



# Habitat modeling of Tetraonidae in the Jura massif: contribution of LiDAR airborne remote sensing

Anouk Glad

## ► To cite this version:

Anouk Glad. Habitat modeling of Tetraonidae in the Jura massif: contribution of LiDAR airborne remote sensing. Ecosystems. Université Grenoble Alpes, 2018. English. NNT : 2018GREAV070 . tel-02609188v2

**HAL Id: tel-02609188**

**<https://theses.hal.science/tel-02609188v2>**

Submitted on 7 May 2019

**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**

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES**

Spécialité : **Biodiversité, écologie, environnement**

Arrêté ministériel : 25 mai 2016

Présentée par

**Anouk GLAD**

Thèse dirigée par **Björn REINEKING**, Directeur de  
Recherche, à IRSTEA Grenoble

Préparée au sein du **Laboratoire IRSTEA**  
Dans l'**École Doctorale Chimie et Sciences du Vivant**

### **Modélisation de l'habitat des tétraonidés dans le massif du Jura : apport de la télédétection LiDAR aéroportée.**

Thèse soutenue publiquement le « **14 décembre 2018** »  
devant le jury composé de :

**Mme Ilse STORCH**

Professeure, Université de Fribourg, Rapportrice

**M. Aurélien BESNARD**

Maitre de Conférences, CEFE Montpellier, Rapporteur

**Mme Anne LOISON**

Directrice de Recherche, Laboratoire d'Ecologie Alpine (CNRS),  
Présidente du Jury

**M. Frédéric ARCHAUX**

Ingénieur-Chercheur, IRSTEA Nogent-sur-Vernisson, Examineur

**Mme Alexandra DEPAZ**

Coordinatrice, Groupe Tétras Jura, Invitée

**M. Jean-Matthieu MONNET**

Ingénieur de Recherche, IRSTEA Grenoble, Co-encadrant

**M. Björn REINEKING**

Directeur de Recherche, IRSTEA Grenoble, Directeur de thèse



# TABLE OF CONTENT

<b>TABLE OF CONTENT .....</b>	<b>2</b>
<b>LIST OF FIGURES.....</b>	<b>6</b>
<b>LIST OF TABLES .....</b>	<b>8</b>
<b>REMERCIEMENTS.....</b>	<b>9</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>12</b>
<b>1 CONTEXT .....</b>	<b>13</b>
1.1 BIODIVERSITY EROSION .....	13
1.2 LOCAL CONTEXT .....	15
1.3 SPECIES DISTRIBUTION MODELS: A TOOL FOR MANAGEMENT .....	16
1.4 SYNTHESIS.....	21
<b>2 REMOTE SENSING: THE RISE OF LIDAR.....</b>	<b>22</b>
2.1 REMOTE SENSING FOR SDMS .....	22
2.2 LIDAR: HISTORY AND DEFINITIONS .....	22
2.3 LIDAR IN ECOLOGY AND CONSERVATION BIOLOGY .....	24
<b>3 TWO EMBLEMATIC SPECIES FROM THE JURA MASSIF: A NEED FOR RELIABLE SPATIAL PREDICTIONS.....</b>	<b>26</b>
3.1 THE GALLIFORM ORDER.....	26
3.2 CAPERCAILLIE.....	26
3.3 HAZEL GROUSE .....	38
<b>4 OBJECTIVES.....</b>	<b>47</b>
4.1 GENERAL QUESTIONS .....	47
4.2 MANUSCRIPT ORGANIZATION .....	48
<b>CHAPTER 2: MATERIALS .....</b>	<b>50</b>
<b>1 THE JURA MASSIF STUDY AREA.....</b>	<b>51</b>
<b>2 CAPERCAILLIE AND HAZEL GROUSE DATASETS .....</b>	<b>52</b>
2.1 CAPERCAILLIE.....	53
2.2 HAZEL GROUSE .....	53
<b>3 LIDAR DATASETS.....</b>	<b>54</b>
3.1 LIDAR CAMPAIGNS.....	54
3.2 LIDAR EXTRACTED VARIABLES: DESCRIPTIONS .....	55
3.3 SYNTHESIS.....	65
<b>4 BEECH REGENERATION DATASETS.....</b>	<b>65</b>
4.1 GROUPE TETRAS JURA DATASET .....	65
4.2 ONF DATASET.....	67

