots: multiple variables from local to landscape-scale to explain the spatial patterns of hoverfly species richness.

Changes in woodlot area and connectivity were computed by retrospective analysis (i.e. looking backwards). For the current time period t , woodlots were always equal to one single spatial object. However, for past time periods, the current woodland could be matched to several older forest fragments because of the disappearance or aggregation processes from one period to another. In this case, when one-to-many or many-to-many matching relationships appeared between time periods, the fragmented woodlot was viewed as a complex object (or “meta-object”) composed of simple objects. The area of the complex object was defined as the sum of all the areas of the simple objects. For connectivity, the convex hull containing all the simple objects was created to represent the extent of the meta-object. A buffer of 4.8 km radius was formed around this convex hull from its boundaries and the proportion of forest within this neighbourhood was calculated, excluding the convex hull area. Finally, relative changes (in %) in woodlot area and connectivity between each consecutive time periods and between 1850 and 2010 were estimated, in addition to historical area and connectivity at each date. A semi-quantitative variable was also defined to provide information about whether forest continuity had been maintained over time since the first historical map. More precisely, four groups of woodlots were identified: the older ones present in all cartographic data since 1850 (continuity variable value = 1, patches > 160 years old), since 1900 (continuity variable value = 2, patches ~ 110-160 years old), since 1954 (continuity variable value = 3, patches ~ 56-110 years old) and since 1979 (continuity variable value = 4, patches ~ 31-56 years old).

).

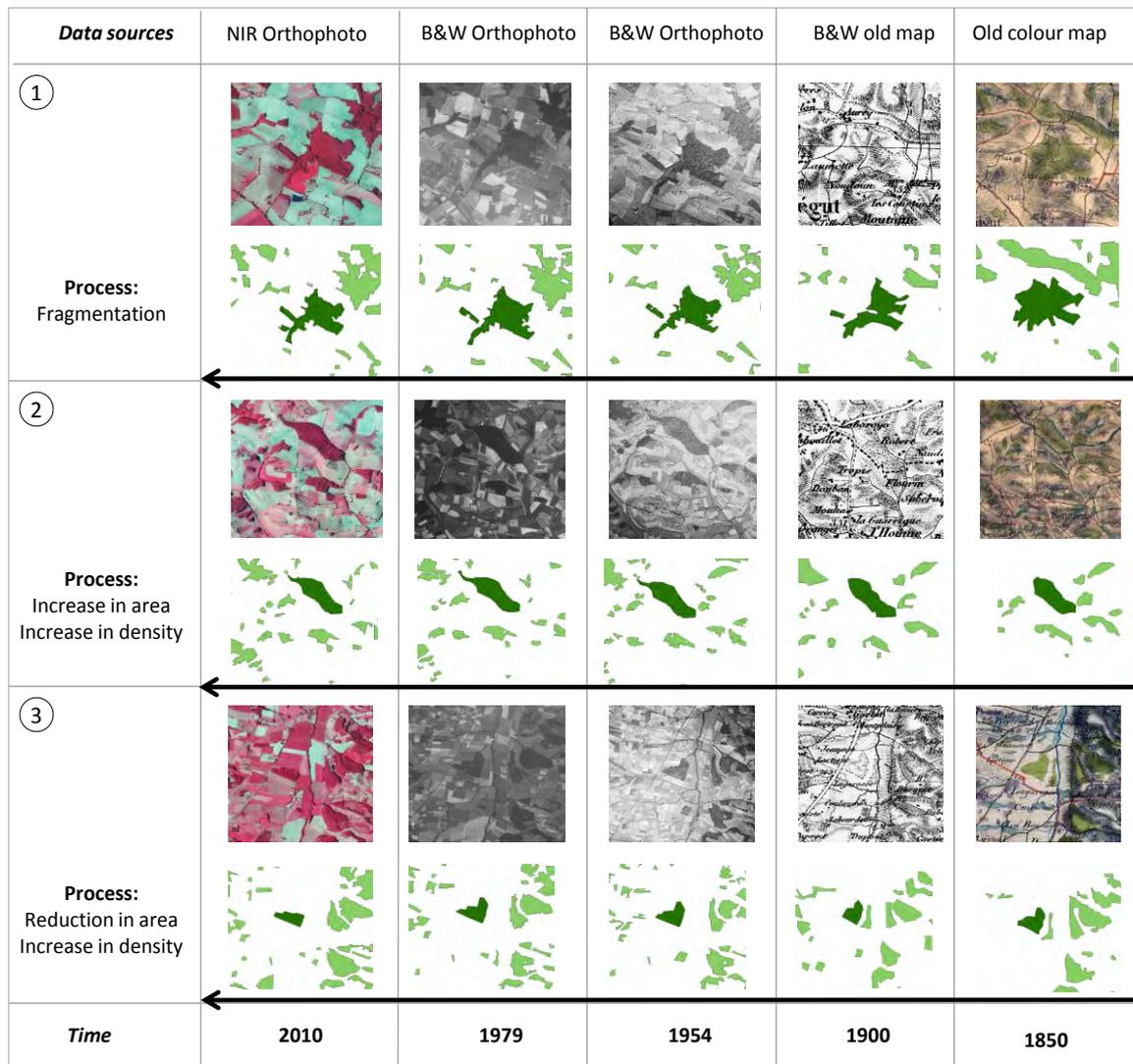


Fig.3 Excerpts from the spatial data sources with examples of forest evolution for some of the woodlots sampled (in dark green).

Statistical analyses

Statistical analyses were performed using R 2.3.0.1 software (R Development Core Team, 2014) and the “vegan” R package (Oksanen et al. 2007).

Because difference in species richness of woodlot may be caused by difference in sampling effort, species accumulation curves were calculated for each woodlot having more than one trap (i.e. with an area > 5ha, n=19) by randomly sampling traps with 100 permutations. Based on the species accumulation curves (Appendix 1), the expected average species richness for each woodlot was estimated to an equivalent sampling effort of one trap. By this way, we removed the potential bias related to the sampling effort before statistical analyses.

Relationships between species richness and explanatory variables were explored using generalised linear models (GLM). Poisson family distribution was used to fit the species richness models. Two main analyses were explored and associated models were tested (Table 2). The contribution was measured of each explanatory variable on the current forest species richness of hoverfly per wood (Model 1). Current area log₁₀-transformed, connectivity and structural heterogeneity variables were incorporated in addition to the historical woodland continuity (thus a total of 11 factors). Interaction

terms between historical continuity and (i) current connectivity, (ii) average number of microhabitats per wood, and (iii) average number of large standing deadwood per wood respectively (thus 3 additional factors) were also tested, with the hypothesis that forest historical continuity influences positively these three structural variables. For the two latter, historical continuity does not necessarily imply maturity (Cateau et al. 2015), but a very recent wood cannot be mature. This first analysis included all the woodlots surveyed (n=48).

Then, the role of past changes in area and connectivity of woods was investigated to explain forest species richness. Firstly, all types of woodlands changes (e.g. reduction of size, increase in isolation) since 1850 were incorporated into the model, in addition to current area and connectivity of woods (Model 2a). The analysis was performed with woodlots existing at each date, i.e. having historical continuity over time (n=28). Secondly, we focused the analysis on the most recent period (1979-2010). The role of current area, change in area during this period 1979-2010 and the effect of interaction between change in area and current connectivity were tested (Model 2b). This analysis included all the woodlots (n=48). Finally, we compared the effect of both area 1979 and area 2010 on the current species richness for isolated woodlots (current connectivity < 15%; n=16) to highlight a potential extinction debt or colonization credit (Models 2c).

Variance inflation factors (VIF) were checked using the 'VIF' R package (Lin 2009) to prevent multicollinearity between covariates. For the first model, all VIF values were below two, suggesting that collinearity of variables did not affect the model (Zuur et al. 2010). For the other models (2a and 2b), variables were more strongly correlated because of the historical change variables (VIF < 6; correlation coefficients $|r| < 0.62$), but the colinearity did not exceed the recommended thresholds of diagnostic indices (VIF < 10; correlation coefficients $|r| < 0.7$) (Dormann et al. 2013).

The most parsimonious models were also assessed on the basis of the Akaike information criterion (AIC) (Burnham and Anderson). A stepwise backward selection of variables was carried out to select the "best" model with the lowest AIC. Variance partitioning was also calculated from the best model to measure the independent contribution of each variable to the explained variance of the response variable.

The goodness of fit of the models was quantified by examining the amount of the explained deviance (D^2) and the adjusted explained deviance (% *adj-D*²), which takes into account the number of observations and explanatory variables (Guisan and Zimmermann 2000). This adjusted measure enables to compare models that include different combinations of variables.

Explanatory variables	Description	
<i>Model (1) with present area, connectivity, structural heterogeneity et historical continuity</i>		
Model form: $E(y) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \beta_9 + \beta_{10} + \beta_{11} + \beta_{12} + \beta_{13} + \beta_{14}$		
Log(Area 2010)	β_1	Present area of woodlot patch (orthophoto 2010)
Connect 2010	β_2	Present forest density in a 4.8km buffer radius (orthophoto 2010)
AvgNb_ATS	β_3	Present average number of indigenous tree species per wood
AvgNb_STW	β_4	Present average number of large standing deadwood per wood
AvgNb_LDW	β_5	Present average number of large lying deadwood per wood
AvgNb_VLLT	β_6	Present average number of very large living trees per wood
AvgNb_MH	β_7	Present average number of microhabitats per wood
AvgNb_TreeMH	β_8	Present average number of trees supporting microhabitats per wood
AvgOpen_Area	β_9	Present average size of open areas per wood
AvgNb_Aquatic	β_{10}	Present average number of aquatic habitat per wood
Histo_continuity	β_{11}	Historical forest continuity over time (since 1850 = 1; since 1900 = 2; since 1954 = 3 ; since 1979 = 4)
Histo_continuity:Connect2010	β_{12}	Interaction term between Histo_continuity et Connect2010
Histo_continuity:AvgNb_STW	β_{13}	Interaction term between Histo_continuity et AvgNb_STW
Histo_continuity:AvgNb_MH	β_{14}	Interaction term between Histo_continuity et AvgNb_MH
<i>Model (2a) with present area and connectivity with their past dynamics since 1850</i>		
Model form: $E(y) = \beta_0 + \beta_1 + \beta_2 + \beta_{21} + \beta_{22} + \beta_{23} + \beta_{24} + \beta_{25} + \beta_{26} + \beta_{27} + \beta_{28} + \beta_{29} + \beta_{30}$		
Log(Area 2010)	β_1	Present area of woodlot patch (orthophoto 2010)
Connect 2010	β_2	Present forest density in a 4.8km buffer radius (orthophoto 2010)
Δ Area 2010-1979	β_{21}	Change in area (%) between 2010 et 1979
Δ Con 2010-1979	β_{22}	Change in connectivity (%) between 2010 et 1979
Δ Area 1979-1954	β_{23}	Change in area (%) between 1979 et 1954
Δ Con 1974-1954	β_{24}	Change in connectivity (%) between 1979 et 1954
Δ Area 1954-1900	β_{25}	Change in area between (%) 1954 et 1900
Δ Con 1954-1900	β_{26}	Change in connectivity (%) between 1954 et 1900
Δ Area 1900-1850	β_{27}	Change in area between (%) 1900 et 1850
Δ Con 1900-1850	β_{28}	Change in connectivity (%) between 1900 et 1850
Δ Area 2010-1850	β_{29}	Change in area between (%) 2010 et 1850
Δ Con 2010-1850	β_{30}	Change in connectivity (%) between 2010 et 1850
<i>Model (2b) with present area and the most recent dynamics since 1979</i>		
Model form: $E(y) = \beta_0 + \beta_1 + \beta_2 + \beta_{21} + \beta_{22} + \beta_{31}$		
Log(Area 2010)	β_1	Present area of woodlot patch (orthophoto 2010)
Connect 2010	β_2	Present forest density in a 4.8km buffer radius (orthophoto 2010)
Δ Area 2010-1979	β_{21}	Change in % area between 2010 et 1979
Δ Con 2010-1979	β_{22}	Change in connectivity (%) between 2010 et 1979
Δ Area 2010-979:Connect 2010	β_{31}	Interaction term between Δ Area 2010-1979 et Connect 2010
<i>Models (2c) with present area or past area 1979 for isolated woodlots (Connect2010 < 15%)</i>		
Model form: $E(y) = \beta_0 + \beta_1$ or $E(y) = \beta_0 + \beta_{32}$		
Log(Area 2010)	β_1	Present area of woodlot patch (orthophoto 2010)
Log(Area 1979)	β_{32}	Past area of woodlot patch (orthophoto1979)

Table 2. GLMs tested for explaining hoverfly species richness of forest specialists.

3. Results

Changes in woodland spatial patterns

The current and historical spatial data showed that of the 48 woodlots surveyed, 28 have historical continuity over time since 1850, 4 since 1900, 14 since 1954 and 2 since 1979. No patch appearing after 1979 were found. It was also observed that the average woodlot area decreased over time, from 1850 to 1954, when all woodlots were considered (Fig 4a), whereas connectivity increased as the number of patches increased (Fig 4b). For old forests having an historical continuity since 1850 (n=28), only connectivity significantly increased from 11.2 % in 1850 to 16.5 % in 2010 (Fig 4d), but not their area (Fig 4c).

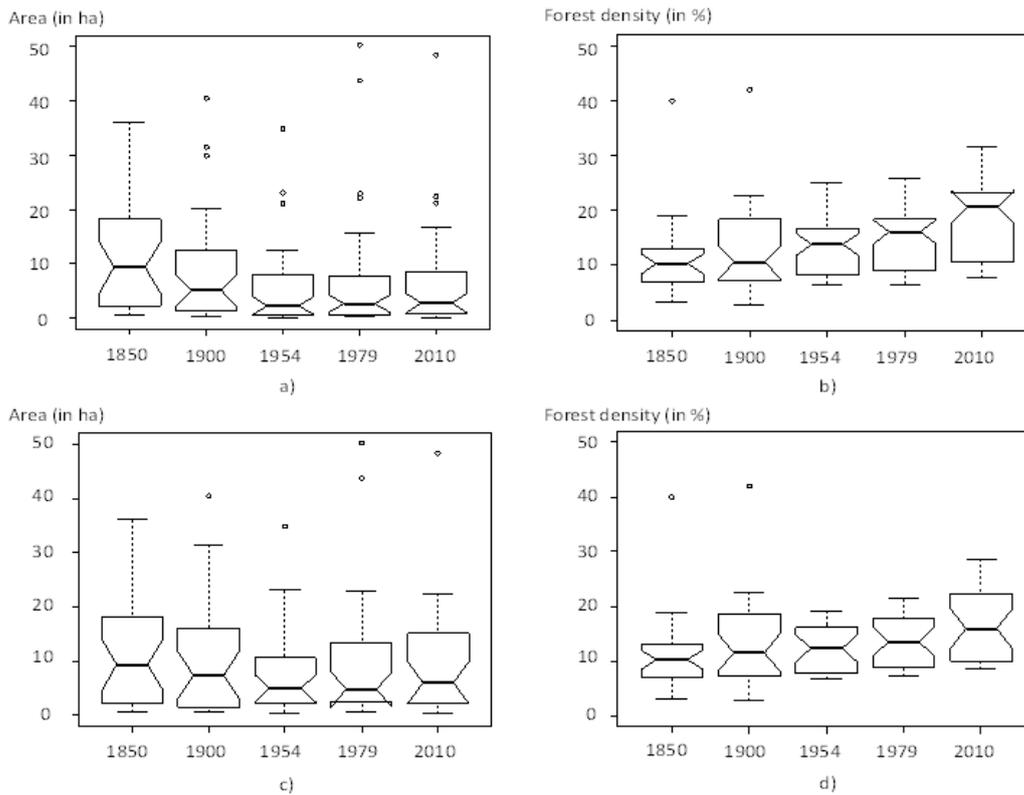


Fig.4 Distribution of woodlot area and forest density (i.e. connectivity) from 1850 to 2010 including all woodlots at each date (plots (a) and (b)) or only ones having an historical continuity over time since 1850 (plots (c) and (d)). The notched box whisker plots show medians (line) and the interquartile range (from 1st to 3rd quartiles; 50% of data). The notch displays a confidence interval around the median. If two boxes' notches do not overlap, there is strong evidence their median differ significantly (95% confidence).

Response of species richness to current area, connectivity, structural heterogeneity and historical continuity

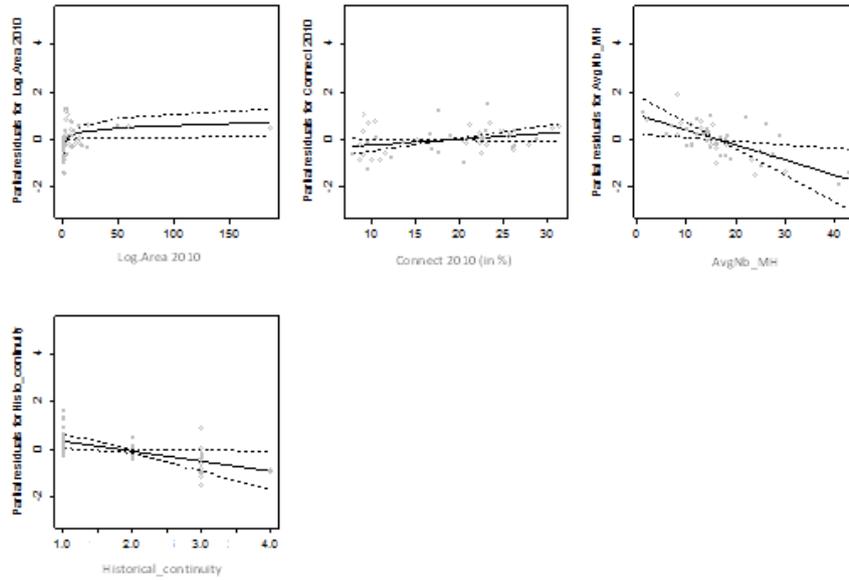
A total of 3044 individuals belonging to 27 hoverfly species were captured in the 99 Malaise traps set up in the 48 woodlands. The average number of species per wood (\pm standard deviation) estimated from species accumulation curves to an equivalent sampling effort of one trap equals to 2.81 ± 1.73 .

Five predictors of species richness were retained after the analysis of AIC: current woodlot area, current connectivity, historical continuity over time, average number of microhabitats per wood (AvgNb_MH) and the interaction term between AvgNb_MH and historical continuity. These variables

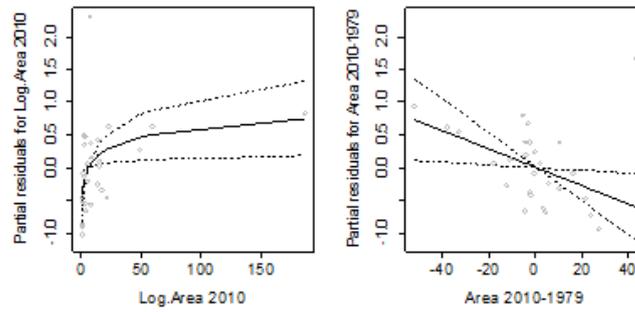
composed the “best model” (*i.e.* the model with the lowest AIC) for the first analysis (Model 1). A strong significant positive relationship (p -value < 0.01) was found between woodlot area and species richness as well as a positive effect of the interaction between AvgNb_MH and historical continuity (Table 3). AvgNb_MH was negatively related to species richness only for the oldest woodlots (Appendix 2). In addition, historical continuity showed a strong significant negative relationship on species richness indicating that the number of species is higher in old woodlots (continuity =1) than in recent ones (continuity > 1). AvgNb_MH variable also showed a weak but significant (p -value < 0.05) negative effect. Finally, a weak positive effect of current connectivity was found to explain species richness of forest hoverflies. Partial residual plots between each predictor of the model and the response variable are provided in Fig. 5. The model explained a high proportion of deviance ($\%D^2 = 35.57$; $\%adj\text{-}D^2 = 27.90$) (Table 3).

Explanatory variables	Estimates	p-value	Sign. level	AIC	$\%D^2$	$\%D^2$ adjusted	$\%Relative$ Contribution
<i>Model (1) with present area, connectivity, structural heterogeneity and historical continuity (n = 48)</i>							
Log(Area 2010)	0.177	0.008	**	176.76	35.57	27.90	36.37
Connect 2010	0.024	0.097	(*)				21.84
Histo_continuity	-0.063	0.033	**				14.99
AvgNb_MH	-0.432	0.009	*				11.74
Histo_continuity:AvgNb_MH	0.031	0.003	**				15.04
<i>Model (2a) with present area and connectivity with their past dynamics since 1850 (n = 28)</i>							
Log(Area 2010)	0.210	0.008	**	100.63	34.48	29.24	56.52
Δ Area 1979-2010	-0.01	0.017	*				43.47
<i>Model (2b) with present area and the most recent dynamics since 1979 (n=48)</i>							
Log(Area 2010)	0.160	0.002	**	179.37	23.69	18.49	58.95
Δ Area 1979-2010	-0.04	0.020	*				19.20
Δ Area 1979-2010:Connect2010	0.001	0.017	*				21.83
<i>Models (2c)</i>							
<i>- with present area for isolated woodlots (Connect2010 < 15%) (n=16)</i>							
Log(Area 2010)	0.209	0,063	(*)	55.37	21.50	15.89	100
<i>- with past area 1979 for isolated woodlots (Connect2010 < 15%) (n=16)</i>							
Log(Area 1979)	0.228	0.049	*	54.93	24.18	18.77	100

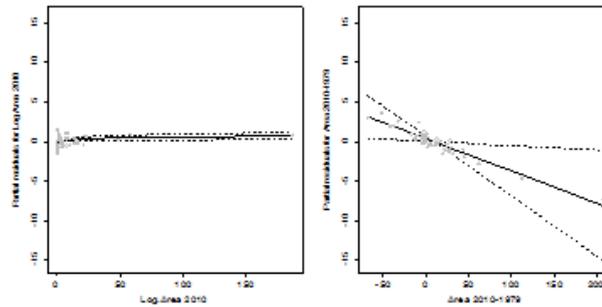
Table 3. Results from GLMs explaining hoverfly species richness of forest specialists. Explanatory variables were selected on the basis of the Akaike information criterion (AIC). Due to different sample size between the models ($n=48$ for model (1) and model (2b); $n=28$ for model (2a); $n=16$ for models (2c)), AIC values are not directly comparable.



Model (1) with present area, connectivity, structural heterogeneity and historical continuity (n = 48)



Model (2a) with present area and connectivity with their past dynamics since 1850 (n = 28)



Model (2b) with present area and the most recent dynamics since 1979 (n = 48)

Fig.5 Partial residual plots of the GLM models explaining hoverfly species richness of forest specialists.

Response of species richness to current area and connectivity with their past dynamics since 1850

The AIC-based stepwise selection showed that the model with the lowest AIC included two explanatory variables to explain species richness when past spatial dynamics were integrated: area 2010 and change in area between 1979 and 2010 (Table 3, Model 2a). As previously, the current area showed a significant positive relationship with species richness, as illustrated by the partial residual plots (Fig. 5). In contrast, a significant negative effect (p -value < 0.05) of change in area between 1979 and 2010 was observed. The amount of deviance explained by this model was high ($\%D^2 = 34.48$; $\%adj-D^2 = 29.24$).

Because average area of woodlots increased from 1979 to 2010 (+6.48%) (Fig. 4), a potentially positive effect of change in area during this period was expected. The negative effect we observed suggested a possible time-lagged response of hoverflies to change in area. An in-depth analysis was therefore conducted with a third model (Model 2b, see below), focused on the most recent time period 1979-2010, and including all the woodlots existing since 1979 ($n=48$).

Response of species richness to current area and the most recent dynamics since 1979

The most parsimonious model based on AIC selection identified three main predictors when only the most recent spatial dynamics were incorporated into the model (Model 2b): current woodlot area, change in area during 1979-2010 and interaction between change in area during 1979-2010 and current connectivity (Table 3). Significant positive and negative effects were observed with the first two predictors respectively, in accordance with the previous model (Model 2a). A significant positive relationship was also found for the interaction between change in area during 1979-2010 and current connectivity (Fig. 5) indicating that the effect of change in woodlot area must be interpreted in relation to their current connectivity. The histogram of current connectivity revealed that two groups of woodlots with distinct connectivity could be distinguished: a first one with current connectivity lower than 15% ($n=16$), and a second one with current connectivity higher than 15% ($n=32$) (Fig. 6a). A negative relationship was observed between change in area during 1979-2010 and the species richness of forest hoverflies when only woodlots of the first group (current connectivity $< 15\%$) were considered (Fig. 6b). In contrast, a positive relationship was measured when only woodlots of the second group (current connectivity $> 15\%$) were considered (Fig. 6b). In other words, when woodlots are well connected, the increase in area during 1979-2010 had a positive effect on species richness, as expected. For isolated woodlots, the increase in area did not show an increase in species richness. The goodness-of-fit of Model 2b was quite high ($\%D^2 = 23.69$; $\%adj-D^2 = 18.49$, Table 3).

Finally, since current species richness of the set of woodlots with current connectivity $< 15\%$ ($n=16$) was better explained by woodlot area in 1979 than in 2010 ($\%adj-D^2 = 18.77$, p -value = 0.049 and $\%adj-D^2 = 15.89$, p -value = 0.06, respectively; Model 2c in Table 3), we assumed the coexistence of extinction debt and colonization credit in these isolated woodlots, depending on the direction of change in area from 1979 to 2010 (negative or positive). In other words, observed species richness in these woodlots would be higher or lower than expected at equilibrium.

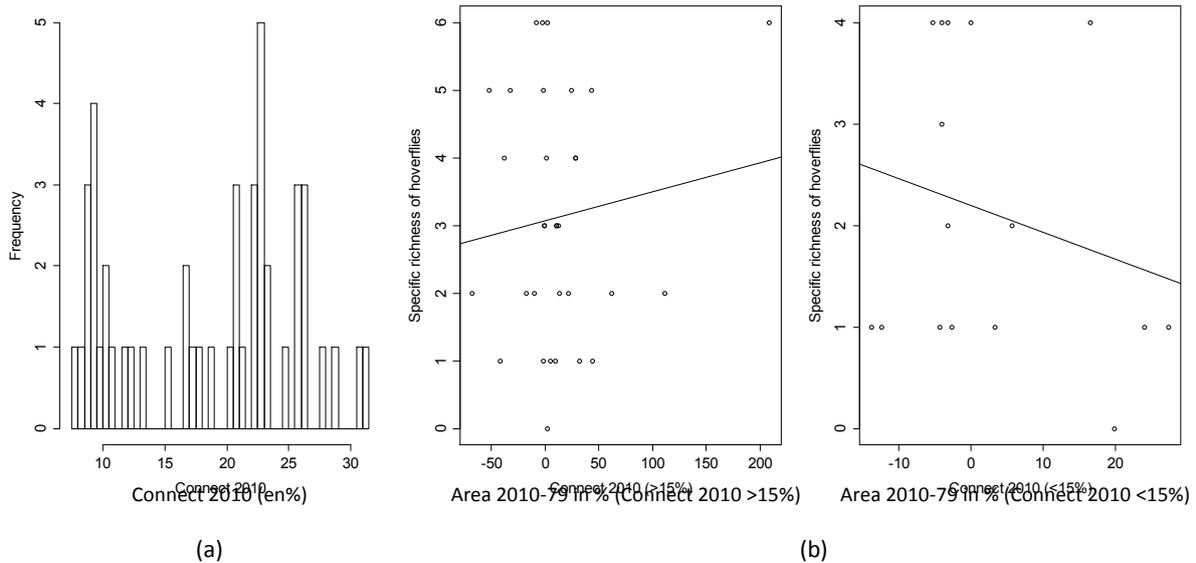


Fig.6 Based on the frequency distribution of current forest density in (a), two groups of woodlots were distinguished: the isolated ones (density < 15%) and the connected ones (density > 15%). Inverted relationships between change in area 2010-1979 and species richness of forest hoverflies were observed in (b) according to belonging to a particular connectivity group.

4. Discussion

Woodlot area, connectivity, historical continuity and average density of microhabitats mainly explained the species richness of forest specialist hoverflies

Our results showed that the species richness of forest specialist hoverflies was mainly explained by a combination of four factors operating at different scales of analysis: the average density of microhabitats (local effect), the current area of the woodlot and its connectivity (patch-scale effect) and the historical continuity of wood state (temporal effect).

Surprisingly, the density of microhabitats was found to be negatively related to species richness, especially in the oldest woodlots. Nevertheless, larvae of saproxylic species, which composed the majority ($72.4\% \pm 21.6$) of the forest specialist hoverflies, are strongly dependent on microhabitats associated with old and senescent trees (Speight 1989; Speight and Good 2003; Speight 2013). In addition, no correlation was found between deadwood and hoverfly richness. This result might indicate that the procedure used to count microhabitats was not relevant for hoverflies or that variables related to microhabitats and deadwood were probably not the only key factors explaining saproxylic hoverfly richness. This is also suggested by Fayt et al. (2006) who found a high number of saproxylic species in woods with an open structure and large trees, providing ecological conditions for a species-rich herb layer, resources of pollen and nectar for imagoes.

The current woodlot area was the main driving factor showing a positive effect on species richness (part of the total explained variation = 36.37 %). This area effect is well-known in the habitat patch framework (Connor and McCoy 1979; Andren 1994) and consistent with previous studies about forest hoverfly specialists (e.g. Ouin et al. 2006) as well as grassland specialists and generalists (Meyer et al. 2009; Bommarco et al. 2014). The increase of species richness with patch area is predicted by the equilibrium theory of island biogeography (MacArthur, 1967). The habitat diversity hypothesis (Root 1973; Huston 1994) is one of the hypotheses proposed to explain this positive relationship. Larger wood fragments are often more heterogeneous than small ones and therefore are more likely to provide critical resources for different species.

Interestingly, this study's results showed that forest hoverflies may be sensitive to "habitat continuity" that may be considered to represent connectivity over time (Hanski 2005), while "habitat connectivity" represents connectivity only in space. Although hoverflies are very skilful flying insects, more or less vagrant depending on the species (Speight 2013), and that some adult specimens are sometimes trapped after having travelled several hundred meters away from their emergence habitat (Larrieu 2005; Sarthou and Sarthou 2007), we found that ancient woodlots host more species than the recent ones. Some species such as *Criorhina floccosa*, *Doros destillatorius* or *Brachylapoides lentus*, whose microhabitat requirements of larvae are strictly linked to trees in stand (Speight 2013), were observed far more frequently in ancient woodlots than in recent ones, and outside forests very rarely (Speight 2013). They might be unwilling to cross an unfavourable matrix such as large fields suggesting that they are not able to colonise favourable recent woodlots easily.

Finally, current level of connectivity played also a positive role in the species richness of forest hoverflies, as observed in previous study (Meyer et al. 2009). Generally, habitat connectivity has a positive effect on species diversity, in particular for specialist species with a low dispersal ability (Hanski 2005). Nevertheless, connectivity could also impose negative indirect effects on certain species through the spread of invasive species (Atobe et al. 2014) or pathogens (Hess 1996). Moreover, De Groot (2003) demonstrates in a case study in the surroundings of Ljubljana (Slovenia) that hoverfly species richness responds negatively to an invasive plant species, namely *Solidago canadensis*, since it homogenises plant layers by supplanting the other flowering plants. This plant species has already been observed locally in the "Vallées et Coteaux de Gascogne" area and has been described as having the ability to colonise habitat patches in fragmented landscapes quickly (BAlent). This plant was not surveyed in the present study but since the relative contribution of current connectivity in the model was high (part of the total explained variation = 21.84%), it assumed that this invasive plant species did not affect the species richness.

Changes in woodlot area between 1979 and 2010 influenced differently current species richness of forest hoverflies depending on woodlot connectivity

Several studies have analysed the effects of history on plant or insect diversity in fragmented forests and grasslands, but they often estimate this effect by comparing diversity between patches having distinct dynamics (Piqueray et al. 2011; Guardiola et al. 2013). When complex dynamics occur (including area loss and gain combined with an increase in connectivity and isolation), grouping patches according to their temporal trajectory may be a challenging task. In addition, it may bring to consider several typologies of change, leading to increase analysis (Piqueray et al. 2011). In our case, because several processes acting jointly (area and connectivity increase or decrease over time), changes in spatial patterns were directly incorporated into the models, in addition to interaction terms between processes. Results showed contrasting effect of change in area during the most recent period 1979-2010 depending on the level of connectivity. For well-connected woodlots, a positive effect of changes in area on species richness was observed, indicating that both increase and decrease in area might be compensated by colonisation-extinction dynamics (MacArthur, 1967). For isolated woodlots, a negative effect was found revealing a possible delayed response of the hoverfly community to landscape changes.

These findings reinforce the conclusions of previous studies demonstrating that not only are past or current spatial patch characteristics important for explaining present biodiversity patterns, but so are changes of patch area and potentially connectivity (Metzger et al. 2009). It also underlines the importance of analysing these spatial patterns dynamics jointly.

A meta-analysis comparing bird response to forest fragments vs true islands has concluded that forest fragments do not function as true islands (Brotons et al. 2003). The authors interpret this result as being a compensatory effect of the surrounding matrix in terms of availability of resources and enhanced connectivity. However, in our case, since only forest specialist hoverflies were

considered, no resource was available in the matrix for larvae (but flower resources could be available for adults). Therefore, connectivity between forest patches may contribute to enable the persistence of the hoverfly community in the landscape through the dispersion of insects between ancient and recent patches (and vice versa).

Are hoverfly forest communities concerned by extinction debt or colonization credit?

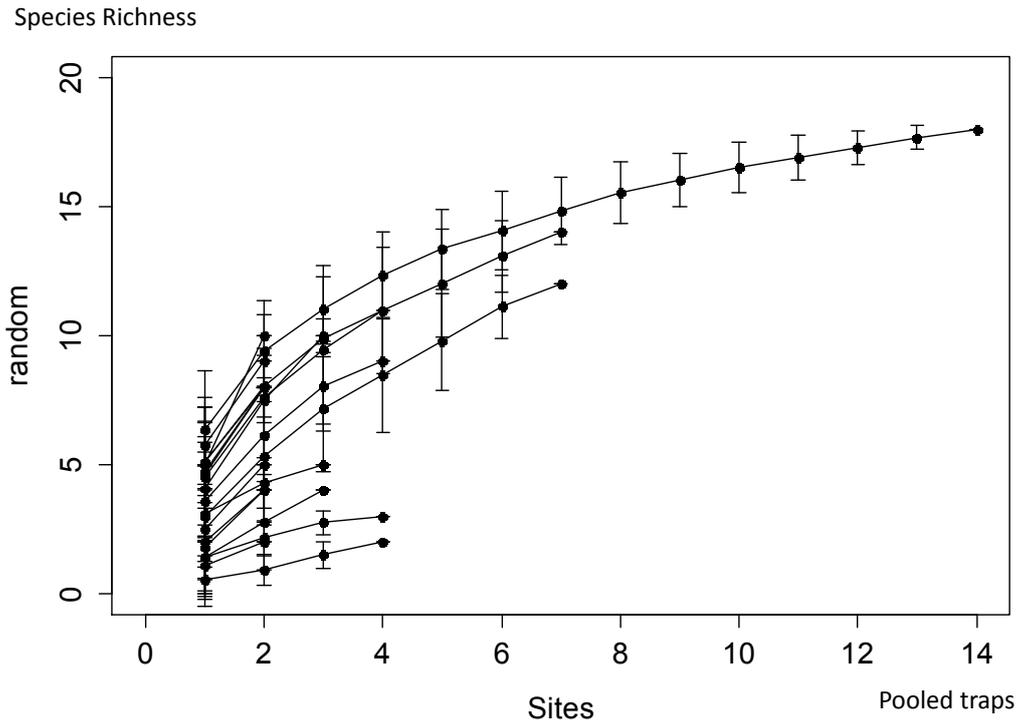
Studies which examined potential extinction debt mainly focused on vascular plants, saproxylic cryptogams or birds, and rarely on short-lived animals (Kuussaari et al. 2009). Despite hoverflies have potential good dispersal capacity (Speight et al. 2013) and high turnover rate, indication of delayed response to change in habitat area has already been observed in a recent study for hoverfly generalists and specialists of semi-natural grasslands (Bommarco et al. 2014). Moreover, forest specialists, which need trees absolutely and whose larvae have very specific microhabitat requirements, are assumed to be more sensitive to environmental changes than generalists and thus to show more easily extinction debt (Kuussaari et al. 2009). In a landscape context of low connectivity, we highlighted both extinction debt for hoverflies in woodlots which have lost area between 1979 and 2010 (10 woodlots with connectivity < 15%), and colonization credit in woodlots which increased their area during the same period (6 woodlots with connectivity < 15%). Detecting extinction debt by using past and current habitat characteristics did not allow estimate its magnitude (Kuussaari et al. 2009). Nevertheless, our results highlighted that 30 years are not enough for assemblages of forest hoverfly to reach new equilibrium.

5. Conclusions

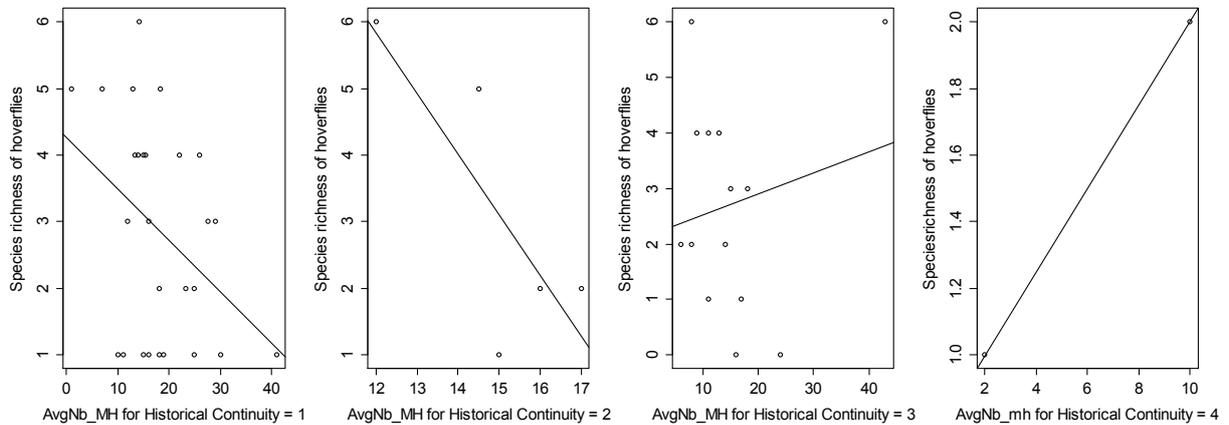
This study demonstrated that the current species richness of forest hoverflies was influenced by a combination of factors operating at a different scale of analysis from plot to landscape. These findings also revealed the interest of adopting a change-oriented approach to explain the current distribution pattern of species.

Incorporating historical changes in landscape ecological studies would appear to be a very promising way of obtaining a better understanding of all the factors affecting current species diversity. However, this would require an extensive spatiotemporal database incorporating a wide variety of historical data sources, the creation of which would be a challenge in itself. The availability of historical maps is often limited and the reconstruction of patch history can take a long time to complete. Recent advances in the automatic processing of digital historical maps and aerial photographs (Leyk et al. 2006; Baily et al. 2011; Herrault et al. 2013) should enable the effects of history to be considered more systematically in future. With this in mind, a closer collaboration between historical ecologists and geo-information scientists should be encouraged.

Appendix



Appendix 1. Species accumulation curves estimated for each woodlot having more than one trap (i.e. with an area > 5ha, n=19). These curves show the average pooled species richness of forest hoverfly specialists when all traps per wood are combined by randomly sampling traps with 100 permutations.



Appendix 2. Relationships between current average number of microhabitats per wood (AvgNb_MH) and species richness of forest hoverflies for each historical continuity category (i.e. since 1850 = 1; since 1900 = 2; since 1954 = 3; since 1979 = 4)

BILAN

Ce travail collaboratif a permis de tirer plusieurs enseignements à la fois thématiques et méthodologiques sur les études visant à étudier les effets du paysage sur les populations de syrphes forestiers :

- Les syrphes sont sensibles à l'histoire à des paysages. Les résultats présentés dans l'article soumis à *Landscape Ecology* montrent premièrement que l'ancienneté des habitats forestiers favorise une plus grande richesse. Deuxièmement, les syrphes peuvent exprimer un délai de réponse plus de 30 ans après avoir connu une perturbation de leur habitats.
- Il est impératif d'adopter des modèles statistiques permettant d'évaluer les effets de plusieurs variables en interaction. Notre étude a notamment mis en lumière le fait que la population de syrphe suivie était particulièrement sensible aux changements de surface de leurs habitats, pour un certain niveau de connectivité. Il était par conséquent indispensable de tester ces deux variables (surface et connectivité) simultanément afin d'évaluer leurs réels effets sur la présence actuelle de syrphes.
- Enfin, les données spatiales anciennes utilisées ont montré toute leur adéquation à ce type de travaux. Il est toutefois important d'adopter des démarches méthodologiques rigoureuses pour éviter d'introduire de potentiel biais dans la modélisation statistique. Dans cette logique, le chapitre qui suit (chapitre 11), a pour objectif d'évaluer les potentiels effets des variables d'entrée construites à partir de données extraites automatiquement sur la réponse des modèles écologiques établis dans cette étude.

CHAPITRE 11. MODELISATION ESPECE-HABITAT : EFFET DES INCERTITUDES SPATIALES

SOMMAIRE

11.1 ORIGINE ET NATURE DES INCERTITUDES SPATIALES

11.2 TRAVAUX TENANT COMPTE DES INCERTITUDES SPATIALES DANS LES
MODELES ESPECES-HABITAT

11.3 MESURER LES EFFETS LIES AU PROCESSUS D'EXTRACTION DES FORETS :
DEMARCHE ET MISE EN ŒUVRE

11.4 BILAN

Cet ultime chapitre a pour objectif d'évaluer les effets potentiels de l'incertitude spatiale sur la réponse des modèles espèces-habitat. Un premier point permettra de rappeler l'origine et la nature des incertitudes spatiales puis nous discuterons brièvement des travaux intégrant cette dimension dans l'analyse des modèles espèces-habitats. L'évaluation d'incertitude menée ci-après portera exclusivement sur l'impact du processus d'extraction des forêts sur la réponse des modèles écologiques définis dans le chapitre 10.

11.1. ORIGINE ET NATURE DES INCERTITUDES DANS L'INFORMATION GEOGRAPHIQUE

11.1.1 DEFINITIONS

Il existe aujourd'hui un certain nombre de concepts structurant les travaux sur la qualité de l'information géographique (Deville et Goodchild 2010). S'il est reconnu que les données géographiques présentent des imperfections (Goodchild 1995; Hunter 1999; Girres 2013) de nombreuses taxonomies intègrent également d'autres notions telles que l'incertitude (Fisher 1999; Fisher et al. 2005), l'imprécision (Worboys 1998) ou l'incomplétude (Schneider 1999) Aucune définition standard de ces termes n'a été adoptée, leur sens étant défini par les différentes communautés d'appartenance, les différents domaines d'application ou points de vue.