<b>5</b>	<b>ADDITIONAL DATASETS .....</b>	<b>69</b>
5.1	SATELLITE IMAGES.....	69
5.2	ANTHROPOGENIC DISTURBANCES.....	69

### **CHAPTER 3: MODELING METHODS ..... 70**

<b>1</b>	<b>FOREWORD.....</b>	<b>71</b>
<b>2</b>	<b>SDMs: PRINCIPLES AND ASSUMPTIONS .....</b>	<b>72</b>
2.1	ECOLOGICAL NICHE AND HABITAT DEFINITIONS.....	72
2.2	CONCEPT.....	73
2.3	ASSUMPTIONS.....	75
2.4	LINK BETWEEN THE NICHE CONCEPT AND SDMS.....	76
2.5	THE CHOICE OF THE STATISTICAL METHOD .....	77
<b>3</b>	<b>ARTICLE 1: DISTRIBUTION MODELS OF DIMORPHIC SPECIES: CORRECTING SPATIAL SAMPLING BIAS IMPROVES ESTIMATES OF HABITAT ASSOCIATION AND SEX RATIO .....</b>	<b>80</b>
3.1	ABSTRACT.....	80
3.2	INTRODUCTION.....	81
3.3	MATERIAL AND METHODS .....	84
3.4	RESULTS.....	92
3.5	DISCUSSION .....	95
3.6	CONCLUSION.....	98
3.7	AUTHORS' CONTRIBUTIONS .....	98
3.8	ACKNOWLEDGEMENTS.....	99
3.9	DATA ACCESSIBILITY.....	99
3.10	BIBLIOGRAPHY.....	99
<b>4</b>	<b>THE ROLE OF SAMPLING DESIGN: CAN WE REDUCE SAMPLING BIAS EFFECT? .....</b>	<b>105</b>
4.1	INTRODUCTION.....	105
4.2	METHODS.....	106
4.3	RESULTS .....	107
4.4	DISCUSSION .....	111
<b>5</b>	<b>THE CHOICE OF SPATIAL SCALE.....</b>	<b>114</b>
5.1	FROM MULTIPLE LEVELS TO MULTI-SCALE SELECTIONS .....	116
5.2	SELECTION OF SCALES AND METHODS .....	117
<b>6</b>	<b>CONCLUSION .....</b>	<b>118</b>

### **CHAPTER 4: HAZEL GROUSE DISTRIBUTION MODELS ..... 120**

<b>1</b>	<b>INTRODUCTION .....</b>	<b>121</b>
<b>2</b>	<b>ARTICLE 2: EFFECT OF INTERACTION TERMS ON SPECIES DISTRIBUTION MODELS PREDICTIONS INTERPRETATION AT MULTIPLE SCALES. ....</b>	<b>122</b>
2.1	ABSTRACT.....	122
2.2	INTRODUCTION.....	123
2.3	MATERIAL AND METHODS .....	127
2.4	RESULTS.....	136



2.5	DISCUSSION .....	142
2.6	CONCLUSION .....	144
2.7	ACKNOWLEDGEMENT .....	145
2.8	BIBLIOGRAPHY .....	145
<b>3</b>	<b>UNIVARIATE RESPONSE CURVES ANALYSIS .....</b>	<b>151</b>
3.1	METHODS .....	151
3.2	RESULTS.....	152
3.3	DISCUSSION.....	158
<b>4</b>	<b>RELATIVE SUITABILITY MAP PREDICTIONS .....</b>	<b>159</b>
4.1	RESULTS.....	159
4.2	DISCUSSION.....	160
<b>5</b>	<b>CONCLUSION .....</b>	<b>160</b>

## **CHAPTER 5: CAPERCAILLIE DISTRIBUTION MODELS .....** **161**

<b>1</b>	<b>INTRODUCTION .....</b>	<b>162</b>
<b>2</b>	<b>ARTICLE 3: MAPPING AND MANAGING BIRD HABITAT WITH LiDAR-BASED MULTI-SCALE OBJECT-ORIENTED INDICATORS .....</b>	<b>164</b>
2.1	ABSTRACT .....	164
2.2	INTRODUCTION .....	165
2.3	MATERIAL .....	168
2.4	METHODS .....	170
2.5	RESULTS.....	175
2.6	DISCUSSION.....	181
2.7	CONCLUSION .....	183
2.8	ACKNOWLEDGEMENTS .....	184
2.9	DATA ACCESSIBILITY .....	184
2.10	BIBLIOGRAPHY .....	184
<b>3</b>	<b>CAPERCAILLIE DISTRIBUTION MODELS .....</b>	<b>188</b>
3.1	MATERIAL .....	188
3.2	METHODS .....	190
3.3	WINTER MODELS RESULTS.....	191
3.4	SUMMER MODELS RESULTS .....	206
<b>4</b>	<b>DISCUSSION .....</b>	<b>216</b>
<b>5</b>	<b>CONCLUSION .....</b>	<b>218</b>