Olteanu (2008) désigne l'imprécision, l'incertitude et l'incomplétude comme les trois grands types d'imperfection des données géographiques (Olteanu 2008). L'*imprécision* exprime la difficulté à retranscrire précisément une information (e.g. la maison est **approximativement** à 1 km de l'exutoire du fleuve). L'*incertitude* exprime le doute de l'observateur à propos d'une connaissance (e.g. **je crois** que la maison est 1 km de l'exutoire). L'*incomplétude* désigne l'absence de connaissances sur la donnée (la distance qui sépare la maison de l'exutoire n'est pas renseignée dans la table attributaire).

Pour Zhang et Goodchild (2002), la notion d'incertitude est elle-même un concept général englobant plusieurs termes (Zhang et Goodchild 2002). Pour illustrer cela, la typologie présentée par Fisher et al. (2005) est particulièrement utilisée pour formaliser l'incertitude des données géographiques (Figure 11-1). Les auteurs suggèrent que tous les processus s'appuyant sur les données géographiques sont entachés d'incertitude et que cette incertitude est principalement liée au processus **d'abstraction du monde réel** (voir section 11.1.3.1). Il distingue les objets bien définis (un bâtiment) des objets mal définis (une forêt). Les objets bien définis sont concernés par des erreurs qu'il est possible de modéliser à l'aide d'approches statistiques (ex : probabilités). Les objets mal définis disposent de limites vagues et peuvent être modélisés par les ensembles flous par exemple (Zadeh, 2005). Ces objets peuvent également être entachés d'ambiguïté qu'il est possible de gérer avec des théories traitant du domaine de l'incertain.

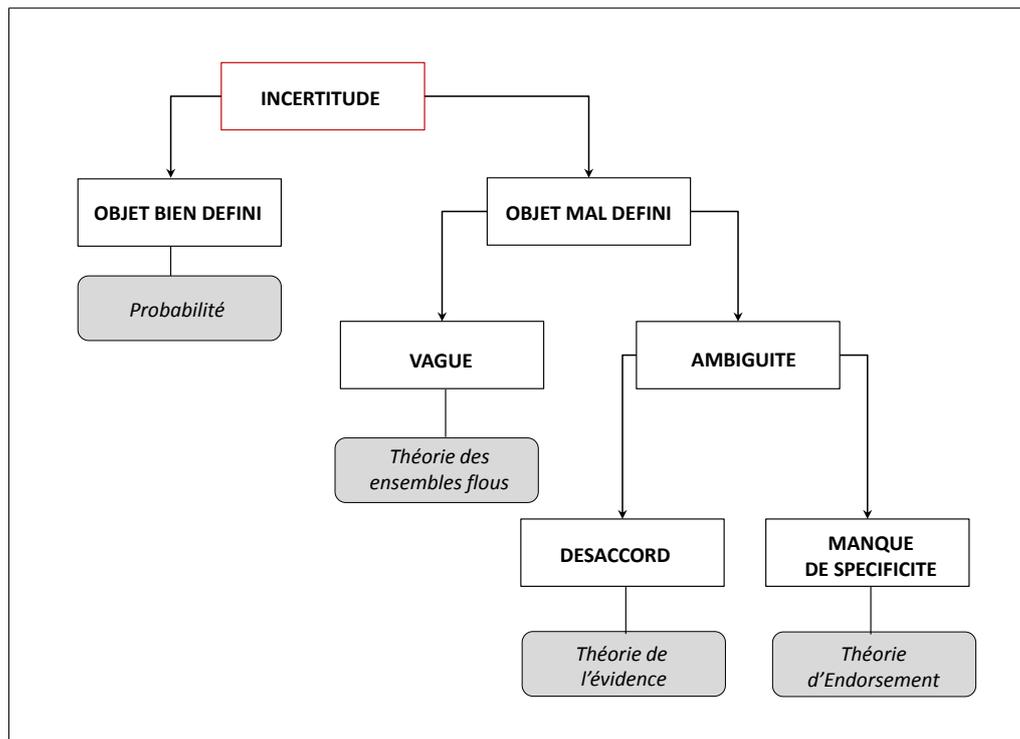


Figure 11-1. Typologie de l'incertitude (source : Fisher et al. 2005)

11.1.2 LA QUALITE DES DONNEES GEOGRAPHIQUES

Pour communiquer sur l'incertitude liée à la production et à l'usage des données spatiales, on fait référence à la notion de qualité des données. Deux concepts sont généralement admis : la qualité interne et la qualité externe (Devilleers 2004). La qualité interne est destinée aux producteurs. Elle désigne la conformité du produit à ses spécifications. Lorsqu'une « erreur » est détectée, elle indique donc la violation des règles fixées dans les **spécifications** (voir section 11.1.3.1) (Vauglin 1997; Devillers et Jeansoulin 2005). Le contrôle de la qualité interne vise à déterminer l'écart existant entre les données produites et les données « convenues » (telles qu'elles devraient l'être d'après les spécifications).

La qualité externe est destinée aux utilisateurs. Elle concerne l'adéquation du produit aux besoins et aux attentes des utilisateurs (« *fitness for use* ») (Chrisman 1991; Devillers et Jeansoulin 2005). Le contrôle de la qualité externe consiste alors à estimer l'écart existant entre le modèle proposé par le producteur et le modèle requis pour l'application. Par exemple, une carte au 1/25 000^e est susceptible d'être mieux adaptée qu'une carte au 1/200 000^e pour conduire une étude écologique à l'échelle des patchs d'habitats forestiers.

Dans ce travail de thèse, nous nous intéresserons principalement aux problèmes de qualité interne. En effet, la qualité externe des données a déjà été traitée dans le cadre de la sélection des données spatiales (voir chapitre 4). Nous avons alors considéré que les données sélectionnées étaient en adéquation avec nos objectifs d'étude et qu'elles permettaient de répondre à nos différents besoins.

11.1.3 EVALUATION DE LA QUALITE INTERNE

11.1.3.1. ABSTRACTION, SPECIFICATIONS ET TERRAIN NOMINAL

La constitution d'une base de données géographiques visa à représenter le monde réel (ou univers) selon un modèle choisi simplifiant la réalité. L'opération qui consiste à simplifier le monde réel pour en extraire les phénomènes à représenter est appelée **abstraction** du monde réel (Figure 11-2). Elle consiste à transformer la réalité (les feux de forêts par exemple) en un modèle intégrant seulement les informations nécessaires à son analyse. Les règles utilisées pour définir le niveau d'abstraction sont généralement décrites au sein d'un document appelé « spécifications ». Selon l'Organisation de Standardisation Internationale (ISO), les spécifications sont *les exigences auxquelles le produit ou le service doit se conformer* (ISO8402, 1994). Elles sont conçues pour que le produit délivré soit conforme aux attentes des utilisateurs (à la suite d'une analyse des besoins réalisée par le producteur). Ainsi **les spécifications de contenu** renseignent sur ce qui doit apparaître dans la base de données tandis que les **spécifications de saisie** indiquent les méthodes et techniques à utiliser pour représenter l'information (Vauglin 1997). Lorsqu'une base de données est produite, par exemple une BD « Forêt », son producteur cherche à tendre vers **le terrain nominal**. Le terrain nominal est l'image du monde réel à travers le filtre des spécifications (Girres 2013). C'est en quelque sorte la version « parfaite » du contenu de la base de données, celle qui devrait être produite d'après les spécifications, en l'absence d'erreurs de saisie.

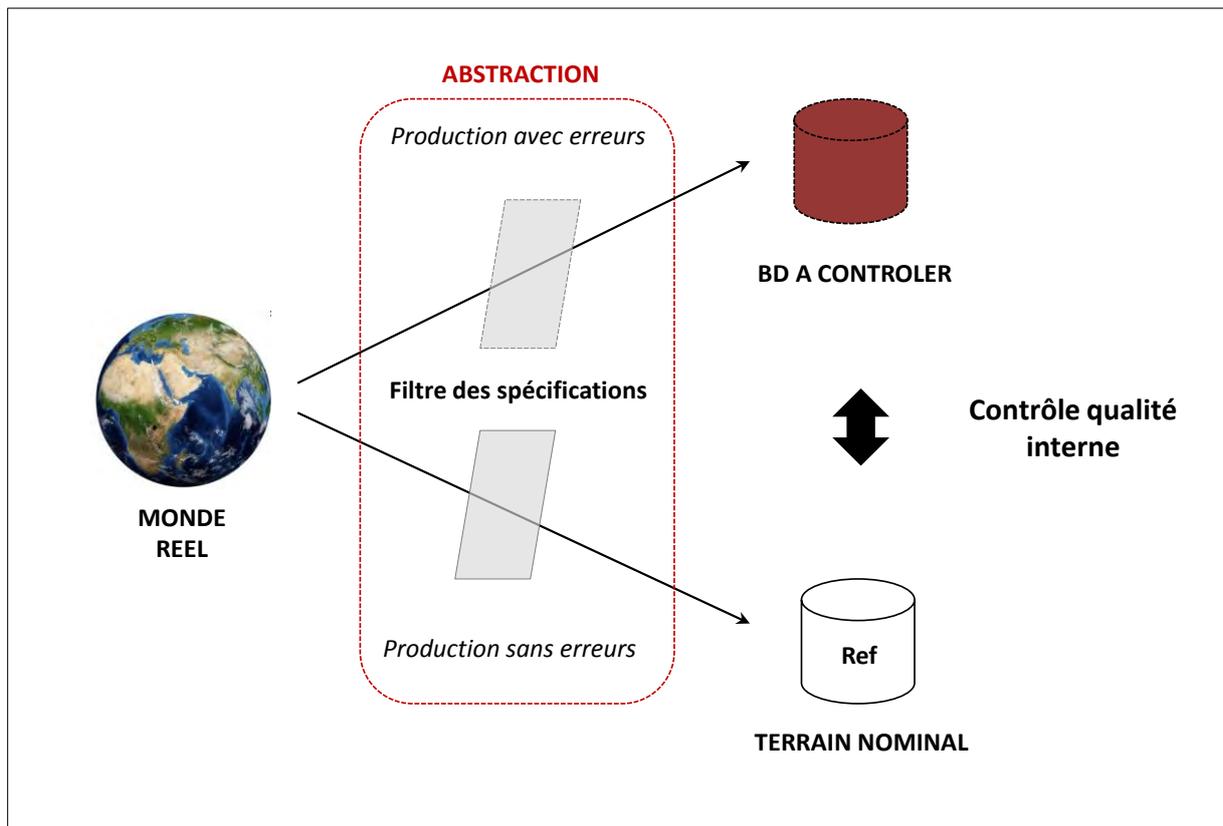


Figure 11-2. Le principe de la qualité interne

11.1.3.2. LES SOURCES D'ERREURS DE LA QUALITE INTERNE

Les sources d'erreurs pouvant entrainer une baisse de la qualité interne des produits sont nombreuses et sont principalement liés aux processus de production de la base mais pas uniquement.

Premièrement, il peut exister **une incertitude du terrain nominal** (Vauglin et Ali 1998). Cela peut concerner par exemple une mauvaise définition des règles de saisie, une absence de règles ou encore, une difficulté à clairement définir ces règles. Le cas des dessins-minutes d'Etat Major est un exemple parlant pour illustrer ce propos. Les règles de spécifications étant mal définies voir inconnues, chaque producteur a tendu vers son propre terrain nominal délimité par l'interprétation qu'il a faite des spécifications. Ceci a ainsi conduit à une faible homogénéité de la carte à l'échelle nationale avec une grande diversité de représentation pour un même thème.

L'erreur peut être due également à la **mauvaise qualité des sources de saisie**. Chacune d'entre elles dispose de son propre niveau de qualité, ceci concernant principalement l'échelle des données (Cartes anciennes, images satellites, photographies aériennes). Par exemple, sur une carte papier au 1/40 000^e, les objets ont une exactitude de position de 4 mètres au mieux. Ce chiffre correspond à l'erreur graphique de la carte (0.1 mm ou 0.01 cm) et ne peut être amélioré compte tenu du pouvoir discriminant de l'œil humain (Sheeren 2005).

L'incertitude peut être provoquée par la **qualité des modes de saisie**. Cela peut concerner les levés topographiques, la numérisation des éléments ou encore le scannage des sources d'informations.

On dénombre aussi un certain nombre d'erreurs liés **aux traitements numériques**. Ceux-ci peuvent concerner :

- les processus de calcul : arrondis, limites décimales
- les interpolations : en plus de la méthode elle-même, l'erreur peut surtout porter sur la définition du paramètre de densité
 - la généralisation cartographique : simplifications, type d'information retenue
 - la topologie : arcs mal connectés, adjacence inter-polygonales non respectée
 - l'algorithmie : erreurs de convergence
 - le géoréférencement : erreur géométrique élevée en raison d'une méthode non adaptée ou d'un trop faible nombre de points d'amer, etc

Enfin, l'incertitude peut être directement reliée à l'actualité des données. Ceci porte par exemple sur une base de données qui n'est plus à jour.

11.1.4 UN CADRE POUR EVALUER L'INCERTITUDE DANS LES DONNEES SPATIALES ANCIENNES

L'usage des données spatiales anciennes est réputée entrainer un plus haut degré d'incertitude que les données récentes (Plewe 2002).

Dans cette logique, plusieurs auteurs ont tenté de définir un cadre conceptuel permettant de gérer l'incertitude liée aux données, aux processus de production ou aux types d'application (Plewe 2002; Leyk et al. 2005). Ils visent à fournir des approches systématiques pour les utilisateurs afin qu'ils puissent gérer eux-mêmes les problèmes d'incertitude.

La typologie de Leyk (2005) est particulièrement bien adaptée à notre travail car définie dans un contexte de changements paysagers sur le temps long. Ce dernier distingue trois approches distinctes pour évaluer l'incertitude. :

(1) L'incertitude orientée production fait référence aux incertitudes liées aux sources de données utilisées. Elle porte globalement sur les mécanismes de production de ces données ainsi que les méthodes employées pour les reproduire.

(2) L'incertitude orientée transformation désigne exclusivement l'incertitude provoquée par les processus de production visant à rendre manipulables les informations retranscrites au sein des sources d'informations. Parmi eux, on retrouve les processus d'extraction des éléments (digitalisation manuelle, extraction automatique) ou la correction géométrique des données.

(3) L'incertitude dépendante de l'application porte sur les différences entre les sources utilisées ou l'incapacité des données à retranscrire l'information à l'échelle d'analyse définie.

Dans ce travail de thèse, nous nous concentrerons principalement sur le deuxième type d'incertitude à savoir l'incertitude orientée « transformation ». Nous évaluerons précisément les effets du processus de numérisation des forêts sur la validité des modèles écologiques définis. Par conséquent, la section qui suit a pour objectif de présenter les travaux tenant compte des incertitudes spatiales au sein des modèles espèces-habitat.

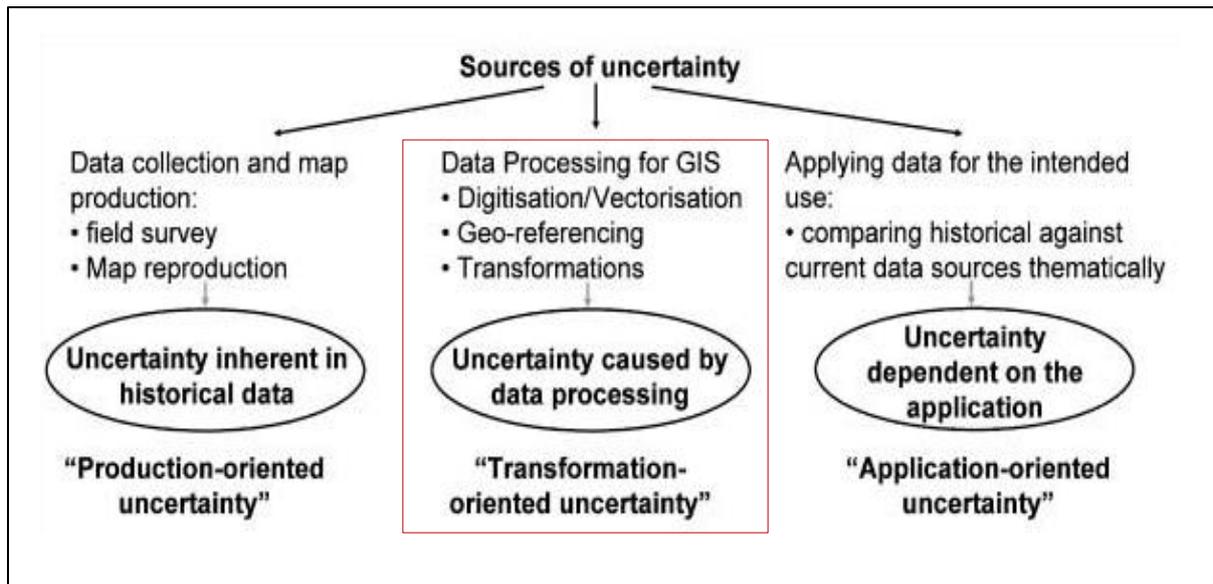


Figure 11-3. Cadre applicatif de l'évaluation de l'incertitude spatiale (source : Leyk et al. 2005). Le cadre rouge indique le type d'incertitude évalué dans ce travail de thèse.

11.2. TRAVAUX TENANT COMPTE DES INCERTITUDES SPATIALES DANS LES MODELES ESPECES-HABITAT

Alors que la question de la fragmentation des habitats et ses effets sur la biodiversité est un sujet phare de l'écologie du paysage, l'influence de l'incertitude des données spatiales sur les modèles écologiques est rarement traitée (Rocchini et al. 2011; Lechner et al. 2012; Moudry et Šimová 2012). Les effets potentiels des erreurs spatiales sont pourtant bien connus des écologues (Jager et King 2004; Barry et Elith 2006) mais souvent ignorés en sortie de modèle. Par conséquent, il existe aujourd'hui un appel à la prise en compte de l'incertitude (Rocchini et al. 2011; Lechner et al. 2012).

La majorité des travaux, peu nombreux, traitant de l'incertitude spatiale dans la modélisation écologique s'intéresse aux erreurs géométriques, sémantiques ou de complétude des données d'entrée (Jager et King 2004). Ces données peuvent concerner les variables explicatives (Hines et al. 2005) (mesures environnementales ou paysagères) ou la variable réponse (Graham et al. 2008) (mesures de piégeage biologiques). D'une manière générale, ces travaux portent principalement sur les erreurs de position des mesures de présence d'espèces intégrées au sein des modèles de distribution spatiale d'espèces (SDM). Les distances inter-pièges peuvent être erronées

et baisser les prédictions de modèles de distribution d'espèces (Graham et al. 2008). Ces erreurs peuvent provenir du processus de correction géométrique de la source utilisée pour localiser les pièges ou bien directement par l'instrument de mesure employé pour mesurer leurs coordonnées géographiques sur le terrain par exemple (Visscher 2006). Ces travaux peuvent aussi porter sur les données d'habitat (Hines et al. 2005) mesurées à partir des données spatiales. La correction géométrique entraîne des imprécisions de position et par conséquent de taille ou de forme des objets numérisés pouvant ensuite causer potentiellement des biais au sein des modèles espèces-habitats.

La seconde grande catégorie de travaux concerne l'impact des erreurs de classification des images utilisées pour caractériser les habitats. Les habitats sont en effet très souvent cartographiés au moyen de la télédétection (Lechner et al. 2012). Les erreurs de classification peuvent impacter la géométrie des objets extraits, leur sémantique ou bien directement en omettre certains (Langford et al. 2006). De nombreux outils et méthodes existent pour mesurer ces erreurs et les caractériser spatialement (Congalton 1994; Stehman 1997; Edwards et al. 1998; Kyriakidis et Dungan 2001; Remmel et Perera 2002). Peu de travaux intègrent toutefois directement ces erreurs dans les modèles espèces-habitats (Ruckelshaus et al. 1997; Beissinger et Westphal 1998). Les travaux de Lechner (2012) ont d'ailleurs clairement mis en évidence la faible considération des écologues du paysage à prendre en compte de l'incertitude spatiale dans les études conduites. Ce dernier a démontré que seuls 69% des articles publiés dans *Landscape Ecology* en 2007 avaient mentionné des informations sur la qualité des données spatiales utilisées et leur potentiel impact sur la validité des interprétations écologiques.

11.3. MESURER L'IMPACT DE L'INCERTITUDE DES DONNEES SPATIALES LIEE AU PROCESSUS DE VECTORISATION AUTOMATIQUE : DEMARCHE ET MISE EN OEUVRE

Cette section vise à évaluer l'impact du processus d'extraction automatique des forêts (chapitre 7) sur la réponse des modèles espèces-habitats définis précédemment. L'approche choisie consiste à comparer les réponses des modèles à l'aide des BD Forêts produites manuellement à celles des modèles intégrant les BD Forêts extraites automatiquement, c'est-à-dire les BD AUTO 1850 et BD AUTO 2010.

L'évaluation de l'incertitude a été menée à travers deux modèles écologiques distincts. Le premier est le modèle 1 de l'article soumis à *Landscape Ecology* et vise à déterminer quelles variables aux différentes échelles d'analyse influent sur la richesse actuelle de syrphes forestiers. Le second est le modèle 2a du même article et a pour objectif d'évaluer si les dynamiques spatio-temporelles des habitats forestiers sont en mesure d'expliquer une part de la variable réponse. La forme de ces modèles est rappelée en tableau 11-1.

MODELE	DESCRIPTION
N°1	$RS \sim \log(\text{Surface}_{2010}) + \text{Connectivité}_{2010} + nb_ess + nb_ess + nb_ch$ $et + nb_bms + nb_tgb + div_mh + nb_mh + open + nb_aqua + Age + Age:dens + Age:nb_ch$ $et + Age:nb_mh$
N°2	$RS \sim \log(\text{Surface}_{2010}) + \text{Connectivité}_{2010} + \Delta surf_{1979-2010} + \Delta con_{1979-2010} + \Delta surf_{1954-1979} + \Delta con_{1954-1979} + \Delta surf_{1900-1954} + \Delta con_{1900-1954} + \Delta surf_{1850-1900} + \Delta con_{1850-1900} + \Delta surf_{1850-2010} + \Delta con_{1850-2010}$

Tableau 11-1. Description des modèles écologiques faisant l'objet d'une évaluation de l'incertitude spatiale. Les « : » désignent les termes d'interaction.

Au total, quatre comparaisons seront effectuées (Figure 11-4).

(1) La comparaison 1 évalue les différences de réponse du modèle 1 en fonction des BD Forêts 2010 introduites. L'une a été produite manuellement (BD MANUEL 2010) et a permis de construire le

Modèle 1 MANUEL tandis que l'autre a été produite automatiquement (BD AUTO 2010) et incorporée dans le Modèle 1 AUTO.

(2) La comparaison 2 évalue les différences de réponse du modèle 2 en fonction des BD AUTO 2010 et BD MANUEL 2010

(3) La comparaison 3 évalue les différences de réponse du modèle 2 en fonction des BD AUTO 1850 et BD MANUEL 1850 (modèle gaussien, cf. section 7.2.5).

(4) La comparaison 4 vise à déterminer les effets combinés des BD AUTO 1850 et BD AUTO 2010 sur la réponse du modèle 2. Elles ont été par conséquent introduites au sein du « modèle 2 AUTO1850+2010 ». Nous comparerons sa réponse à celle du « modèle 2 MANUEL ».

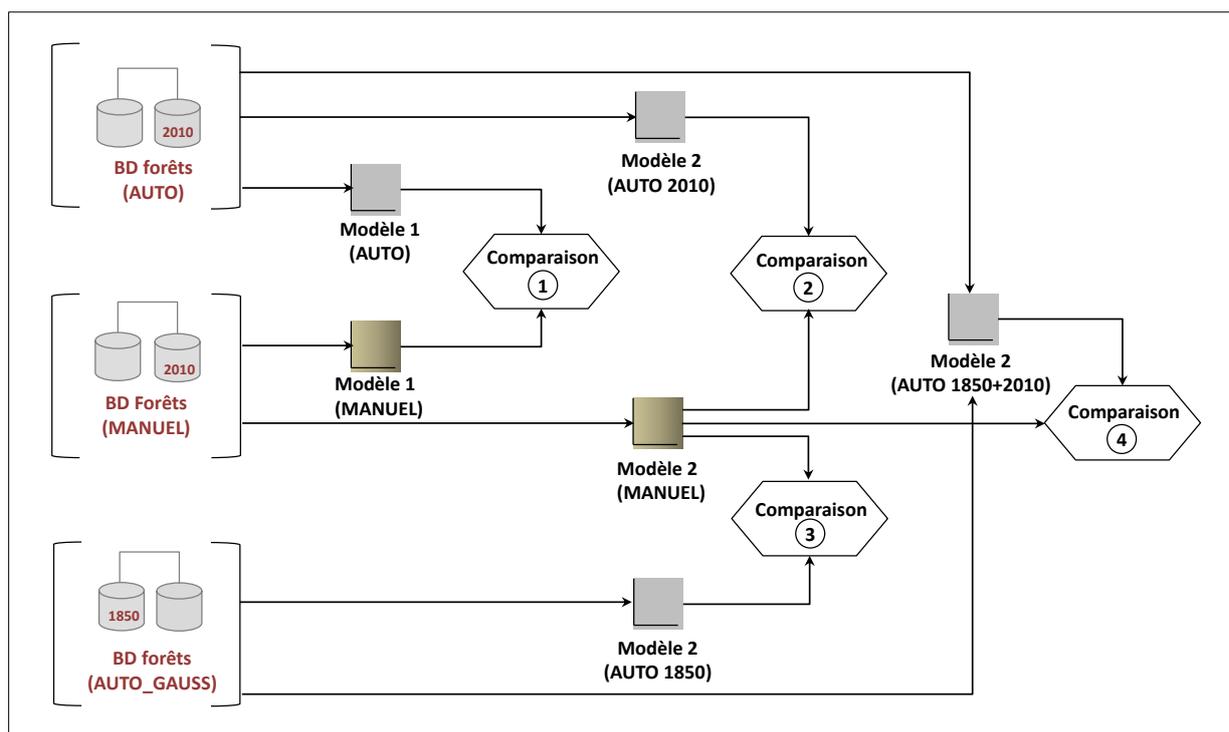


Figure 11-4. Schéma méthodologique de la stratégie d'évaluation de l'incertitude de classification

A noter enfin que la procédure de modélisation statistique est exactement la même que celle présentée en chapitre 10. Pour ne pas alourdir le manuscrit, nous avons fait le choix de ne pas la redétailler ici.

11.3.1 RESULTATS DE COMPARAISON

Nous présentons maintenant les résultats des quatre comparaisons effectuées (Figure 11-4).

Comparaison 1 :

Seules 26 forêts sur les 48 échantillonnées ont été détectées par la méthode développée pour extraire automatiquement les forêts de la BD Ortho® IRC. Pour rappel, cette méthode a fourni des résultats encourageants ($\kappa = 0.72$) mais a tout de même entraîné beaucoup d'erreurs grossières ayant fortement altéré les petits objets notamment (cf. section 7.3.3). Par conséquent, elle a aussi eu des effets importants sur la complétude du jeu de forêts échantillonnés.

Le modèle 1 a donc été conduit avec ces 26 individus détectées automatiquement et comparé à un modèle basé sur ces 26 mêmes bois cette fois ci extraits manuellement. Les résultats de la sélection du modèle le plus parcimonieux montrent que 5 variables identiques ont été retenues pour ces deux modèles (*connectivité2010*, *Nb_ch* et *Nb_mh*, *Nb_aqua* et *continuité*). Le pouvoir explicatif de ces deux modèles est très proche ($D2\% = 44.35$ contre $D2\% = 43.71$) tandis que l'écart d'AIC indique que les modèles sont équivalents ($\Delta AIC=0,36$). Pour rappel, on peut considérer que deux modèles diffèrent si la valeur de l'AIC est supérieure à 2. Le niveau de significativité associé aux variables retenues est lui aussi identique. On peut donc conclure qu'il n'y a à priori aucun impact des erreurs d'extraction sur les effets mis en évidence par le modèle. Toutefois, il existe un risque d'avoir un modèle moins robuste car basé sur un plus petit nombre d'observations (réduction de 46%). Plus exactement, on s'aperçoit que les effets mis en évidence en se basant sur un échantillon de taille plus réduite sont différents de ceux observés avec le jeu d'échantillons complet. La réduction de l'effectif a bouleversé l'échantillonnage des bois. Ceci a eu pour conséquence d'« effacer » l'effet de la surface des bois qui n'apparaît plus dans les modèles restreints à 26 individus. Le gradient de surface parmi les forêts échantillonnées n'existe plus, il n'est donc plus possible d'observer un effet de ce paramètre sur la richesse actuelle de syrphes forestiers.

Comparaison 2 :

Nous analysons maintenant l'effet de la BD AUTO 2010 sur la réponse du modèle 2.

A nouveau, les erreurs de classification entraînées par la méthode nous ont amené à conduire un modèle avec seulement 20 bois (Modèle 2 AUTO2010) étant donné que le modèle initial a été construit à partir de seulement 28 bois (les forêts anciennes). Ce dernier modèle a été comparé au modèle 2 MANUEL basé sur ces 20 mêmes forêts.

Les différences entre ces deux modèles sont importantes. Premièrement, les variables retenues par la procédure de sélection diffèrent fortement. Le modèle 2 AUTO2010 inclut quatre variables (*log(Surface2010 auto)*, *ΔCon1954-1900*, *ΔCon1900-1850* et *ΔCon2010-1850auto*) dont une seulement affiche un effet significatif (*ΔCon2010-1850auto*). Le modèle 2 MANUEL ne comporte que deux variables différentes (*ΔCon2010-1979* et *ΔCon1900-1850*) ayant toutes les deux des effets significatifs sur la variable réponse. Le pouvoir explicatif (ou la qualité de l'ajustement) de ce dernier modèle est bien plus important que le premier ($D2\% = 41.28$ pour le modèle basé MANUEL contre $D2\% = 22.78\%$ pour le modèle basé AUTO) et l'écart d'AIC enregistré ($\Delta AIC = 5.58$) affiche une différence significative. On note d'ailleurs que deux variables retenues dans le cadre du modèle 2 AUTO2010 ont été construites sur la base des fragments extraits automatiquement (*log(Surface2010 auto)* et *ΔCon2010-1850auto*), ce qui indique que la BD AUTO2010 a fortement biaisé la réponse de ce modèle. Une nouvelle fois, la réduction de l'effectif a considérablement impacté les relations observées puisqu'aucune des deux variables retenues dans le cadre du modèle 2 TEMOIN (avec 28 individus) ont été sélectionnées au sein du modèle 2 MANUEL (avec 20 individus).

Comparaison 3 :

La comparaison 3 vise à évaluer l'effet de l'incertitude introduite par la BD AUTO 1850 dans le modèle 2 (2010 étant uniquement issue de la BD MANUEL).

Cette fois, la totalité des bois nécessaires à la construction du modèle 2 ont été détectés par la méthode d'extraction automatique à partir des dessins-minutes d'Etat-Major. Le nombre d'observations n'est donc pas différent du modèle 2 initial présenté en chapitre 10. Pour ce modèle (modèle témoin), l'effectif des forêts était de 28. Premièrement, les deux variables retenues au sein du modèle 2 TEMOIN l'ont été également dans le cadre du modèle 2 AUTO1850 à savoir, la surface des patchs présents en 2010 et l'évolution de la surface entre 1979 et 2010. Néanmoins, ce dernier modèle intègre également trois autres variables (*Connectivité2010*, *ΔCon1979-1954* et *ΔCon1900-*

1850auto) participant à accroître de manière erronée le pouvoir explicatif du modèle ($D2\% = 35.26$ contre $D2\% = 22.78$). En effet, plus le nombre de variables explicatives augmente au sein d'un modèle, plus son pouvoir explicatif augmente (compromis biais/variance, chapitre 10). Parmi elles, seule une variable a été construite à partir des patches extraits automatiquement en 1850 tandis que les deux autres (*Connectivité2010*, $\Delta Con1979-1954$) n'ont à priori aucuns liens avec ces patches. Ceci montre que l'erreur introduite par la BD AUTO 1850 a non seulement biaisé l'effet des variables construites à partir de cette base mais s'est aussi propagé en bouleversant les interactions potentielles qui peuvent exister entre variables explicatives. Pourtant, la qualité de la vectorisation automatique de 1850 sur le site d'étude est bonne ($\kappa = 0.89$) mais des erreurs minimes peuvent entraîner des modifications importantes dans la réponse des modèles notamment lorsque l'effectif est peu important ($N = 28$).

Comparaison 4 :

Cette dernière comparaison a pour objectif d'évaluer les effets combinés des BD AUTO 1850 et 2010 sur la réponse du modèle 2. Le modèle construit à partir de ces deux bases intègre 20 forêts du fait des erreurs introduites par la méthode d'extraction automatique des forêts présentes en 2010. Un modèle 2 MANUEL a par conséquent été construit avec ces vingt mêmes bois. Il est d'ailleurs identique à celui défini dans le cadre de la comparaison 2, ce qui indique que les observations concernant le biais introduit par la réduction de l'effectif sur la réponse du modèle sont similaires.

Concernant les modèles basés sur vingt forêts, leur structure diffère. Le modèle 2 « AUTO 1850+2010 » contient une seule variable ($\Delta Con2010-1850auto$) tandis que le modèle 2 « MANUEL » a retenu deux autres variables explicatives différentes ($\Delta Con2010-1979$ et $\Delta Con1900-1850$). La différence de déviance expliquée est importante ($D2\% = 29$ contre $D2\% = 41.28$) tandis que les scores d'AIC associées aux deux modèles sont très proches et n'affichent pas de différence significatives entre les modèles ($\Delta AIC = 0.24$). Il est intéressant de voir que la variable retenue dans le modèle 2 AUTO est construite exclusivement à partir des patches extraits automatiquement en 1850 et en 2010. La combinaison des erreurs de classification a impacté le calcul de l'évolution de la connectivité des bois entre 1850 et 2010 entraînant de fait un effet erroné de cette variable explicative sur la richesse des syrphes forestiers. Le calcul de variables basées sur deux BD incertaines est doublement impacté par l'incertitude de classification et peut avoir des effets non-négligeables sur la réponse des modèles espèces-habitat.

MODELES	EFFECTIF	BASE DE DONNEES EN ENTREE	VARIABLES EXPLICATIVES	COEFFICIENTS	NIVEAU DE SIGNIFICATIVITE	AIC	D2AJUSTEE%	CONTRIBUTION RELATIVE %
COMPARAISON 1								
Modèle 1 (MANUEL) - TEMOIN	N = 48	BD2010(MANU) BDHétérogénéité	<i>log(Surface2010)</i> <i>Connectivité2010</i> <i>Continuité</i> <i>nb_mh</i> <i>nb_mh:Age</i>	0.1766 0.0244 -0.0631 -0.4325 0.0310	** . ** * **	176.76	27.90	36.38 21.84 14.99 11.74 15.05
Modèle 1 (AUTO2010)	N = 26	BD2010(AUTO) BDHétérogénéité	<i>Connectivité2010auto</i> <i>Nb_chand</i> <i>Nb_mh</i> <i>Nb_aqua</i> <i>Continuité</i>	0.0492 0.2495 -0.0362 0.7200 -0.4236	* . . . *	97.49	44.35	38.37 14.96 22.82 11.63 12.22
Sous-Modèle 1 (MANUEL)	N = 26	BD2010(MANU) BD Hétérogénéité	<i>Connectivité2010</i> <i>Nb_chand</i> <i>Nb_mh</i> <i>Nb_aqua</i> <i>Continuité</i>	0.0421 0.2895 -0.0345 0.6702 -0.4611	* * . . *	97.13	43.71	35.89 18.43 22.07 10.45 13.15
COMPARAISON 2								
Modèle 2 (MANUEL) - TEMOIN	N = 28	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>log(Surface2010)</i> <i>Δsurf2010-1979</i>	0.2149 -0.0140	** *	100.63	29.24	56.53 43.47
Modèle 2 (AUTO2010)	N = 20	BD2010(AUTO) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>log(Surface2010auto)</i> <i>ΔCon1954-1900</i> <i>ΔCon1900-1850</i> <i>ΔCon2010-1850auto</i>	0.0043 -0.0086 -0.0073 0.0060	. . . *	78.04	22.78	93.71 1.87 3.57 0.84
Sous-Modèle 2 (MANUEL)	N = 20	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>ΔCon2010-1979</i> <i>ΔCon1900-1850</i>	0.0527 -0.0066	** *	72.46	41.28	82.81 17.19

MODELES	EFFECTIF	BASE DE DONNEES EN ENTREE	VARIABLES EXPLICATIVES	COEFFICIENTS	NIVEAU DE SIGNIFICATIVITE	AIC	D2AJUSTEE%	CONTRIBUTION RELATIVE %
COMPARAISON 3								
Modèle 2 (MANUEL) TEMOIN	N = 28	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>log(Surface2010)</i> <i>Δsurf2010-1979</i>	0.2149 -0.0140	** *	100.63	29.24	56.53 43.47
Modèle 2 (AUTO1850)	N = 28	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(AUTO)	<i>log(Surface2010auto)</i> <i>Connectivité2010</i> <i>ΔSurf2010-1979</i> <i>ΔCon1979-1954</i> <i>ΔCon1900-1850auto</i>	0.2116 0.0569 -0.0152 -0.0252 -0.0060	** * * * *	103.31	35.26	36.27 18.75 22.33 9.72 12.93
COMPARAISON 4								
Modèle 2 (MANUEL) – TEMOIN	N = 28	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>log(Surface2010)</i> <i>Δsurf2010-1979</i>	0.2149 -0.0140	** *	100.63	29.24	56.53 43.47
Modèle 2 (AUTO1850+2010))	N = 20	BD2010(AUTO) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(AUTO)	<i>ΔCon20101850auto</i>	0.0065	*	72.22	29.00	100.00
Sous-Modèle 2 (MANUEL)	N = 20	BD2010(MANU) BD1979(MANU) BD1954(MANU) BD1900(MANU) BD1850(MANU)	<i>ΔCon2010-1979</i> <i>ΔCon1900-1850</i>	0.05268 -0.0066	** *	72.46	41.28	82.80698 17.19302

Tableau 11-2. Résultats de comparaison des modèles espèces-habitat

11.4. BILAN

Nous avons évalué l'impact du processus d'extraction automatique des forêts à partir des dessins-minutes d'Etat-major et de la BD Ortho®2010 de l'IGN. Les résultats montrent un impact important de ces erreurs sur la réponse des modèles, ce qui pourrait conduire à interpréter différemment les relations espèces-habitats.

Le premier impact notable porte sur la complétude des modèles écologiques, notamment en ce qui concerne la BD AUTO 2010. Les modèles intégrant ces variables ont été réduit à 26 individus dans le cas du modèle 1 et à 20 individus dans le cadre du modèle 2. La réduction de l'effectif entraîne une modification de l'échantillonnage initial ayant tendance à effacer les effets de plusieurs paramètres. La comparaison 1 a notamment mis en évidence la disparation de l'effet de la surface des bois après la réduction de l'effectif. De manière similaire, aucune des deux variables retenues au sein du modèle 2 TEMOIN n'a été sélectionnée au sein du sous-modèle 2 MANUEL. Ces observations sont d'une importance considérable car elles pourraient conduire à interpréter différemment la réponse des modèles. Dans le premier cas (le modèle 1), nous pourrions conclure que la taille des bois ne joue aucun rôle sur la richesse de syrphes alors qu'elle est la variable qui contribue le plus à l'explication de la variable réponse dans le modèle témoin. Dans le second cas (modèle 2), nous aurions pu nous focaliser de manière erronée sur d'autres transitions paysagères (e.g. $\Delta Con2010-1979$ $\Delta Con1900-1850$) en mettant de côté celle qui a joué un rôle crucial dans la présence actuelle de syrphes forestiers ($\Delta surf2010-1979$).

Deuxièmement, les erreurs d'extraction automatique peuvent faire varier les interactions potentielles entre variables explicatives et modifier la sélection du modèle le plus parcimonieux. Basé sur un effectif identique à celui du modèle TEMOIN, le modèle 2 AUTO1850 comprend 5 variables dont deux seulement ont été retenus dans le modèle TEMOIN. Parmi les trois autres variables, deux sont construites à partir des BD forêts produites manuellement et ne sont pas concernées a priori par les incertitudes de la BD AUTO 1850. De fait, on comprend qu'une légère fluctuation d'une ou plusieurs variables intégrées au sein des modèles peut masquer ou amplifier les effets des autres variables mobilisées même si elle a évidemment des effets directs sur les variables construites à partir des bois extraits automatiquement.

Enfin, la comparaison 4 a permis de démontrer que la combinaison de plusieurs BD incertaines pouvait accentuer le bais au sein des modèles. Le modèle 2 (AUTO 1850+2010) n'a retenu qu'une seule variable définie exclusivement à l'aide des BD extraites automatiquement alors que celle-ci n'avait pas été sélectionné dans le cadre du modèle TEMOIN ou du « sous-modèle 2 MANUEL ». Le calcul des dynamiques paysagères doit par conséquent s'appuyer sur des bases de données spatiales les plus précises possibles pour éviter la propagation des erreurs de classification.

CONCLUSION DE LA PARTIE D

La dernière partie de cette thèse a concerné l'étude des effets des évolutions des forêts sur la richesse actuelle de syrphes forestiers dans les coteaux de Gascogne.

L'intérêt du chapitre 10 a été premièrement d'explicitier la démarche de modélisation utilisée. Nous nous sommes tournés vers une approche corrélative et avons adopté des modèles linéaires généralisés (GLM) pour modéliser les relations entre l'histoire des habitats et notre variable de présence biologique. La stratégie de modélisation a ensuite fait l'objet d'un travail collaboratif entre plusieurs membres de l'UMR DYNAFOR et du Swiss Federal Research Institute (WSL). Les résultats ont alors démontré l'influence de l'ancienneté des forêts sur la richesse de syrphes et ont permis de mettre en évidence à la fois la présence d'une dette d'extinction et d'un crédit de colonisation pour la population suivie.

Le second volet de cette partie (chapitre 11) a concerné l'impact de l'incertitude spatiale sur la réponse des modèles écologiques définis dans le chapitre 10. Les bases de données extraites automatiquement ont été directement introduites au sein des modèles écologiques pour lesquels nous avons ensuite comparé leur réponse à celle des modèles construits à l'aide des BD produites manuellement. Nous avons constaté un fort impact des erreurs de classification qui pourrait conduire à interpréter différemment les modèles écologiques.

Ce chapitre a fait l'objet de travaux de valorisation scientifique :

HERRAULT P.-A., LARRIEU L., OUIN A., SHEEREN D. 2015. (Poster) Influence of map accuracy on modeling species diversity of hoverflies, *9th International Association of Landscape Ecology World Congress (IALE'2015)*, Portland (USA), 5-10 July 2015, accepted

HERRAULT P.-A., OUIN A., SHEEREN D. 2014. Spatial uncertainty effects on a species-landscape relationship model in ecology, *11th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Spatial Accuracy'2014)*, Michigan State University (USA), 8-11 July 2014.