## **CHAPTER 6: BEECH REGENERATION DISTRIBUTION MODELS .....** **220**

<b>1</b>	<b>INTRODUCTION .....</b>	<b>221</b>
1.1	HABITAT LOSS .....	221
1.2	BEECH .....	222
1.3	LiDAR AND UNDERSTORY MODELING.....	222
1.4	OBJECTIVES .....	223
<b>2</b>	<b>MATERIALS AND METHODS.....</b>	<b>223</b>



2.1 ENVIRONMENTAL VARIABLES.....	223
2.2 AIN STUDY AREA .....	223
2.3 JURA STUDY AREA .....	224
2.4 VISUALIZATION .....	224
2.5 MODELING .....	224
<b>3 RESULTS.....</b>	<b>225</b>
3.1 FIND POTENTIAL VARIABLES .....	225
3.2 MODELS.....	231
<b>4 DISCUSSION .....</b>	<b>234</b>
<b>5 CONCLUSION .....</b>	<b>237</b>
 <b>CHAPTER 7: SYNTHESIS AND PERSPECTIVES .....</b>	 <b>238</b>
 <b>1 IMPROVING SDMs PERFORMANCES AND UNDERSTAND THEIR LIMITS .....</b>	 <b>239</b>
1.1 SPATIAL SAMPLING BIAS .....	239
1.2 SCALE EVALUATION .....	241
1.3 PREDICTIONS.....	243
1.4 ENVIRONMENTAL VARIABLES.....	244
<b>2 IMPROVING TRANSMISSION TO MANAGERS .....</b>	<b>247</b>
<b>3 LOCAL PERSPECTIVES .....</b>	<b>249</b>
<b>4 CONCLUSION .....</b>	<b>250</b>
 <b>BIBLIOGRAPHY.....</b>	 <b>251</b>
 <b>ANNEXES .....</b>	 <b>264</b>
 <b>1 CONTENT.....</b>	 <b>265</b>