CONCLUSION GENERALE

L'objectif principal poursuivi dans cette thèse est le développement d'une **démarche méthodologique** visant à **reconstruire les évolutions spatio-temporelles des habitats forestiers pour évaluer leurs effets** sur la richesse spécifique de **syrrhes forestiers**. Nous avons combiné pour cela à la fois des méthodes et techniques employées en géomatique et des approches écologiques. Par conséquent, contrairement aux travaux abordant l'une ou l'autre discipline, cette thèse constitue une réelle originalité en proposant une chaîne de traitement méthodologique globale, de la sélection des données spatiales à l'évaluation des effets du paysage sur la biodiversité.

Principaux résultats :

Le premier volet de cette thèse a concerné la production d'une base de données spatiale historique forestière. Différentes sources ont été sélectionnées, dont deux cartes anciennes, permettant d'obtenir des informations sur la configuration spatiale des forêts depuis le milieu du 19^{ème} siècle à cinq moments de temps : 1850, 1900, 1954, 1979, 2010. Nous nous sommes ensuite concentrés sur le processus de correction géométrique des cartes anciennes, étape indispensable pour intégrer ces documents dans un référentiel cartographique actuel et permettre leur comparaison avec des données plus récentes. Une nouvelle méthode de correction géométrique basée sur la régression *ridge* à noyau a été proposée. Cette approche s'est avérée être une alternative possible aux modèles globaux conventionnels (les méthodes polynomiales) et a fourni des résultats très proches de la méthode de correction locale fondée sur la Triangulation de Delaunay, qui s'est avéré être la plus efficace. Notre attention s'est ensuite tournée vers le problème de la vectorisation automatique des cartes anciennes et plus spécifiquement les dessins-minutes d'Etat-Major du 19^{ème} siècle au 1/40 000^e. En nous appuyant sur différentes techniques de traitement d'images, nous avons développé une nouvelle méthode permettant d'identifier automatiquement les surfaces boisées dans ces dessins-minutes avec un niveau de performance élevé. Cette méthode s'est révélée très efficace sur les cartes les moins surchargées et très encourageante sur les cartes souffrant d'un « bruit » important lié à la superposition des éléments. Pour ces dernières, l'intervention de l'utilisateur semble requise pour corriger les artéfacts persistants entraînés par notre méthode. Enfin, le dernier point de cette partie a porté sur le développement d'une méthode d'extraction automatisée des surfaces boisées à partir de la BD Ortho IRC de 2010. Pour cela, nous avons combiné une approche orientée objet d'analyse d'image et un modèle de classification par apprentissage supervisé basé sur les arbres de décision. Les résultats obtenus, bien qu'exploitables, ont souligné les limites de cette méthode pour les forêts de petite taille.

Le second volet du travail a consisté à définir un cadre méthodologique pour caractériser les évolutions spatio-temporelles des forêts. Nous avons premièrement procédé à un appariement manuel basé sur un examen visuel des objets homologues aux différentes dates pour reconstituer la généalogie de chaque forêt échantillonnée. Puis, nous avons caractérisé la structure spatiale des bois appariés à l'aide de métriques paysagères. Elles visent à qualifier deux aspects clés des habitats forestiers susceptibles d'affecter la biodiversité forestière: leur surface et leur isolement. Enfin, deux approches ont été proposées pour caractériser à proprement parler les évolutions spatio-temporelles des forêts suivies. L'une a consisté à calculer les taux de changements des métriques calculées et permet d'obtenir une vision claire des évolutions de chaque forêt entre deux dates, nous l'avons baptisé d'approche « individu ». L'autre a visé plutôt à faire ressortir des patrons d'évolutions parmi l'ensemble des forêts suivies, c'est l'approche « population ». Une analyse statistique multivariée a ensuite été mise en œuvre pour représenter l'évolution de chaque forêt sous la forme d'une trajectoire sur lesquelles nous nous sommes ensuite appuyés pour établir une typologie. D'une

manière générale, les résultats de l'une et l'autre approche ont fait ressortir des tendances d'évolutions très hétérogènes qui montrent la complexité de mener des analyses à l'échelle de l'objet. En effet, bien qu'on puisse tirer certaines conclusions au sujet des évolutions à l'échelle des paysages, telle que la fragmentation des forêts, ceci ne se vérifie pas toujours pour l'ensemble des forêts de ce paysage.

Enfin, le dernier aspect de cette thèse a eu pour objectif de modéliser les effets des évolutions spatio-temporelles des forêts sur la présence actuelle de syrphes forestiers et plus exactement, sur la richesse spécifique. Le développement de l'approche proposée s'est inspirée des travaux récents en écologie historique visant à vérifier la présence d'une dette d'extinction ou d'un crédit de colonisation pour des habitats ayant connus une perturbation. Une approche statistique fondée sur l'emploi de modèles linéaires généralisés (GLM) et intégrant des variables d'évolutions spatio-temporelles a été mise en œuvre et a permis de mettre en évidence l'effet de l'histoire sur la richesse spécifique en syrphes forestiers, en plus d'autres effets tels que la taille et la connectivité des forêts de même que le nombre de types de micro-habitats. La richesse spécifique est apparue plus importante dans les forêts anciennes que dans les forêts récentes. Nous avons également mis en lumière la présence à la fois d'une dette d'extinction et d'un crédit de colonisation sur la dernière période de temps étudiée (entre 1979 et 2010) et ceci, pour les forêts faiblement connectées. Enfin, le dernier volet de cette partie a évalué les effets potentiels des erreurs d'extraction automatique sur la réponse des modèles espèces-habitats établis. Les résultats de cette étude préliminaire ont fait ressortir un impact relativement important de ce type d'incertitude sur la validité des modèles.

Contributions en géomatique :

La méthode de correction géométrique proposée permet de résoudre deux problèmes majeurs dans l'emploi des modèles de transformation géométrique. Premièrement, elle introduit un paramètre de régularisation (pénalisation *ridge*) qui pénalise les solutions non-régulières et permet de s'affranchir des problèmes liés à l'estimation des paramètres lorsque le nombre d'observations est faible. Deuxièmement, l'utilisation des noyaux et particulièrement le noyau gaussien permet de prendre en compte les distorsions locales de la zone d'intérêt tout en fournissant un seul modèle global de transformation. Elle permet donc de s'affranchir des éventuels problèmes liés à la partition de Delaunay qui peut avoir tendance à déformer grossièrement certains objets à cheval sur plusieurs partitions.

La méthode d'extraction automatique des surfaces boisées à partir des dessins-minutes d'Etat-major dispose aussi de nombreux avantages. Premièrement, elle possède une réelle portée opérationnelle comme notre participation au projet OCS ANCIEN en témoigne. Les tests effectués dans ce cadre ont aussi permis de montrer qu'elle pouvait s'appliquer aux différents thèmes présents sur ces cartes et pas seulement aux forêts. Deuxièmement, elle s'appuie sur des techniques connues, facilement paramétrables et indépendantes les unes des autres permettant à l'utilisateur de l'adapter en fonction de ses besoins. Troisièmement, un outil a été implémenté au sein de la plateforme ENVI et est aujourd'hui en libre accès, ce qui facilite son utilisation à l'échelle nationale. Il a d'ailleurs déjà été utilisé dans le Programme Interdisciplinaire de Recherche sur l'Environnement de la Seine (PIREN Seine).

Contributions en écologie :

D'un point de vue thématique, cette thèse a permis de fournir des résultats importants en écologie du paysage. Notre étude a confirmé l'effet de l'histoire sur la richesse spécifique de syrphes forestier en mettant en lumière la présence d'une dette d'extinction et d'un crédit de colonisation pour les forêts isolées. Cela n'avait jamais été démontré en écologie pour les syrphes forestiers. Nous avons mis en évidence que la richesse en espèce dépend de différents facteurs agissant à différentes échelles : la taille des bois, l'isolement, l'hétérogénéité interne des peuplements et la dimension

temporelle. Ces résultats confirment l'importance de l'effet de l'histoire pour expliquer la diversité biologique actuelle. Elles soulignent aussi l'intérêt de s'intéresser aux espèces supposées moins sensibles aux effets de l'ancienneté des forêts car plus mobiles. Nous avons montré qu'une durée de 30 ans n'était visiblement pas suffisante pour que certaines espèces de syrphes soient en mesure de s'adapter en réponse à une modification de leur habitat, lorsque celui-ci est peu connecté. Deuxièmement, il est nécessaire de mettre en place des mesures de conservation à différentes échelles spatiales pour favoriser au maximum la diversité des syrphes. En effet, le maintien d'une grande superficie, d'une connectivité importante et le développement des stades de maturité au sein des forêts apparaissent comme trois composantes clés du maintien de la diversité des syrphes dans les paysages agricoles.

Perspectives :

Tout d'abord, nous réfléchissons à la possibilité d'automatiser la méthode d'appariement. Notre approche manuelle a pu être mise en application car nous suivions un groupe de 48 forêts relativement petit mais il est clair qu'elle montrera ses limites dès lors que le nombre d'individus augmentera.

Concernant la méthode d'extraction automatique à partir des cartes anciennes, nous pensons qu'elle pourrait être améliorée en adaptant plus localement la première étape de nettoyage au relief de la zone à transformer. En s'appuyant sur un modèle numérique de terrain, il serait par exemple possible de cibler les zones de hachures les plus denses de manière implicite pour ensuite mieux calibrer les différents filtres utilisés. De plus, l'espace colorimétrique CIELab employé dans cette approche pourrait permettre de régler les problèmes de non-uniformité de la teinte pour différents types de données. Les photographies aériennes anciennes parfois impactées par les effets de vignettage ou de luminosité variable sont un des exemples possibles où le CIELab pourrait démontrer toute son efficacité.

L'approche automatique proposée pour extraire les surfaces boisées de la BD Ortho® IRC est aussi largement perfectible. Pour l'améliorer, nous envisageons deux possibilités. La première est d'appliquer plusieurs niveaux de segmentation pour tenir compte de la variabilité de la taille des espaces boisés. La seconde est d'adapter les paramètres des opérateurs de post-traitements en fonction de la taille des objets. Cela pourrait ainsi éviter de tronquer trop fortement certains objets et de fait, améliorer les résultats de cette approche.

La méthode de correction géométrique proposée pourrait s'avérer très intéressante dès lors que l'on souhaite corriger une image où peu de solutions existent pour récolter des points d'amers. Cela peut être le cas des cartes anciennes mais pas seulement. La correction d'images satellitaires ou photographies aériennes acquises sur des zones souffrant du même type de problèmes, dans les zones de relief par exemple, pourrait profiter d'une telle application. Nous prévoyons également d'implémenter la méthode dans la plateforme Monteverdi/OTB en langage python afin de fournir un outil libre d'accès aux différents utilisateurs susceptibles de l'utiliser.

L'approche de caractérisation des évolutions spatio-temporelles est potentiellement applicable à différents types de milieux. Les études d'écologie historique publiées ces dernières années montrent que les prairies sont elles aussi fortement impactées par la fragmentation et peuvent abriter des espèces susceptibles d'être affectés par ces bouleversements.

Nous prévoyons d'aborder plus en détail la représentation spatiale des incertitudes. Bien que plusieurs études aient abordé ce problème depuis plusieurs années, peu d'entre elles sont capables d'estimer l'incertitude sur les changements et encore moins de les spatialiser. Les recherches dans ce domaine pourraient ainsi permettre aux écologues d'avoir une idée plus claire des incertitudes existantes sur la zone qu'il traite et de les alerter sur les potentiels biais que cela pourrait représenter

pour les modèles espèces-habitat. Deux types d'erreurs pourraient faire l'objet d'une spatialisation : les erreurs de classification et les erreurs de correction géométriques.

D'un point de vue écologique, nous n'avons testé qu'une partie des approches présentées par Kuussaari et al. (2009) pour vérifier la présence d'une dette d'extinction ou d'un crédit de colonisation. Par conséquent, les autres méthodes proposées pourraient être appliquées et fournir d'avantage de connaissances, notamment sur la magnitude de la dette ou du crédit d'extinction.

Enfin, nous nous sommes spécifiquement concentrés sur la biodiversité alpha. Une autre possibilité serait d'évaluer la biodiversité bêta en comparant les compositions en espèces des différents bois échantillonnés. Cela permettrait de vérifier les effets de l'histoire sur l'hétérogénéité de composition à l'échelle du groupe de forêts pour mieux identifier localement les espèces concernées.

BIBLIOGRAPHIE

CHAPITRE 1

Alard D., Poudevigne I, 2002, Biodiversity in changing landscapes, from species or patch assemblages to system organization, in *Application of Geographic Information Systems and Remote Sensing in River Studies*, Leiden, The Netherlands 9-24.

Allen TF, Starr TB, 1982 *Hierarchy, perspectives for ecological complexity*, University of Chicago Press Chicago, 310 p

Arrhenius O, 1921 Species and area, *Journal of Ecology*, 9, pp 95-99

Beese WJ, Bryant AA, 1999, Effect of alternative silvicultural systems on vegetation and bird communities in coastal montane forests of British Columbia, *Canada. Forest Ecology and Management*, 115, pp 231-242.

Benamammar H, 2014, Caractérisation de la faune ornithologique des Monts de l'Ourit dans le parc national de Tlemcen, Mémoire de Magister en foresterie, Université Abou Bekr Belkaid Tlemcen

Bertrand G, 1975, Pour une histoire écologique de la France rurale, *Histoire de la France rurale*, 1, pp 37-113.

Bolger DT, Suarez AV, Crooks KR, 2000, Arthropods in urban habitat fragments in southern California, area, age, and edge effects, *Ecological Applications*, 10, 1230-1248.

Boorman SA, Levitt PR, 1973, Group selection on the boundary of a stable population. Theoretical Population, *Biology* 4, pp 85-128.

Brooks TM, Pimm SL, Oyugi JO, 1999, Time lag between deforestation and bird extinction in tropical forest fragments, *Conservation Biology*, 13, pp 1140-1150.

Brunet R, Ferras R, Théry H, 1993, *Les mots de la géographie, dictionnaire critique*, La documentation française, 518p

Burel F, Baudry J, 1999, *Écologie du paysage. Concepts, méthodes et applications*, Paris Editions TEC et DOC, 359p

Burnett MR, August PV, Brown JH, Killingbeck KT, 1998, The influence of geomorphological heterogeneity on biodiversity I, A patch-scale perspective, *Conservation Biology*, 12, pp 363-370.

Caughley G, Gunn A, Ralls K, 1996, *Conservation biology in theory and practice*, Blackwell Science Cambridge, Wiley, 459p

Clarke B, Murray J, Johnson MS, 1984, The extinction of endemic species by a program of biological control, *Pacific Science*, 38, pp 97-104.

Dale MR, 1999, Spatial pattern analysis in plant ecology, *Ecology*, 88, pp 366-370.

De Boeck HJ, Nijs I, Lemmens CM, Ceulemans R, 2006, Underlying effects of spatial aggregation, clumping, in relationships between plant diversity and resource uptake, *Oikos*, 113, pp 269-278.

- Décamps H, Décamps O, 2004**, *Au printemps des paysages*, Buchet-Chastel, 240p.
- Deléglise C, Loucougaray G, Alard D, 2011**, Effects of grazing exclusion on the spatial variability of subalpine plant communities, multiscale approach, *Basic and Applied Ecology*, 12, pp 609-619.
- Dennis RL, Shreeve TG, Van Dyck H, 2003**, Towards a functional resource-based concept for habitat, a butterfly biology viewpoint, *Oikos*, pp 417-426.
- Drakare S, Lennon JJ, Hillebrand H, 2006**, The imprint of the geographical, evolutionary and ecological context on species-area relationships, *Ecology Letters*, 9, pp 215-227.
- Duflot R, 2013**, *Hétérogénéité fonctionnelle et biodiversité, quel est le rôle des interfaces ou lisières dans les paysages agricoles?*, Thèse en sciences du vivant/Sciences agricoles, Université Rennes 1.
- Dufour A, Gadallah F, Wagner HH, 2006**, Plant species richness and environmental heterogeneity in a mountain landscape, effects of variability and spatial configuration, *Ecography*, 29, pp 573-584.
- Dufrêne M, 2003**, Méthodes d'analyse des données écologiques et biogéographiques, [http, pp //old.biodiversite.wallonie.be/outils/methodo/](http://old.biodiversite.wallonie.be/outils/methodo/)
- Dufrêne M, 1992**, *Biogéographie et écologie des communautés de Carabidae en Wallonie*, Thèse en écologie des communautés, Université Catholique de Louvain, 194p.
- Dutilleul P, 1993**, Spatial heterogeneity and the design of ecological field experiments, *Ecology*, pp 1646-1658.
- Elton CS, 1927**, *Animal ecology*. University of Chicago Press, 254 p
- Ewers RM, Didham RK, Wratten SD, Tylianakis JM, 2005**, Remotely sensed landscape heterogeneity as a rapid tool for assessing local biodiversity value in a highly modified New Zealand landscape. *Biodiversity & Conservation*, 14, pp 1469-1485.
- Fahrig L, 2003**, Effects of habitat fragmentation on biodiversity, *Annual Review of Ecology, Evolution, and Systematics*, pp 487-515.
- Fahrig L, 2005**, When is a landscape perspective important?, *Issues and Perspectives in Landscape Ecology*, pp 3-10.
- Forman RT, 1995**, Some general principles of landscape and regional ecology. *Landscape Ecology*, 10, pp 133-142.
- Forman RT, Godron M, 1981**, Patches and structural components for a landscape ecology, *BioScience*, 31, pp 733-740.
- Fraser RH, 1998**, Vertebrate species richness at the mesoscale, relative roles of energy and heterogeneity, *Global Ecology & Biogeography Letters*, 7, pp 215-220.
- Freemark KE, Merriam HG, 1986**, Importance of area and habitat heterogeneity to bird assemblages in temperate forest fragments, *Biological Conservation*, 36, pp 115-141.
- Frontier S, Pichod-Viale D, 1995**, *Théorie des écosystèmes, structure, fonctionnement, évolution*, Collection d'Ecologie, Dunod, 558p.
- Grinnell J, 1917**, The niche-relationships of the California Thrasher, *The Auk*, pp 427-433.

- Guisan A, Thuiller W, 2005**, Predicting species distribution, offering more than simple habitat models, *Ecology Letters*, 8, pp. 993-1009.
- Guisan A, Zimmermann NE, 2000**, Predictive habitat distribution models in ecology, *Ecological Modelling*, 135, pp 147-186.
- Hanski I, Gilpin M, 1991**, Metapopulation dynamics, brief history and conceptual domain, *Biological Journal of the Linnean Society*, 42, pp 3-16.
- Harrison S, Taylor AD, 1997**, Empirical evidence for metapopulation dynamics, in *Metapopulation Biology, Ecology, Genetics, and Evolution*, Academic Press, pp 27-42.
- Helaouët P, Beaugrand G, 2009**, Physiology, ecological niches and species distribution, *Ecosystems*, 12, pp 1235-1245.
- Hutchinson, G.E., 1957**, Concluding remarks, *Cold Spring Harbor Symposium Quantitative Biology*, 22, pp 415-427.
- Kati V, Poirazidis K, Dufrière M, 2010**, Towards the use of ecological heterogeneity to design reserve networks, a case study from Dadia National Park, Greece, *Biodiversity and Conservation*, 19, pp 1585-1597.
- Kerr JT, Packer L, 1997**, Habitat heterogeneity as a determinant of mammal species richness in high-energy regions, *Nature*, 385, pp 252-254.
- Kruuk H, Parish T, Brown CAJ, Carrera J, 1979**, The use of pasture by the European badger, *Meles meles*, *Journal of Applied Ecology*, pp 453-459.
- Levins R, 1970**, Extinction, *Lectures on Mathematics in the Life Sciences*, 2, pp 75-107.
- Li H, Reynolds JF, 1995**, On definition and quantification of heterogeneity, *Oikos*, pp 280-284.
- Lopez S, Rousset F, Shaw FH, 2009**, Joint effects of inbreeding and local adaptation on the evolution of genetic load after fragmentation, *Conservation Biology*, 23, pp 1618-1627.
- MacArthur RH, 1967**, *The theory of island biogeography*, Princeton University Press, 224p.
- McGarigal K, Marks BJ, 1995**, Spatial pattern analysis program for quantifying landscape structure.
- McNab BK, 1994**, Energy conservation and the evolution of flightlessness in birds, *American Naturalist*, pp 628-642.
- McNeill SE, Fairweather PG, 1993**, Single large or several small marine reserves? An experimental approach with seagrass fauna, *Journal of Biogeography*, pp 429-440.
- Morin E, 1990**, *Introduction à la pensée complexe*, Points, 158p.
- Noss RF, 1999**, Assessing and monitoring forest biodiversity, a suggested framework and indicators, *Forest ecology and Management*, 115, pp 135-146.
- Pacala SW, Levin SA, 1997**, Biologically generated spatial pattern and the coexistence of competing species, in *Spatial ecology, the role of space in population dynamics and interspecific interactions*, Princeton University, pp 204-232.
- Pausas JG, Lavorel S, 2003**, A hierarchical deductive approach for functional types in disturbed ecosystems, *Journal of Vegetation Science*, 14, pp 409-416.

- Pereboom V, 2006**, Mode d'utilisation du milieu fragmenté par une espèce forestière aux habitudes discrètes, Thèse en écologie, Université de Rennes I, 75p.
- Pickett ST, Cadenasso ML, 1995**, Landscape ecology, spatial heterogeneity in ecological systems. *Science, AAAS-Weekly Paper Edition*, 269, pp 331-333.
- Pimm SL, 1998**, Ecology, The forest fragment classic, *Nature*, 393, pp 23-24.
- Pulliam HR, 2000**, On the relationship between niche and distribution, *Ecology Letters*, 3, pp 349-361.
- Pulliam HR, 1988**, Sources, sinks, and population regulation, *American Naturalist*, pp 652-661.
- Redon M, 2012**, Biodiversité potentielle dans les forêts du Vercors, une approche hiérarchique pour la conservation des espaces forestiers, Thèse en écologie, Université de Grenoble, 194p
- Rescia AJ, Schmitz MF, Martin de Agar P, 1994**, Influence of landscape complexity and land management on woody plant diversity in northern Spain, *Journal of Vegetation Science*, 5, pp 505-516.
- Ricketts TH, 2001**, The matrix matters, effective isolation in fragmented landscapes, *The American Naturalist*, 158, pp 87-99.
- Risser PG, 1987**, Landscape ecology, state of the art, in *Landscape Heterogeneity and Disturbance*. Springer, pp 3-14.
- Rosignol N, 2006**, Heterogeneity of the vegetation and of the grazing, functional consequences in natural grassland, Thèse en écologie, Université Rennes I, 198p.
- Ruas A, 2004**, Le changement de niveau de détail dans la représentation de l'information géographique, in *Symposium on Spatial Data Handling, SDH'98*, Canada/Vancouver, 235p.
- Ryan P, Glass JP, da Cunha T, 2001**, Inaccessible Island nature reserve management plan. Government of Tristan da Cunha Edinburgh, Tristan da Cunha, [en ligne, URL: http://jncc.defra.gov.uk/pdf/OT_TristandaC.pdf].
- Skov F, Svenning J-C, 2003**, Predicting plant species richness in a managed forest, *Forest Ecology and Management*, 180, pp 583-593.
- Smith FE, 1972**, Spatial heterogeneity, stability, and diversity in ecosystems, *Academy Arts Sciences*, 44, 335
- Soulé ME, Simberloff D, 1986**, What do genetics and ecology tell us about the design of nature reserves?, *Biological Conservation*, 35, pp 19-40.
- Steiner NC, Köhler W, 2003**, Effects of landscape patterns on species richness—a modelling a roach, *Agriculture, Ecosystems & Environment*, 98, pp 353-361.
- Tamme R, Hiiesalu I, Laanisto L, 2010**, Environmental heterogeneity, species diversity and co-existence at different spatial scales. *Journal of Vegetation Science*, 21, pp 796-801.
- Tilman D, May RM, Lehman CL, Nowak MA, 1994**, Habitat destruction and the extinction debt, *Nature*, 371, pp 65-66.

Troll C, 1939, Luftbildplan und ökologische Bodenforschung, ihr zweckmäßiger Einsatz für die wissenschaftliche Erforschung und praktische Erschließung wenig bekannter Länder. *Ges. für Erdkunde*.

Turnbull LA, Coomes DA, Purves DW, Rees M, 2007, How spatial structure alters population and community dynamics in a natural plant community, *Journal of Ecology*, 95, pp 79-89.

Turner MG, 2005, Landscape ecology, what is the state of the science?, *Annual Review of Ecology, Evolution, and Systematics*, 36, 319-344.

Turner MG, 1989, Landscape ecology, the effect of pattern on process, *Annual Review of Ecology and Systematics*, 20, pp 171-197.

Virgos E, Casanovas JG, 1999, Badger *Meles meles* sett site selection in low density Mediterranean areas of central Spain, *Acta Theriologica*, 44, pp 173-182.

Virolainen KM, Suomi T, Suhonen J, Kuitunen M, 1998, Conservation of vascular plants in single large and several small mires, pp species richness, rarity and taxonomic diversity, *Journal of Applied Ecology*, 35, pp 700-707.

Wu JJ, 2013, Landscape ecology, in *Ecological Systems*, Springer, pp 179-200.

Wu J, Loucks OL, 1995, From balance of nature to hierarchical patch dynamics, a paradigm shift in ecology, *Quarterly Review of Biology*, pp 439-466.

CHAPITRE 2

Baudry J, Merriam H.G 1988, Connectivity and connectedness: functional versus structural patterns in landscapes, in *Connectivity and Landscape Ecology*, Ed Schreiber K-F, pp 23-28.

Bellemare J, Motzkin G, Foster DR, 2002, Legacies of the agricultural past in the forested present, an assessment of historical land-use effects on rich mesic forests, *Journal of Biogeography*, 29, pp 1401-1420.

Bommarco R, Lindborg R, Marini L, Öckinger E, 2014, Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history, *Diversity and Distributions*, 20, pp 591-599.

Bossuyt B, Hermy M, Deckers J, 1999, Migration of herbaceous plant species across ancient-recent forest ecotones in central Belgium, *Journal of Ecology*, 87, 629-638.

Bouget C, Larrieu L, Nusillard B, Parmain G, 2013, In search of the best local habitat drivers for saproxylic beetle diversity in temperate deciduous forests, *Biodiversity and Conservation*, 22, pp 2111-2130.

Cateau E, Larrieu L, Vallauri D, Brustel H, 2015, Ancienneté et maturité, deux qualités complémentaires d'un écosystème forestier, *Comptes Rendus Biologies*, 338, pp 58-73.

Cinotti B, 1996, Evolution des surfaces boisées en France : proposition de reconstitution depuis le début du XIXe siècle, *Revue Forestière Française*, 6, pp 512-562.

Collin E, Lefèvre F, Muratorio SO, 2012, Indicateurs de la diversité intra-spécifique chez les arbres forestiers. Les indicateurs de biodiversité forestière, *Synthèse des réflexions issues du programme de recherche " Biodiversité, gestion forestière et politiques publiques "*, pp 79-81.

Compton JE, Boone RD, 2000, Long-term impacts of agriculture on soil carbon and nitrogen in New England forests, *Ecology*, 81, pp 2314-2330.

- Dambrine E, Dupouey J-L, Laüt L, 2007**, Present forest biodiversity patterns in France related to former Roman agriculture, *Ecology*, 88, pp 1430-1439.
- Decocq G, Vieille V, Racinet P, 2002**, Influence des facteurs historiques sur la végétation actuelle, le cas des mottes castrales en milieu forestier, Picardie, France, *Acta Botanica Gallica*, 149, pp 197-215.
- Deconchat M, Balent G, 1996**, Biodiversité et forêt dans un paysage agricole, *Étude bibliographique. Études et Recherches sur les Systèmes Agraires et le Développement*, pp 15-36.
- Donohue K, Foster DR, Motzkin G, 2000**, Effects of the past and the present on species distribution, land-use history and demography of wintergreen, *Journal of Ecology*, 88, pp 303-316.
- Dupouey J-L, Dambrine E, Laffite J-D, Moares C, 2002**, Irreversible impact of past land use on forest soils and biodiversity, *Ecology*, 83, pp 2978-2984.
- Fahrig L, Merriam G, 1985**, Habitat patch connectivity and population survival, *Ecology*, pp 1762-1768.
- Fairbanks DH, McGwire KC, 2004**, Patterns of floristic richness in vegetation communities of California, regional scale analysis with multi-temporal NDVI, *Global Ecology and Biogeography*, 13, pp 221-235.
- Fischer J, Lindenmayer DB, 2002**, Small patches can be valuable for biodiversity conservation, two case studies on birds in southeastern Australia, *Biological Conservation*, 106, pp 129-136.
- Flinn KM, Vellend M, 2005**, Recovery of forest plant communities in post-agricultural landscapes. *Frontiers in Ecology and the Environment*, 3, pp 243-250.
- Forman RT, Baudry J, 1984**, Hedgerows and hedgerow networks in landscape ecology. *Environmental management*, 8, pp 495-510.
- Foster DR, 1988**, Disturbance history, community organization and vegetation dynamics of the old-growth Pisgah Forest, south-western New Hampshire, USA, *The Journal of Ecology*, pp 105-134.
- Foster DR, 1992**, Land-use history, 1730-1990, and vegetation dynamics in central New England, USA, *Journal of Ecology*, pp 753-771.
- Fuller RJ, Warren M, 1991**, Conservation management in ancient and modern woodlands, responses of fauna to edges and rotations, in *The Scientific Management of Temperate Communities for Conservation. 31st Symposium of British Ecological Society*, Ed Spellerberg, I.F., Goldsmith, F.B. & Morris, M.G, pp 445-471.
- Gilg O, 2004**, Old-growth forests. Characteristics, *Conservation and Monitoring*, 74, pp 1-96.
- Gillespie TW, Foody GM, Rocchini D, 2008**, Measuring and modelling biodiversity from space, *Progress in Physical Geography*, 32, pp 203-221.
- Gosselin F, Bouget C, Gosselin M, 2009**, L'Etat et les enjeux de biodiversité forestière, in *Biomasse et Biodiversité forestières. Augmentation de l'utilisation de la biomasse forestière : implications pour la biodiversité et les ressources naturelles*, Ed Landmann, G., Gosselin, F., Bonhême, I, Chap 6.
- Gossner MM, Lachat T, Brunet J, et al, 2013**, Current near-to-nature forest management effects on functional trait composition of saproxylic beetles in beech forests, *Conservation Biology*, 27, pp 605-614.

Hermey M, Honnay O, Firbank L, et al, 1999, An ecological comparison between ancient and other forest plant species of Europe, and the implications for forest conservation, *Biological Conservation*, 91, pp 9-22.

Hermey M, Verheyen K, 2007, Legacies of the past in the present-day forest biodiversity: a review of past land-use effects on forest plant species composition and diversity, in *Sustainability and Diversity of Forest Ecosystems*, Springer, 361-371.

Honnay O, Degroote B, Hermey M, 1998, Ancient-forest plant species in Western Belgium: a species list and possible ecological mechanisms, *Belgian Journal of Botany*, pp 139-154.

Honnay O, Hermey M, Coppin, 1999, Effects of area, age and diversity of forest patches in Belgium on plant species richness, and implications for conservation and reforestation, *Biological Conservation*, 87, pp 73-84.

Jamoneau A, Sonnier G, Chabrierie O, 2011, Drivers of plant species assemblages in forest patches among contrasted dynamic agricultural landscapes, *Journal of Ecology*, 99, pp 1152-1161.

Larrieu L, Cabanettes A, Brin A, 2014, Tree microhabitats at the stand scale in montane beech-fir forests, practical information for taxa conservation in forestry, *European Journal of Forest Research*, 133, pp 355-367.

Larrieu L, Cabanettes A, Delarue A, 2012, Impact of silviculture on dead wood and on the distribution and frequency of tree microhabitats in montane beech-fir forests of the Pyrenees, *European Journal of Forest Research*, 131, pp 773-786.

Larrieu L, Gonin P, 2008, L'Indice de biodiversité Potentielle, IBP : une méthode simple et rapide pour évaluer la biodiversité potentielle des peuplements forestier, *Revue Forestière Française*, 60, pp 727-748.

Laurance WF, Yensen E, 1991, Predicting the impacts of edge effects in fragmented habitats. *Biological Conservation*, 55, pp 77-92.

Luyssaert S, Schulze E-D, Börner A, 2008, Old-growth forests as global carbon sinks, *Nature*, 455, pp 213-215.

Matlack GR, 1994, Plant species migration in a mixed-history forest landscape in eastern North America, *Ecology*, 75, pp 1491-1502.

Nivet C, Gosselin F, Gosselin M, 2012, Utilité des indicateurs taxonomiques de biodiversité forestière. Les indicateurs de biodiversité forestière, *Synthèse des réflexions issues du programme de recherche " Biodiversité, Gestion Forestière et Politiques Publiques "*, pp 59-72.

Normandin D, Cinotti B, 2002, Exploitants agricoles et propriété forestière : où est passée la "forêt paysanne"?, *Revue Forestière Française*, 54, pp 311-328.

Nunery JS, Keeton WS, 2010, Forest carbon storage in the northeastern United States: Net effects of harvesting frequency, post-harvest retention, and wood products, *Forest Ecology and Management* 259, pp 1363-1375.

Oindo BO, Skidmore AK, 2002, Interannual variability of NDVI and species richness in Kenya, *International Journal of Remote Sensing*, 23, pp 285-298.

ONF, 2010, Quels indicateurs pour la biodiversité forestières, Synthèse du programme de recherche « Biodiversité, gestion forestière et politiques publiques », n°2/3.

Opdam P, 1990, Dispersal in fragmented populations: the key to survival, *Species Dispersal in Agricultural Habitats*, pp 3-17.

Paillet Y, Bergès L, 2010, Naturalité des forêts et biodiversité : une comparaison par méta-analyse de la richesse spécifique des forêts exploitées et non exploitées en Europe, chap 4.

Palmer MW, Earls PG, Hoagland BW, 2002, Quantitative tools for perfecting species lists, *Environmetrics*, 13, pp 121-137.

Peterken GF, 1977, Habitat conservation priorities in British and European woodlands. *Biological Conservation*, 11, pp 223-236.

Peterken GF, Game M, 1984, Historical factors affecting the number and distribution of vascular plant species in the woodlands of central Lincolnshire, *The Journal of Ecology*, pp 155-182.

Raymond R, 2012, L'identification ordinaire de la biodiversité. En dehors du Codex, des indicateurs ordinaires de biodiversité, *Les Indicateurs de Biodiversité Forestière*, pp 103-107.

Redon M, 2012, Biodiversité potentielle dans les forêts du Vercors, une approche hiérarchique pour la conservation des espaces forestiers, Thèse en écologie, Université de Grenoble, 194p

Rocchini D, Ricotta C, Chiarucci A, 2007, Using satellite imagery to assess plant species richness: role of multispectral systems, *A lied Vegetation Science*, 10, pp 325-331.

Saunders DA, Hobbs RJ, Margules CR, 1991, Biological consequences of ecosystem fragmentation: a review, *Conservation Biology*, 5, pp 18-32.

Savoie J.M., Brustel H, 2011 Vieilles forêts et coléoptères saproxyliques, *Insectes*, n°162, pp 15-20

Sciama D, Augusto L, Dupouey J-L, 2009, Floristic and ecological differences between recent and ancient forests growing on non-acidic soils, *Forest Ecology and Management*, 258, pp 600-608.

Sharitz RR, Boring LR, Van Lear DH, Pinder III JE, 1992, Integrating ecological concepts with natural resource management of southern forests, *Ecological Applications*, 2(5), pp 226-237.

Taylor PD, Fahrig L, Henein K, Merriam G, 1993, Connectivity is a vital element of landscape structure, *Oikos*, pp 571-573.

Vallauri D, 2002 Le bois mort, un attribut vital de la biodiversité de la forêt naturelle, une lacune des forêts gérées, *Ra ort Scientifique du GIP ECOFOR*, 31p

Vallauri D., Grel A., Granier E., Dupouey J.L. 2012. Les forêts de Cassini. Analyse quantitative et comparaison avec les forêts actuelles, *Ra ort WWF/INRA*, Marseille, 64 p.

Vallauri D, Neyroumande E, 2009, Les forêts françaises : une biodiversité à la fois riche et menacée, *in Annales des Mines-Responsabilité et Environnement*, ESKA, pp 75-81.

CHAPITRE 3

Andrieu E, 2013, Reconstitution de l'histoire méconnue de la gestion de la petite forêt privée, Les cahiers de Framespa, [en ligne], 13. URL : <http://framespa.revues.org/2287>

Bommarco R, Lindborg R, Marini L, Öckinger E, 2014, Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history, *Diversity and Distributions*, 20, pp 591-599.

Bonthoux S, 2011, Les relations spatiales et temporelles entre les communautés d'oiseaux et les paysages agricoles, Thèse en écologie, 155p.

Chambers RJ, Adams THL, 1986, Quantification of the impact of hoverflies (Diptera Syrphidae) on cereal aphids in winter wheat: an analysis of field populations, *Journal of Applied Ecology*, 23, pp 895-904.

De Warnaffe GDB, Deconchat M, Ladet S, Balent G, 2006, Variability of cutting regimes in small private woodlots of south-western France, *Annals of Forest Science*, 63, pp 915-927.

Girel J, 2006, Quand le passé éclaire le présent : écologie et histoire du paysage, *Géocarrefour : Revue de Géographie de Lyon*, 81, pp 249-264.

Good JA, Speight MCD, 1996, Saproxyllic invertebrates and their conservation throughout Europe, *Nature and Environment Series*, 42, 82p.

Groot M, Bevk D, 2012, Ecosystem services and phenology of hoverflies (Diptera Syrphidae) in a slovenian forest stand, *Les Ljubljana*, 64, pp 123-128.

IFN, 2000, Inventaire forestier national pour le département de la Haute Garonne.

Joly D, Brossard T, Cardot H, 2010, Les types de climats en France, une construction spatiale.

Keil P, Konvicka M, 2005, Local species richness of Central European hoverflies (Diptera Syrphidae): a lesson taught by local faunal lists, *Diversity and Distributions*, 11, pp 417-426.

Ouin A, Sarthou JP, Bouyjou B, 2006, The species-area relationship in the hoverfly (Diptera, Syrphidae) communities of forest fragments in southern France, *Ecography*, 29, pp 183-190.

Raymond L, 2013, *Stratégies d'hivernation et d'exploitation du paysage par les syrphes aphidiphages : implications pour le contrôle biologique des pucerons*, Thèse en écologie, Agrocampus ouest / INP-ENSAT, 125p.

Sarthou JP, Fromage P, Genet B, Vinauger A, Heintz W et Monteil C, 2010. *SYRFID vol. 4: Syrphidae of France Interactive Data* [On-Line URL: syrfid.ensat.fr].

Speight MCD, 2010, Species accounts of European (Diptera Syrphidae), *Syrph the Net, the Database of European Syrphidae*, 59, 285p.

Speight MCD, Castella E, Obrdlik P, 2000, Use of the Syrph the Net database 2000, *Syrph the Net, the Database of European Syrphidae*, 25, 99p.

Speight MR, Wainhouse D, 1989, *Ecology and management of forest insects*, Clarendon Press, 384p.

Tenhumberg B, Poehling H-M, 1995, Syrphids as natural enemies of cereal aphids in Germany: aspects of their biology and efficacy in different years and regions, *Agriculture, Ecosystems & Environment*, 52, pp 39-43.

CHAPITRE 4

Berthaut H, 1898, La Carte de France 1750-1898, 2 vols. Paris : Imprimerie du Service Géographique de l'Armée, 9, pp 48-70.

Bret F, 1980, A roche du problème de l'équilibre agriculture-forêt en région de moyenne montagne. Le cas du Haut-Beaujolais, *Des arbres et des hommes, op-cit*, pp 57-79.

Cinotti B, 1996, Evolution des surfaces boisées en France : proposition de reconstitution depuis le début du XIXe siècle, *Revue Forestière Française*, 6, pp 512-562.

Costa L, Robert S, 2009, Guide de lecture des cartes anciennes, Ed Errances, 104p.

Delambre JBJ, Legendre AM, 1798, *Méthodes analytiques pour la détermination d'un arc du méridien : précédées d'un mémoire sur le même sujet*, Imprimerie de Crapelet.

Dupouey JL, Bachacou J, Cosserat R, 2007, Vers la réalisation d'une carte géoréférencée des forêts anciennes de France, *Revue du Comité Français de Cartographie*, 191, pp 85-98.

Fabre E, Vernet C, 2006, Évolution de l'occupation du sol dans les Alpes-de-Haute-Provence, début du XIXe siècle-fin du XXe siècle, *Méditerranée Revue Géographique des Pays Méditerranéens/Journal of Mediterranean Geography*, 458, 35-42.

Favre C, Granier E, Cosserat-Mangeot R, Dupouey JL 2012, La digitalisation des cartes anciennes, Manuel pour la vectorisation de l'usage des sols et le géo-référencement de la carte d'Etat-major, [en ligne]. URL : <http://www.foretsanciennes.fr/wp-content/uploads/2012/12/Fabvre-al-2011-digitalisation.pdf>

Franchomme M, 2008, Du cadastre napoléonien à la trame verte, le devenir des petites zones humides périurbaines en région Nord-Pas de Calais, Thèse en géographie/aménagement, Université des Sciences et Technologie de Lille I, 406p.

Koerner W, Cinotti B, Jussy J-H, Benoît M, 2000, Evolution des surfaces boisées en France depuis le début du XIXe siècle, identification et localisation des boisements des territoires agricoles abandonnés, *Revue Forestière Française*, 52, pp 249-270.

Nadal F, 2011, La cartographie militaire des Pyrénées françaises et la guerre civile espagnole, *Sud-Ouest Européen*, 1, 169-182.

Pinoteau C, Di Pietro F, 2003, Croisement des a roches en écologie du paysage et en archéogéographie. Le bassin-versant de l' Aubrière (Indre-et-Loire), *Evaluation des risques environnementaux pour une gestion durable des espaces*, 2003, Cemagref, pp 257-259.

Nicolas Poirier, 2006, Des plans terriers au cadastre ancien : Mesurer l'évolution de l'occupation du sol grâce au SIG », *Le Médiéviste et l'ordinateur*, 44 [En ligne]. URL : <http://lmo.irht.cnrs.fr/44/plans-terriers.htm>

Rudel TK, Coomes OT, Moran E, 2005, Forest transitions: towards a global understanding of land use change, *Global Environmental Change*, 15, pp 23-31.

Vallauri D., Grel A., Granier E., Dupouey J.L. 2012. Les forêts de Cassini. Analyse quantitative et comparaison avec les forêts actuelles, *Rapport WWF/INRA*, Marseille, 64 p.

Dupouey J-L., 2011, Etude de faisabilité d'une carte des forêts anciennes de France, Présentation orale-Manifestation GIP ECOFOR [En ligne]. <http://www.gip-ecofor.org/?q=node/760>

Plumejeaud C, 2012, Présentation des données, *Rapport ANR Géopeuple*, 20p.

IGN 2011. Scan Etat-major 40K descriptif de contenu, version 1, IGN, Saint-Mandé.

CHAPITRE 5

Laffly D, 2005, *A roche numérique du paysage : formalisation, enjeux et pratiques de recherche*, Besançon, Mémoire Habilitation à diriger des recherches, Université de Franche-Comté, 423p.

- Dupouey J-L., 2011**, Etude de faisabilité d'une carte des forêts anciennes de France, Présentation orale-Manifestation GIP ECOFOR [En ligne]. URL : <http://www.gip-ecofor.org/?q=node/760>
- Akima H,1978**, A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points, *ACM Transactions on Mathematical Software*, 4, pp 148-159.
- Ardizzone E, Gallea R, Gambino O, Pirrone R, 2009**, Multi-modal image registration using fuzzy kernel regression, in *proceedings of Image Processing, ICIP, 16th IEEE International Conference on. IEEE*, 193-196.
- Baily B, Riley M, Aucott P, Southall H, 2011**, Extracting digital data from the First Land Utilisation Survey of Great Britain-Methods, issues and potential, *A lied Geography*, 31, pp 959-968.
- Baiocchi V, Lelo K, 2005**, Georeferencing the historical maps of Rome between the seventeenth and eighteenth centuries, in *International Symposium-International cooperation to save the world's cultural heritage, Torino*, pp 114-118.
- Balletti C, 2006**, Georeference in the analysis of the geometric content of early maps, *e-Perimetron*, 1, pp 32-42.
- Balletti C, Guerra F, 2009**, Image matching for historical maps comparison, *e-Perimetron*, 4, pp 180-186.
- Benavides J, Koster E, 2006**, Identifying surviving landmarks on historical maps, *e-Perimetron*, 1, pp 194-208.
- Billen R, Cornélis B, Sheeren D, 1999**, Influence de la précision absolue des points d'amer sur la qualité de la géoréférenciation des images à très haute résolution. La télédétection en francophonie : analyse critique et perspective, *Huitième journées scientifiques du " Réseau Télédétection " de l'Aupelf-Uref*, pp 15-22.
- Bitelli G, Cremonini S, Gatta G, 2009**, Ancient map comparisons and georeferencing techniques: a case study from the Po river delta, Italy, *e-Perimetron*, 4, pp 221-228.
- Bitelli G, Gatta G, 2011**, Experiences on georeferencing of maps from the XIX century Gregorian Cadastre of Bologna, Italy, *e-Perimetron*, 6, pp 270-275.
- Boutoura C, Livieratos E, 2006**, Some fundamentals for the study of the geometry of early maps by comparative methods, *e-Perimetron*, 1, pp 60-70.
- Boutoura C, Livieratos E, 1986**, Strain analysis for geometric comparisons of maps, *The Cartographic Journal*, 23, pp 27-34.
- Buiten HJ, Van Putten B, 1997**, Quality assessment of remote sensing image registration—analysis and testing of control point residuals, *ISPRS journal of photogrammetry and remote sensing*, 52, pp 57-73.
- Caloz R, Collet C, 2001**, *Traitements numériques d'images de télédétection*, Presses de l'Université du Québec, 386p.
- Chen LC, Lee LH, 1992**, Progressive generation of control frameworks for image registration, *Photogrammetric Engineering and Remote Sensing*, 58, pp 1321-1328.
- Dupouey JL, Bachacou J, Cosserat R, 2007**, Vers la réalisation d'une carte géoréférencée des forêts anciennes de France, *Revue du Comité Français de Cartographie*, 191, pp 85-98.

- Fauvel M, Arbelot B, Benediktsson JA, 2013**, Detection of hedges in a rural landscape using a local orientation feature: from linear opening to path opening, *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal*, 6, pp 15-26.
- Fonseca LM, Manjunath BS, 1996**, Registration techniques for multisensor remotely sensed imagery, *Photogrammetric Engineering & Remote Sensing*, 62, pp 1049-1056.
- Ford GE, Algazi VR, Agee BG, 1978**, On the correction of geometric distortion in satellite-acquired images, in Annual Asilomar Conference on Circuits, Systems, and Computers 12th, 6(8).
- Ford GE, Zanelli CI, 1985**, Analysis and quantification of errors in the geometric correction of satellite images, *Photogrammetric Engineering and Remote Sensing*, 51, pp 1725-173.
- Fuse T, 1998**, A study on geometric correction of historical maps, *Photogrammetry and Remote Sensing*, 32, pp 543-548.
- Goshtasby A, 1986**, Piecewise linear mapping functions for image registration, *Pattern Recognition*, 19, pp 459-466.
- Goshtasby A, 1987**, Piecewise cubic mapping functions for image registration, *Pattern Recognition*, 20, pp 525-533.
- Haworth J, Shawe-Taylor J, Cheng T, Wang J, 2014**, Local online kernel ridge regression for forecasting of urban travel times, *Emerging Technologies*, 46, pp 151-178.
- Huang C, Davis LS, Townshend JRG, 2002**, An assessment of support vector machines for land cover classification, *International Journal of Remote Sensing*, 23, pp 725-749.
- Jenny B, 2006**, MapAnalyst-A digital tool for the analysis of the planimetric accuracy of historical maps, *e-Perimtron*, 1, pp 239-245.
- Jenny B, Hurni L, 2011**, Studying cartographic heritage, Analysis and visualization of geometric distortions, *Computers & Graphics*, 35, pp 402-411.
- Labovitz ML, Marvin JW, 1986**, Precision in geodetic correction of TM data as a function of the number, spatial distribution, and success in matching of control points: a simulation, *Remote sensing of environment*, 20, pp 237-252.
- Leyk S, 2005**, *Computing the Past-Utilizing Historical Data Sources for Map-Based Retrospective Landscape Research*, Thèse en sciences de l'information géographique, Université de Zurich, 172p
- Liu B, Zhang J, Liao X, 2011**, Projection Kernel Regression for Image Registration and Fusion in Video-Based Criminal Investigation, in *proceedings of Multimedia and Signal Processing (Conference on IEEE)*, pp 348-352.
- Mather PM, 1995**, Map-image registration accuracy using least-squares polynomials, *International Journal of Geographical Information Systems*, 9, pp 543-554.
- Moré G, Pons X, 2012**, Influence of the nature and number of ground control points to the quality of remote sensing geometric corrections, *Geoscience and Remote Sensing Symposium, IGARSS, 2012 IEEE International, IEEE*, pp 2356-2359.
- Orti F, 1981**, Optimal distribution of control points to minimize the Landsat image mean square registration errors, *Photogrammetric Engineering and Remote Sensing*, 47, pp 101-110.

- Orun AB, Natarajan K, 1994**, A modified bundle adjustment software for SPOT imagery and photography, *Photogrammetric Engineering and Remote Sensing*, 60, pp 1431-1437.
- Saalfeld A, 1985**, A fast rubber-sheeting transformation using simplicial coordinates, *The American Cartographer*, 12, pp 169-173.
- Servigne S, Ubada T, Puricelli A, Laurini R, 2000**, A methodology for spatial consistency improvement of geographic databases, *Geoinformatica*, 4, pp 7-34.
- Shawe-Taylor J, Cristianini N, 2004**, Kernel methods for pattern analysis, Cambridge university press, 478p.
- Shimizu E, Fuse T, 2003**, Rubber-sheeting of historical maps in GIS and its application to landscape visualization of old-time cities: focusing on Tokyo of the past, in Proceedings of the 8th international conference on computers in urban planning and urban management, 11(3).
- Shlien S, 1979**, Geometric correction, registration, and resampling of Landsat imagery, *Canadian Journal of Remote Sensing*, 5, pp 74-89.
- Sprague DS, Iwasaki N, Takahashi S, 2007**, Measuring rice paddy persistence spanning a century with Japan's oldest topographic maps: georeferencing the Rapid Survey Maps for GIS analysis, *International Journal of Geographical Information Science*, 21, pp 83-95.
- Takeda H, Farsiu S, Milanfar P, 2007, Kernel regression for image processing and reconstruction, *Image Processing, IEEE Transactions*, 16, pp 349-366.
- Ubada T, Egenhofer MJ, 1997**, Topological error correcting in GIS, in *Advances in Spatial Databases*, Springer, pp 281-297.
- Wang J, Ge Y, Heuvelink GB, 2012**, Effect of the sampling design of ground control points on the geometric correction of remotely sensed imagery, *International Journal of A lied Earth Observation and Geoinformation*, 18, pp 91-100.
- White Jr MS, Griffin P, 1985**, Piecewise linear rubber-sheet map transformation, *The American Cartographer*, 12, pp 123-131.
- Wieczorek J, Guo Q, Hijmans R, 2004**, The point-radius method for georeferencing locality descriptions and calculating associated uncertainty, *International journal of geographical information science*, 18, pp 745-767.
- Yang G, Jiao WL, 2011**, An Optimization Method on Ground Control Point Distribution, *Advanced Materials Research*, 268, pp 1092-1095.
- Zhang J, Cheng B, 2009**, Distribution optimization of ground control points in remote sensing image geometric rectification based on cluster analysis, *Sixth International Symposium on Multispectral Image Processing and Pattern Recognition, International Society for Optics and Photonics*, 749707(7).
- Zhang Y, Zhang D, Gu Y, Tao F, 2006**, Impact of GCP distribution on the rectification accuracy of Landsat TM imagery in a coastal zone, *Acta oceanologica sinica*, 4, pp 14-22.
- Zitova B, Flusser J, 2003**, Image registration methods: a survey, *Image and vision computing*, 21, pp 977-1000.

CHAPITRE 6

Favre C, Granier E, Cosserat-Mangeot R, Dupouey JL 2012, La digitalisation des cartes anciennes, Manuel pour la vectorisation de l'usage des sols et le géo-référencement de la carte d'Etat-major, [en ligne]. URL: <http://www.foretsanciennes.fr/wp-content/uploads/2012/12/Fabvre-al-2011-digitalisation.pdf>

Jiang B, 2014, Geospatial analysis requires a different way of thinking: the problem of spatial heterogeneity, *GeoJournal*, 80(1), 1-13.

Robert S, Autran J, 2012, Décrire à grande échelle l'occupation des sols urbains par photo-interprétation. Réflexion méthodologique et expérimentation en Provence, *Sud-Ouest européen Revue Géographique des Pyrénées et du Sud-Ouest*, 33, 25-40.

Ruas A, 1999, *Modèle de généralisation de données urbaines à base de contraintes et d'autonomie*, Thèse en sciences de l'information géographiques, Université de Marne La Vallée, 57p.

Sheeren D, 2005, *Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales. Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique*, Thèse en sciences de l'information géographique, Université Pierre et Marie Curie-Paris VI, 294p.

CHAPITRE 7

Angulo J, Serra J, 2003, Mathematical morphology in color spaces applied to the analysis of cartographic images, in *Proceedings of GEOPRO*, 3, pp 59-66.

Ansoult M, Soille P, Loodts J, 1990, Mathematical morphology- A tool for automated GIS data acquisition from scanned thematic maps, *Photogrammetric Engineering and Remote Sensing*, 56, pp 1263-1271.

Arellano O, Moctezuma M, Parmiggiani F, 1998, Segmentation of remote sensing imagery based on quadtree structures, in *Proceedings Geoscience and Remote Sensing Symposium IGARSS'98*, pp 1058-1060.

Arias JF, Lai CP, Chandran S, et al, 1993, Interpretation of telephone system manhole drawings. Document Analysis and Recognition, 1993, in *Proceedings of the Second International Conference on IEEE*, pp 365-368.

Arrighi P, Soille P, 1999, From scanned topographic maps to digital elevation models, in *Proceedings of Geovision 99*, pp 1-4.

Baatz M, Schäpe A, 2000, Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation, *Angewandte Geographische Informationsverarbeitung*, 7, pp 12-23.

Bacher U, Mayer H, 2005, Automatic road extraction from multispectral high resolution satellite images, in *CMRT05. IAPRS*, 36(3), pp 29-34.

Benoit-Cattin H, Zouagi T, Odet C, 2000, Une vision fonctionnelle de la segmentation d'images,

Beucher S, 1992, The watershed transformation applied to image segmentation, in *Scanning Microscopy Supplement*, 299p.

Bindel M, Hese S, Berger C, Schmullius C, 2011, Feature selection from high resolution remote sensing data for biotope mapping, *High-Resolution Earth Imaging for Geospatial Information Hannover, June 14-17*.

- Blaschke T, 2010**, Object based image analysis for remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65, pp 2-16.
- Blaschke T, Hay GJ, Kelly M, 2014**, Geographic object-based image analysis-towards a new paradigm, *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp 180-191.
- Boesch R, 1996**, Detection and extraction of complex map symbols, *International Archives of Photogrammetry and Remote Sensing*, 31, pp 65-69.
- Boureau J-G, 2008**, Manuel d'interprétation des photographies aériennes infrarouges, application aux milieux forestiers et naturels. Inventaire Forestier National.
- Boyd DS, Danson FM, 2005**, Satellite remote sensing of forest resources: three decades of research development, *Progress in Physical Geography*, 29, pp 1-26.
- Breiman L, 1996**, Technical note: Some properties of splitting criteria, *Machine Learning*, 24, pp 41-47.
- Bulletin d'information scientifique et technique de l'IGN, 2006**, 75,
- Burnett C, Blaschke T, 2003**, A multi-scale segmentation/object relationship modelling methodology for landscape analysis, *Ecological modeling*, 168, pp 233-249.
- Cao R, Tan CL, 2002**, Text/graphics separation in maps, in *Graphics Recognition Algorithms and Applications*, Springer, pp 167-177.
- Carreira-Perpinan MA, 2007**, Gaussian mean-shift is an EM algorithm, *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 29, pp 767-776.
- Centeno JS, 1998**, Segmentation of thematic maps using colour and spatial attributes, *Graphics Recognition Algorithms and Systems*. Springer 221-230.
- Chiang Y-Y, Leyk S, Knoblock CA, 2013**, Efficient and robust graphics recognition from historical maps, in *Graphics Recognition. New Trends and Challenges*, Springer, pp 25-35.
- Chi Z, Yan H, 1993**, Map image segmentation based on thresholding and fuzzy rules, *Electronics Letters*, 29, pp 1841-1843.
- Cocquerez JP, Phili -Foliguet S, 1995**, *Analyse d'images : filtrage et segmentation*, Dunod, 457p.
- Collins JB, Woodcock CE, 1996**, An assessment of several linear change detection techniques for mapping forest mortality using multitemporal Landsat TM data, *Remote Sensing of Environment*, 56, pp 66-77.
- Coops NC, Catling PC, 1997**, Predicting the complexity of habitat in forests from airborne videography for wildlife management, *International Journal of Remote Sensing*, 18, pp 2677-2682.
- Daniil M, Tsioukas V, Papadopoulos K, Livieratos E, 2003**, Scanning options and choices in digitizing historic maps, in *Proceedings of CIPA 2003 International Symposium, Antalya, Turkey*.
- Deng JS, Wang K, Deng YH, Qi GJ, 2008**, PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data, *International Journal of Remote Sensing*, 29, pp 4823-4838.
- Deseilligny MP, Mariani R, Labiche J, Mullot R, 1998**, Topographic maps automatic interpretation, Some proposed strategies, in *Graphics Recognition Algorithms and Systems*, Springer 175-193

Devillers R, Jeansoulin R, 2005, *Qualité de l'information géographique*, Hermes Science Publications, 453p.

Dorff P, Kowalewski J, Doehler SAL, Krabbenhoeft UJ, 1998, Method for generating smooth color corrections in a color space, particularly a CIELAB color space, *U.S. Patent*, 5, pp 751-845.

Drăguț L, Tiede D, Levick SR, 2010, ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data, *International Journal of Geographical Information Science*, 24, pp 859-871.

Fauvel M, Sheeren D, Chanussot J, Benediktsson JA, 2012, Hedges detection using local directional features and su ort vector data description, in *Proceedings of Geoscience and Remote Sensing Symposium, IGARSS'12*, pp 2320-2323.

Favre C, Granier E, Cosserat-Mangeot R, Dupouey JL 2012, La digitalisation des cartes anciennes, Manuel pour la vectorisation de l'usage des sols et le géo-référencement de la carte d'Etat-major, [en ligne]. URL: <http://www.foretsanciennes.fr/wp-content/uploads/2012/12/Fabvre-al-2011-digitalisation.pdf>

Fortier A, Ziou D, Armenakis C, Wang S, 1999, Survey of work on road extraction in aerial and satellite images, Technical Report 241.

Fukunaga K, Hostetler L, 1975, The estimation of the gradient of a density function, with applications in pattern recognition, *Information Theory, IEEE Transactions*, 21, pp 32-40.

Ganesan P, Rajini V, Rajkumar RI, 2010, Segmentation and edge detection of color images using CIELAB color space and edge detectors, in *Proceedings of Emerging Trends in Robotics and Communication Technologies, INTERACT 2010*, pp 393-397.

Gruen A, Li H, 1995, Road extraction from aerial and satellite images by dynamic programming, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50, pp 11-20.

Guyomard J, 2013, *Vectorisation automatiques de symboles*, Rapport ANR GéoPeuple, 51p [en ligne]. URL: <http://geopeuple.ign.fr/doc/rapports/L3.1-VectorisationAutomatiqueSymboles.pdf>

Haralick RM, 1979, Statistical and structural a roaches to texture, in *Proceedings of the IEEE*, 67, pp 786-804.

Hay GJ, Blaschke T, Marceau DJ, Bouchard A, 2003, A comparison of three image-object methods for the multiscale analysis of landscape structure, *ISPRS Journal of Photogrammetry and Remote Sensing* 57, pp 327-345.

Henderson TC, Linton T, Potupchik S, Ostanin A, 2009, Automatic segmentation of semantic classes in raster map images, in *Proceedings of the Eighth IAPR International Workshop on Graphics Recognition*, pp 253-262.

Holopainen M, Wang G, 1998, *The calibration of digitized aerial photographs for forest stratification*. *International Journal of Remote Sensing*, 19, pp 677-696.

Iwanowski M, Kozak J, 2012, Automatic detection of forest regions on scanned old maps. *Electronical Review*, 88, pp 249-252.

Jackson PL, Gaston GG, 1994, Digital enhancement as an aid to detecting patterns of vegetation stress using medium-scale aerial photography, *Remote Sensing*, 15, pp 1009-1018.

- Kaneko T, 1992**, Line structure extraction from line-drawing images, *Pattern Recognition*, 25, pp 963-973.
- Kayitakire F, Giot P, Defourny P, 2002**, Discrimination automatique de peuplements forestiers à partir d'orthophotos numériques couleur : un cas d'étude en Belgique, *Canadian Journal of Remote Sensing*, 28, pp 629-640.
- King DJ, 1995**, Airborne multispectral digital camera and video sensors: a critical review of system designs and applications, *Canadian Journal of Remote Sensing*, 21, pp 245-273.
- Kollár S, Vekerdy Z, Márkus B, 2013**, Aerial Image Classification for the Mapping of Riparian Vegetation Habitats, *Acta Silvatica et Lignaria Hungarica*, 9, pp 119-133.
- Lefebvre A, 2011**, *Contribution de la texture pour l'analyse d'images à très haute résolution spatiale : application à la détection de changement en milieu périurbain*, Thèse en géographie, Université Rennes 2, 284p.
- Lefsky MA, Harding D, Cohen WB, et al, 1999**, Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA, *Remote Sensing of Environment*, 67, pp 83-98.
- Lévesque J, King DJ, 2003**, Spatial analysis of radiometric fractions from high-resolution multispectral imagery for modelling individual tree crown and forest canopy structure and health, *Remote Sensing of Environment* 84, pp 589-602.
- Leyk S, Boesch R, 2010**, Colors of the past, color image segmentation in historical topographic maps based on homogeneity, *GeoInformatica*, 14, pp 1-21.
- Leyk S, Boesch R, Weibel R, 2005**, A Conceptual Framework for Uncertainty Investigation in Map-based Land Cover Change Modelling, *Transactions in GIS*, 9, pp 291-322.
- Lichtner W, 1985**, Pattern recognition procedures for automatic digitizing of cadastral maps, in *Proceedings Auto Carto 7*, Washington DC, 41, pp 337
- Lillesand TM, Kiefer RW, Chipman JW, 2004**, *Remote sensing and image interpretation*, John Wiley & Sons Ltd, 768p.
- Loussert P, Guerec K, 2014**, *Historical maps objects extraction toolbox for ENVI-IDL*, [en ligne], [http, pp //fauvel.mathieu.free.fr/software.html](http://fauvel.mathieu.free.fr/software.html).
- Lu D, Weng Q, 2007**, A survey of image classification methods and techniques for improving classification performance, *International journal of Remote sensing*, 28, pp 823-870.
- Mahmoud MI, 2012**, *Information extraction from paper maps using object oriented analysis (OOA)*, Thèse en gestion de l'information géographique, Université de Twente, 78p
- Mather P, Koch M, 2011**, *Computer processing of remotely-sensed images: an introduction*, John Wiley & Sons, 460p.
- Morgan JL, Gergel SE, Coops NC, 2010**, Aerial photography: a rapidly evolving tool for ecological management, *BioScience*, 60, pp 47-59.
- Myers GK, Mulgaonkar PG, Chen C-H, et al, 1996**, Verification-based a roach for automated text and feature extraction from raster-scanned maps, in *Graphics Recognition Methods and Applications*, Springer, pp 190-203.

- Palandro D, Andréfouët S, Dustan P, Muller-Karger FE, 2003**, Change detection in coral reef communities using Ikonos satellite sensor imagery and historic aerial photographs, *International Journal of Remote Sensing*, 24, pp 873-878.
- Pasqualini V, Pergent-Martini C, Clabaut P, Pergent G, 1998**, Mapping of Posidonia oceanica using Aerial Photographs and Side Scan Sonar: Application off the Island of Corsica, France, *Estuarine, Coastal and Shelf Science*, 47, pp 359-367.
- Péteri R, Celle J, Ranchin T, 2003**, Detection and extraction of road networks from high resolution satellite images, in *Proceedings of Image Processing, 2003 (ICIP 2003)*, pp 301-304.
- Puissant A, 2003**, Information géographique et images à très haute résolution : utilité et applications en milieu urbain, Thèse en géographie, Université Louis Pasteur-Strasbourg I, 447p.
- Quinlan JR, 1986**, Induction of decision trees, *Machine learning*, 1, pp 81-106.
- Quinlan JR, 2014**, *C4. 5: programs for machine learning*, Morgan Kaufmann, 302p.
- Radoux J, Bogaert P, 2014**, Accounting for the area of polygon sampling units for the prediction of primary accuracy assessment indices, *Remote Sensing of Environment*, 142, pp 9-19.
- Rochon BF, Bonn F, 1992**, *Précis de télédétection-Volume 1-Principes et méthodes*, Presses de l'Université du Québec, 485p.
- Rose J-L, 2008**, Croissance de région variationnelle et contraintes géométriques tridimensionnelles pour la segmentation d'image, Thèse en sciences de l'information, des dispositifs et des systèmes, INSA Lyon, 160p.
- Salvatore S, Guitton P, 2004**, Contour line recognition from scanned topographic maps, *Journal of WSCG*, 12(3), 8p.
- Samet H, Soffer A, 1998**, Magellan: Map acquisition of geographic labels by legend analysis, *International Journal on Document Analysis and Recognition*, 1, pp 89-101.
- Samet R, Askerbeyli IA, Varol C, 2010**, An implementation of automatic contour line extraction from scanned digital topographic maps, *Application Computer Mathematics*, 9, pp 116-127.
- Shang D, Waite M-L, 1991**, *Utilization of textural and spectral features from Landsat TM imagery for stratifying forest sample plots*, University of Helsinki, 25p.
- Shaw TW, Bajcsy P, 2011**, Automation of digital historical map analyses, *IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics*, 10, pp 78690A-78690A.
- Sheeren D, 2005**, *Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales. Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique*, Thèse en sciences de l'information géographique, Université Pierre et Marie Curie-Paris VI, 294p.
- Sheeren D, Lefèvre S, Weber J, 2007**, La morphologie mathématique binaire pour l'extraction automatique des bâtiments dans les images THRS, *Revue Internationale de Géomatique*, 17, pp 333-352.
- Shoshany M, Degani A, 1992**, Shoreline detection by digital image processing of aerial photography, *Journal of Coastal Research*, pp 29-34.

- Skupinski G, BinhTran D, Weber C, 2009**, Les images satellites Spot multi-dates et la métrique spatiale dans l'étude du changement urbain et suburbain-Le cas de la basse vallée de la Bruche, Bas-Rhin, France, Cybergeog, 22p.
- Soille P, 1999**, *Morphological image analysis*, Springer, 392p.
- Sonka M, Hlavac V, Boyle R, 2014**, Image processing, analysis, and machine vision, *Brooks*, 829p.
- Stehman SV, Wickham JD, 2011**, Pixels, blocks of pixels, and polygons : Choosing a spatial unit for thematic accuracy assessment, *Remote Sensing of Environment*, 115, pp 3044-3055.
- St-Onge BA, Cavayas F, 1997**, Automated forest structure ma ing from high resolution imagery based on directional semivariogram estimates, *Remote Sensing of Environment*, 61, pp 82-95.
- Tuominen S, Pekkarinen A, 2005**, Performance of different spectral and textural aerial photograph features in multi-source forest inventory, *Remote Sensing of Environment*, 94, pp 256-268.
- Vincent L, Soille P, 1991**, Watersheds in digital spaces: an efficient algorithm based on immersion simulations, *IEEE transactions on pattern analysis and machine intelligence*, 13, pp 583-598.
- Wei Y, Zhao Z, Song J, 2004**, Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection, in Proceedings of Geoscience and Remote Sensing Symposium, 2004 (IGARSS'04).
- Willhauck G, Schneider T, De Kok R, Ammer U, 2000**, Comparison of object oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos, in *Proceedings of XIX ISPRS Congress*, pp 35-42.
- Wilson EH, Sader SA, 2002**, Detection of forest harvest type using multiple dates of Landsat TM imagery, *Remote Sensing of Environment*, 80, pp 385-396.
- Wise S, 1999**, Extracting raster GIS data from scanned thematic maps, *Transactions in GIS*, 3, pp 221-237.
- Witten IH, Frank E, Trigg LE, 1999**, Weka: Practical machine learning tools and techniques with Java implementations, *Computer Science Working Papers*, 99, 11p.
- Woodcock CE, Strahler AH, Jupp DL, 1988**, The use of variograms in remote sensing, Scene models and simulated images, *Remote Sensing of Environment*, 25, pp 323-348.
- Wuest B, Zhang Y, 2009**, Region based segmentation of QuickBird multispectral imagery through band ratios and fuzzy comparison, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, pp 55-64.
- Yang X, Lo CP, 2002**, Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area, *International Journal of Remote Sensing*, 23, pp 1775-1798.
- Yu L, Liang L, Wang J, 2014**, Meta-discoveries from a synthesis of satellite-based land-cover ma ing research, *International Journal of Remote Sensing*, 35, pp 4573-4588.
- Zammit O, 2008**, Détection de zones brûlées après un feu de forêt à partir d'une seule image satellitaire SPOT 5 par techniques SVM, Thèse en traitement du signal et des images, Université Nice Sophia Antipolis, 170p.

Zharikov Y, Skilleter GA, Loneragan NR, 2005, Mapping and characterising subtropical estuarine landscapes using aerial photography and GIS for potential application in wildlife conservation and management, *Biological Conservation*, 125, pp 87-100.

CHAPITRE 8

Alignier A, 2010, *Distribution des communautés végétales sous l'influence des lisières forestières dans des bois fragmentés*, Thèse en écologie, Université de Toulouse, 239p.

Alignier A, Deconchat M, 2011, Variability of forest edge effect on vegetation implies reconsideration of its assumed hypothetical pattern, *Landscape Vegetation Science*, 14, pp 67-74.

Andren H, 1994, Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review, *Oikos*, pp 355-366.

Badard T, 2000, Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques, Thèse en science de l'information géographique, Université de Marne La Vallée, 114p.

Baker WL, Cai Y, 1992, The role programs for multiscale analysis of landscape structure using the GRASS geographical information system, *Landscape Ecology*, 7, pp 291-302.

Bel Hadj Ali A, 2001, *Qualité géométrique des entités géographiques surfaciques : Application à l'ariement et définition d'une typologie des écarts géométriques*, Thèse en sciences de l'information géographique, Université de Marne-la-Vallée, 210p.

Bender DJ, Contreras TA, Fahrig L, 1998, Habitat loss and population decline, a meta-analysis of the patch size effect. *Ecology* 79, pp 517-533.

Bettinger P, Bradshaw GA, Weaver GW, 1996, Effects of geographic information system vector-raster-vector data conversion on landscape indices, *Canadian Journal of Forest Research*, 26, pp 1416-1425.

Bittner T, Donnelly M, Smith B, 2004, Endurants and perdurants in directly depicting ontologies, *AI Communications*, 17, pp 247-258.

Bordin P, 2006, *Méthode d'observation multi-niveaux pour le suivi de phénomènes géographiques avec un SIG*, Thèse en science de l'information géographique, Université de Marne la Vallée, 283p.

Bregt AK, Denneboom J, Gesink HJ, Van Randen Y, 1991, Determination of rasterizing error a case study with the soil map of The Netherlands, *International Journal of Geographical Information System*, 5, pp 361-367.

Buard E, 2013, Dynamiques des interactions espèces-espace : Mise en relation des pratiques de déplacement des populations d'herbivores et de l'évolution de l'occupation du sol dans le parc de Hwange, Zimbabwe, Thèse en science de l'information géographique, Université Panthéon-Sorbonne-Paris I, 397p.

Claramunt C, Thériault M, 1995, Managing time in GIS an event-oriented approach, in *Recent Advances in Temporal Databases*, Springer, pp 23-42.

- Connor EF, McCoy ED, 1979**, The statistics and biology of the species-area relationship, *American Naturalist*, 113(6), pp 791-833.
- Crooks KR, 2002**, Relative sensitivities of mammalian carnivores to habitat fragmentation, *Conservation Biology*, 16, pp 488-502.
- Crooks KR, Sanjayan M, 2006**, Connectivity conservation: maintaining connections for nature, *Conservation Biology Series Cambridge*, 14, pp 1-20.
- Deconchat M, Balent G, 1996**, Biodiversité et forêt dans un paysage agricole, Étude bibliographique, *Études et Recherches sur les Systèmes Agraires et le Développement*, pp 15-36.
- Del Mondo G, 2011**, *Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques*, Thèse en informatique/géomatique, Université de Bretagne occidentale-Brest, 163p.
- Devogele T, 1997**, Processus d'intégration et d'appariement de bases de données géographiques; a lication à une base de données routières multi-échelles, Thèse en informatique, Université de Versailles-Saint Quentin en Yvelines, 206p.
- ESRI, 2013**, help/arcgis.com
- Ewers RM, Thorpe S, Didham RK, 2007**, Synergistic interactions between edge and area effects in a heavily fragmented landscape, *Ecology*, 88, pp 96-106.
- Fahrig L, 2003**, Effects of habitat fragmentation on biodiversity, *Annual Review of Ecology, Evolution, and Systematics*, pp 487-515.
- Fahrig L, 2013**, Rethinking patch size and isolation effects: the habitat amount hypothesis, *Journal of Biogeography*, 40, pp 1649-1663.
- Fischer DT, Church RL, 2003**, Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model, *Forest Science*, 49, pp 555-565.
- Foltête J-C, Clauzel C, Vuidel G, Tournant P, 2012**, Integrating graph-based connectivity metrics into species distribution models, *Landscape Ecology*, 27, pp 557-569.
- Forman RT, Godron M, 1981**, Patches and structural components for a landscape ecology, *BioScience*, 31, pp 733-740.
- Gascon C, Lovejoy TE, Bierregaard Jr RO, 1999**, Matrix habitat and species richness in tropical forest remnants, *Biological Conservation* 91, pp 223-229.
- Gesbert N, 2005**, *Etude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration*, Thèse en informatique, Université Marne La Vallée, 134p.
- Grenon P, Smith B, 2004, SNAP and SPAN: Towards dynamic spatial ontology, *Spatial Cognition and Computation*, 4, pp 69-104.
- Gustafson EJ, Parker GR, 1992**, Relationships between landcover proportion and indices of landscape spatial pattern, *Landscape Ecology*, 7, pp 101-110.
- Hanski I, 1994**, Patch-occupancy dynamics in fragmented landscapes, *Trends in Ecology & Evolution*, 9, pp 131-135.
- Hanski I, Gaggiotti OE, 2004**, *Ecology, evolution, and genetics of metapopulations*, Academic Press, San Diego, 512p.

- Harbelot B, Arenas H, Cruz C, 2015**, LC3, un modèle spatial et sémantique pour découvrir la connaissance dans les jeux de données géospatiaux, in *Actes de l'atelier GAST-Gestion et Analyse de données Spatiales et temporelles*, pp 9-21.
- Haydon DT, Friar JK, Pianka ER, 2000**, Fire-driven dynamic mosaics in the Great Victoria Desert, Australia-I, Fire geometry. *Landscape Ecology*, 15, pp 373-382.
- Hornsby K, Egenhofer MJ, 2000**, Identity-based change: a foundation for spatio-temporal knowledge representation, *International Journal of Geographical Information Science*, 14, pp 207-224.
- Hornsby K, Egenhofer MJ, 2002**, Modeling moving objects over multiple granularities, *Annals of Mathematics and Artificial Intelligence*, 36, pp 177-194.
- Huh Y, Kim J, Lee J, 2014**, Identification of multi-scale corresponding object-set pairs between two polygon datasets with hierarchical co-clustering, *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp 60-68.
- Jaeger JA, 2000**, Landscape division, splitting index, and effective mesh size, new measures of landscape fragmentation, *Landscape Ecology*, 15, pp 115-130.
- Jiang J, Worboys M, 2009**, Event-based topology for dynamic planar areal objects, *International Journal of Geographical Information Science*, 23, pp 33-60.
- Jiao L, Liu Y, Li H, 2012**, Characterizing land-use classes in remote sensing imagery by shape metrics, *ISPRS Journal of Photogrammetry and Remote Sensing*, 72, pp 46-55.
- Jolliffe I, 2002**, *Principal component analysis*, Springer, 448p.
- Langran G, Chrisman NR, 1988**, A framework for temporal geographic information, *Cartographica : The International Journal for Geographic Information and Geovisualization*, 25, pp 1-14.
- Lindemann JD, Baker WL, 2001**, Attributes of blowdown patches from a severe wind event in the Southern Rocky Mountains, USA, *Landscape Ecology*, 16, pp 313-325.
- Magle SB, Theobald DM, Crooks KR, 2009**, A comparison of metrics predicting landscape connectivity for a highly interactive species along an urban gradient in Colorado, USA, *Landscape Ecology*, 24, pp 267-280.
- Malcolm JR, 1994**, Edge effects in central Amazonian forest fragments, *Ecology*, 75, pp 2438-2445.
- Mathian H, Sanders L, 2014**, *Objets géographiques et processus de changements*, Ed ISTE, 178p.
- McGarigal K, 2002**, Landscape pattern metrics, [en ligne]. URL : <http://www.umass.edu/landeco/pubs/mcgarigal.2013.pdf>
- McGarigal K, Cushman SA, Neel MC, Ene E, 2002**, FRAGSTATS: spatial pattern analysis program for categorical maps [en ligne]. URL : <http://www.umass.edu/landeco/research/fragstats/fragstats.html>
- McGarigal K, Marks BJ, 1995**, Spatial pattern analysis program for quantifying landscape structure.
- Mladenoff DJ, White MA, Pastor J, Crow TR, 1993**, Comparing spatial pattern in unaltered old-growth and disturbed forest landscapes, *Ecological Applications*, 3, pp 294-306.
- Mustière S, Devogele T, 2008**, Matching networks with different levels of detail, *Geoinformatica*, 12, pp 435-453.

- Olteanu AM, 2008**, *Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance*, Thèse en informatique, Université Paris-Est, 269p.
- O'Neill RV, 1986**, *A hierarchical concept of ecosystems*, Princeton University Press, 253p.
- Ouin A, Sarthou JP, Bouyjou B, 2006**, The species-area relationship in the hoverfly, Diptera, Syrphidae, communities of forest fragments in southern France, *Ecography*, 29, pp 183-190.
- Peuquet DJ, Duan N, 1995**, An event-based spatiotemporal data model, ESTDM, for temporal analysis of geographical data, *International Journal of Geographical Information Systems*, 9, pp 7-24.
- Plumejeaud C, 2011**, Modèles et méthodes pour l'information spatio-temporelle évolutive, Thèse en informatique, Université de Grenoble, 331p.
- Prugh LR, Hodges KE, Sinclair AR, Brashares JS, 2008**, Effect of habitat area and isolation on fragmented animal populations, in *Proceedings of the National Academy of Sciences*, 105, pp 20770-20775.
- Rand TA, Tylianakis JM, Tscharrntke T, 2006**, Spillover edge effects, the dispersal of agriculturally subsidized insect natural enemies into adjacent natural habitats, *Ecology Letters*, 9, pp 603-614.
- Ranta P, Blom TOM, Niemela J, 1998**, The fragmented Atlantic rain forest of Brazil, size, shape and distribution of forest fragments, *Biodiversity & Conservation*, 7, pp 385-403.
- Riitters KH, O'Neill RV, Hunsaker CT, 1995**, A factor analysis of landscape pattern and structure metrics, *Landscape ecology*, 10, pp 23-39.
- Safra E, Kanza Y, Sagiv Y, Doytsher Y, 2013**, Ad hoc matching of vectorial road networks, *International Journal of Geographical Information Science*, 27, pp 114-153.
- Saura S, Carballal P, 2004**, Discrimination of native and exotic forest patterns through shape irregularity indices, an analysis in the landscapes of Galicia, Spain, *Landscape Ecology*, 19, pp 647-662.
- Sheeren D, 2002**, L'appariement pour la constitution de bases de données géographiques multirésolutions, *Revue internationale de géomatique*, 12, pp 151-168.
- Sheeren D, 2005**, *Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales. Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique*, Thèse en sciences de l'information géographique, Université Pierre et Marie Curie-Paris VI, 294p.
- Shortridge AM, 2004**, Geometric variability of raster cell class assignment, *International Journal of Geographical Information Science*, 18, pp 539-558.
- Theobald DM, Norman JB, Sherburne MR, 2006**, FunConn v1 user's manual, ArcGIS tools for functional connectivity modeling [en ligne]. URL: <http://www.stat.colostate.edu/~nsu/starmap/s/Technical%20Reports/FunConn.Users.Guide.pdf>
- Turner MG, 1990**, Spatial and temporal analysis of landscape patterns, *Landscape Ecology*, 4, pp 21-30.
- Turner MG, Gardner RH, 1991**, *Quantitative methods in landscape ecology*, Springer, 536p.

Vauglin F, 1997, *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*, Thèse en sciences de l'information géographique, Université de Marne La Vallée, 294p.

Worboys M, 2005, Event-oriented approaches to geographic phenomena, *International Journal of Geographical Information Science*, 19, pp 1-28.

CHAPITRE 9

Deconchat M, Gibon A, Cabanettes A, 2007, How to set up a research framework to analyze social-ecological interactive processes in a rural landscape, *Ecology and Society*, 12, pp 15.

De Warnaffe GDB, Deconchat M, Ladet S, Balent G, 2006, Variability of cutting regimes in small private woodlots of south-western France, *Annals of Forest Science*, 63, pp 915-927.

CHAPITRE 10

Akaike H, 1981, Likelihood of a model and information criteria, *Journal of econometrics*, 16, pp 3-14.

Akaike H, 1987, Factor analysis and AIC, *Psychometrika*, 52, pp 317-332.

Andrieu E, De Warnaffe G, Ladet S, 2008, Cartographie l'historique des coupes forestières dans les petits bois, *Revue Forestière Française*, 60, pp 667-676.

Atobe T, Osada Y, Takeda H, 2014, Habitat connectivity and resident shared predators determine the impact of invasive bullfrogs on native frogs in farm ponds, in *Proceedings of the Royal Society B, Biological Sciences*, 281, pp 21-26.

Baily B, Riley M, Aucott P, Southall H, 2011, Extracting digital data from the First Land Utilisation Survey of Great Britain-Methods, issues and potential, *A lied geography*, 31, pp 959-968.

Bommarco R, Lindborg R, Marini L, Öckinger E, 2014, Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history, *Diversity and Distributions*, 20, pp 591-599.

Bossuyt F, Milinkovitch MC, 2001, Amphibians as indicators of early Tertiary " Out-of-India" Dispersal of vertebrates, *Science*, 292, pp 93-95.

Breslow NE, Clayton DG, 1993, Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, 88, pp 9-25.

Brooks TM, Pimm SL, Oyugi JO, 1999, Time lag between deforestation and bird extinction in tropical forest fragments, *Conservation Biology*, 13, pp 1140-1150.

Brottons L, Mönkkönen M, Martin JL, 2003, Are fragments islands? landscape context and density-area relationships in boreal forest birds, *The American Naturalist*, 162, pp 343-357.

Bulman CR, Wilson RJ, Holt AR, 2007, Minimum viable metapopulation size, extinction debt, and the conservation of a declining species, *Ecological Applications*, 17, pp 1460-1473.

Burnham KP, Anderson DR, 2002, *Model selection and multimodel inference: a practical information-theoretic approach*, Springer Science & Business Media, 488p.

Chatfield C, Goodhardt GJ, 1976, The beta-binomial model for consumer purchasing behavior, in *Mathematical Models in Marketing*, Springer, pp 53-57.

- Cousins SA, 2009**, Extinction debt in fragmented grasslands: paid or not? *Journal of Vegetation Science*, 20, pp 3-7.
- Cousins SA, Vanhoenacker D, 2011**, Detection of extinction debt depends on scale and specialization, *Biological Conservation*, 144, pp 782-787.
- Cozzi G, Müller CB, Krauss J, 2008**, How do local habitat management and landscape structure at different spatial scales affect fritillary butterfly distribution on fragmented wetlands? *Landscape Ecology*, 23, pp 269-283.
- Cristofoli S, Mahy G, 2010**, Restauration écologique : contexte, contraintes et indicateurs de suivi, *Biotechnologie, Agronomie, Société et Environnement*, 14, pp 203-211.
- Cristofoli S, Piqueray J, Dufrêne M, 2010**, Colonization credit in restored wet heathlands, *Restoration Ecology*, 18, pp 645-655.
- De Warnaffe GDB, Deconchat M, Ladet S, Balent G, 2006**, Variability of cutting regimes in small private woodlots of south-western France, *Annals of Forest Science*, 63, pp 915-927.
- Dormann CF, Elith J, Bacher S, 2013**, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, pp 27-46.
- Dullinger S, Gatttringer A, Thuiller W, 2012**, Extinction debt of high-mountain plants under twenty-first-century climate change, *Nature Climate Change*, 2, pp 619-622.
- Elith J, Leathwick JR, 2009**, Species distribution models: ecological explanation and prediction across space and time, *Annual Review of Ecology, Evolution, and Systematics*, 40, pp 677-697.
- Ewers RM, Didham RK, Pearse WD, 2013**, Using landscape history to predict biodiversity patterns in fragmented landscapes, *Ecology Letters*, 16, pp 1221-1233.
- Fahrig L, 2003**, Effects of habitat fragmentation on biodiversity, *Annual Review of Ecology, Evolution, and Systematics*, pp 487-515.
- Fahrig L, Merriam G, 1985**, Habitat patch connectivity and population survival, *Ecology*, 1762-1768.
- Favre C, Granier E, Cosserat-Mangeot R, Dupouey JL 2012**, La digitalisation des cartes anciennes, Manuel pour la vectorisation de l'usage des sols et le géo-référencement de la carte d'Etat-major, [en ligne]. URL: <http://www.foretsanciennes.fr/wp-content/uploads/2012/12/Fabvre-al-2011-digitalisation.pdf>
- Fayt P, Dufrêne M, Branquart E, 2006**, Contrasting responses of saproxylic insects to focal habitat resources, the example of longhorn beetles and hoverflies in Belgian deciduous forests, *Journal of Insect Conservation*, 10, pp 129-150.
- F Dormann C, M McPherson J, B Araújo M, 2007**, Methods to account for spatial autocorrelation in the analysis of species distributional data: a review, *Ecography*, 30, pp 609-628.
- Foley JA, DeFries R, Asner GP, 2005**, Global consequences of land use, *Science*, 309, pp 570-574.
- Ford HA, Walters JR, Cooper CB, 2009**, Extinction debt or habitat change?-Ongoing losses of woodland birds in north-eastern New South Wales, *Australia. Biological Conservation*, 142, pp 3182-3190.