# LIST OF FIGURES

FIGURE 1 : LIDAR SYSTEM MEASUREMENT .....	23
FIGURE 2 : MALE AND FEMALE CAPERCAILLIE .....	27
FIGURE 3 : CAPERCAILLIE DISTRIBUTION RANGE .....	28
FIGURE 4 : CAPERCAILLIE PRESENCE IN FRANCE WITH A ZOOM IN THE JURA MASSIF STUDY AREA (OGM, 2009). .....	29
FIGURE 5 : HABITAT OF GOOD QUALITY FOR CAPERCAILLIE .....	30
FIGURE 6 : IDEAL HABITAT WITH A DENSE BILBERRY COVER .....	32
FIGURE 7 : MALE HAZEL GROUSE .....	38
FIGURE 8 : HAZEL GROUSE DISTRIBUTION RANGE .....	39
FIGURE 9 : HAZEL GROUSE POPULATION REPARTITION IN FRANCE AND IN THE JURA MASSIF STUDY AREA .....	40
FIGURE 10 : A GOOD HABITAT FOR HAZEL GROUSE .....	42
FIGURE 11 : MANUSCRIPT ORGANIZATION .....	49
FIGURE 12 : COVERAGE OF THE TWO LIDAR CAMPAIGNS IN THE FRENCH JURA MASSIF CONSTITUTING TWO STUDY AREAS. ....	55
FIGURE 13 : THE NORMALIZATION OF THE LIDAR POINT CLOUD GENERATE THE DEFORMATION OF THE TREE CROWN AND A SHIFT OF THE TREE TOP LOCATION .....	56
FIGURE 14 : EXAMPLE OF A DELAUNAY TRIANGULATION APPLIED ON A POINT CLOUD .....	57
FIGURE 15 : TREE TOP DETECTION AND CROWN SEGMENTATION .....	60
FIGURE 16 : THE METHOD USED TO EXTRACT GAP OBJECTS FROM LIDAR DATASETS .....	61
FIGURE 17 : TREES LOCALIZATION IN THE FIELD (RED CIRCLES) AND LIDAR DETECTED TREES (BLACK CROSSES) ON THREE FIELD PLOTS, LARGER SYMBOLS ARE INDICATING HIGHER TREES .....	63
FIGURE 18 : LOCATION OF THE TWO CAMPAIGNS OVERLAPPING AREAS. ....	63
FIGURE 19 : GTJ VEGETATION PLOTS .....	67
FIGURE 20 : SUB-PLOTS LOCATION, PROTOCOLE SUIVI D'ESPACE NATURELS PROTEGES MEDD 2005 .....	67
FIGURE 21 : ONF PERMANENT VEGETATION PLOTS .....	68
FIGURE 22 : SDMS CONCEPT ILLUSTRATED WITH TWO ENVIRONMENTAL VARIABLES .....	74
FIGURE 23 : STUDY AREAS WITH LIDAR SURVEYS, CAPERCAILLIE SAMPLING DESIGNS AND LOCALIZATIONS FOR VIRTUAL SPECIES CASE STUDY. ....	85
FIGURE 24 : OVERVIEW OF THE THREE-SAMPLING DESIGNS FOR THE VIRTUAL SPECIES CASE STUDY OVER THE INTENSITY DISTRIBUTION MAP. A) REPRESENTS THE RANDOM SAMPLING DESIGN AND RANDOM BACKGROUND POINT, B) REPRESENTS THE TRANSECTS SAMPLING DESIGN AND TARGETED BACKGROUND POINT SAMPLING, C) REPRESENT THE SUBJECTIVE SAMPLING DESIGN AND TARGETED BACKGROUND POINTS. ....	90
FIGURE 25 : PARAMETER ESTIMATIONS (100 MODELS) FOR SUB-GROUP 1 (FEMALE) .....	93
FIGURE 26 : SEX RATIO ESTIMATED IN THE VIRTUAL CASE STUDY FROM THE NUMBER OF OBSERVATIONS AND FROM PREDICTIONS OF SDMS THAT USE DIFFERENT SAMPLING DESIGNS AND CORRECTION METHODS. THE RED LINE REPRESENTS THE MEAN SEX RATIO CALCULATED FROM ALL SIMULATED PRESENCES. ....	94
FIGURE 27 : THE PARAMETERS ESTIMATION OF CAPERCAILLIE FOR EACH SEX IN FUNCTION OF THE SAMPLING DESIGN AND APPLIED CORRECTIONS. ....	95
FIGURE 28 : LOCATION OF FIVE CANDIDATE TRANSECTS .....	107
FIGURE 29 : DENSITY CURVES FOR THE THREE ENVIRONMENTAL VARIABLES ALONG SUBJECTIVE TRAJECTORIES, CANDIDATE TRANSECTS AND WITHIN THE LANDSCAPE. ....	108
FIGURE 30 : PARAMETERS ESTIMATION OF 100 MODELS FOR THE VIRTUAL SPECIES (SUB-GROUP 2) IN DIFFERENT SAMPLING CASES. ....	110
FIGURE 31 : PROPORTION OF SUITABLE HABITAT IN THE LANDSCAPE AND ALONG SYSTEMATIC TRANSECT AND SUBJECTIVE TRAJECTORIES, BLUE IS MALE AND RED IS FEMALE .....	112
FIGURE 32 : FRAMEWORK USED TO PRODUCE AND COMPARE MODELS USING DIFFERENT SCALES .....	118
FIGURE 33 : LOCATION OF THE STUDY AREA WITHIN THREE FRENCH DEPARTMENTS, (AIN, JURA, AND DOUBS). TWO LIDAR CAMPAIGNS WERE CONDUCTED IN TWO DIFFERENT PARTS OF THE STUDY AREA. ....	128