- Franklin J, Wejnert KE, Hathaway SA, 2009**, Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California, *Diversity and Distributions*, 15, pp 167-177.
- Gelman A, Park DK, Ansolabehere S, 2001**, Models, assumptions and model checking in ecological regressions, *Journal of the Royal Statistical Society, Series A, Statistics in Society*, 164, pp 101-118.
- George EI, 2000**, The variable selection problem, *Journal of the American Statistical Association*, 95, pp 1304-1308.
- Good JA, Speight MCD, 1996, Saproxylic invertebrates and their conservation throughout Europe.
- Gosselin M, Laroussinie O, 2004**, *Biodiversité et gestion forestière : connaître pour préserver, synthèse bibliographique*, Ed Cemagref, 111p.
- Guardiola M, Pino J, Rodà F, 2013**, Patch history and spatial scale modulate local plant extinction and extinction debt in habitat patches, *Diversity and Distributions*, 19, pp 825-833.
- Guisan A, Zimmermann NE, 2000**, Predictive habitat distribution models in ecology, **Ecological Modelling**, 135, pp 147-186.
- Gutiérrez D, León-Cortés JL, Menéndez R, 2001**, Metapopulations of four lepidopteran herbivores on a single host plant, *Lotus corniculatus*, *Ecology*, 82, pp 1371-1386.
- Hanski I, 2000**, Extinction debt and species credit in boreal forests, modelling the consequences of different a roaches to biodiversity conservation, *Annales Zoologici Fennici*, 271-280.
- Hanski I, 2005**, Landscape fragmentation, biodiversity loss and the societal response, *EMBO reports*, 6, pp 388-392.
- Hanski I, Ovaskainen O, 2002**, Extinction debt at extinction threshold, *Conservation Biology*, 16, pp 666-673.
- Hanski I, Thomas CD, 1994**, Metapopulation dynamics and conservation: a spatially explicit model a lied to butterflies, *Biological Conservation*, 68, pp 167-180.
- Hastie T, Tibshirani R, 1986**, Generalized additive models, *Statistical Science*, pp 297-310.
- Helm A, Hanski I, Pärtel M, 2006**, Slow response of plant species richness to habitat loss and fragmentation, *Ecology Letters*, 9, pp 72-77.
- Herrault P-A, Sheeren D, Fauvel M, Paegelow M, 2013**, Automatic extraction of forests from historical maps based on unsupervised classification in the CIELab color space, *in Geographic Information Science at the Heart of Europe*, Springer 95-112.
- Hess G, 1996**, Disease in metapopulation models: implications for conservation, *Ecology*, pp 1617-1632.
- Jackson ST, Sax DF, 2010**, Balancing biodiversity in a changing environment : extinction debt, immigration credit and species turnover, *Trends in Ecology & Evolution*, 25, pp 153-160.
- Jacquemyn H, Butaye J, Hermy M, 2003**, Influence of environmental and spatial variables on regional distribution of forest plant species in a fragmented and changing landscape, *Ecography*, 26, pp 768-776.

- Keil P, Konvicka M, 2005**, Local species richness of Central European hoverflies, Diptera Syrphidae: a lesson taught by local faunal lists, *Diversity and Distributions*, 11, pp 417-426.
- Krauss J, Bommarco R, Guardiola M, 2010**, Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels, *Ecology Letters*, 13, pp 597-605.
- Kuussaari M, Bommarco R, Heikkinen RK, 2009**, Extinction debt: a challenge for biodiversity conservation, *Trends in Ecology & Evolution*, 24, pp 564-571.
- Laird NM, Ware JH, 1982**, Random-effects models for longitudinal data, *Biometrics*, 38(4), pp 963-974.
- Lancelot R, Lesnoff M, 2005**, Sélection de modèles avec l'AIC et critères d'information dérivés [en ligne]. URL : ftp://ftp.cirad.fr/pub/group-r/groupe-r/Fiches/AIC_v3.pdf
- Larrieu L, Gonin P, 2008**, L'Indice de biodiversité Potentielle, IBP : une méthode simple et rapide pour évaluer la biodiversité potentielle des peuplements forestiers, *Revue Forestière Française*, 60, pp 727-748.
- Larsson TB, 2001**, *Biodiversity evaluation tools for European forests, Criteria and Indicators for Sustainable Forest Management at the Forest Management Unit Level 75*, Wiley Blackwell, 240p.
- Leprieur F, Rubin A, 2011**, INQUAVA : Prédiction de l'établissement des espèces exotiques dans les milieux aquatiques : vers une anticipation des invasions biologiques, *rapport de la Convention Onema-MNHN 2010 Dynamique des biocénoses aquatiques (Action 7)*, [en ligne]. URL: http://www.gt-ibma.eu/wp-content/uploads/2013/01/INVAQUA_rapport_Octobre2011.pdf
- Le Rest K, 2013**, *Méthodes statistiques pour la modélisation des facteurs influençant la distribution et l'abondance de populations : Application aux rapaces diurnes nichant en France*, Thèse en Statistiques, Écologie, Informatique, Université de Poitiers, 156p.
- Leyk S, Boesch R, Weibel R, 2006**, Saliency and semantic processing: Extracting forest cover from historical topographic maps, *Pattern recognition*, 39, pp 953-968.
- Lindborg R, Eriksson O, 2004**, Historical landscape connectivity affects present plant species diversity, *Ecology*, 85, pp 1840-1845.
- Lin D, Lin MD, 2009**, Package "VIF." [en ligne]. URL : <https://cran.r-project.org/web/packages/VIF/index.html>
- MacArthur RH, 1967**, *The theory of island biogeography*, Princeton University Press, 224p.
- Magle SB, Theobald DM, Crooks KR, 2009**, A comparison of metrics predicting landscape connectivity for a highly interactive species along an urban gradient in Colorado, USA, *Landscape Ecology*, 24, pp 267-280.
- McFadden D, 1973**, Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, Ed P. Zarembka 105-142, Academic Press: New York, 1974.
- Metzger JP, Martensen AC, Dixo M, 2009**, Time-lag in biological responses to landscape changes in a highly dynamic Atlantic forest region. *Biological Conservation* 142, pp 1166-1177.
- Münzbergová Z, Mildén M, Ehlén J, Herben T, 2005**, Population viability and reintroduction strategies: a spatially explicit landscape-level approach, *Ecological Applications*, 15, pp 1377-1386.

- Nagelkerke NJ, 1991**, A note on a general definition of the coefficient of determination, *Biometrika* 78, pp 691-692.
- Oksanen J, Kindt R, Legendre P, 2007**, The vegan package [en ligne]. URL: <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Ouin A, Sarthou J-P, Bouyjou B, 2006**, The species-area relationship in the hoverfly, Diptera, Syrphidae, communities of forest fragments in southern France, *Ecography*, 29, pp 183-190.
- Peterken GF, Game M, 1984**, Historical factors affecting the number and distribution of vascular plant species in the woodlands of central Lincolnshire, *The Journal of Ecology*, 155-182.
- Piqueray J, Cristofoli S, Bisteau E, 2011**, Testing coexistence of extinction debt and colonization credit in fragmented calcareous grasslands with complex historical dynamics, *Landscape Ecology*, 26, pp 823-836.
- Raymond L, 2013**, *Stratégies d'hivernation et d'exploitation du paysage par les syrphes aphidiphages : implications pour le contrôle biologique des pucerons*, Thèse en écologie, Agrocampus ouest / INP-ENSAT, 125p.
- Raymond L, Sarthou J-P, Plantegenest M, 2014**, Immature hoverflies overwinter in cultivated fields and may significantly control aphid populations in autumn, *Agriculture, Ecosystems & Environment*, 185, pp 99-105.
- Sang A, Teder T, Helm A, Pärtel M, 2010**, Indirect evidence for an extinction debt of grassland butterflies half century after habitat loss, *Biological Conservation*, 143, pp 1405-1413.
- Sarthou J-P, Speight M, 2005**, Les Diptères Syrphidés, peuple de tous les espaces, *Insectes*, 137, pp 3-8.
- Saunders DA, Hobbs RJ, Margules CR, 1991**, Biological consequences of ecosystem fragmentation: a review, *Conservation Biology*, 5, pp 18-32.
- Seal HL, 1967**, Studies in the history of probability and statistics. XV The historical development of the Gauss linear model, *Biometrika*, 54, pp 1-24.
- Segurado P, Araujo MB, 2004**, An evaluation of methods for modelling species distributions, *Journal of Biogeography*, 31, pp 1555-1568.
- Shibata R, 1981**, An optimal selection of regression variables, *Biometrika*, 68, pp 45-54.
- Smith GF, Gittings T, Wilson M, 2008**, Identifying practical indicators of biodiversity for stand-level management of plantation forests, *Biodiversity and Conservation*, 17, pp 991-1015.
- Sommaggio D, 1999**, Syrphidae: can they be used as environmental bioindicators? *Agriculture, Ecosystems & Environment*, 74, pp 343-356.
- Speight MCD, 2010**, Species accounts of European (Diptera Syrphidae), *Syrph the Net, the Database of European Syrphidae*, 59, 285p.
- Speight MCD, Good JA, 2003**, Development of eco-friendly forestry practices in Europe and the maintenance of saproxylic biodiversity, in Proceedings of the International Symposium "Dead wood a key to biodiversity", 73-77.
- Tilman D, May RM, Lehman CL, Nowak MA, 1994**, Habitat destruction and the extinction debt, *Nature*, 371, pp 65-66.

- Tischendorf L, Bender DJ, Fahrig L, 2003**, Evaluation of patch isolation metrics in mosaic landscapes for specialist vs. generalist disperser, *Landscape Ecology*, 18, pp 41-50.
- Vellend M, Verheyen K, Jacquemyn H, 2006**, Extinction debt of forest plants persists for more than a century following habitat fragmentation, *Ecology* 87, pp 542-548.
- Verheyen K, Honnay O, Motzkin G, 2003**, Response of forest plant species to land-use change: a life-history trait-based approach, *Journal of Ecology*, 91, pp 563-577.
- Wedderburn RW, 1974**, Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method, *Biometrika*, 61, pp 439-447.
- White H, 1980**, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica: Journal of the Econometric Society*, 817-838.
- Young AG, Clarke GM, 2000**, Genetics, demography and viability of fragmented populations, Cambridge University Press, 460p.
- Zuur AF, Ieno EN, Elphick CS, 2010**, A protocol for data exploration to avoid common statistical problems, *Methods in Ecology and Evolution*, 1, pp 3-14.
- Andren H, 1994**, Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat, a review, *Oikos*, 71, pp 355-366.
- Andrieu E, De Warnaffe G, Ladet S, 2008**, Cartographie l'historique des coupes forestières dans les petits bois, *Revue Forestière Française*, 60, pp 667-676.
- Atobe T, Osada Y, Takeda H, 2014**, Habitat connectivity and resident shared predators determine the impact of invasive bullfrogs on native frogs in farm ponds, in *Proceedings of the Royal Society B: Biological Sciences*, 281, pp 21-26.
- Baily B, Riley M, Aucott P, Southall H, 2011**, Extracting digital data from the First Land Utilisation Survey of Great Britain-Methods, issues and potential, *Applied Geography*, 31, pp 959-968.
- Bommarco R, Lindborg R, Marini L, Öckinger E, 2014**, Extinction debt for plants and flower-visiting insects in landscapes with contrasting land use history, *Diversity and Distributions*, 20, pp 591-599.
- Bouget C, Parmain G, Gilg O, Noblecourt T, Nusillard B, Paillet Y, Pernot C, Larrieu L, Gosselin F, 2014**, Does a set aside conservation strategy help restore old-growth attributes and conserve saproxylic beetles in temperate forests?, *Animal Conservation*, 17, pp 342-353.
- Brooks TM, Pimm SL, Oyugi JO, 1999**, Time lag between deforestation and bird extinction in tropical forest fragments, *Conservation Biology*, 13, pp 1140-1150.
- Brotens L, Mönkkönen M, Martin JL, 2003**, Are fragments islands? landscape context and density-area relationships in boreal forest birds, *The American Naturalist*, 162, pp 343-357.
- Burnham KP, Anderson DR, 1998**, Model selection and inference: a practical information-theoretic approach, Springer, New York.
- Cateau E, Larrieu L, Vallauri D, Savoie JM, Touroult J, Brustel H, 2015**, Ancienneté et maturité : deux qualités complémentaires d'un écosystème forestier, *Comptes Rendus Biologies*, 338, pp 58-73.
- Connor EF, McCoy ED, 1979**, The statistics and biology of the species-area relationship, *American Naturalist*, 113(6), pp 791-833.

- Cousins SA, 2009**, Extinction debt in fragmented grasslands: paid or not? , *Journal of Vegetation Science*, 20, pp 3-7.
- Cousins SA, Vanhoenacker D, 2011**, Detection of extinction debt depends on scale and specialization, *Biological Conservation*, 144, pp 782-787.
- Decaëns T, Margerie P, Aubert M, 2008**, Assembly rules within earthworm communities in North-Western France- a regional analysis, *Applied Soil Ecology*, 39, pp 321-335.
- Deconchat M, Balent G, 2002**, Effets de la sylviculture et de l'exploitation forestière sur la Diversité végétale, *Revue Forestière Française*, 54, pp 559-566.
- De Groot M, 2003**, Insect responses to invasive plant species, Mémoire de Master, University of Ljubljana, 50p.
- De Warnaffe GDB, Deconchat M, Ladet S, Balent G, 2006**, Variability of cutting regimes in small private woodlots of south-western France, *Annals of forest science*, 63, pp 915-927.
- Dormann CF, Elith J, Bacher S, 2013**, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, pp 27-46.
- Ewers RM, Didham RK, Pearse WD, 2013**, Using landscape history to predict biodiversity patterns in fragmented landscapes. *Ecology Letters* 16, pp 1221-1233.
- Fahrig L, 2003**, Effects of habitat fragmentation on biodiversity, *Annual review of ecology, evolution, and systematics*, 487-515.
- Favre C, Granier E, Cosserat-Mangeot R, Dupouey JL 2012**, La digitalisation des cartes anciennes, Manuel pour la vectorisation de l'usage des sols et le géo-référencement de la carte d'Etat-major, [en ligne]. URL: <http://www.foretsanciennes.fr/wp-content/uploads/2012/12/Fabvre-al-2011-digitalisation.pdf>
- Fayt P, Dufrêne M, Branquart E, 2006**, Contrasting responses of saproxylic insects to focal habitat resources, the example of longhorn beetles and hoverflies in Belgian deciduous forests, *Journal of Insect Conservation*, 10, pp 129-150.
- Foley JA, DeFries R, Asner GP, 2005**, Global consequences of land use, *Science*, 309, pp 570-574.
- Ford HA, Walters JR, Cooper CB, 2009**, Extinction debt or habitat change?-Ongoing losses of woodland birds in north-eastern New South Wales, *Australia, Biological Conservation*, 142, pp 3182-3190.
- Good JA, Speight MCD, 1996**, Saproxylic invertebrates and their conservation throughout Europe
- Gosselin M, Laroussinie O, 2004**, *Biodiversité et gestion forestière : connaître pour préserver, synthèse bibliographique*, Ed Cemagref, 111p.
- Gossner MM, Lachat T, Brunet J, 2013**, Current near-to-nature forest management effects on functional trait composition of saproxylic beetles in beech forests, *Conservation Biology*, 27, pp 605-614.
- Groot M, Bevk D, 2012**, Ecosystem services and phenology of hoverflies (Diptera Syrphidae) in a slovenian forest stand, *Les Ljubljana*, 64, pp 123-128.
- Guisan A, Zimmermann NE, 2000**, Predictive habitat distribution models in ecology, *Ecological Modelling*, 135, pp 147-186.

- Hanski I, 2005**, Landscape fragmentation, biodiversity loss and the societal response, *EMBO reports*, 6, pp 388-392.
- Hanski I, Ovaskainen O, 2002**, Extinction debt at extinction threshold, *Conservation Biology*, 16, pp 666-673.
- Hanski I, Thomas CD, 1994**, Metapopulation dynamics and conservation: a spatially explicit model applied to butterflies. *Biological Conservation* 68, pp 167-180.
- Harris LD, 1984**, *The fragmented forest: island biogeography theory and the preservation of biotic diversity*, University of Chicago press, 230p.
- Helm A, Hanski I, Pärtel M, 2006**, Slow response of plant species richness to habitat loss and fragmentation, *Ecology Letters*, 9, pp 72-77.
- Herrault P-A, Sheeren D, Fauvel M, Paegelow M, 2013**, Automatic extraction of forests from historical maps based on unsupervised classification in the CIELab color space, in *Geographic Information Science at the Heart of Europe*, Springer 95-112.
- Hess G, 1996**, Disease in metapopulation models, implications for conservation, *Ecology*, 72(5), 1617-1632.
- Huston MA, Huston MA, 1994**, *Biological diversity: the coexistence of species*, Cambridge University Press, 704p.
- IGN 2011**. Scan Etat-major 40K descriptif de contenu, version 1, IGN, Saint-Mandé.
- Keil P, Konvicka M, 2005**, Local species richness of Central European hoverflies, Diptera Syrphidae: a lesson taught by local faunal lists, *Diversity and Distributions*, 11, pp 417-426.
- Koerner W, Cinotti B, Jussy J-H, Benoît M, 2000**, Evolution des surfaces boisées en France depuis le début du XIXe siècle : identification et localisation des boisements des territoires agricoles abandonnés, *Revue Forestière Française*, 52, pp 249-270.
- Krauss J, Bommarco R, Guardiola M, 2010**, Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels, *Ecology letters*, 13, pp 597-605.
- Kuussaari M, Bommarco R, Heikkinen RK, 2009**, Extinction debt: a challenge for biodiversity conservation, *Trends in Ecology & Evolution*, 24, pp 564-571.
- Larrieu L, 2005**, *Etude de certains aspects de la diversité biologique de la forêt des montagnes particulières de Hèches, Vallée d'Aure, Hautes-Pyrénées, en vue d'une gestion sylvicole compatible avec sa conservation*, CRPF de Midi-Pyrénées, rapport, 188p.
- Larrieu L, Cabanettes A, Delarue A, 2012**, Impact of sylviculture on dead wood and on the distribution and frequency of tree microhabitats in Montane Beech-Fir forests of the Pyrenees, *European journal of Forest Research*, 131(3), pp 773-786.
- Larrieu L, Gonin P, 2008**, L'Indice de Biodiversité Potentielle, IBP : une méthode simple et rapide pour évaluer la biodiversité potentielle des peuplements forestiers, *Revue Forestière Française*, 60, pp 727-748.
- Larsson TB, 2001**, *Biodiversity evaluation tools for European forests, Criteria and Indicators for Sustainable Forest Management at the Forest Management Unit Level* 75, Wiley Blackwell, 240p.

- Leyk S, Boesch R, Weibel R, 2006**, Saliency and semantic processing, Extracting forest cover from historical topographic maps, *Pattern Recognition*, 39, pp 953-968.
- Lindborg R, Eriksson O, 2004**, Historical landscape connectivity affects present plant species diversity, *Ecology*, 85, pp 1840-1845.
- Lin D, Lin MD, 2009**, Package "VIF."
- MacArthur RH, 1967**, *The theory of island biogeography*, Princeton University Press, 224p.
- Magle SB, Theobald DM, Crooks KR, 2009**, A comparison of metrics predicting landscape connectivity for a highly interactive species along an urban gradient in Colorado, USA, *Landscape Ecology*, 24, pp 267-280.
- Metzger JP, Martensen AC, Dixo M, 2009**, Time-lag in biological responses to landscape changes in a highly dynamic Atlantic forest region. *Biological Conservation* 142, pp 1166-1177.
- Meyer B, Jauker F, Steffan-Dewenter I, 2009**, Contrasting resource-dependent responses of hoverfly richness and density to landscape structure, *Basic Applied Ecology*, 10, pp 178-186.
- Münzbergová Z, Mildén M, Ehrlén J, Herben T, 2005**, Population viability and reintroduction strategies: a spatially explicit landscape-level approach, *Ecological Applications*, 15, pp 1377-1386.
- Nadal F, 2011**, La cartographie militaire des Pyrénées françaises et la guerre civile espagnole, *Sud-Ouest Européen*, 31, pp 169-182.
- Oksanen J, Kindt R, Legendre P, 2007**, The vegan package [en ligne]. URL: <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Ouin A, Sarthou J-P, Bouyjou B, 2006**, The species-area relationship in the hoverfly, Diptera, Syrphidae, communities of forest fragments in southern France, *Echography*, 29, pp 183-190.
- Piqueray J, Cristofoli S, Bisteau E, 2011**, Testing coexistence of extinction debt and colonization credit in fragmented calcareous grasslands with complex historical dynamics, *Landscape Ecology*, 26, pp 823-836.
- Reemer M, 2005**, Saproxyllic hoverflies benefit by modern forest management, Diptera: Syrphidae, *Journal of Insect Conservation*, 9, pp 49-59.
- R Development Core Team, 2014**. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna.
- Ricklefs RE, 1987**, Community diversity, relative roles of local and regional processes, *Science* 235, pp 167-171.
- Sarthou J-P, Speight M, 2005**, Les Diptères Syrphidés, peuple de tous les espaces, *Insectes*, 137, pp 3-8.
- Sarthou V, Sarthou J-P, 2007**, *Evaluation écologique d'écosystèmes forestiers de Réserves Naturelles de Haute Savoie à l'aide des Diptères Syrphidés. Zone Natura 2000 Arve-Giffre, Réserve Naturelle des Contamines-Montjoie, Réserve Naturelle des Aiguilles Rouges*, Rapport d'études pour ASTERS, pp 93.
- Sarthou JP., Fromage P, Genet B, Vinauger A, Heintz W et Monteil C, 2010**. SYRFID vol. 4: Syrphidae of France Interactive Data [On-Line URL: syrfid.ensat.fr].

Smith GF, Gittings T, Wilson M, 2008, Identifying practical indicators of biodiversity for stand-level management of plantation forests, *Biodiversity Conservation* 17, pp 991-1015.

Sommaggio D, 1999, Syrphidae: can they be used as environmental bioindicators?, *Agriculture, Ecosystems & Environment*, 74, pp 343-356.

Speight MCD, 1989, Saproxyllic invertebrates and their conservation, *Concil of Europe, Nature and Environment Series*, 42, pp 1-79.

Speight MCD, 2004, An « expert system” a roach to development of decision tools for use in maintenance of invertebrate biodiversity forests, *Pan-European Ecological Network in Forests: Conservation of Biodiversity an Sustainable Management ; Joint United Nations Environnemental Programme/Council of Europe Symposium*.

Speight MCD, 2013, Species accounts of European Syrphidae, Diptera, 2010, *Syrph Net Database European Syrphidae*, 72, pp 316.

Speight MCD, Castella E, Obrdlik P, 2000, Use of the Syrph the Net database 2000, *Syrph Net Database European Syrphidae*, 25, 99p.

Speight MCD, Good JA, 2003, Development of eco-friendly forestry practices in Europe and the maintenance of saproxyllic biodiversity, in Proceedings of. International Symposium, “Dead Wood Key Biodiversity”, pp 73-77.

Speight MR, Wainhouse D, 1989, Ecology and management of forest insects, Clarendon Press,

Tilman D, May RM, Lehman CL, Nowak MA, 1994, Habitat destruction and the extinction debt, *Nature*, 371, pp 65-66.

Tischendorf L, Bender DJ, Fahrig L, 2003, Evaluation of patch isolation metrics in mosaic landscapes for specialist vs. generalist dispersers, *Landscape Ecology*, 18, pp 41-50.

Turner MG, 2001, *Landscape ecology in theory and practice*, Springer, 406p.

Winter S, Möller GC, 2008, Microhabitats in lowland beech forests as monitoring tool for nature conservation, *Forest Ecological Management*, 255, pp 1251-1261.

Zuur AF, Ieno EN, Elphick CS, 2010, A protocol for data exploration to avoid common statistical problems, *Methods in Ecology and Evolution*, 1, pp 3-14.

CHAPITRE 11

Barry S, Elith J, 2006, Error and uncertainty in habitat models, *Journal of Applied Ecology*, 43, pp 413-423.

Beissinger SR, Westphal MI, 1998, On the use of demographic models of population viability in endangered species management, *The Journal of Wildlife Management*, 62, 821-841.

Chrisman NR, 1991, *The error component in spatial data*, Geographical Information Systems, 1, pp 165-174.

Congalton RG, 1994, Green, K., 1999, Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, CRC Press, 200p.

- Devillers R, 2004**, *Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales*, Thèse en sciences de l'information géographique, Université de Marne la Vallée, 167p.
- Devillers R, Goodchild H, 2010**, *Spatial data quality: from process to decisions*, CRC Press, 226p.
- Devillers R, Jeansoulin R, 2005**, *Qualité de l'information géographique*, Hermes, 343p.
- Edwards TC, Moisen GG, Cutler DR, 1998**, Assessing map accuracy in a remotely sensed, ecoregion-scale cover map, *Remote Sensing of Environment*, 63, pp 73-83.
- Fisher P, Comber A, Wadsworth R, Richard, 2005**, Nature de l'incertitude pour les données spatiales, in *Qualité de l'information géographique*, chap3.
- Fisher PF, 1999**, Models of uncertainty in spatial data, *Geographical Information Systems*, 1, pp 191-205.
- Girres J-F, 2013**, Modèle d'estimation de l'imprécision des mesures géométriques de données géographiques, Thèse en sciences de l'information géographique, Université Paris-Est, 373p.
- Goodchild MF, 1995**, Sharing imperfect data, [en ligne, URL: <http://www.geog.ucsb.edu/~good/papers/228.pdf>]
- Graham CH, Elith J, Hijmans RJ, 2008**, The influence of spatial errors in species occurrence data used in distribution models, *Journal of Applied Ecology*, 45, pp 239-247.
- Hines EM, Franklin J, Stephenson JR, 2005**, Estimating the effects of map error on habitat delineation for the California spotted owl in southern California, *Transactions in GIS*, 9, pp 541-559.
- Hunter GJ, 1999**, Managing uncertainty in GIS, *Geographical Information Systems*, 2, pp 633-641.
- Jager HI, King AW, 2004**, Spatial uncertainty and ecological models, *Ecosystems*, 7, pp 841-847.
- Kyriakidis PC, Dungan JL, 2001**, A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions, *Environmental and Ecological Statistics*, 8, pp 311-330.
- Langford WT, Gergel SE, Dietterich TG, Cohen W, 2006**, Map misclassification can cause large errors in landscape pattern indices, examples from habitat fragmentation, *Ecosystems*, 9, pp 474-488.
- Lechner AM, Langford WT, Bekessy SA, Jones SD, 2012**, Are landscape ecologists addressing uncertainty in their remote sensing data?, *Landscape Ecology*, 27, pp 1249-1261.
- Leyk S, Boesch R, Weibel R, 2005**, A Conceptual Framework for Uncertainty Investigation in Map-based Land Cover Change Modelling, *Transactions in GIS*, 9, pp 291-322.
- Moudrý V, Šímová P, 2012**, Influence of positional accuracy, sample size and scale on modelling species distributions, a review, *International Journal of Geographical Information Science*, 26, pp 2083-2095.
- Olteanu AM, 2008**, *Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance*, Thèse en informatique, Université Paris-Est, 269p.
- Plewe B, 2002**, The nature of uncertainty in historical geographic information, *Transactions in GIS*, 6, pp 431-456.

- Remmel TK, Perera AH, 2002**, Accuracy of discontinuous binary surfaces: a case study using boreal forest fires, *International Journal of Geographical Information Science*, 16, pp 287-298.
- Rocchini D, Hortal J, Lengyel S, 2011**, Accounting for uncertainty when mapping species distributions, the need for maps of ignorance, *Progress in Physical Geography*, 35, pp 211-226.
- Ruckelshaus M, Hartway C, Kareiva P, 1997**, Assessing the data requirements of spatially explicit dispersal models, *Conservation Biology*, 11, pp 1298-1306.
- R Visscher D, 2006**, GPS measurement error and resource selection functions in a fragmented landscape, *Ecography*, 29, pp 458-464.
- Schneider M, 1999**, Uncertainty management for spatial data in databases: Fuzzy spatial data types, *Advances in Spatial Databases*, Springer, pp 330-351.
- Sheeren D, 2005**, *Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales. Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique*, Thèse en sciences de l'information géographique, Université Pierre et Marie Curie-Paris VI, 294p.
- Stehman SV, 1997**, Selecting and interpreting measures of thematic classification accuracy, *Remote sensing of Environment*, 62, pp 77-89.
- Vauglin F, 1997**, *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*, Thèse en sciences de l'information géographique, Université de Marne La Vallée, 294p.
- Vauglin F, Ali ABH, 1998**, Geometric matching of polygonal surfaces in GISs, in Proceedings of ASPRS Annual Meeting.
- Worboys M, 1998**, Imprecision in finite resolution spatial data, *GeoInformatica*, 2, pp 257-279.
- Zhang J, Goodchild MF, 2002**, Uncertainty in geographical information, CRC press, 288p.

ANNEXES

Annexe A. Résultats des tests de significativité relatifs au protocole 1 (cf. section 5.6.3)

	EFFECTIF DE POINTS D'AMER	25	50	75	100	150
MODELES DE TRANSFORMATION	SIGNIFICATIVITE					
Polynôme d'ordre 1	W	453	473	632	376	634
	p	0.9707	0.7412	0.0073	0.2772	0.0067
Polynôme d'ordre 1	W	507	530	408.5	387.5	457.5
	p	0.4063	0.2398	0.5444	0.3593	0.9176
Polynôme d'ordre 1	W	424	356	487	365.5	379.5
	p	0.7082	0.1681	0.5920	0.2143	0.3007
Triangulation de Delaunay	W	413	499	545	553	549.5
	p	0.5920	0.4761	0.1624	0.1304	0.1433
Noyau Gaussien	W	122	234	461	486	749.5
	p	0.0000	0.0012	0.8766	0.5997	0.0000
Noyau polynomial d'ordre 1	W	422	363	531	412	517.5
	p	0.6865	0.2009	0.2340	0.5793	0.3219
Noyau polynomial d'ordre 2	W	476	556	330	469.5	428
	p	0.7062	0.1194	0.0772	0.7788	0.7524
Noyau polynomial d'ordre 3	W	392	396	342	405.5	226
	p	0.3980	0.4290	0.1120	0.5153	0.0010

Impact du type de points d'amer : Bâtiments/Croisements de route (test de Wilcoxon)

	EFFECTIF DE POINTS D'AMER	ENTRE 25 ET 50	ENTRE 50 ET 75	ENTRE 75 ET 100	ENTRE 100 ET 150
MODELES DE TRANSFORMATION	SIGNIFICATIVITE				
Polynôme d'ordre 1	W	645	464.5	697	398
	p	0.0035	0.8360	0.0003	0.4464
Polynôme d'ordre 2	W	744	710	497.5	571
	p	0.0000	0.0001	0.4871	0.0748
Polynôme d'ordre 3	W	698	491.5	623	491
	p	0.0002	0.5444	0.0100	0.5493
Triangulation de Delaunay	W	429.5	563	605	732.5
	p	0.7675	0.0965	0.0224	0.0000
Noyau Gaussien	W	669.5	460	524	517
	p	0.0012	0.8883	0.2772	0.3255
Noyau polynomial d'ordre 1	W	618	446.5	539.5	561.5
	p	0.0125	0.9646	0.1882	0.1008
Noyau polynomial d'ordre 2	W	536.5	701	341.5	479
	p	0.2036	0.0001	0.1103	0.6735
Noyau polynomial d'ordre 3	W	508.5	619	489	541.5
	p	0.3912	0.0127	0.5692	0.1785

Impact du nombre de points d'amer (test de Wilcoxon)

	EFFECTIF DE POINTS D'AMER	25	50	75	100	150
MODELES DE TRANSFORMATION	SIGNIFICATIVITE					
Polynôme d'ordre 1	W	424.5	483.5	413.5	108.5	34
	p	0.0000	0.0002	0.0000	0.0000	0.0000
Polynôme d'ordre 2	W	470	312	91	85	0
	p	0.0001	0.0000	0.0000	0.0000	0.0000
Polynôme d'ordre 3	W	580	378.5	528.5	249	127.5
	p	0.0027	0.0000	0.0006	0.0000	0.0000
Triangulation de Delaunay	W	378	338	179.5	21	0
	p	0.0000	0.0000	0.0000	0.0000	0.0000
Noyau Gaussien	W	276	274.5	209.5	117	12
	p	0.0000	0.0000	0.0000	0.0000	0.0000
Noyau polynomial d'ordre 1	W	475.5	310	404	115	2
	p	0.0001	0.0000	0.0000	0.0000	0.0000
Noyau polynomial d'ordre 2	W	284	374.5	68	42	0
	p	0.0000	0.0000	0.0000	0.0000	0.0000
Noyau polynomial d'ordre 3	W	117	141	63.5	59.5	0
	p	0.0000	0.0000	0.0000	0.0000	0.0000

Impact de la répartition : aléatoire/régulière (test de Wilcoxon)

Annexe B. Résultats des tests de significativité relatifs au protocole OCS ANCIEN (cf. section 5.7.3)

MODELES DE TRANSFORMATION	CARTES	239SE		239SO		253NO		257SE	
	EFFECTIF DE POINTS D'AMER	W	p	W	p	W	p	W	p
Polynôme d'ordre 1	Entre 25 et 50	856.0000	0.0000	900.0000	0.0000	731.0000	0.0000	675.0000	0.0007
	Entre 50 et 100	711.0000	0.0001	796.0000	0.0000	587.0000	0.0430	566.0000	0.0878
	Entre 100 et 250	596.0000	0.0307	693.0000	0.0002	542.0000	0.1774	525.0000	0.2729
	Entre 250 et 500	505.0000	0.4232	532.0000	0.2301	484.0000	0.6228	482.0000	0.6438
	Entre 500 et 1000	486.0000	0.6022	509.0000	0.3898	474.0000	0.7302	473.0000	0.7412
Polynôme d'ordre 2	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	897.0000	0.0000
	Entre 50 et 100	855.0000	0.0000	899.0000	0.0000	739.0000	0.0000	711.0000	0.0001
	Entre 100 et 250	726.0000	0.0000	856.0000	0.0000	606.0000	0.0207	583.0000	0.0496
	Entre 250 et 500	553.0000	0.1304	628.0000	0.0080	509.0000	0.3898	509.0000	0.3898
	Entre 500 et 1000	509.0000	0.3898	546.0000	0.1591	486.0000	0.6022	490.0000	0.5619
Polynôme d'ordre 3	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 50 et 100	900.0000	0.0000	900.0000	0.0000	892.0000	0.0000	899.0000	0.0000
	Entre 100 et 250	865.0000	0.0000	890.0000	0.0000	760.0000	0.0000	750.0000	0.0000
	Entre 250 et 500	633.0000	0.0063	660.0000	0.0016	560.0000	0.1058	537.0000	0.2025
	Entre 500 et 1000	498.0000	0.4853	556.0000	0.1194	506.5000	0.4077	497.0000	0.4946
Triangulation de Delaunay	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 50 et 100	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 100 et 250	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 250 et 500	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 500 et 1000	900.0000	0.0000	900.0000	0.0000	899.0000	0.0000	899.0000	0.0000
Noyau Gaussien	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	899.0000	0.0000	900.0000	0.0000
	Entre 50 et 100	900.0000	0.0000	884.0000	0.0000	818.0000	0.0000	878.0000	0.0000
	Entre 100 et 250	844.0000	0.0000	900.0000	0.0000	849.0000	0.0000	899.0000	0.0000
	Entre 250 et 500	819.0000	0.0000	900.0000	0.0000	876.0000	0.0000	848.0000	0.0000
	Entre 500 et 1000	754.0000	0.0000	893.0000	0.0000	767.0000	0.0000	690.0000	0.0003
Noyau polynomial d'ordre 1	Entre 25 et 50	825.0000	0.0000	896.0000	0.0000	728.0000	0.0000	737.0000	0.0000
	Entre 50 et 100	663.0000	0.0014	783.0000	0.0000	584.0000	0.0479	593.0000	0.0345
	Entre 100 et 250	597.0000	0.0296	681.0000	0.0005	521.0000	0.2996	536.0000	0.2078
	Entre 250 et 500	516.0000	0.3354	537.0000	0.2025	481.0000	0.6543	503.0000	0.4404
	Entre 500 et 1000	498.0000	0.4853	493.0000	0.5325	475.0000	0.7191	488.0000	0.5819
Noyau polynomial d'ordre 2	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	896.0000	0.0000	894.0000	0.0000
	Entre 50 et 100	840.0000	0.0000	896.0000	0.0000	741.0000	0.0000	750.0000	0.0000
	Entre 100 et 250	694.0000	0.0002	831.5000	0.0000	615.0000	0.0143	611.0000	0.0169
	Entre 250 et 500	540.0000	0.1872	600.5000	0.0266	508.0000	0.3980	509.0000	0.3898
	Entre 500 et 1000	495.0000	0.5133	529.0000	0.2478	483.0000	0.6333	491.0000	0.5520
Noyau polynomial d'ordre 3	Entre 25 et 50	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000
	Entre 50 et 100	900.0000	0.0000	900.0000	0.0000	900.0000	0.0000	889.0000	0.0000
	Entre 100 et 250	865.0000	0.0000	899.0000	0.0000	778.0000	0.0000	780.0000	0.0000
	Entre 250 et 500	623.0000	0.0100	720.0000	0.0000	559.0000	0.1091	582.5000	0.0510
	Entre 500 et 1000	532.0000	0.2301	601.0000	0.0253	509.0000	0.3898	512.0000	0.3659

Annexe C. Matrices de confusion relatives à la procédure d'extraction automatisée appliquée aux dessin-minutes d'Etat-Major

1 A partir des trois extraits sélectionnés

Modèle gaussien		Extraction Manuelle			
Extrait 1		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	686655	24091	710746	7.95
	Non-Forêts	10345	278909	289254	1.48
	Total	697000	303000	1000000	
	Déficit (en %)	3.58	3.39		Kappa = 0.91

K-moyennes		Extraction Manuelle			
Extrait 1		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	692117	51435	743552	16.98
	Non-Forêts	4883	251565	256448	0.7
	Total	697000	303000	1000000	
	Déficit (en %)	1.9	6.92		Kappa = 0.86

Modèle gaussien		Extraction Manuelle			
Extrait 2		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	1669474	65799	1735273	11.54
	Non-Forêts	10587	504140	514727	0.63
	Total	1680061	569939	2250000	
	Déficit (en %)	2.06	3.79		Kappa = 0.90

K-moyennes		Extraction Manuelle			
Extrait 2		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	1668957	63726	1732683	11.18
	Non-Forêts	11104	506213	517317	0.66
	Total	1680061	569939	2250000	
	Déficit (en %)	2.15	3.68		Kappa = 0.90

Modèle gaussien		Extraction Manuelle			
Extrait 3		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	2762098	79817	2841915	6.5
	Non-Forêts	9071	1149014	1158085	0.33
	Total	2771169	1228831	4000000	
	Déficit (en %)	0.78	2.81		Kappa = 0.94

K-moyennes		Extraction Manuelle			
Extrait 3		Forêts	Non-Forêts	Total	Excédent (en %)
Extraction Automatique	Forêts	2766742	46485	2813227	3.78
	Non-Forêts	4427	1182346	1186773	0.16
	Total	2771169	1228831	4000000	
	Déficit (en %)	0.37	1.65		Kappa = 0.96

2 A partir du site d'étude

Modèle gaussien		Extraction Manuelle			
Site d'étude		Forêts	Non-Forêts	Total	Commission(en %)
Extraction Automatique	Forêts	198329832	2167674	200497506	13.29
	Non-Forêts	737374	14023334	14760708	0.37
	Total	199067206	16191008	215258214	
	Omission(en %)	5	1.08		Kappa = 0.89

Annexe D. Matrices de confusion relatives à la procédure d'extraction automatisée appliquée aux minutes d'Etat-Major dans le cadre du Projet OCS ANCIEN

1. La feuille 191NE

191NE		Extraction INRA			
		Forêt	Non-Forêt	Total	Commission(en %)
Extraction DYNAFOR	Forêt	54483	6782280	6836763	0.8
	Non-Forêt	33996649	1460983	35457632	4.12
	Total	34051132	8243263	42294395	
	Omission(en %)	17.72	0.16		Kappa = 0.87

191NE		Extraction INRA			
		Forêt	Non-Forêt	Total	Commission(en %)
Extraction IGN	Forêt	125433	7140853	7266286	1.73
	Non-Forêt	33925699	1102410	35028109	3.15
	Total	34051132	8243263	42294395	
	Omission(en %)	13.37	0.37		Kappa = 0.90

191NE			Extraction INRA		
		Terres-H	Non-Terres-H	Total	Commission(en %)
Extraction DYNAFOR	Terres-H	113969	1164510	1278479	8.91
	Non-Terres-H	40398621	617295	41015916	1.51
	Total	40512590	1781805	42294395	
	Omission(en %)	34.64	0.28		Kappa = 0.75

191NE		Extraction INRA			
		Terres-H	Non-Terres-H	Total	Commission(en %)
Extraction IGN	Terres-H	82511	1333929	1416440	5.83
	Non-Terres-H	40430079	447876	40877955	1.1
	Total	40512590	1781805	42294395	
	Omission(en %)	25.14	0.2		Kappa = 0.82

191NE		Extraction INRA			
		Vignes	Non-vignes	Total	Commission(en %)
Extraction DYNAFOR	Vignes	174273	1376496	1550769	11.24
	Non-vignes	40360188	383438	40743626	0.94
	Total	40534461	1759934	42294395	
	Omission(en %)	21.79	0.43		Kappa = 0.82

		Extraction INRA			
		Vignes	Non-vignes	Total	Commission(en %)
Extraction IGN	Vignes	297651	1325870	1623521	18.33
	Non-vignes	40236810	434064	40670874	1.07
	Total	40534461	1759934	42294395	
	Omission(en %)	24.66	0.73		Kappa = 0.77

2. La feuille198NO

198NO		Extraction INRA			
		Forêt	Non-Forêt	Total	Commission(en %)
Extraction DYNAFOR	Forêt	1154063	13801358	14955421	7.72
	Non-Forêt	24911088	1707005	26618093	6.41
	Total	26065151	15508363	41573514	
	Omission(en %)	11.01	4.43		Kappa = 0.85

198NO		Extraction INRA			
		Terres-H	Non-Terres-H	Total	Commission(en %)
Extraction DYNAFOR	Terres-H	337404	860095	1197499	28.18
	Non-Terres-H	40112692	263323	40376015	0.65
	Total	40450096	1123418	41573514	
	Omission(en %)	23.44	0.83		Kappa = 0.73

198NO		Extraction INRA			
		Forêt	Non-Forêt	Total	Commission(en %)
Extraction DYNAFOR	Forêt	2135194	9551966	11687160	18.27
	Non-Forêt	23929957	5956397	29886354	19.93
	Total	26065151	15508363	41573514	
	Omission(en %)	38.41	8.19		Kappa = 0.56

3. 257NO

257NO		Extraction INRA			
		Forêt	Non-Forêt	Total	Commission(en %)
Extraction DYNAFOR	Forêt	871183	1389454	2260674	29.79
	Non-Forêt	38706926	780889	39487852	1.3
	Total	39578146	2170380	41748563	
	Omission(en %)	31.24	1.8		Kappa = 0.68

257NO		Extraction INRA			
		Terres-H	Non-Terres-H	Total	Commission(en %)
Extraction DYNAFOR	Terres-H	871220	1389491	2260711	38.54
	Non-Terres-H	38706963	780926	39487889	1.98
	Total	39578183	2170417	41748600	
	Omission(en %)	35.98	2.2		Kappa = 0.60

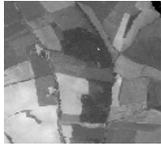
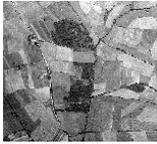
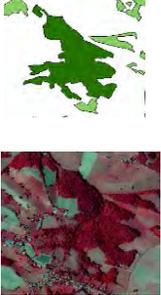
Annexe E : matrice de confusion relative à la procédure d'extraction automatisée appliquée à la BD Ortho® IRC de 2010 (appliquée et validée sur l'ensemble du site d'étude)

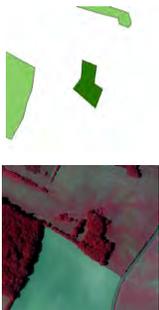
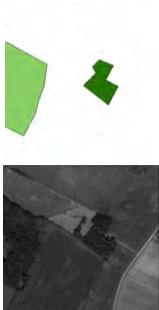
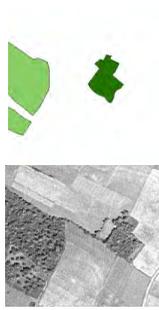
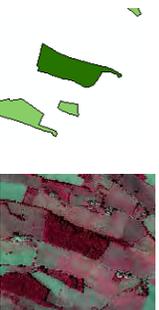
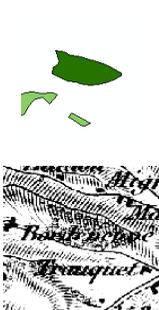
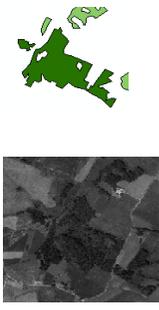
		Extraction manuelle			
		Forêt	Non-Forêt	Total	Excédent(en %)
Extraction automatique	Forêt	2938353325	102655907	3041009232	17.26
	Non-Forêt	48473122	232384046	280857168	3.38
	Total	2986826447	335039953	3321866400	
	Déficit(en %)	30.64	1.62		Kappa = 0.72

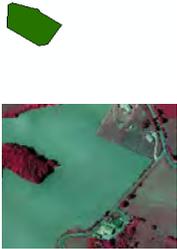
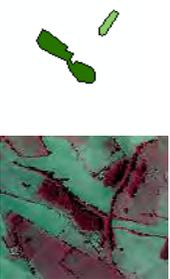
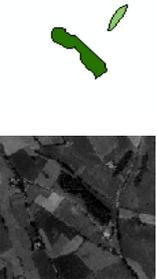
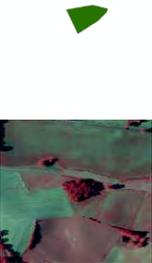
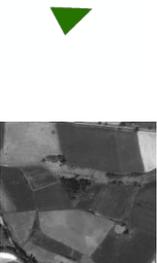
Annexe F : Table d'appariement des forêts étudiées : les encadrés rouges indiquent que l'appariement a entraîné la prise en compte d'une entité multi-partie.