FIGURE 34 : SCHEMATIC REPRESENTATION OF MODEL SELECTION FRAMEWORK. ONLY THREE VARIABLES AND FOUR SCALES ARE REPRESENTED, AS AN EXAMPLE. ....	136
FIGURE 35 : THREE DIMENSIONAL RESPONSE CURVES AT 7 HA SCALE: SUITABILITY PREDICTIONS AS FUNCTION OF THE MEAN PROPORTION OF OPEN AREAS AND OF THE STANDARD DEVIATION OF 2-5 M PENETRATION RATIO, DEPENDING ON THE INCLUSION OF INTERACTION TERMS IN THE MODELS (TOP ROW : AIN STUDY AREA, AND BOTTOM ROW : JURA STUDY AREA) ..	139
FIGURE 36 : THREE DIMENSIONAL RESPONSE CURVES OF EACH OF THE THREE CASES AT 27.5 HA SCALE (TOP ROW: AIN STUDY AREA, AND BOTTOM ROW: JURA STUDY AREA) .....	139
FIGURE 37 : CONTRIBUTIONS OF COMMON VARIABLES FOR BOTH STUDY AREAS.....	142
FIGURE 38 : RESPONSES CURVES IN THE AIN STUDY AREAS (PART 1) .....	154
FIGURE 39 : RESPONSES CURVES IN THE AIN STUDY AREAS (PART 2) .....	155
FIGURE 40 : RESPONSES CURVES IN THE JURA STUDY AREAS (PART 1) .....	156
FIGURE 41 : RESPONSES CURVES IN THE JURA STUDY AREAS (PART 2) .....	157
FIGURE 42 : PREDICTED MAPS OF TWO BEST MODELS WITHIN THE TWO STUDY AREAS.....	159
FIGURE 43 : THE STUDY AREA AND THE TWO LIDAR CAMPAIGNS.....	169
FIGURE 44 : VISUALIZATION OF CALCULATED LIDAR METRICS .....	173
FIGURE 45 : CONTRIBUTION OF THE VARIABLES TO EACH MODEL .....	177
FIGURE 46 : MALE RESPONSE CURVES .....	179
FIGURE 47 : FEMALE RESPONSE CURVES .....	180
FIGURE 48 : CONTRIBUTION OF THE VARIABLES IN THE AIN STUDY AREA FOR BOTH SEXES. ....	194
FIGURE 49 : CONTRIBUTION OF THE VARIABLES IN THE JURA STUDY AREA FOR BOTH SEXES.....	195
FIGURE 50 : CONTRIBUTION FOR THE SCALE 0.0625HA, IN THE AIN STUDY AREA.....	196
FIGURE 51 : CONTRIBUTION FOR THE SCALE 0.0625HA, IN THE JURA STUDY AREA .....	197
FIGURE 52 : RESPONSE CURVES FOR FEMALE MODELS IN THE AIN STUDY AREA. ....	199
FIGURE 53 : RESPONSE CURVES FOR MALE MODELS IN THE AIN STUDY AREA. ....	200
FIGURE 54 : RESPONSE CURVES FOR FEMALE MODELS IN THE JURA STUDY AREA.....	202
FIGURE 55 : RESPONSE CURVE FOR MALE MODELS IN THE JURA STUDY AREA. ....	203
FIGURE 56 : PREDICTION AT TWO DIFFERENT SCALES AND MIXED MODELS FOR FEMALE .....	205
FIGURE 57 : COMPARISON OF THE PREDICTED MALE AND FEMALE HABITAT SUITABILITY AT THE SCALE OF 0.81 HA (AIN STUDY AREA, CHAMPFROMIER FOREST.....	206
FIGURE 58 : COMPRISON OF SUMMER MODELS IN JURA : THE PROBLEM OF OPEN AREA AND WINDFALL .....	207
FIGURE 59 : VARIABLE CONTRIBUTION FOR THE AIN STUDY AREA MODELS .....	209
FIGURE 60 : VARIABLE CONTRIBUTION IN THE JURA STUDY AREA FOR SUMMER MODELS .....	210
FIGURE 61 : RESPONSE CURVES FOR THE AIN STUDY AREA MODELS .....	211
FIGURE 62 : RESPONSE CURVE FOR THE JURA STUDY AREA MODELS.....	213
FIGURE 63 : PREDICTIONS AT TWO DIFFERENT SCALES AND FOR MIXED MODELS IN SUMMER .....	215
FIGURE 64 : EXAMPLE OF UNSUITABLE HABITAT DUE TO BEECH INTENSE REGENERATION (GROUPE TÉTRAS JURA) .....	221
FIGURE 65 : EXAMPLE OF IDEAL SUMMER HABITAT WITH BILBERRY GROUND LAYER (GROUPE TÉTRAS JURA.....	221
FIGURE 66 : BOXPLOTS AND P-VALUES (WILCOXON TEST) FOR EACH PC AREA BASED LIDAR VARIABLE AND TOPOGRAPHIC VARIABLE IN THE AIN STUDY AREA. ....	226
FIGURE 67 : CIRCULAR HISTOGRAMS FOR EXPOSITION .....	227
FIGURE 68 : BOXPLOTS AND P-VALUES (WILCOXON TEST) FOR EACH OO LIDAR VARIABLE IN THE AIN STUDY AREA.....	227
FIGURE 69 : BOXPLOTS AND P-VALUES (WILCOXON TEST) FOR EACH PC AREA BASED LIDAR VARIABLE AND TOPOGRAPHIC VARIABLE IN THE JURA STUDY AREA.....	229
FIGURE 70 : BOXPLOTS AND P-VALUES (WILCOXON TEST) FOR EACH OO LIDAR VARIABLE IN THE JURA STUDY AREA. ....	230
FIGURE 71 : PREDICTIONS FROM THE GLM POISSON DISTRIBUTION MODEL (MONT NOIR FOREST, JURA STUDY AREA) .....	233
FIGURE 72 : PREDICTIONS FROM THE GLM POISSON DISTRIBUTION MODEL (CHAMPFROMIER FOREST, AIN STUDY AREA) .....	234
FIGURE 73 : DIFFERENCES BETWEEN THE AIN AND JURA STUDY AREA WHEN CALCULATING THE VALUES AT DIFFERENT SCALES .....	242