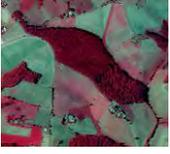
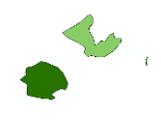
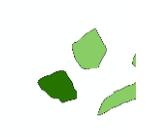
ID FORET	BD2010	BD1979	BD1954	BD1900	BD1850		ID FORET	BD2010	BD1979	BD1954	BD1900	BD1850
1	43A	39B	17C	107D	189E			SUITE				
2	50A	48B	15C	113D	180E		48	112A	312B	204C	143D	87E
2	50A	48B	15C	115D	180E		48	112A	313B	205C	150D	90E
2	50A	48B	16C	0D	0E		40	1265A	1206B	1452C	853D	643E
2	50A	50B	1579C	0D	0E		40	1265A	1206B	1452C	848D	647E
3	111A	393B	26C	203D	262E		40	1265A	1206B	1452C	855D	0E
22	22A	74B	30C	127D	103E		40	1265A	1206B	1452C	858D	645E
5	115A	338B	29C	155D	165E		40	1265A	1206B	1452C	858D	646E
6	177A	306B	21C	143D	199E		40	1265A	1250B	1451C	853D	649E
6	177A	306B	20C	143D	199E		40	1265A	1250B	1451C	853D	650E
6	177A	306B	19C	143D	199E		41	1267A	1248B	714C	849D	688E
6	177A	295B	165C	143D	199E		42	1342A	959B	822C	937D	668E
7	185A	294B	18C	137D	216E		17	290A	213B	363C	326D	0E
8	187A	292B	167C	132D	215E		19	305A	199B	401C	313D	0E
8	187A	292B	168C	132D	215E		30	1115A	976B	1249C	0D	0E
10	253A	420B	32C	297D	340E		30	1115A	978B	0C	0D	0E
11	260A	214B	399C	328D	311E		30	1115A	977B	1248C	0D	0E
11	260A	214B	400C	328D	311E		35	1222A	963B	1270C	750D	0E
12	269A	226B	388C	332D	330E		32	1137A	938B	799C	708D	0E
13	271A	228B	33C	298D	333E		32	1137A	938B	801C	0D	0E
14	278A	187B	37C	305D	306E		4	113A	315B	27C	0D	0E
15	279A	188B	36C	304D	329E		9	241A	422B	35C	0D	0E
16	289A	208B	22C	329D	297E		20	314A	147B	72C	0D	0E
18	298A	198B	368C	312D	300E		24	362A	165B	1577C	0D	0E
18	298A	198B	374C	312D	301E		25	403A	101B	100C	0D	0E
21	316A	152B	1C	17D	110E		29	957A	1506B	632C	0D	0E
21	316A	152B	69C	18D	110E		31	1128A	956B	813C	0D	0E
23	340A	150B	3C	21D	1077E		34	1147A	1036B	1217C	0D	0E
26	798A	850B	1146C	574D	514E		34	1147A	1039B	1217C	0D	0E
27	870A	1504B	1152C	629D	773E		36	1224A	960B	1578C	0D	0E
27	870A	1503B	1153C	629D	773E		38	1344A	1569B	1575C	0D	0E
27	870A	867B	1158C	0D	0E		44	1345A	989B	1272C	0D	0E
27	870A	867B	1159C	629D	773E		47	1348A	1566B	814C	0D	0E
28	889A	922B	780C	693D	629E		38	1244A	1565B	1567C	0D	0E
33	1146A	950B	819C	736D	634E							
33	1146A	951B	820C	737D	634E							
37	1228A	988B	1271C	753D	661E							
39	1252A	1183B	1299C	836D	651E							

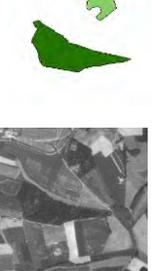
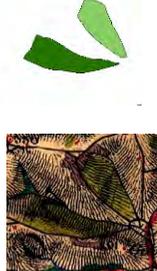
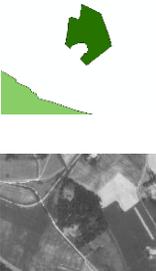
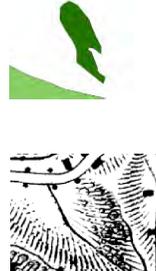
Annexe G Aperçu des forêts appariées à chaque date. Les forêts en vert foncé désignent les forêts partageant la généalogie de la forêt échantillonnée. Les autres (vert clair) sont les forêts dans le voisinage des forêts appariées.

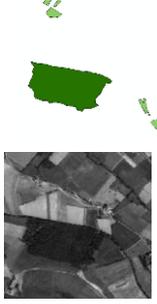
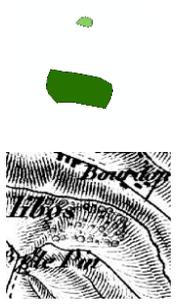
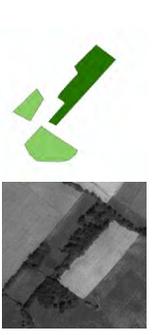
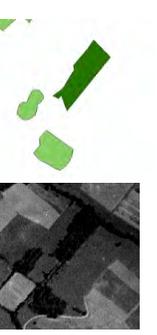
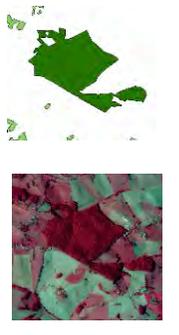
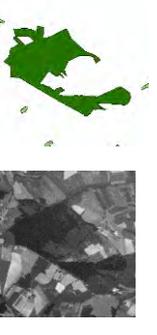
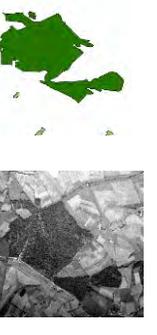
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
1					
2					
3					

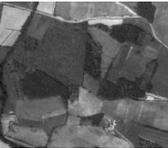
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
4					
5					
6					

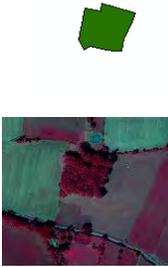
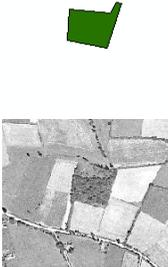
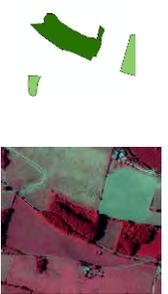
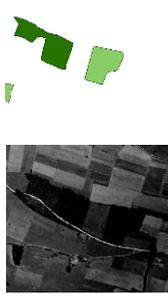
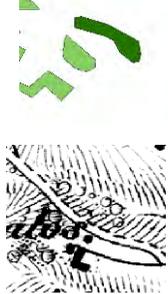
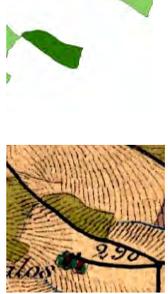
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
7					
8					
9					

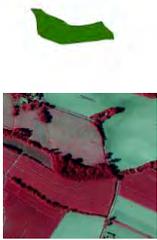
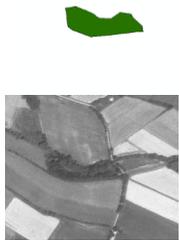
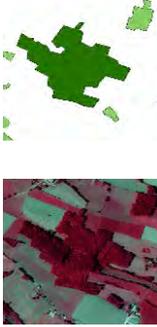
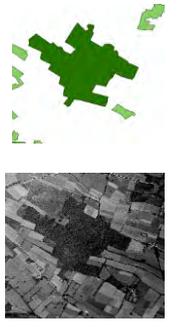
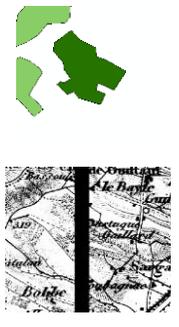
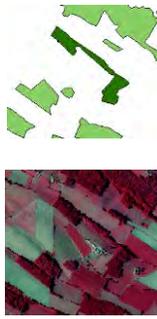
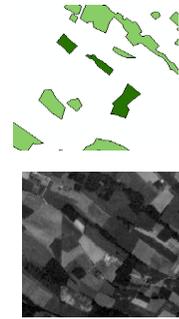
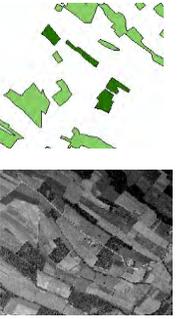
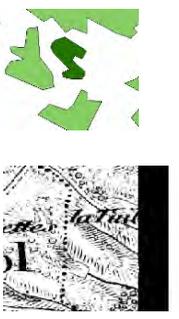
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
10	 	 	 	 	 
11	 	 	 	 	 
12	 	 	 	 	 

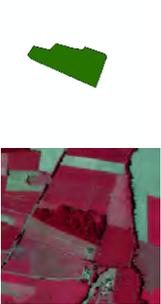
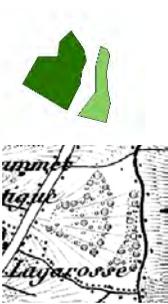
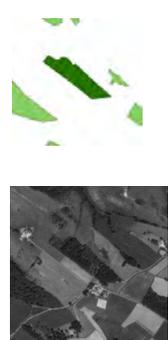
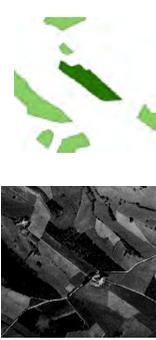
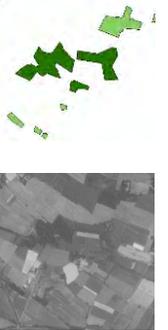
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
13					
14					
15					

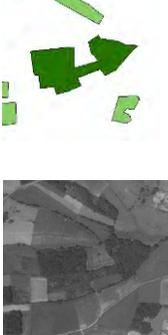
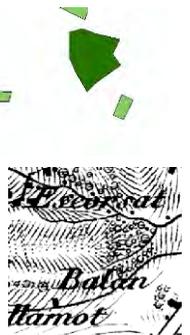
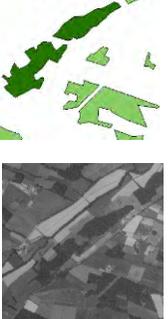
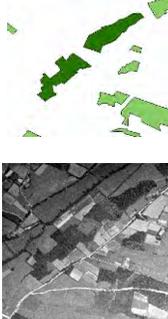
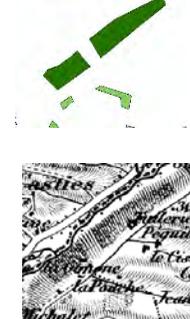
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
16					
17					
18					

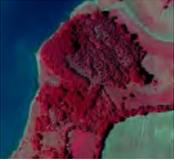
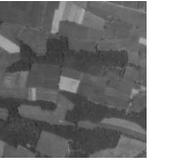
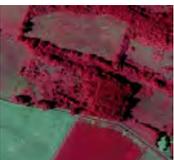
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
19					
20					
21					

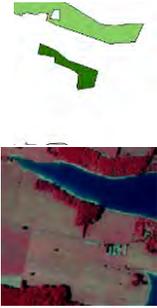
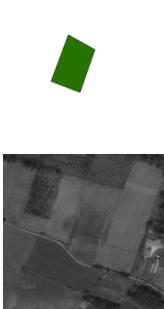
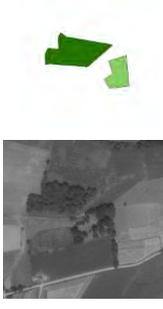
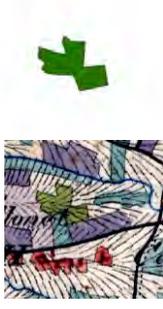
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
22					
23					
24					

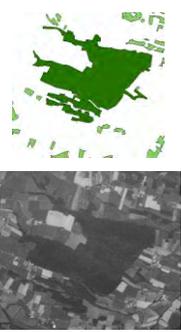
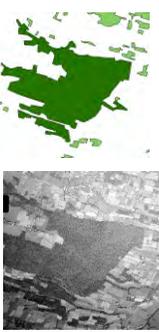
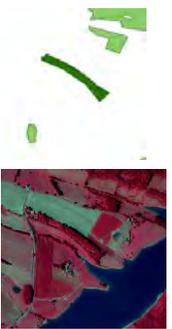
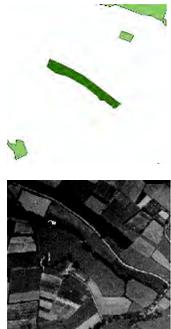
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
25					
26					
27					

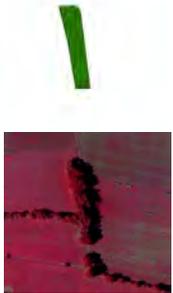
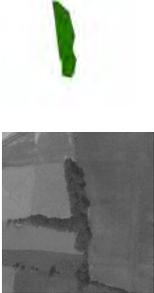
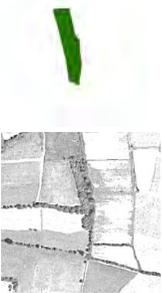
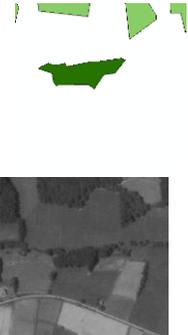
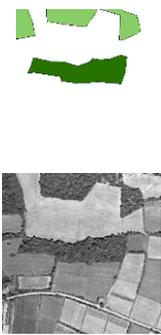
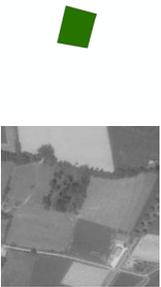
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
28					
29					
30					

ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
31					
32					
33					

ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
34	 	 	 		
35	 	 	 	 	
36	 	 	 		

ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
37					
38					
39					

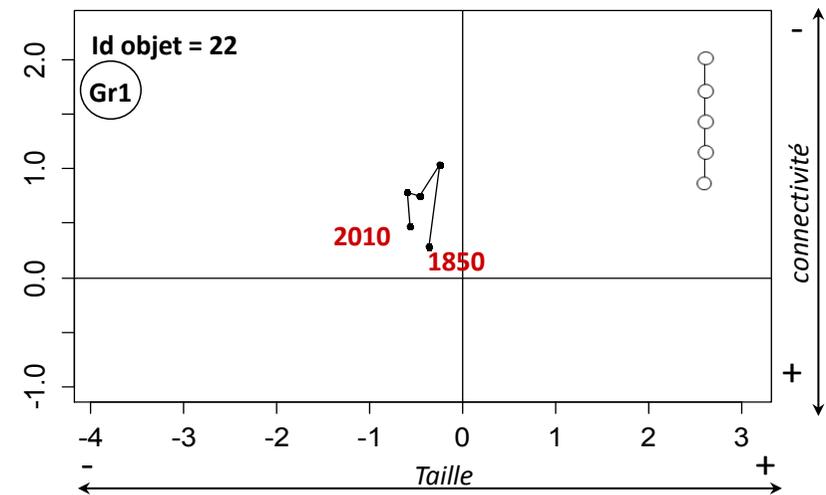
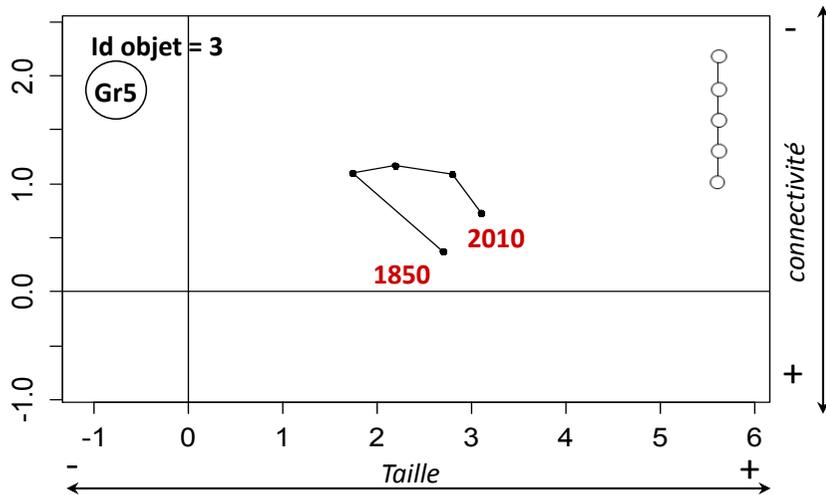
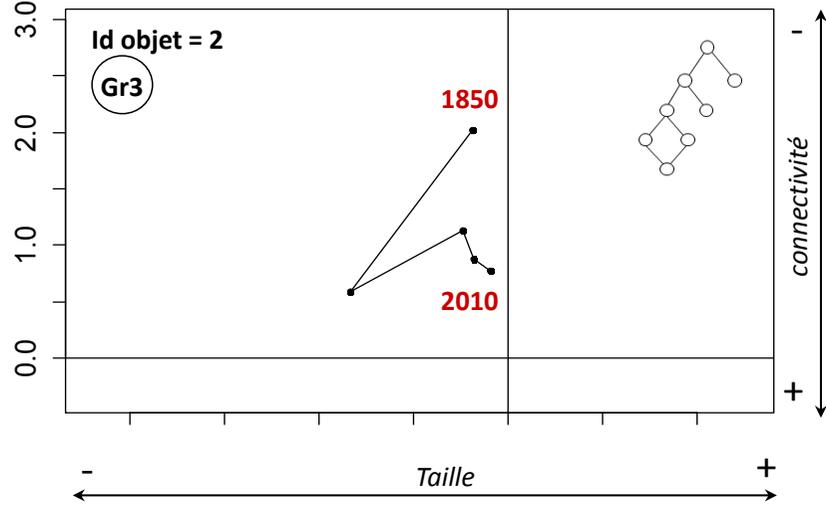
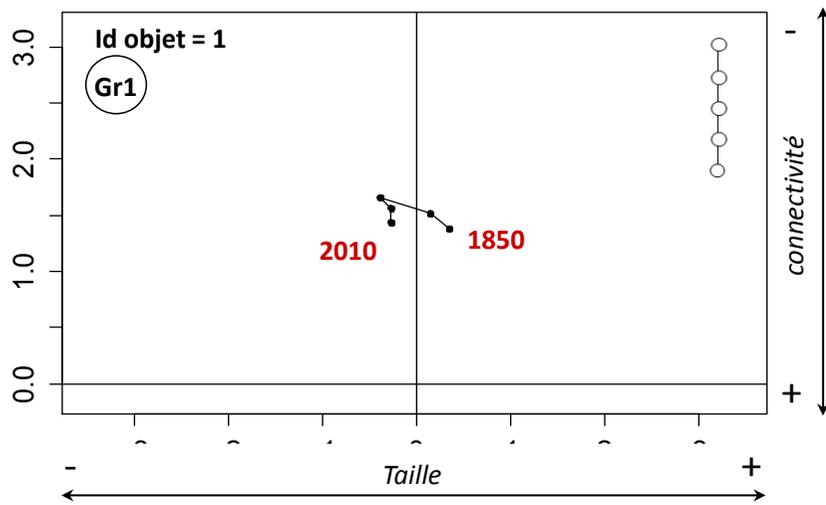
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
40					
41					
42					

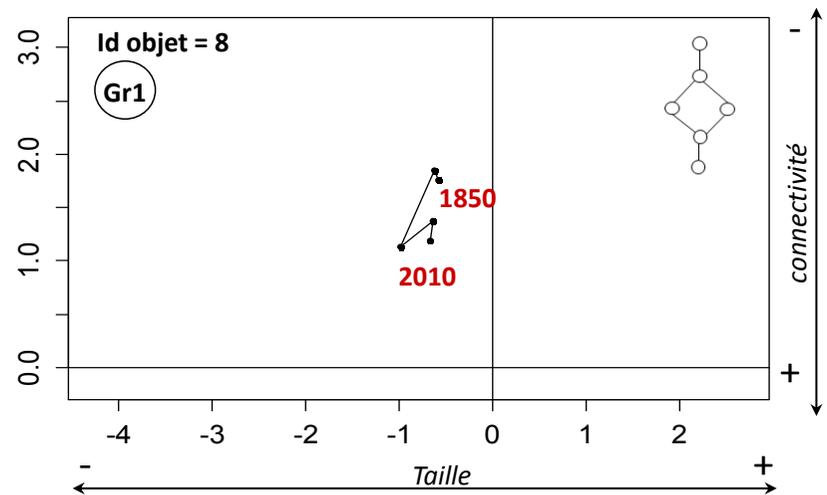
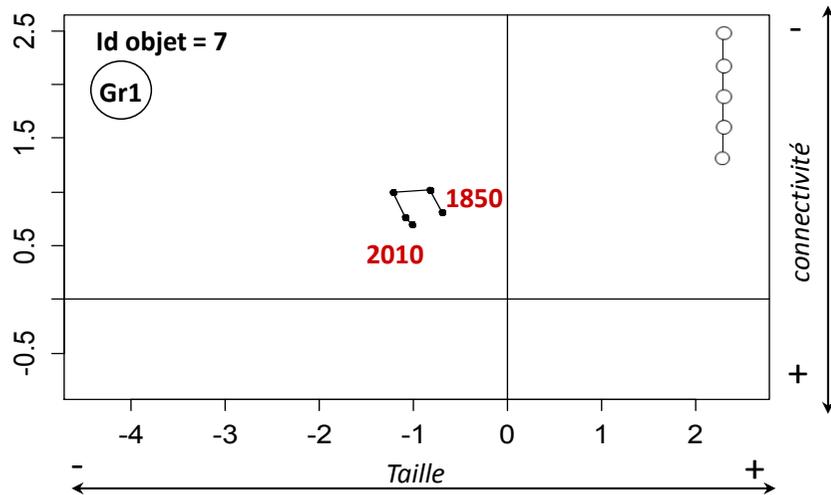
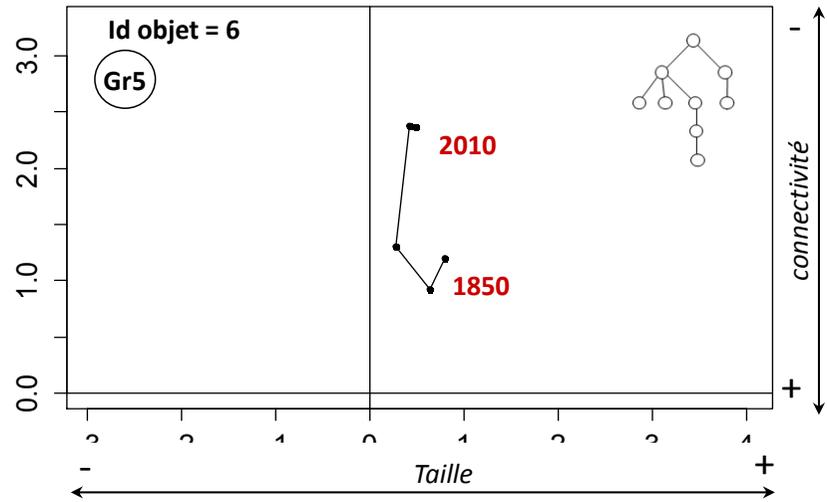
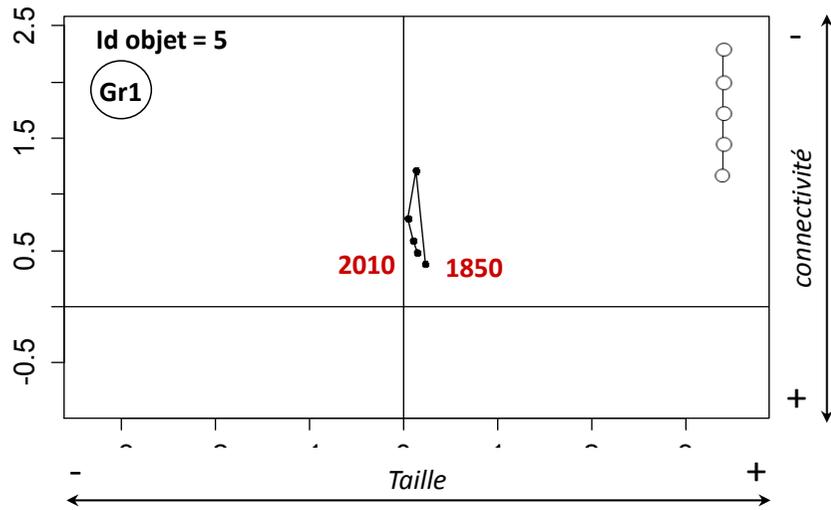
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
43					
44					
45					

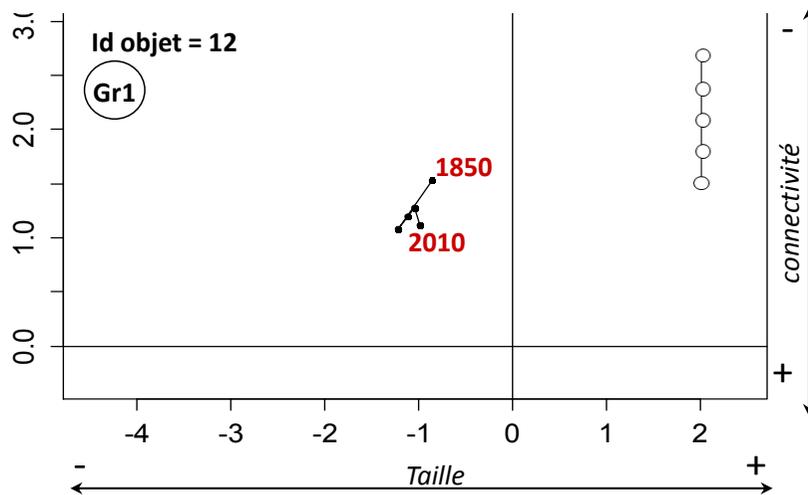
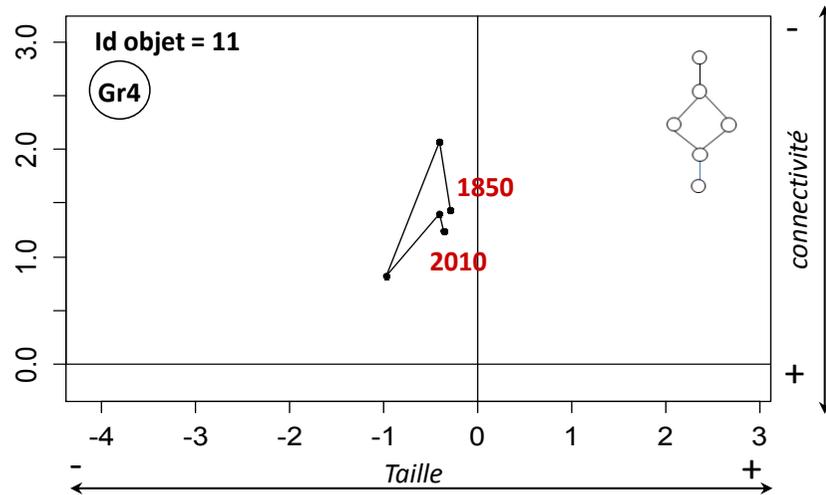
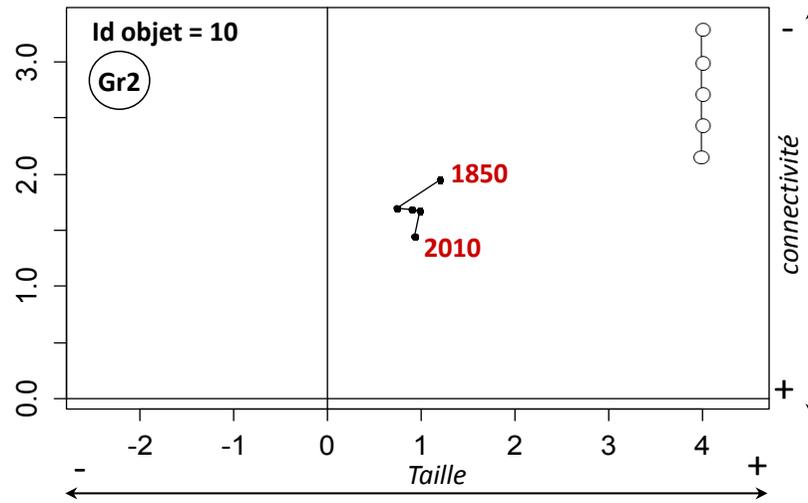
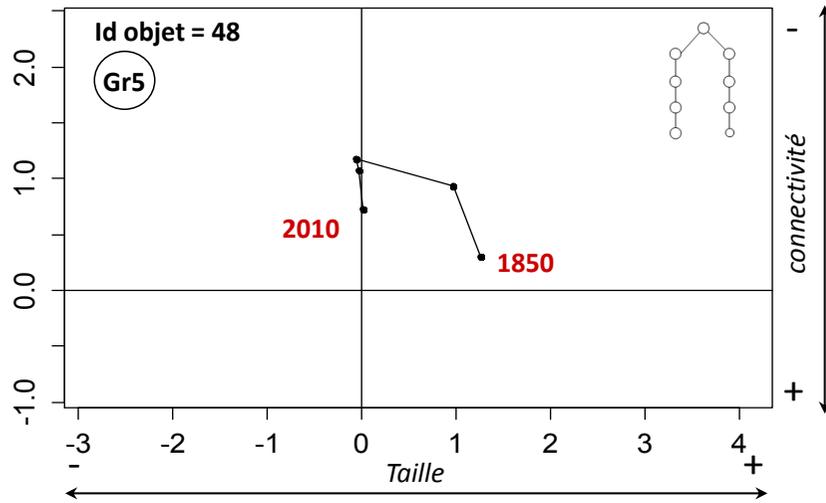
ID	BD ortho IRC 2010	Photographies aériennes 1979	Photographies aériennes 1954	Révision EM Type 1889	Dessins-Minutes EM 1850
46					
47					
48					

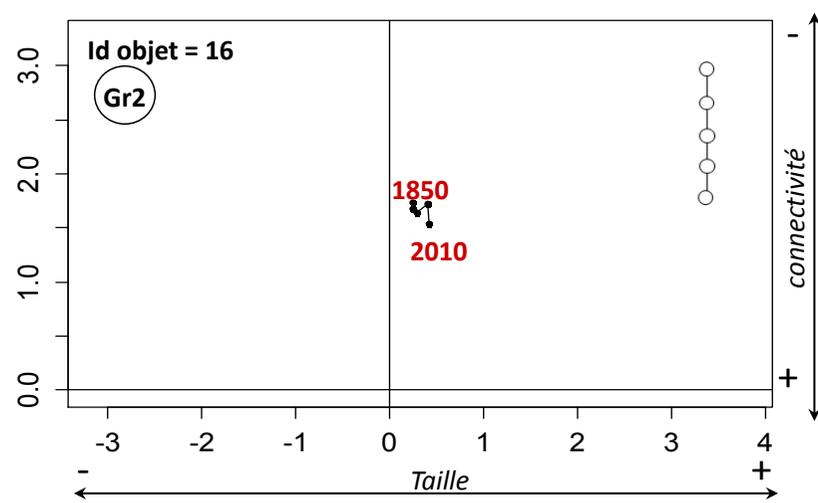
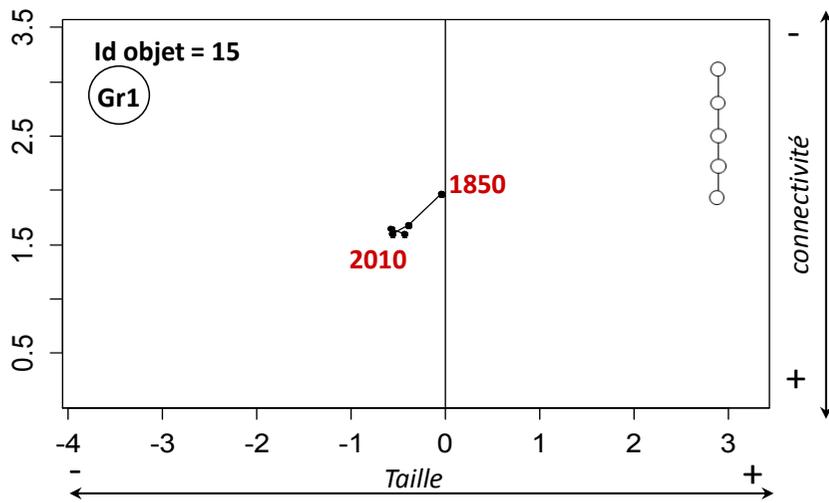
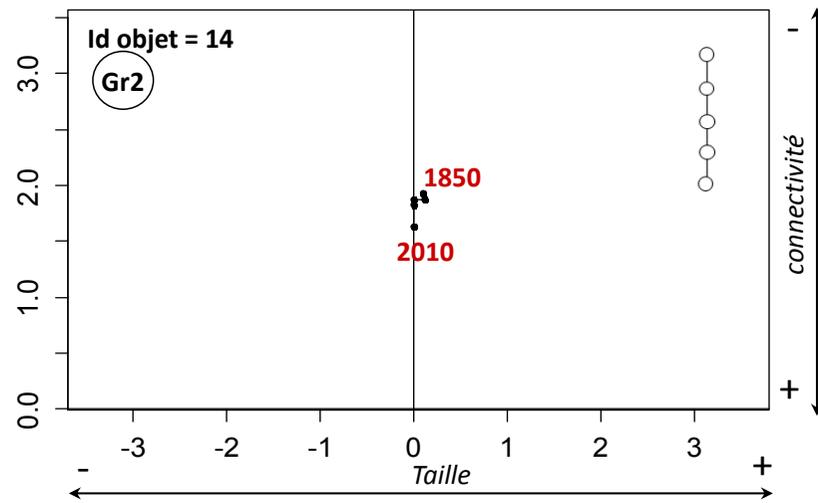
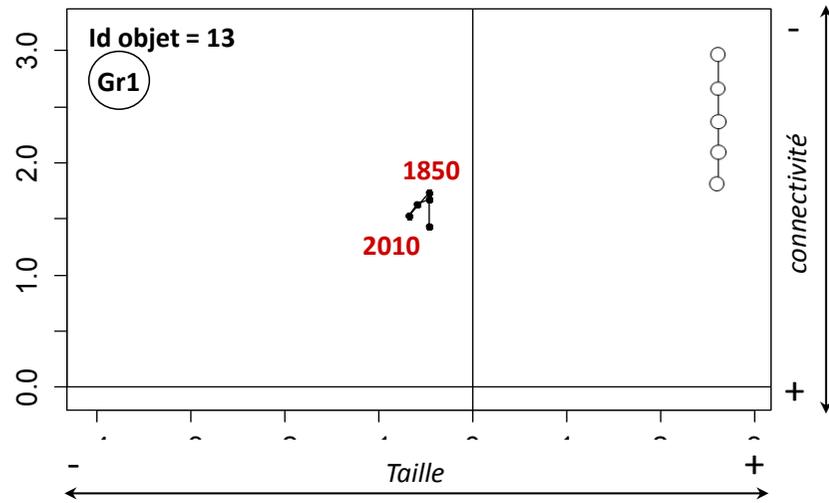
Annexe H : Trajectoires d'évolution spatio-temporelles des forêts étudiées classées selon leur continuité historique. Les numéros de groupe indiqués sur les trajectoires des forêts présentes depuis 1850 (Gr) sont les résultats de la procédure de regroupement effectué en section 9-4-3

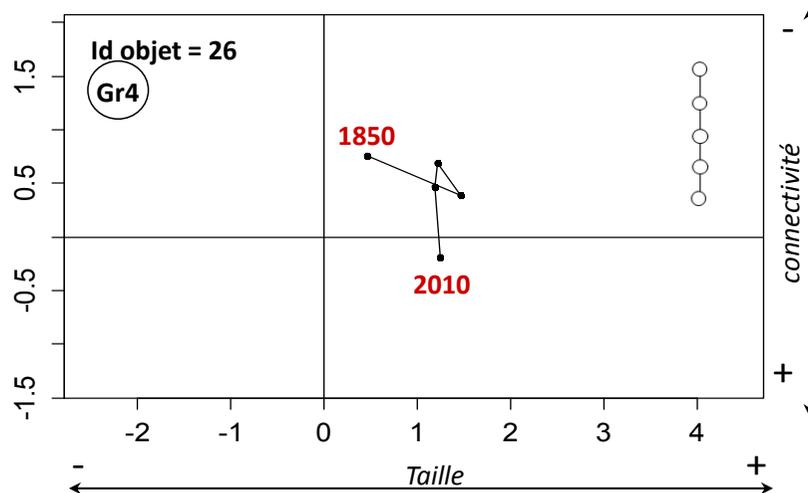
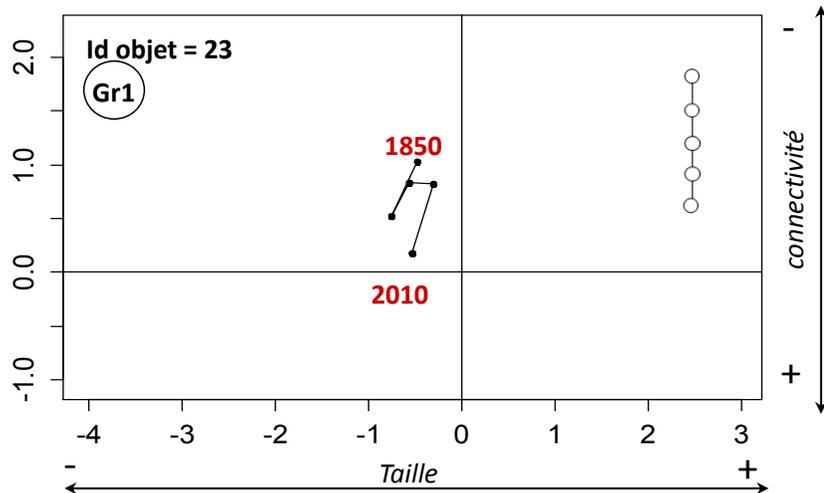
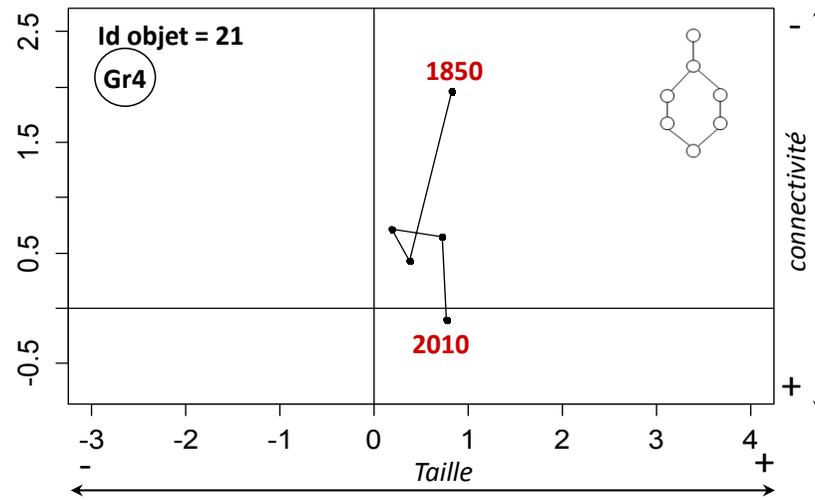
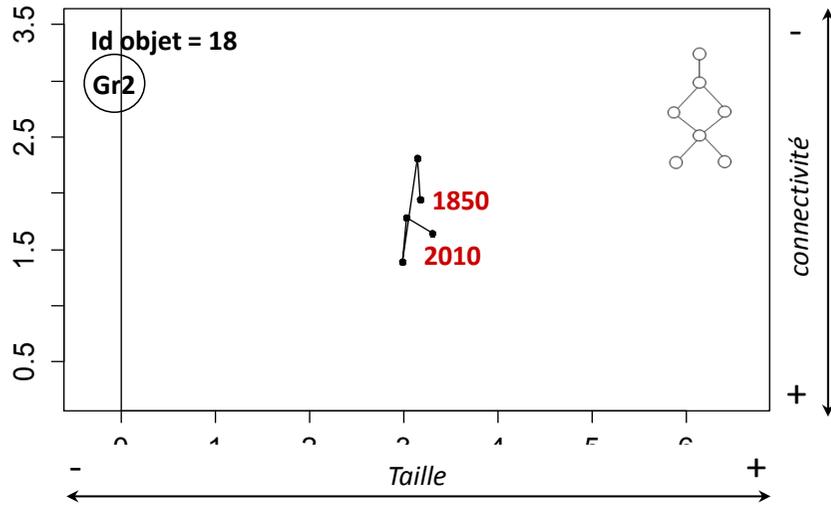
TRAJECTOIRES DES FORETS PRESENTES DEPUIS 1850

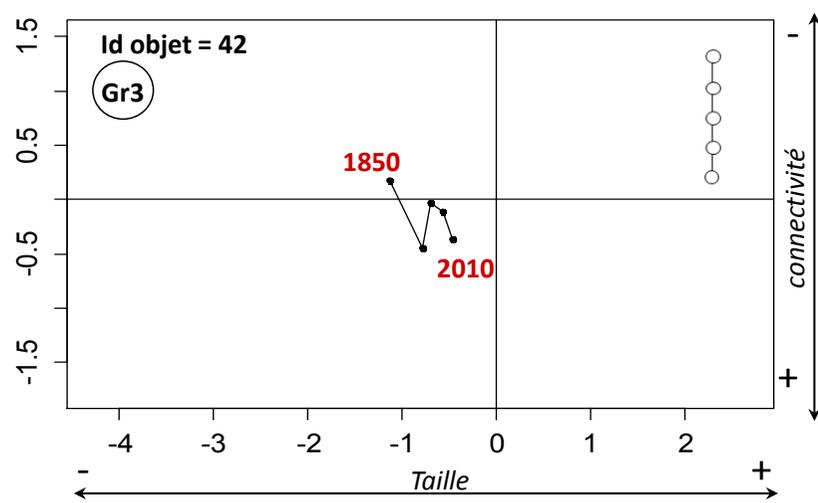
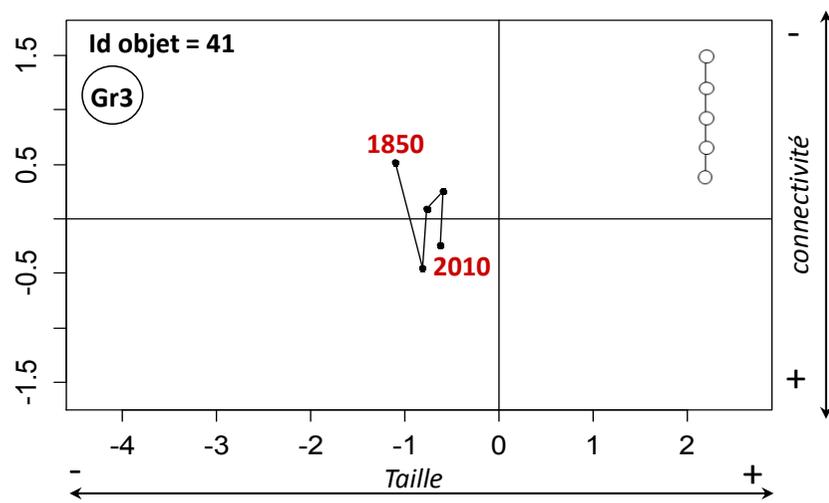
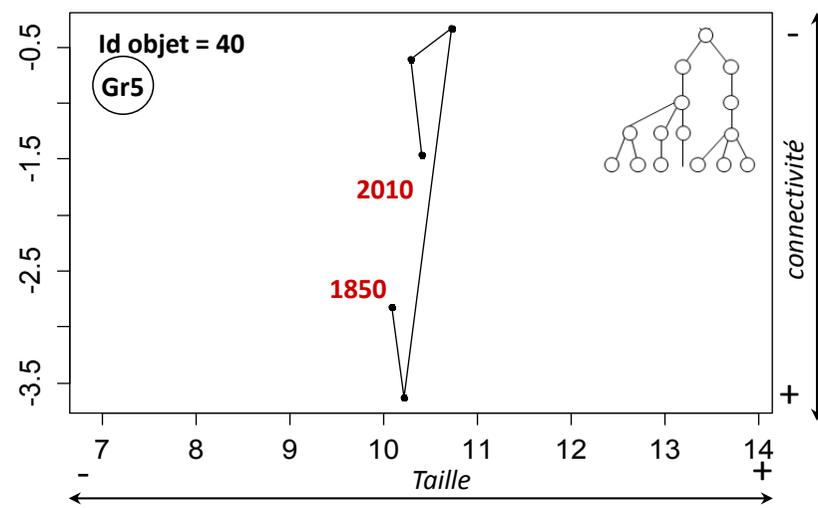
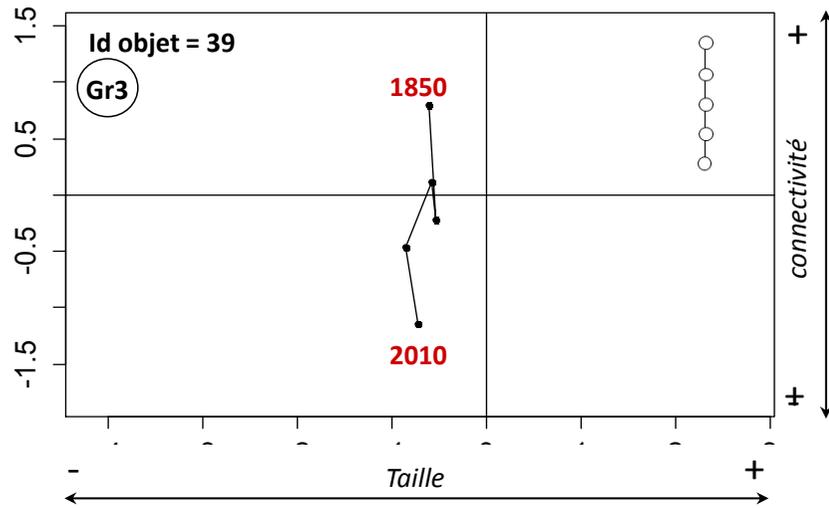


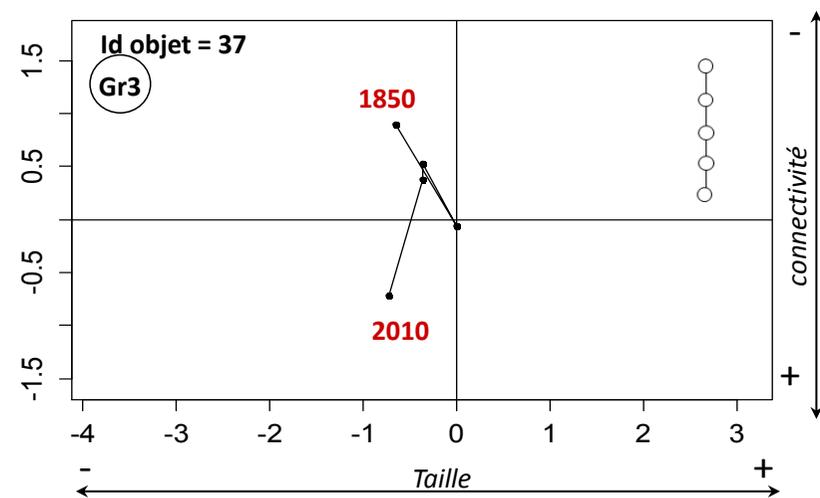
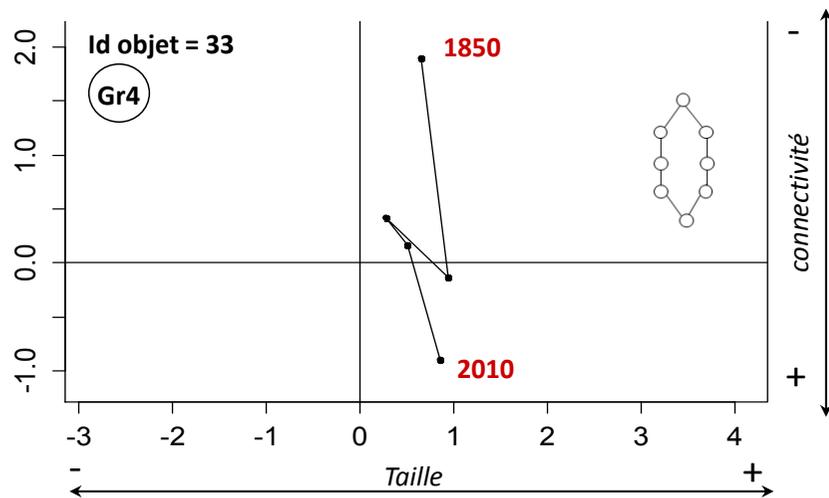
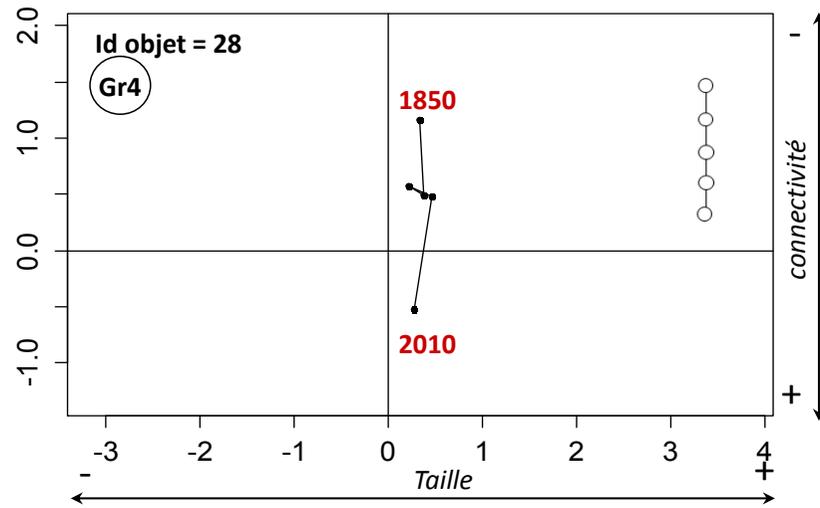
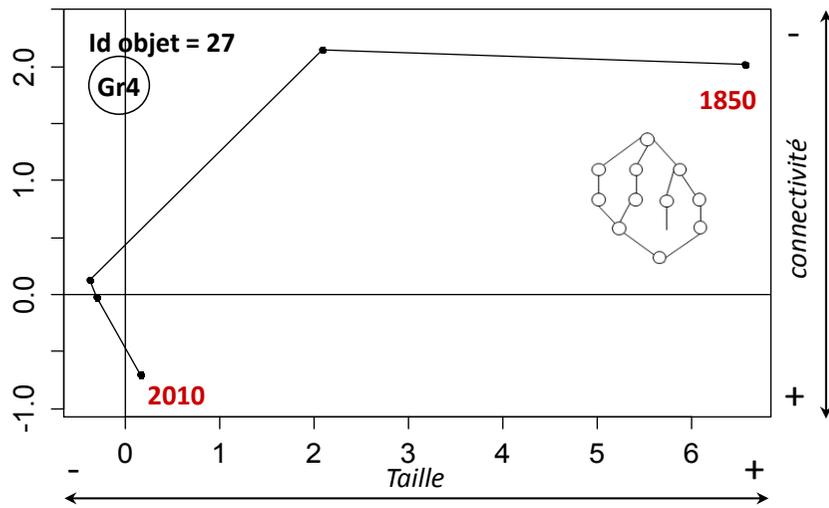




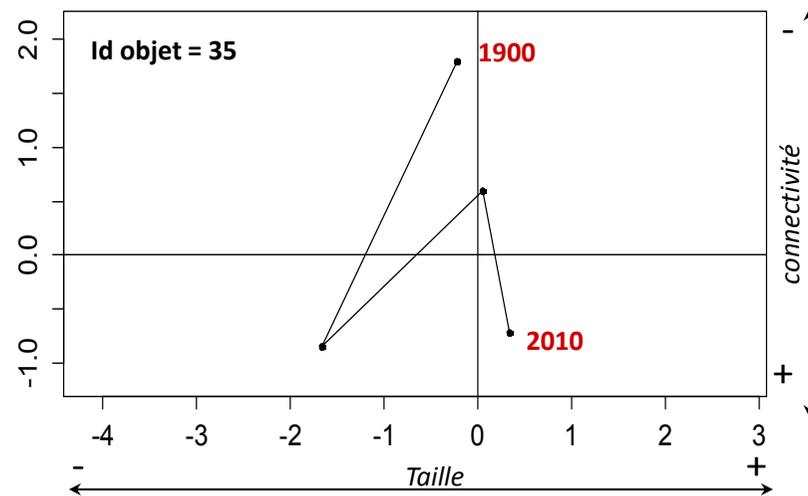
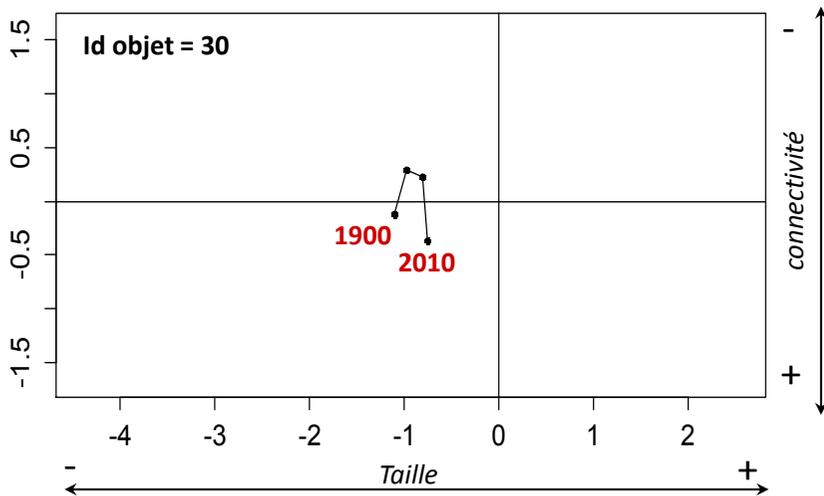
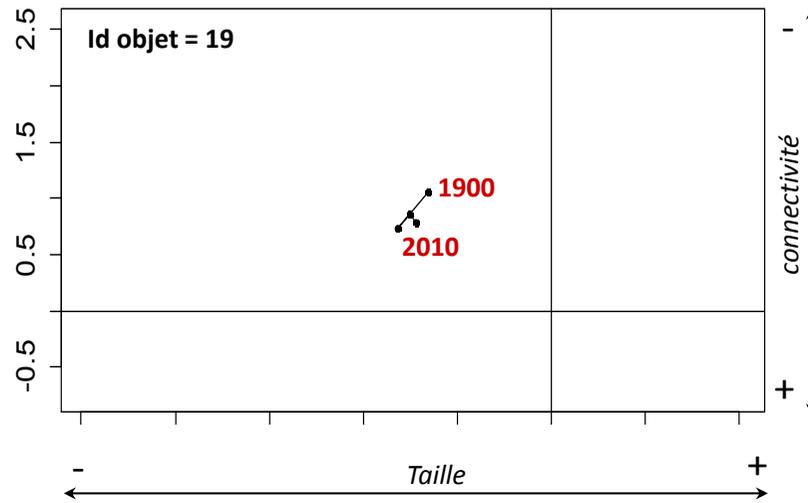
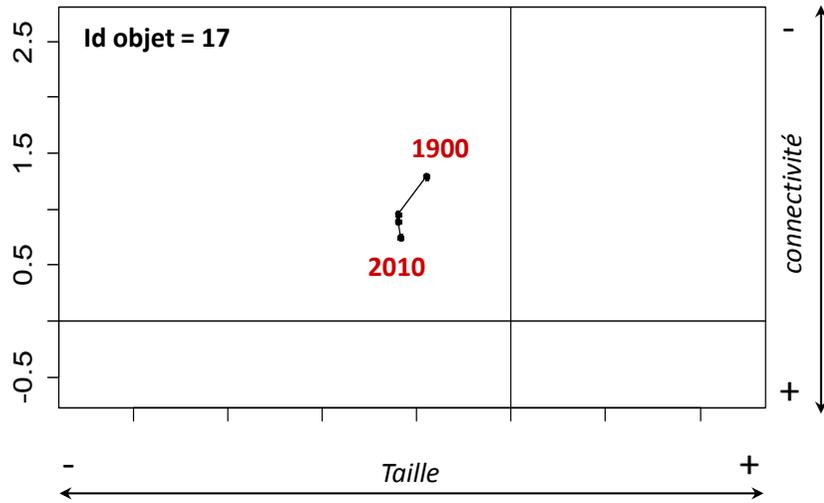


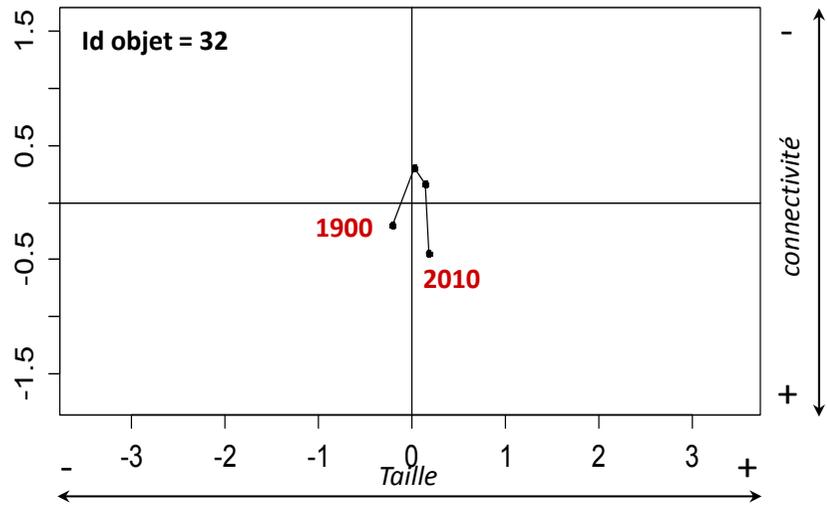




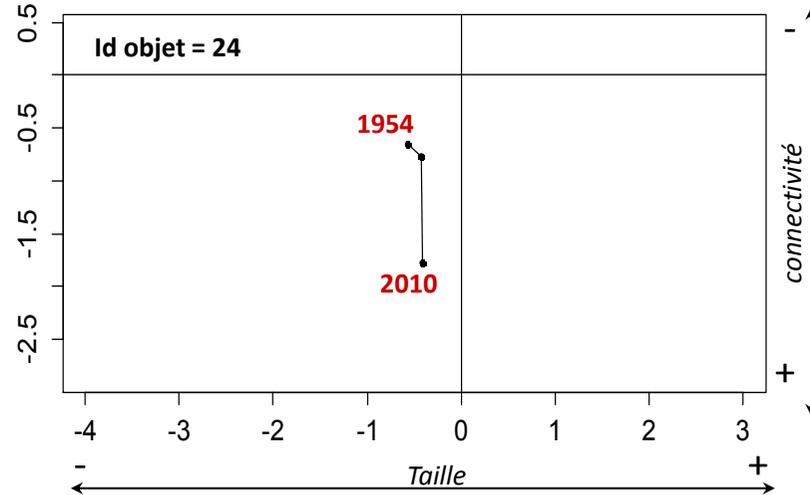
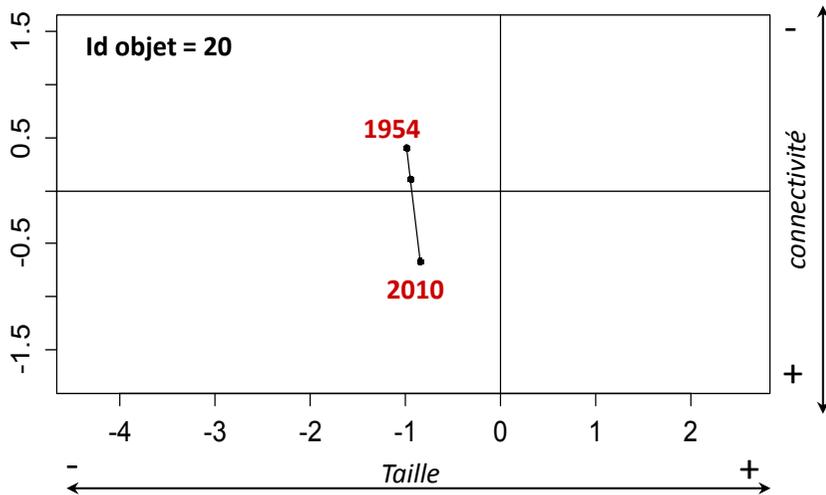
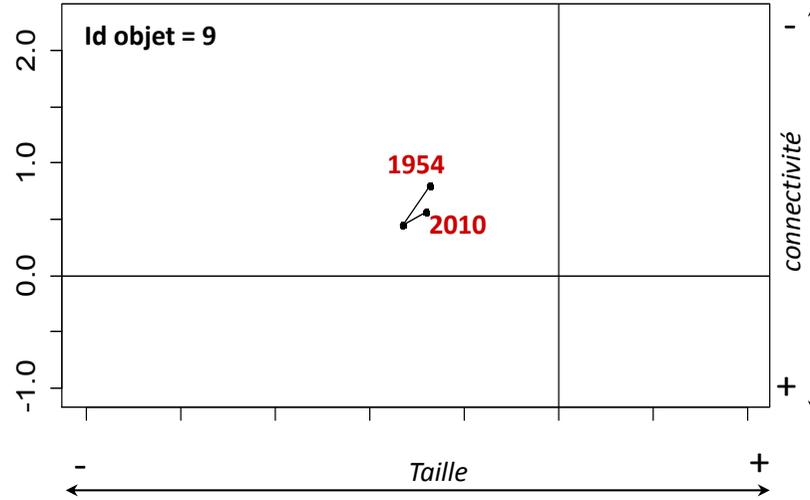
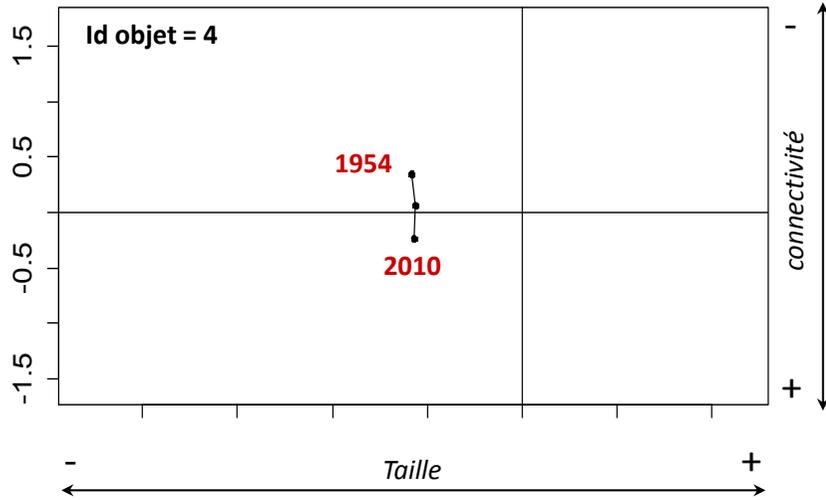


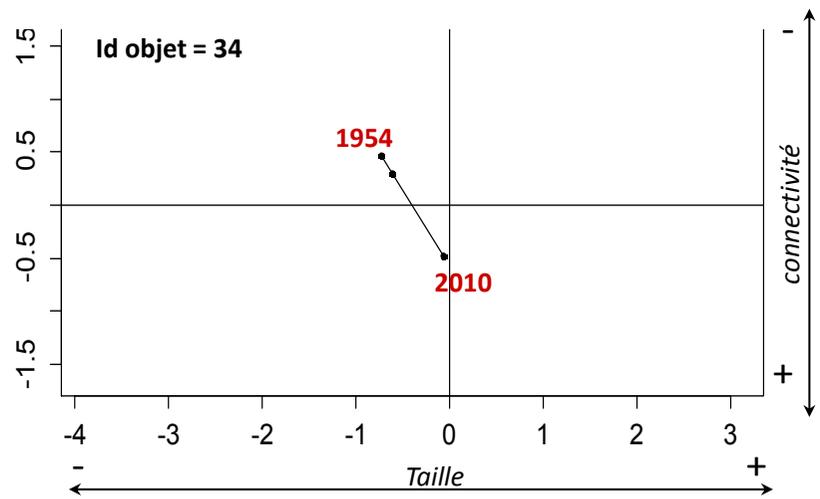
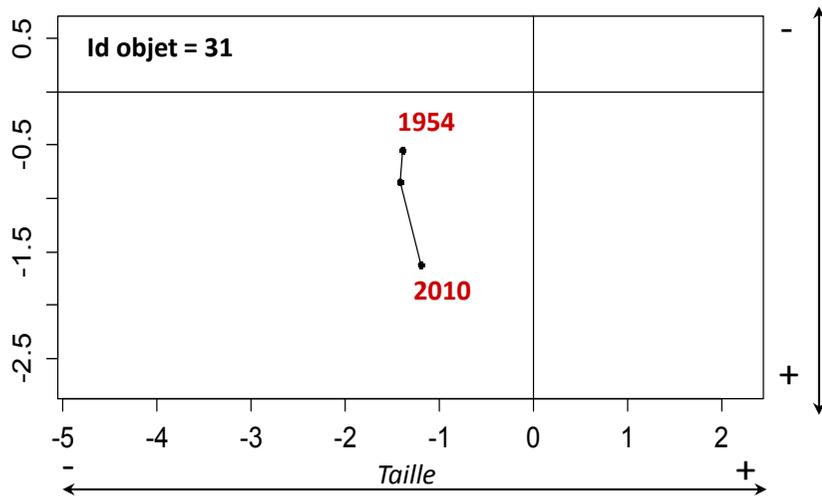
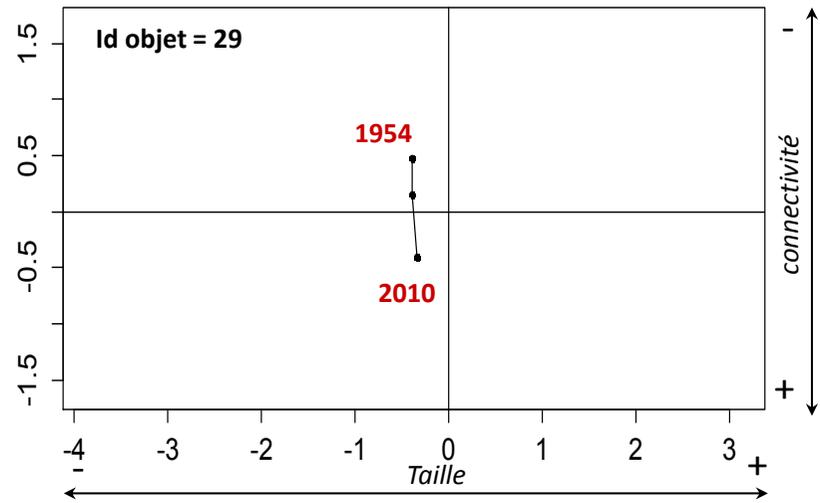
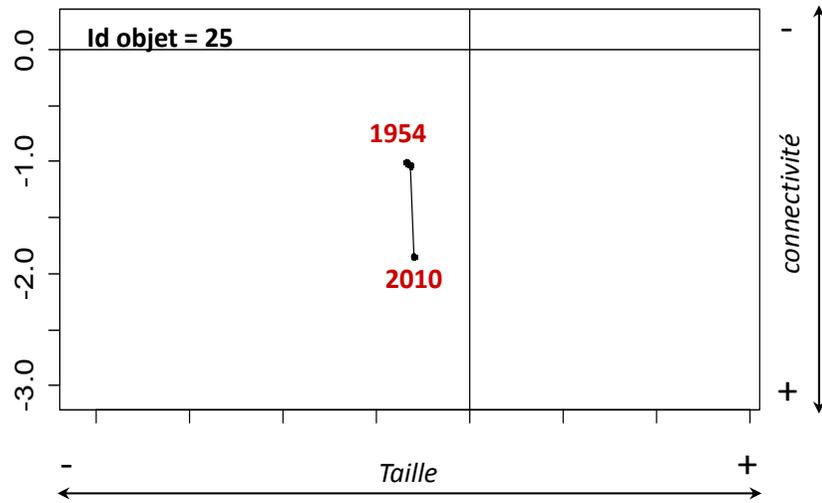
TRAJECTOIRES DES FORETS PRESENTES DEPUIS 1900

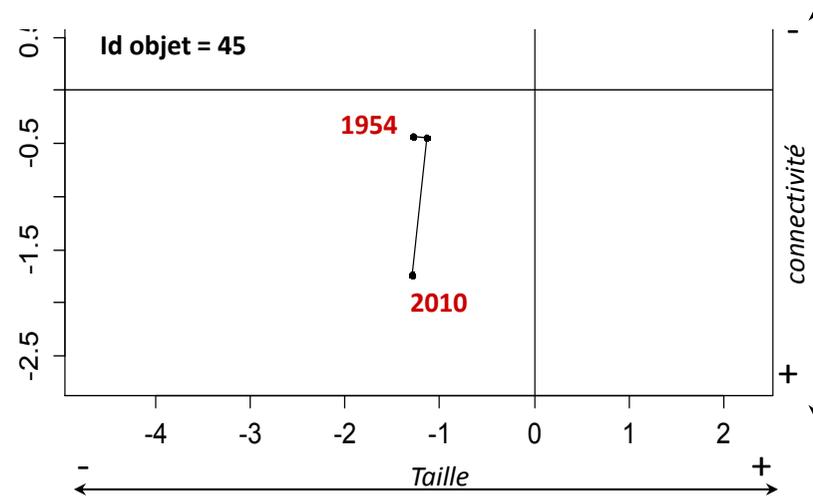
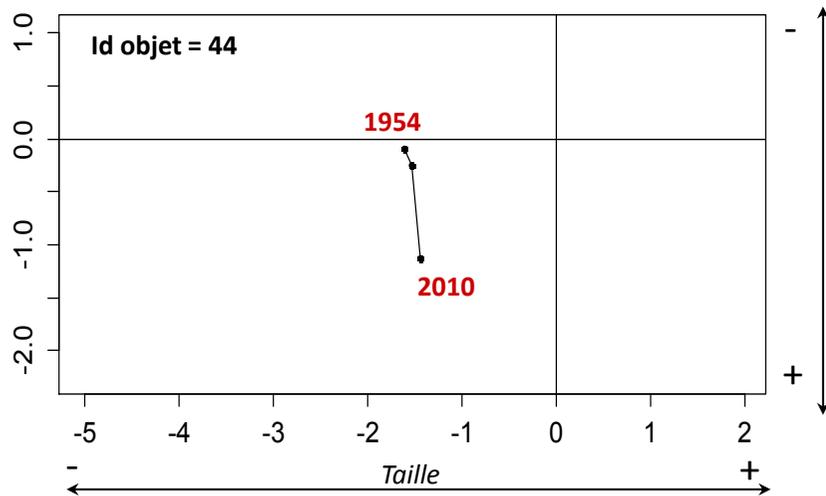
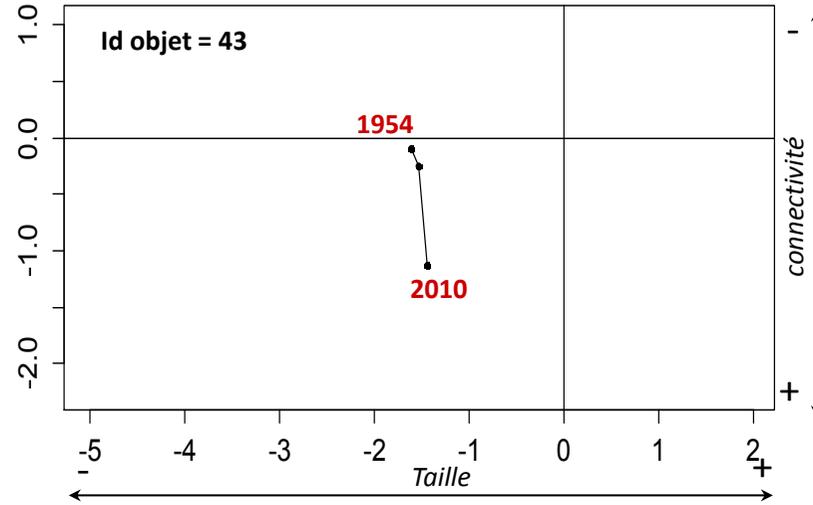
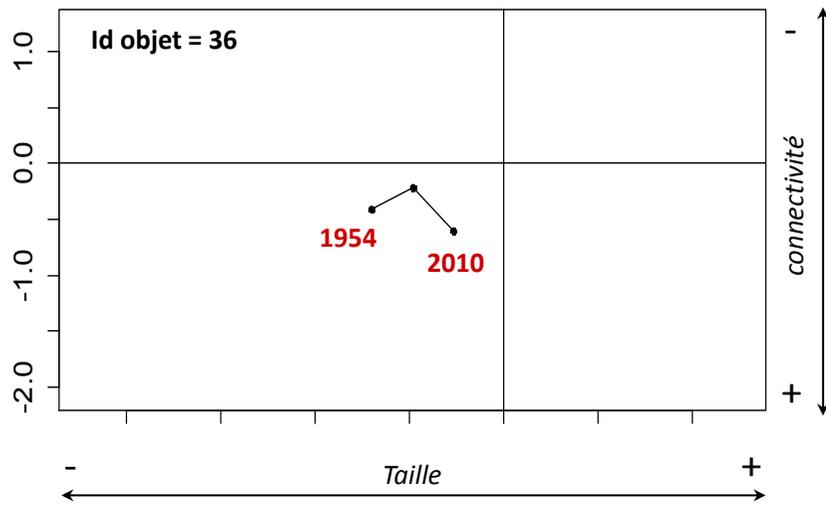


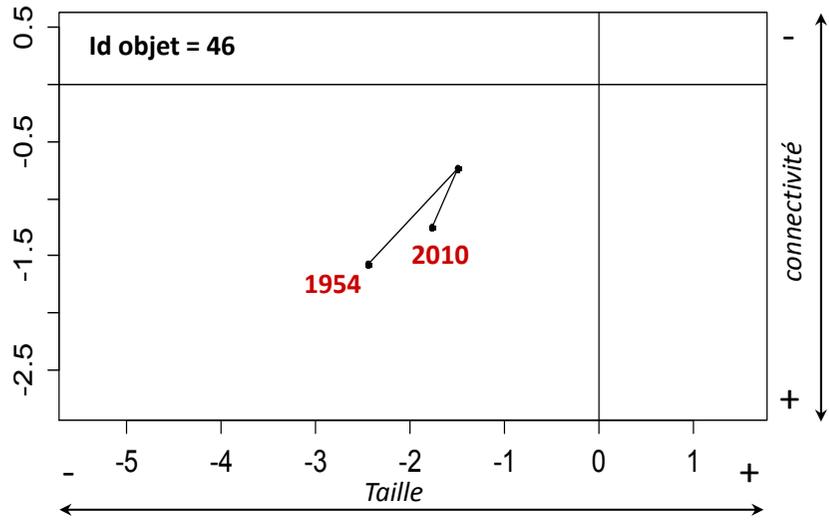


TRAJECTOIRES DES FORETS PRESENTES DEPUIS 1954









TRAJECTOIRES DES FORETS PRESENTES DEPUIS 1979

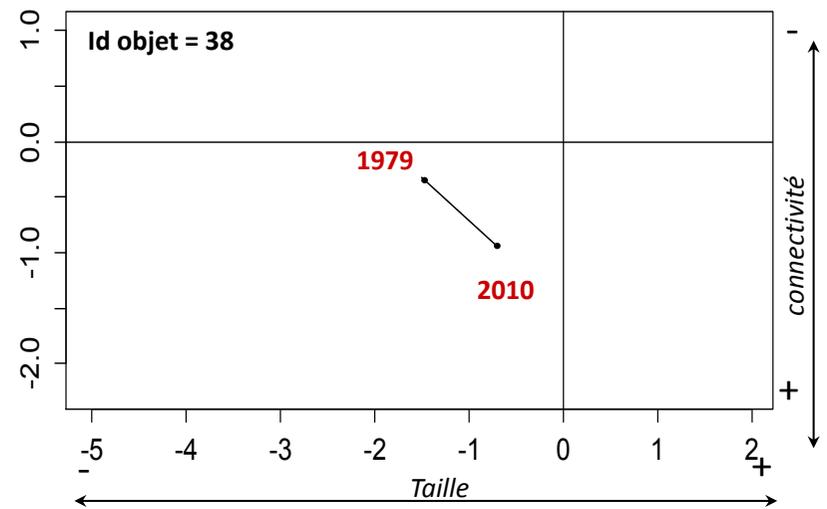
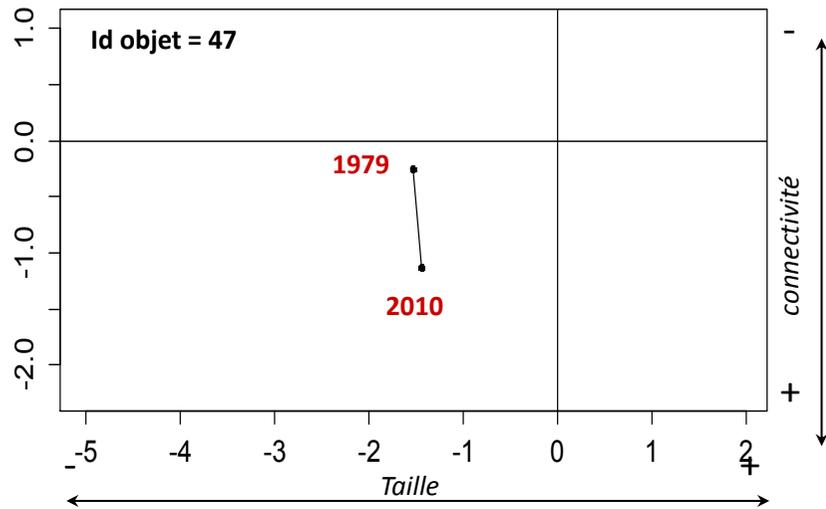


TABLE DES FIGURES

CHAPITRE 1 :

Figure 1-1. <i>L'écologie du paysage structurée autour de trois aspects fondamentaux (Source : McGarigal et Marks 1995)</i>	13
Figure 1-2. <i>L'écologie traditionnelle et l'écologie du paysage (Source: Fahrig 2003)</i>	14
Figure 1-3. <i>Hétérogénéité de composition et de configuration (Source: Fahrig 2005)</i>	15
Figure 1-4. <i>Le modèle Tache-Corridor-Matrice adopté en écologie du paysage</i>	16
Figure 1-5 <i>Les effets de la fragmentation sur la composition et configuration spatiale des patches d'habitats (source : Fahrig 2003)</i>	17
Figure 1-6. <i>Fragmentation d'habitats forestiers au nord d'Alberta, Canada.</i>	18
Figure 1-7. <i>Evolution de l'hétérogénéité perçue en fonction du grain et de l'étendue : (A) l'hétérogénéité diminue avec la taille du grain ; (B) l'hétérogénéité augmente avec l'augmentation de l'étendue (Turner, 1989)</i>	20
Figure 1-8. <i>L'organisation hiérarchique d'une communauté d'espèces (source : http://www.com.univ-mrs.fr/~poggiale/ens/M1/UE209/IntroModelM1.pdf, adapté)</i>	21
Figure 1-9. <i>Représentation du principe général de la théorie de la hiérarchie (source : Redon 2012 d'après Wu, 1999)</i>	22
Figure 1-10. <i>Le modèle de la niche écologique selon Hutchinson (1957). L'enveloppe F désigne la niche fondamentale tandis que l'enveloppe R désigne la niche réalisée</i>	24
Figure 1-11. <i>Relation aire espèce pour une population d'escargots forestiers. La relation est représentée en logarithmes ce qui en fait une droite (Dufrêne, 2003 d'après Blondel 1995)</i>	25
Figure 1-12. <i>Le modèle de l'équilibre extinction/colonisation à la base de la théorie de la biogéographie insulaire (d'après MacArthur et Wilson, (1967). Lorsque le taux de colonisation est égal au taux d'extinction, la richesse en espèces reste constante.</i>	26
Figure 1-13. <i>Illustration des effets de la taille et de l'isolement des habitats dans le modèle de la biogéographie insulaire (Mac Arthur et Wilson 1967). A taux d'immigration équivalent, le taux d'extinction diminue quand la taille de l'habitat augmente : l'équilibre des dynamiques colonisation/extinction montre que la richesse sera plus élevée dans l'habitat de plus grande taille. A surface égale, le taux de colonisation est plus faible quand l'habitat est plus éloigné de la source : la richesse sera donc plus élevée pour l'habitat le moins isolé.</i>	27
Figure 1-14. <i>Les modèles de métapopulation</i>	30

CHAPITRE 2 :

Figure 2-1 (A) <i>Taux d'accroissement des surfaces boisées par département de 1980 à 2010 ; (B) Taux de boisement par département en 2010</i>	33
Figure 2-2. (A) <i>Evolution de la surface des forêts paysannes par département de 1988 à 2000 (en ha) ; (B) Part de la forêt paysanne dans la forêt privée en 2000 (en %)</i>	34
Figure 2-3. <i>Comment mesurer la biodiversité ? (Source : GIP ECOFOR)</i>	39
Figure 2-4. <i>Les facteurs affectant la biodiversité forestière (clichés : Larrieu et Gonin, 2008)</i>	41