# LIST OF TABLES

TABLE 1 : LIDAR VARIABLES AND THEIR DESCRIPTION (CAPERCAILLIE MODELS) .....	37
TABLE 2: LIDAR VARIABLES AND THEIR DESCRIPTION (HAZEL GROUSE MODELS) .....	46
TABLE 3 : NUMBER OF OBSERVATION BY PROTOCOLS AND SEX FOR CAPERCAILLIE .....	53
TABLE 4 : NUMBER OF OBSERVATIONS BY PROTOCOL AND SEX FOR HAZEL GROUSE .....	54
TABLE 5 : POINT-CLOUD AREA-BASED METRICS.....	58
TABLE 6 : OBJECT-ORIENTED METRICS .....	61
TABLE 7 : LIDAR PREDICTED METRICS .....	62
TABLE 8 : CORRELATION BETWEEN THE OVERLAPPING AREAS FOR EACH METRICS.....	64
TABLE 9 : PARAMETER VALUES FOR THE VIRTUAL SPECIES SUB-GROUPS 1 (FEMALE) AND 2 (MALE) .....	88
TABLE 10 : WILCOXON TEST P-VALUES ON PARAMETER ESTIMATION MEAN VALUES .....	93
TABLE 11 : P-VALUES FROM THE WILCOXON TEST ON THE RATIO ESTIMATION DISTRIBUTION.....	94
TABLE 12 : SEX RATIO ESTIMATES FOR CAPERCAILLIE .....	95
TABLE 13 : P-VALUES FROM WILCOXON RANK TEST .....	110
TABLE 14 : MULTI-SCALE MODEL STRATEGIES FROM (MCGARIGAL ET AL., 2016) .....	116
TABLE 15 : ENVIRONMENTAL METRICS SELECTED TO PREDICT HAZEL GROUSE SUITABLE HABITAT.....	132
TABLE 16 : HABITAT LEVELS AND CORRESPONDING SCALES FOR THE MODELS.....	133
TABLE 17 : AUC RESULTS FOR ALL SINGLE-SCALE, MULTI-VARIABLE MODELS AND THE THREE INTERACTION CASES: NO INTERACTION, INTERACTION BETWEEN MEAN PROPORTION OF OPEN AREA AND STANDARD DEVIATION OF PENETRATION RATIO 2-5M (INTERACTION 1 CASE), INTERACTION BETWEEN MEAN PROPORTION OF OPEN AREA, STANDARD DEVIATION OF PENETRATION RATIO 2-5M AND STANDARD DEVIATION OF PENETRATION RATIO 2-5M (INTERACTION 2 CASE). .....	137
TABLE 18 : SCALES CHOICE FOR THE MIXED MODELS.....	140
TABLE 19 : MODEL PERFORMANCE (AUC AND BOYCE INDEX) .....	176
TABLE 20 : PEARSON CORRELATION BETWEEN PC AREA-BASED AND OO PREDICTIONS OVER THE AIN STUDY AREA. ....	181
TABLE 21 : PEARSON CORRELATION BETWEEN PREDICTIONS OF EACH MODEL TYPE, CALCULATED WITH TWO LIDAR DATASETS.....	181
TABLE 22 : LIDAR VARIABLES USED FOR CAPERCAILLIE MODELS.....	189
TABLE 23 : AUC VALUES CALCULATED FOR SINGLE-SCALE, SINGLE-VARIABLE WINTER MODELS .....	192
TABLE 24 : AUC VALUES FOR WINTER MODELS .....	193
TABLE 25 : AUC OF SINGLE-SCALE SINGLE-VARIABLE MODELS .....	208
TABLE 26 : AUC VALUES FOR SUMMER MODELS.....	208
TABLE 27 : VARIABLES SELECTED FOR MODELING .....	228
TABLE 28 : VARIABLES SELECTED FOR MODELING .....	230
TABLE 29 : MODELS RESULTS .....	232
TABLE 30 : RANDOM FOREST MODELS RESULTS.....	232