Figure 2-5. (A) La forêt de Bialowieza (Pologne) ; (B) La forêt de Zofinsky (République Tchèque) ; (C) La forêt Tronçais (France); (D) La Forêt de Soignes (Belgique) (clichés : http://www.gip-ecofor.org/doc/drupal/liens_article/evenements/2011/ForetsAnciennes_VALLAURI_Dec2011.pdf)	44
Figure 2-6. Les conséquences de l'usage passé sur la biodiversité (source : Flinn et Velend, 2005)	46
Figure 2-7. Les enjeux liés aux vieilles forêts	48

CHAPITRE 3 :

Figure 3-1 Carte de localisation du site d'étude	52
Figure 3-2 Les paysages des coteaux de Gascogne	53
Figure 3-3 A gauche, un syrphe au stade adulte, à droite, un syrphe au stade larvaire (source : http://www.naturablog.com)	54
Figure 3-4 Exemples de piégeages de syrphes : l'utilisation d'une tente malaise et d'une bouteille à échantillonnage	55

CHAPITRE 4 :

Figure 4-1. La date du minimum forestier en France (source: Rudel et al. 2005)	66
Figure 4-2. Disponibilité des données spatiales pour le suivi de la dynamique forestière sur le temps long dans les Coteaux de Gascogne	68
Figure 4-3. Boussoles à éclimètre de Rochette et Bichot (Source: Berthaut 1898)	69
Figure 4-4. Tableau d'assemblage reprenant les dates de levés des feuilles d'Etat-major au 1/80 000 ^e (Source : Dupouey, 2011)	70
Figure 4-5. Exemple d'un canevas de Bonne	71
Figure 4-6. La triangulation du 1 ^{er} ordre de la France sur laquelle s'est appuyée la réalisation de la carte de Cassini et de la carte d'Etat-Major	71
Figure 4-7. Exemple d'un cartouche présent sur un dessin-minute d'Etat-major (Feuille 241 NE, Saint Gaudens (31))	72
Figure 4-8. Etendue d'un dessin-minute au 1/40 000 ^e . La combinaison des quatre planches au 1 :40 000e permet de constituer une feuille de la carte d'Etat-Major au 1/80 000 ^e (Source: Plumejeaud 2012, adapté)	72
Figure 4-9. Aperçu des dessins-minutes traités dans ce travail de thèse. 1) Feuille 229SO, 2) Feuille 229SE, 3) Feuille 241NO, d) 241NE	80
Figure 4-10. Structure d'un quart de feuille Type 1889 (Plumejeaud 2012, adapté)	81
Figure 4-11. Révisions des feuilles de l'Etat-major au 1/80 000 ^e (Source : Nadal 2011)	81
Figure 4-12. Exemples de représentation des surfaces boisées dans la révision Type 1889.	85
Figure 4-13. Aperçu des quatre feuilles sélectionnées. 1) Feuille 229SE ; 2) Feuille 229 SO ; 3) Feuille 241 NE ; 4) Feuille 241NO	86
Figure 4-14. Exemple de cliché aérien issu de la mission aérienne de Juillet 1953	87
Figure 4-15. Exemple de cliché aérien issu des missions aériennes de Juin 1979	88
Figure 4-16. Exemple d'une dalle issue du produit BD ortho [®] IRC	88
Figure 4-17. Récapitulatif des sources de données spatiales retenues dans le cadre de cette étude	89
Figure 4-18. Exemple de la simplification de la forme d'une forêt causée par la présence de poncifs et l'absence de délimitation claire dans la Révision Type 1889. 1) Dessin-Minute d'Etat-Major, 2) Révision Type 1889	90

CHAPITRE 5 :

Figure 5-1. Principe de la correction géométrique des données spatiales	93
Figure 5-2. Conséquences de l'augmentation de l'ordre du polynôme sur l'image à corriger (source : Laffly, 2005). Le modèle ajuste de mieux en mieux les points d'amer mais entraîne des déformations de plus en plus marquées.	95
Figure 5-3. Transformation géométrique basée sur une partition de Delaunay et l'application d'une fonction polynomiale d'ordre 1 dans chacune des partitions (Shimizu et Fuse 2003)	97
Figure 5-4. Exemples de déformations introduites par le modèle de Delaunay : l'image de droite est l'image originale, l'image au centre est l'image transformée par l'application d'une fonction polynomiale d'ordre 1 et l'image de droite est l'image transformée par l'application d'une fonction polynomiale d'ordre 5 (source : Shimizu et Fuse 2003)	99
Figure 5-5. Illustration du principe du noyau gaussien (1) et du noyau polynomial (2).	102
Figure 5-6. Protocole de simulation défini pour étudier l'impact des différents paramètres de la correction géométrique sur la qualité de transformation des cartes anciennes.	105
Figure 5-7. Illustration des trois jeux de points d'amer sélectionnés. (a) 200 bâtiments répartis aléatoirement sur la zone à transformer ; (b) 200 croisements de routes répartis aléatoirement sur la zone à transformer ; (c) 200 croisements de routes répartis régulièrement sur la zone à transformer	107
Figure 5-8. Résultats des simulations de transformations géométriques selon la nature des points d'amer. Les points et barres d'erreur représentent respectivement la moyenne et l'écart-type d'EMQ obtenue (avec en bordeaux les bâtiments et en gris avec les croisements de routes). Les résultats de significativité sont issus du test de Wilcoxon visant à comparer deux moyennes (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales.	110
Figure 5-9. Résultats des simulations de transformations géométriques selon la répartition des points d'amer. Les points et barres d'erreur représentent respectivement la moyenne et l'écart-type d'EMQ obtenue (en bordeaux pour la répartition aléatoire, en gris pour la répartition régulière). Les résultats de significativité sont issus du test de Wilcoxon visant à comparer deux moyennes (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales	111
Figure 5-10. Aperçu des quatre minutes d'Etat-major sélectionnées dans le cadre du protocole OCS ANCIEN. (1) 239SO ; (2) 239SE ; (3) 253NO. (4) 257SE	115
Figure 5-11. Protocole de simulation défini dans le projet OCS ANCIEN pour étudier l'impact des différents paramètres de la correction géométrique sur la qualité de transformation des cartes anciennes.	117
Figure 5-12. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H_0 signifie que les deux moyennes sont égales	119
Figure 5-13. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif,	

* = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales	120
Figure 5-14. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales	121
Figure 5-15. Résultats des simulations de transformations géométriques des quatre feuilles sélectionnées pour le polynôme d'ordre 1 et 2. Les tableaux répertorient la significativité des tests de Wilcoxon comparant les moyennes obtenues entre les différents effectifs testés (ns = non significatif, * = pvalue < 0.05 ; ** = pvalue < 0.01 ; *** = pvalue < 0.001). Une p-value significative permet de rejeter l'hypothèse nulle où H0 signifie que les deux moyennes sont égales	122

CHAPITRE 6

Figure 6-1. Démarche de digitalisation manuelle appliquée pour la production des BD Forêts Multi-dates (MANUEL)	126
Figure 6-2 Extraits des photographies aériennes mosaïquées aux différentes dates.	131
Figure 6-3 Aperçu des couvertures forestières numérisées aux différentes dates	135

CHAPITRE 7 :

Figure 7-1. Illustration de la superposition d'éléments planimétriques dans les cartes anciennes. L'image (a) est un extrait d'une minute d'Etat-major de 1850 tandis que l'image (b) est un extrait de la Révision Type 1889. A noter que l'absence de couleur dans l'image (b) complexifie l'extraction automatique des éléments paysagers.	141
Figure 7-2. Principe du filtrage morphologique (érosion) (a) et du filtrage de convolution médian (b) (Source : Sheeren et al. 2007, adapté).	142
Figure 7-3. Démarche de vectorisation automatique pour extraire les forêts des dessins minutes d'Etat-major	146
Figure 7-4. Trois étapes successives de filtrage dans la phase de prétraitements	146
Figure 7-5. Illustration des corrections apportées par les différentes étapes de filtrage	147
Figure 7-6. L'espace colorimétrique CIE Lab: L exprime la clarté (Luminance) qui varie de 0 (noir) à 100 (blanc) ; a et b expriment une gamme de couleurs définies par un mélange de vert (-a) à magenta (+a) et un mélange de bleu (-b) à jaune (+b) ; toutes les couleurs de même luminosité s'inscrivent dans un plan circulaire traversé par les axes a* et b* (source : http://www.bonne-mesure.com/cie_lab.php)	149
Figure 7-7. Illustration des corrections apportées par les processus de prétraitements : les trous enclavés au sein des forêts ont été bouchés et les pixels isolés supprimés.	152
Figure 7-8. Aperçu de l'interface « Historical Maps » disponible sous ENVI/IDL. Cette extension regroupe la totalité des outils nécessaires à la mise en place de la démarche d'extraction proposée.	152
Figure 7-9. Trois extraits sélectionnés pour l'application de la méthode proposée. L'extrait 1 est localisé dans la région de Mende et se caractérise par un relief marqué ainsi que des espaces boisées représentées en vert foncé. ; l'extrait 2 est situé dans la région de Tarbes avec des forêts représentées en jaune et une présence importante de hachures (relief) ; l'extrait 3, dans la région de Saint Gaudens,	

<i>est caractérisé par un relief moins marqué, des forêts colorées en vert pâle et une surcharge planimétrique causé par les toponymes.</i>	153
Figure 7-10. Description des indicateurs de qualité dérivés d'une matrice de confusion	154
Figure 7-11. Comparaison des extractions automatiques et de la référence avec une spatialisation des erreurs d'excédents et de déficits	157
Figure 7-12. Résultats d'extraction automatique des forêts sur le site d'étude :a) et b) présentent respectivement les couches de forêts extraites manuellement et automatiquement ;(c) répertorie trois exemples d'erreurs type observées sur la couche extraite automatiquement ; d) compare les distributions des valeurs de métriques calculées à partir des objets extraits automatiquement et manuellement. Le test de comparaison de moyenne utilisé est le test de Wilcoxon pour lequel une p-value inférieure à 0.05 permet de rejeter HO qui signifie que les deux moyennes sont égales.	159
Figure 7-13. Aperçu des dessins-minutes d'Etat-Major sélectionnés dans le cadre du projet OCS Ancien pour les tests d'extraction. 1) Feuille 191NE ; 2) Feuille 198NE ; 3) Feuille 257NO	162
Figure 7-14. Illustration des conséquences provoquées par le processus de filtrage. A gauche, l'image originale, à droite, l'image après filtrage. Les forêts (en vert) et les terres humides (en bleu) ont été lissés très fortement, le contraste entre le fond de carte (en beige) et ces deux classes est par conséquent nettement moins marqué.	163
Figure 7-15. Résultats d'extraction à partir des trois dessins-minutes d'Etat-major sélectionnés dans le projet OCS Ancien. Les n.s (non-significatif) indiquent que les thèmes en question n'ont pu être extraits suffisamment précisément pour être présentés.	165
Figure 7-16 Démarche générale de la méthode proposée pour extraire automatiquement les forêts de la BD ortho IRC	169
Figure 7-17 Principe général de la segmentation (source : Benoit-Cattin et al. 2000)	171
Figure 7-18 Principe de la segmentation multi-résolution. Différents niveaux de segmentation peuvent être produits et reliés par une structure hiérarchique (Willhauck et al. 2000)	172
Figure 7-19 Interface de l'outil ESP Tool	173
Figure 7-20 Exemples de sorties de l'outil ESP Tool pour deux dalles de la BD Ortho IRC. La courbe de la variance locale n'affiche aucun points d'inflexion tandis que le premier pic apparent sur la courbe du taux de changement est situé à une valeur d'échelle de 180.	173
Figure 7-21 Exemple d'une dalle de la BD Ortho IRC segmentée avec les paramètres retenus	174
Figure 7-22. Localisation des trois zones d'échantillonnage. Les polygones gris représentent la couche de forêts extraite manuellement et vise à faciliter la localisation des zones d'apprentissage. Les polygones verts sont les segments d'exemples labélisés « forêt » et les polygones jaunes sont les segments d'exemples labélisés « non-forêt ».	176
Figure 7-23 Arbre de décision obtenu après application de l'algorithme C5.0	177
Figure 7-24. Application des règles de décision dans la plateforme eCognition à partir d'un arbre de décision construit avec le logiciel WEKA	178
Figure 7-25 Exemple de résultats observés après application des post-traitements : les haies ont été supprimées et les objets « forêt » ont été isolés les uns des autres	179
Figure 7-26. Aperçu des résultats d'extraction à partir de la BD Ortho IRC sur le site d'étude. (1) la couche de forêts extraites automatiquement ; (2) la couche de forêts extraite manuellement (référence); (3) spatialisation des erreurs de déficit et d'excédent.	181

CHAPITRE 8 :

Figure 8-1 Définition de la distance surfacique (source : Sheeren 2005)	193
---	-----

Figure 8-2. Démarche méthodologique pour appairer les espaces forestiers	194
Figure 8-3. Illustration de l'écart géométrique susceptible d'apparaître entre deux homologues	195
Figure 8-4. Illustration de l'écart de position entre deux homologues. Le contexte géographique des objets offre une alternative efficace pour les relier.	196
Figure 8-5. Illustration des effets provoqués par le sens d'appariement sur la somme des objets appariés à chaque date	196
Figure 8-6. Procédure d'établissement de la table d'appariement après l'identification visuelle des homologues	198
Figure 8-7. Présentation des différents liens de filiation pouvant apparaître à l'issue de la phase d'appariement	198
Figure 8-8. Définition de la métrique « effective mesh-size » (source : Jaeger 2000)	203
Figure 8-9. Procédure de traitement pour déterminer la connectivité potentielle d'un patch d'habitat vue comme une densité d'habitat dans son voisinage	204
Figure 8-10. Exemple d'un modèle STC pour la rotation des cultures (1 = Blé, 2 = Maïs)	206
Figure 8-11 Les primitives de base du langage CDL: (1) entité qui n'existe pas; (2) entité qui existe; (3) transition entre deux instants (4) création d'une entité (source : Del Mondo 2011 d'après Hornsby et Egenhofer 2000)	206
Figure 8-12 Représentation d'une géolifeline	207
Figure 8-13. Les types de processus spatio-temporels (source: Claramunt et Thériault 1995)	208
Figure 8-14. Typologie des processus spatio-temporels forestiers ayant potentiellement un impact sur la biodiversité forestière (source : Joint Research Center)	209
Figure 8-15. Liste des processus spatio-temporels susceptibles d'apparaître dans l'analyse d'évolution des surfaces boisées. Les Delta (Δ) indiquent une modification des caractéristiques d'une forêt. Pour rappel, la généalogie fait référence à l'existence de l'objet.	209
Figure 8-16. Description du modèle « snapshot » pour la mise en évidence des évolutions	210
Figure 8-17. Description de l'approche « individu » pour quantifier les évolutions de la structure spatiale des bois à travers les changements d'états. Ce modèle s'applique à chaque couple de dates.	211
Figure 8-18. Description de l'approche « population » pour mettre en évidence des tendances d'évolution à travers une typologie de trajectoires	212

CHAPITRE 9 :

Figure 9-1. Représentation des processus spatio-temporels (entre deux dates) pouvant affecter la généalogie des forêts par des primitives géométriques.	216
Figure 9-2. Représentation de la cardinalité des liens d'appariement pour les 48 forêts suivies	218
Figure 9-3. Distribution des valeurs de métriques pour toutes les forêts identifiées sur chaque source	221
Figure 9-4. Résultats de l'Analyse en Composantes Principale basée sur les forêts de 2010	225
Figure 9-5. Exemples de quatre trajectoires forestières correspondant à quatre durées d'existence différente	227
Figure 9-6. Aperçu des trajectoires caractéristiques des cinq groupes identifiés.	229

CHAPITRE 10 :

- Figure 10-1. *Le principe de la dette d'extinction (source : Kuussaari et al.,2009). (a) Une dette d'extinction peut être mise en évidence en observant une différence entre le nombre d'espèces présentes au début de la perturbation et le nombre d'espèces au nouveau point d'équilibre, (b) Une dette d'extinction peut être mise en évidence en observant une différence entre le nombre d'espèce présentes dans un habitat non-perturbé et le nombre d'espèces présentes dans un habitat perturbé de même taille.* 242
- Figure 10-2. *Les approches pour détecter une dette d'extinction. Ces dernières peuvent être remobilisées également pour détecter un crédit de colonisation.(source : Kuussaari et al. 2009)* 245
- Figure 10-3. *Illustration des deux approches de modélisation permettant d'étudier la distribution des espèces (source : https://www.eccorev.fr/IMG/pdf/leriche_mod2010.pdf, adapté)* 248
- Figure 10-4. *Schéma méthodologique d'une démarche modélisation basée sur l'emploi d'un modèle linéaire généralisé (GLM)* 249
- Figure 10-5 *Illustration du principe de parcimonie (Burnham et Anderson 2002)* 253

CHAPITRE 11:

- Figure 11-1. *Typologie de l'incertitude (source : Fisher et al. 2005)* 277
- Figure 11-2. *Le principe de la qualité interne* 278
- Figure 11-3. *Cadre applicatif de l'évaluation de l'incertitude spatiale (source : Leyk et al. 2005). Le cadre rouge indique le type d'incertitude évalué dans ce travail de thèse.* 280
- Figure 11-4. *Schéma méthodologique de la stratégie d'évaluation de l'incertitude de classification* 282

LISTE DES TABLEAUX

CHAPITRE 3 :

Tableau 3-1 <i>Pour chaque classe de surface de forêts : nombre de pièges, nombre de forêts dans la classe et moyenne de richesse en espèces par forêts</i>	55
---	----

CHAPITRE 4 :

Tableau 4-1. <i>Description des principaux thèmes observés sur les dessins-minutes d'Etat-Major au 1/40 000^e</i>	76
Tableau 4-2. <i>Exemples de représentation des surfaces boisées dans les dessins-minutes d'Etat-Major au 1/40.000^e</i>	79
Tableau 4-3. <i>Caractéristiques des dessins-minutes sélectionnés dans la cadre de cette étude</i>	80
Tableau 4-4. <i>Description des principaux thèmes observés dans la Révision Type 1889. Les encadrés bordeaux indiquent les apports de la Révision Type 1889</i>	84
Tableau 4-5. <i>Caractéristiques des feuilles de la Révision Type 1889 sélectionnés dans la cadre de cette étude</i>	86
Tableau 4-6. <i>Caractéristiques des missions retenues pour les années 1953-1954</i>	87
Tableau 4-7. <i>Caractéristiques des missions aériennes retenues pour l'année 1979</i>	87

CHAPITRE 5:

Tableau 5-1 <i>Détails concernant la situation et le contexte géographique des quatre minutes d'Etat-major sélectionnées dans le cadre du protocole OCS ANCIEN</i>	116
Tableau 5-2. <i>Détails des tirages aléatoires effectués sur chacun des jeux de points pour constituer la base de points d'entraînements et de validation à chaque échantillon testé</i>	117

CHAPITRE 6 :

Tableau 6-1. <i>Base de règles définie pour la vectorisation des forêts dans les minutes de l'Etat-Major au 1/40 000^e (Source : d'après Favre et al.,2012)</i>	128
Tableau 6-2. <i>Base de règles définie pour la vectorisation des forêts dans les feuilles de la Révision Type 1889</i>	130
Tableau 6-3 <i>Base de règles définie pour la vectorisation des forêts dans les collections de photographies aériennes</i>	133
Tableau 6-4 <i>Statistiques élémentaires des forêts extraites manuellement à partir des différentes sources spatiales</i>	136

CHAPITRE 7 :

Tableau 7-1. <i>Résultats des extractions opérées sur les trois extraits sélectionnés</i>	156
Tableau 7-2. <i>Résultats de validation de la couche extraite à partir de la méthode de vectorisation automatique</i>	158

Tableau 7-3. <i>Caractéristiques des feuilles sélectionnées pour les tests d'extraction du projet OCS ANCIEN</i>	161
Tableau 7-4. <i>Résultats des tests d'extraction effectués dans le cadre du Projet OCS ANCIEN. Les acronymes DYN et IGN indiquent respectivement la méthode présentée dans ce travail de thèse (DYNAFOR) et celle développée par l'IGN. Les n.s signifient que les résultats d'extractions ont été jugé trop peu fiables pour les présenter.</i>	164
Tableau 7-5. <i>Description des attributs sélectionnés pour caractériser les segments d'image (eCognition, Reference Guide)</i>	175
Tableau 7-6. <i>Résultats de validation de la BD Forêts 2010 (AUTO)</i>	179
 CHAPITRE 8 :	
Tableau 8-1 <i>Différents rapports au temps des entités : points de vue ontologique et thématique (source : Mathian et Sanders, 2014)</i>	205
 CHAPITRE 9 :	
Tableau 9-1. <i>Résultats des calculs des métriques pour toutes les forêts identifiées sur chaque source. Il s'agit ici des résultats moyennés à l'échelle du groupe de forêt mais chacune de ces métriques a été calculée pour chaque individu à chaque date.</i>	220
Tableau 9-2. <i>Résultats des calculs des taux de changement entre deux dates. N(+) indiquent le nombre d'individus ayant connu une évolution positive tandis que les N(-) indiquent le nombre d'individus ayant connu une évolution négative.</i>	223
Tableau 9-3. <i>Caractéristiques des groupes de forêts identifiés à la suite de la procédure de regroupement</i>	228
 CHAPITRE 10 :	
Tableau 10-1. <i>Récapitulatif des variables explicatives générées à partir des différentes bases de données produites</i>	247
 CHAPITRE 11 :	
Tableau 11-1. <i>Description des modèles écologiques faisant l'objet d'une évaluation de l'incertitude spatiale. Les « : » désignent les termes d'interaction.</i>	281
Tableau 11-2. <i>Résultats de comparaison des modèles espèces-habitat</i>	286

TABLE DES MATIERES

REMERCIEMENTS	
LISTE DES SIGLES ET ACRONYMES	
SOMMAIRE	
INTRODUCTION GENERALE	1
PARTIE A	7
INTRODUCTION DE LA PARTIE A	9
CHAPITRE 1. CADRE DE L'ETUDE : L'ECOLOGIE DU PAYSAGE	11
1.1. UNE DISCIPLINE A LA CROISEE ENTRE L'ECOLOGIE ET LA GEOGRAPHIE	12
1.2. L'HETEROGENEITE COMME CAUSE DE LA BIODIVERSITE	14
1.3. LA FRAGMENTATION DES HABITATS : UN PROCESSUS AFFECTANT POTENTIELLEMENT LA BIODIVERSITE	17
1.4. LES THEORIES ECOLOGIQUES FONDATRICES	19
1.4.1 UN CADRE THEORIQUE POUR ETUDIER LES EFFETS DE L'HETEROGENEITE SPATIALE	20
1.4.1.1. LA THEORIE DE LA HIERARCHIE	20
1.4.1.2. LA THEORIE DE LA NICHE ECOLOGIQUE	23
1.4.2 UN CADRE THEORIQUE POUR ANALYSER LES EFFETS DE LA FRAGMENTATION DES HABITATS	24
1.4.2.1. LA THEORIE DE LA BIOGEOGRAPHIE INSULAIRE	24
1.4.2.2. LA THEORIE DES METAPOPOPULATIONS	28
CHAPITRE 2. OBJET D'ETUDE : LA FORET PAYSANNE FRAGMENTEE ET LA BIODIVERSITE ASSOCIEE	31
2.1. LA FORET PAYSANNE : RESERVOIR DE BIODIVERSITE DANS L'ESPACE RURAL	32
2.1.1 LA FORET PAYSANNE EN DECLIN	32
2.1.2 LA RECONNAISSANCE DE LA RICHESSE DES FORETS PAYSANNES	35
2.1.3 LA FORET PAYSANNE MORCELEE : UN ENVIRONNEMENT PARTICULIER	35
2.2. DE QUELLE BIODIVERSITE PARLE-T-ON ?	36
2.2.1 UNE BIODIVERSITE FORESTIERE MENACEE	36
2.2.2 COMMENT EST-ELLE MESUREE ?	37
2.2.3 LA BIODIVERSITE FORESTIERE DANS CETTE THESE	40
2.3. LES FACTEURS GENERAUX AFFECTANT LA BIODIVERSITE FORESTIERE	40
2.4. L'ANCIENNETE DES FORETS : UN FACTEUR CLE POUR COMPRENDRE LEUR BIODIVERSITE	42
2.4.1 MATURITE VERSUS ANCIENNETE	42
2.4.2 LA SPECIFICITE DES FORETS ANCIENNES	45
2.4.3 LA SPECIFICITE DES VIEILLES FORETS	47
CHAPITRE 3. SITE D'ETUDE ET DONNEES BIOLOGIQUES SELECTIONNEES	51
3.1. LES FORETS DES COTEAUX DE GASCOGNE	52
3.1.1 LE SITE D'ETUDE	52
3.1.2 LE CLIMAT	53
3.1.3 LE PAYSAGE	53
3.2. LES SYRPHEs : UN BIO-INDICATEUR DES PAYSAGES RURAUX	53
CONCLUSION DE LA PARTIE A	57

INTRODUCTION DE LA PARTIE B	63
CHAPITRE 4. SOURCES DE DONNEES SPATIALES RETENUES ET SPECIFICITES	65
4.1. CRITERES DE CHOIX DES DONNEES SPATIALES	66
4.2. LES MINUTES DE LA CARTE D'ETAT-MAJOR AU 1/40 000 ^e	68
4.2.1 Historique de constitution de la carte	68
4.2.2 SPECIFICATIONS DES DESSINS-MINUTES D'ETAT-MAJOR	70
4.2.3 REPRESENTATION DES ESPACES FORESTIERS DANS LES DESSINS-MINUTES	76
4.2.4 LES DESSINS-MINUTES DE LA ZONE D'ETUDE	80
4.3. LA REVISION TYPE 1889 DE LA CARTE D'ETAT-MAJOR AU 1/80 000 ^e	80
4.3.1 HISTORIQUE DE CONSTITUTION DE LA CARTE	80
4.3.2 LES SPECIFICATIONS DE LA REVISION	82
4.3.3 REPRESENTATION DES ESPACES FORESTIERS DANS LA CARTE TYPE 1889	85
4.3.4 LES CARTES TYPE 1889 DE LA ZONE D'ETUDE	86
4.4. LES PHOTOGRAPHIES AERIENNES NOIR ET BLANC NUMERISEES	86
4.4.1 LES MISSIONS AERIENNES DE 1953-1954	87
4.4.2 LES MISSIONS AERIENNES DE 1979	87
4.5. LA BD ORTHO [®] INFRAROUGE 2010 DE L'IGN	88
4.6. BILAN : DES DONNEES DIFFERENTES QUI RESTENT COMPATIBLES	89
CHAPITRE 5. METHODE DE CORRECTION GEOMETRIQUE DES CARTES ANCIENNES	91
5.1. CORRIGER, RECALER : UNE NECESSITE	92
5.2. PRINCIPE GENERAL DE LA CORRECTION GEOMETRIQUE	92
5.3. METHODES DE CORRECTION GEOMETRIQUE EXISTANTES	94
5.3.1 METHODES DITES GLOBALES	95
5.3.2 METHODES DITES LOCALES	96
5.4. METHODES DE CORRECTION SPECIFIQUES AUX CARTES ANCIENNES	97
5.4.1 CHOIX DES POINTS D'AMER	97
5.4.2 CHOIX D'UN MODELE DE TRANSFORMATION	98
5.4.3 BILAN	99
5.5. PROPOSITION D'UNE METHODE DE CORRECTION ALTERNATIVE : LA REGRESSION A NOYAU	99
5.5.1 LA REGRESSION RIDGE A NOYAU	100
5.5.1.1. L'INTRODUCTION DE LA PENALISATION RIDGE DANS LA REGRESSION POLYNOMIALE	100
5.5.1.2. L'UTILISATION DES NOYAUX	101
5.6. SIMULER POUR CHOISIR UNE STRATEGIE OPTIMALE DE CORRECTION DES CARTES ANCIENNES	103
5.6.1 CARTES ANCIENNES A CORRIGER	103
5.6.2 PROTOCOLE D'EXPERIMENTATION	104
5.6.3 MISE EN ŒUVRE ET RESULTATS	107
5.6.4 DISCUSSION	112
5.7. SIMULER POUR AIDER A DECIDER DANS UN CONTEXTE OPERATIONNEL	113
5.7.1 LE PROJET OCS ANCIEN	114
5.7.2 PROTOCOLE D'EXPERIMENTATION DEFINI DANS LE PROJET OCS ANCIEN	114
5.7.3 RESULTATS	118
5.7.4 DISCUSSION	123
5.7.5 Bilan	123

CHAPITRE 6	.VECTORISATION MANUELLE DES ESPACES FORESTIERS	125
6.1.	LA VECTORISATION MANUELLE : PRINCIPE, AVANTAGES ET LIMITES	125
6.2	MISE EN ŒUVRE DANS LES DESSINS-MINUTES D'ETAT-MAJOR AU 1/40 000 ^e	125
6.3	MISE EN ŒUVRE DANS LA REVISION TYPE 1889	125
6.4	MISE EN ŒUVRE DANS LES PHOTOGRAPHIES AERIENNES	125
6.5	RESULTATS DE LA VECTORISATION MANUELLE	125
6.1.	LA VECTORISATION MANUELLE : PRINCIPE, AVANTAGES ET LIMITES	126
6.2.	MISE EN ŒUVRE DANS LES DESSINS-MINUTES D'ETAT-MAJOR AU 1/40 000 ^e	127
6.2.1	LA BASE DE REGLES	127
6.2.2	LA DIGITALISATION	128
6.3.	MISE EN ŒUVRE DANS LA REVISION TYPE 1889	129
6.3.1	LA BASE DE REGLES	129
6.3.2	LA DIGITALISATION	130
6.4.	MISE EN ŒUVRE DANS LES PHOTOGRAPHIES AERIENNES	131
6.4.1	PRETRAITEMENTS ET ORTHORECTIFICATION DES PHOTOGRAPHIES AERIENNES ANCIENNES	131
6.4.2	LA BASE DE REGLES DE DIGITALISATION	132
6.4.3	LA DIGITALISATION	133
6.5.	RESULTATS DE LA VECTORISATION MANUELLE	133
CHAPITRE 7.	VECTORISATION AUTOMATIQUE DES ESPACES FORESTIERS	139
7.1.	ENJEUX DE L'EXTRACTION AUTOMATIQUE	140
7.2.	PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE AUX FORETS DES MINUTES DE L'ETAT-MAJOR	140
7.2.1	L'EXTRACTION AUTOMATIQUE DANS LES CARTES SCANNEES : APPROCHES EXISTANTES	140
7.2.2	DEMARCHE GENERALE PROPOSEE	145
7.2.3	PRETRAITEMENTS ET FILTRAGE DES ELEMENTS SURCHARGEANT LA FORET	146
7.2.4	TRANSFORMATION DE L'ESPACE COLORIMETRIQUE : DU RVB AU CIELAB	147
7.2.5	RECONNAISSANCE AUTOMATIQUE DES FORETS PAR CLASSIFICATION SUPERVISEE ET NON SUPERVISEE	149
7.2.6	TRAITEMENTS POST-CLASSIFICATION	151
7.2.7	IMPLEMENTATION DE LA METHODE AUTOMATIQUE DANS LA PLATEFORME ENVI/IDL	152
7.2.8	EXPERIMENTATION ET MISE EN ŒUVRE SUR LE SITE D'ETUDE	152
7.2.9	RESULTATS ET DISCUSSION	154
	A PARTIR DES TROIS EXTRAITS SELECTIONNES	154
7.2.10	MISE EN ŒUVRE DANS LE CADRE DU PROJET OCS ANCIEN DE L'IGN	160
7.2.11	BILAN	166
7.3.	PROPOSITION D'UNE DEMARCHE AUTOMATIQUE APPLIQUEE A LA BD ORTHO [®] IRC	167
7.3.1	ENJEUX DE L'EXTRACTION AUTOMATIQUE DES FORETS DANS LES PHOTOGRAPHIES AERIENNES	167
7.3.2	DEMARCHE GENERALE PROPOSEE POUR EXTRAIRE LES FORETS DE LA BD ORTHO [®] IRC	168
7.3.3	RESULTATS ET DISCUSSION	179
	CONCLUSION DE LA PARTIE B	183
PARTIE C		187
	INTRODUCTION DE LA PARTIE C	189
CHAPITRE 8.	DEMARCHE METHODOLOGIQUE PROPOSEE	191
8.1.	L'APPARIEMENT DES FORETS	192
8.1.1	L'APPARIEMENT : DEFINITION ET METHODES ACTUELLES	192

8.1.2	LE PROBLEME D'APPARIEMENT DES FORETS	193
8.1.3	DEMARCHE D'APPARIEMENT VISUEL	194
8.1.3.1.	LES CRITERES D'APPARIEMENT	195
8.1.3.2.	LE SENS D'APPARIEMENT	196
8.1.3.3.	LA PROCEDURE D'APPARIEMENT	197
8.2.	LA CARACTERISATION DES ETATS	199
8.2.1	DECRIRE LA GEOMETRIE ET LE CONTEXTE DES OBJETS PAR DES METRIQUES	199
8.2.2	LES METRIQUES PAYSAGERES EN ECOLOGIE DU PAYSAGE	199
8.2.3	CHOIX DES METRIQUES PAYSAGERES POUR CARACTERISER LES FORETS	200
8.2.3.1.	LES METRIQUES POUR MESURER LES EFFETS DE LA MORPHOLOGIE DES FORETS	201
8.2.3.2.	LES METRIQUES POUR MESURER LES EFFETS D'ISOLEMENT DES FORETS	202
8.3.	LA CARACTERISATION DES ENTITES SPATIO-TEMPORELLES	204
8.3.1	ETATS, EVENEMENTS ET PROCESSUS : DEFINITIONS	204
8.3.2	DECRIRE L'EVOLUTION DES OBJETS GEOGRAPHIQUES : PLUSIEURS MODELES	205
8.3.3	LES EVOLUTIONS POSSIBLES DES FORETS	208
8.3.4	CHOIX D'UN MODELE DESCRIPTIF POUR CARACTERISER LES EVOLUTIONS DES FORETS	210
8.3.5	MESURER LES EVOLUTIONS A TRAVERS LES CHANGEMENTS D'ETATS : UNE APPROCHE « INDIVIDU »	210
8.4.	CLASSER LES EVOLUTIONS SPATIO-TEMPORELLES : UNE APPROCHE « POPULATION »	211
8.4.1	CONSTRUIRE, DECRIRE ET CATEGORISER DES TYPES D'EVOLUTIONS	211
8.4.2	DEMARCHE ADOPTEE POUR IDENTIFIER DES TYPES D'EVOLUTIONS : UNE APPROCHE « POPULATION »	211
8.4.3	REALISATION D'UNE ACP SUR L'ANNEE DE REFERENCE	213
8.4.4	PROJECTION DES AUTRES DATES DANS LE PLAN FACTORIEL DE REFERENCE	213
8.4.5	LA CLASSIFICATION DES TRAJECTOIRES	213
CHAPITRE 9. MISE EN ŒUVRE DE LA DEMARCHE METHODOLOGIQUE PROPOSEE		215
9.1.	REPRESENTATION ET INTERPRETATION DES LIENS D'APPARIEMENT	216
9.1.1	DESCRIPTION DES ARBRES D'EVOLUTION	217
9.1.2	DISCUSSION	219
9.2.	CALCUL ET ANALYSE DES METRIQUES PAYSAGERES PAR DATE	219
9.3.	ANALYSE DES CHANGEMENTS D'ETATS	221
9.4.	TYPOLOGIE DES EVOLUTIONS SPATIO-TEMPORELLES	224
9.4.1	ACP SUR L'ANNEE DE REFERENCE	224
9.4.2	LES TRAJECTOIRES D'EVOLUTION	226
9.4.3	LES GROUPES DE TRAJECTOIRES	227
CONCLUSION DE LA PARTIE C		231
PARTIE D		235
INTRODUCTION DE LA PARTIE D		237
CHAPITRE 10. MODELISATION ESPECE-HABITAT : EFFET DE L'EVOLUTION DES FORETS SUR LA BIODIVERSITE		239
10.1.	LA MODELISATION ESPECE-HABITAT EN ECOLOGIE HISTORIQUE : PRINCIPES ET SPECIFICITES	240
10.1.1	LES NOTIONS DE CREDIT DE COLONISATION ET DE DETTE D'EXTINCTION	240
10.1.2	LES DETERMINANTS DE LA DETTE ET DU CREDIT	242
10.1.3	COMMENT MESURER UNE DETTE D'EXTINCTION OU UN CREDIT DE COLONISATION ?	243
10.1.4	BILAN	245
10.2.	DEMARCHE DE MODELISATION PROPOSEE	246

10.2.1	LES VARIABLES EXPLICATIVES ET LA VARIABLE REPONSE	246
10.2.2	LE CHOIX DU MODELE STATISTIQUE	247
10.2.3	LA PROCEDURE DE MODELISATION	249
10.2.3.1.	VERIFICATION DES CORRELATIONS ENTRE VARIABLES EXPLICATIVES (PRE-AJUSTEMENT)	250
10.2.3.2.	CALCUL D'UN GLM DE TYPE « POISSON »	250
10.2.3.3.	VERIFICATION DE L'HOMOSCEDASTICITE (POST-AJUSTEMENT)	251
10.2.3.4.	ESTIMATION DE QUALITE DE L'AJUSTEMENT (POST-AJUSTEMENT)	251
10.2.3.5.	SELECTION DE VARIABLES	252
10.2.3.6.	ANALYSE DES SORTIES DE MODELE	253
10.2.4	DEMARCHE DE MODELISATION	254
10.3.	UNE ETUDE PUBLIEE DANS LANDSCAPE ECOLOGY JOURNAL	254
CHAPITRE 11.	MODELISATION ESPECE-HABITAT : EFFET DES INCERTITUDES SPATIALES	275
11.1.	ORIGINE ET NATURE DES INCERTITUDES DANS L'INFORMATION GEOGRAPHIQUE	276
11.1.1	DEFINITIONS	276
11.1.2	LA QUALITE DES DONNEES GEOGRAPHIQUES	277
11.1.3	EVALUATION DE LA QUALITE INTERNE	278
11.1.3.1.	ABSTRACTION, SPECIFICATIONS ET TERRAIN NOMINAL	278
11.1.3.2.	LES SOURCES D'ERREURS DE LA QUALITE INTERNE	278
11.1.4	UN CADRE POUR EVALUER L'INCERTITUDE DANS LES DONNEES SPATIALES ANCIENNES	279
11.2.	TRAVAUX TENANT COMPTE DES INCERTITUDES SPATIALES DANS LES MODELES ESPECES-HABITAT	280
11.3.	MESURER L'IMPACT DE L'INCERTITUDE DES DONNEES SPATIALES LIEE AU PROCESSUS DE VECTORISATION AUTOMATIQUE : DEMARCHE ET MISE EN OEUVRE	281
11.3.1	RESULTATS DE COMPARAISON	282
11.4.	BILAN	287
	CONCLUSION DE LA PARTIE D	289
	CONCLUSION GENERALE	291
	BIBLIOGRAPHIE	297
	ANNEXE	333
	LISTE DES FIGURES	381
	LISTE DES TABLEAUX	389
	TABLE DES MATIERES	393

RESUME : La biodiversité dans les paysages dépend des caractéristiques de ce paysage mais peut aussi être influencée par son histoire. En effet, certaines espèces ne réagissent pas immédiatement à une perturbation mais peuvent montrer un temps de réponse plus ou moins long. De ce fait, la prise en compte de l'évolution de l'habitat des espèces est devenue un enjeu important en écologie depuis quelques années, pour mieux comprendre la présence ou la diversité biologique actuelle. L'objectif de cette thèse en géomatique s'inscrit dans ce contexte applicatif d'écologie historique. Le sujet que nous traitons porte sur l'extraction automatique d'îlots boisés et la caractérisation de leur évolution spatio-temporelle depuis le milieu du XIXe siècle pour modéliser l'effet de leur trajectoire historique sur la diversité actuelle en syrphes forestiers (Diptera : Syrphidae). Le site d'étude est un paysage agri-forestier des Coteaux de Gascogne. La démarche générale proposée se compose de trois étapes principales : (1) la constitution de la base de données spatiales des îlots boisés intégrant plusieurs sources de données hétérogènes, (2) l'appariement des îlots boisés aux différentes dates avec la caractérisation de leur évolution spatio-temporelle, (3) la modélisation statistique des relations espèces-habitats intégrant l'histoire comme un des facteurs explicatifs de la diversité en syrphes observée. Plusieurs contributions méthodologiques ont été apportées à cette démarche. Nous avons proposé une nouvelle méthode de correction géométrique fondée sur la régression *ridge* à noyau pour rendre compatible les données spatiales anciennes et actuelles mobilisées. Nous avons également développé une approche et un outil de vectorisation automatique des forêts dans les dessins-minutes de la carte d'Etat-Major du XIXe siècle. Enfin, une première évaluation de l'impact de l'incertitude spatiale sur la réponse des modèles espèces-habitats a également été initiée. D'un point de vue écologique, les résultats ont révélé un effet significatif de la continuité temporelle des habitats sur la diversité en syrphes forestiers. Nous avons montré que les forêts les plus isolées présentaient une dette d'extinction ou un crédit de colonisation selon le type d'évolutions apparues au cours de la dernière période étudiée (1979-2010). Il s'est avéré qu'une durée de 30 ans n'était pas suffisante aux syrphes forestiers pour qu'ils retrouvent un état d'équilibre à la suite d'une évolution spatiale de leur habitat isolé.

Mots clés : cartes anciennes, forêts, vectorisation automatique, correction géométrique, régression *ridge* à noyau, évolution spatio-temporelle, modélisation espèce-habitat, écologie historique.

ABSTRACT: Biodiversity in landscapes depends on landscape spatial patterns but can also be influenced by landscape history. Indeed, some species are likely to respond in the longer term to habitat disturbances. Therefore, in recent years, landscape dynamics have become a possible factor to explain current biodiversity. The aim of this thesis in GIS is part of this historical ecology context. We are dealing with automatic extraction of forest patches and characterization of their spatiotemporal evolution. The objective is to evaluate forest dynamics effects on current diversity of forest hoverflies. (Diptera: Syrphidae) in the agri-forestry landscape of Coteaux de Gascogne. The proposed general approach consists of three main steps: (1) the forest spatial database production from heterogeneous sources, (2) forest patches matching and characterization of their spatiotemporal evolution, (3) species-habitat modeling while integrating history as one of the factors likely to explain hoverflies diversity. Several methodological contributions were made. We proposed a new geometric correction approach based on kernel ridge regression to make consistent past and present selected data sources. We also developed an automatic extraction approach of forest from Historical Map of France of the 19th century. Finally, spatial uncertainty effects on ecological models responses have been assessed. From an ecological viewpoint, a significant effect from historical continuity of patches on forest hoverflies diversity was revealed. The most isolated fragments presented an extinction debt or a colonization credit according to area dynamics occurred in the last time-period (1970-2010). As it turns out, 30 years was not sufficient for forest hoverflies to reach new equilibrium after isolated habitat changes.

Keywords: Historical map, forest, automatic extraction, geometric correction, kernel ridge regression, spatio-temporal evolution, species-habitat modeling, historical ecology