# REMERCIEMENTS

De nombreuses personnes se sont impliquées dans ce projet de thèse, et je tiens toutes les à remercier pour leur aide à chacune des étapes de ce travail, en espérant n'avoir oublié personne.

Je voudrais tout d'abord remercier les structures qui ont financé cette thèse, le département de l'Ain, la Région Franche-Comté, le Groupe Tétrás Jura et le centre IRSTEA Grenoble qui m'a accueillie pendant trois ans ainsi que l'ONF qui a partagé les données placettes et des modèles LiDAR avec moi.

Je souhaite aussi remercier mon directeur de thèse Björn Reineking pour son encadrement, pour ces précieux conseils lors des fameux « meeting » et pour la profusion de ses idées pour explorer toujours plus de pistes de recherche. Je remercie aussi Jean-Matthieu Monnet co-encadrant de la thèse, en particulier pour avoir développé le package lidaRtRee, pour son aide et ses recommandations, mais aussi pour l'organisation et l'accompagnement pendant les quelques rares sorties sur le terrain.

Un grand merci à Alexandra Depraz pour sa motivation et son implication sans faille, pour les fréquentes discussions (en particulier ces derniers mois !) et pour les remarques sur les résultats (malgré l'anglais) qui m'ont permis de mieux cerner les problématiques de conservation du Grand Tétrás et les habitudes de la Gélinothe des bois.

Je tiens aussi à remercier Marc Montadert pour son aide et ses remarques tout au long des trois ans lors des réunions trimestrielles organisées à Sévrier, mais aussi lors des comités de thèse où tout simplement par mail. Je remercie aussi Joern Pagel pour ses conseils pour la rédaction de mon premier article et pour sa participation au comité de thèse. J'adresse aussi mes remerciements aux autres membres du comité de thèse, Stéphane Dumas, Thomas Cordonnier, Gwenaél Jacob et Kurt Bollmann pour les discussions annuelles enrichissantes. Je remercie également Benoît Courbaud pour son aide et ses conseils dans la dernière ligne droite de la Thèse.



## Remerciements

Je remercie Damien Buttiglieri et Eric Maldonado, toujours prompt à régler les petits ou gros problèmes informatiques qui sinon nous feraient perdre un temps précieux.

Je souhaite remercier toutes les personnes qui ont participé avec le Groupe Tétras Jura à la collecte des données dont je me suis servie durant la thèse. Sans vous rien n'aurait été possible : le GTJ en particulier mais aussi l'ONF, l'ONCFS, le PNNHCJ et de nombreux bénévoles qui ont apporté leurs contributions.

Je remercie grandement les membres du Jury, Ilse Storch, Aurélien Besnard, Anne Loison et Frédéric Archaux pour avoir accepté d'évaluer mon travail.

Merci à l'équipe jeune (?) et dynamique des thésards (et docteurs !) de l'IRSTEA & Co à la fois au labo et en dehors (les nombreux apéro-balcons!). Merci à Lulu pour les sorties à cheval, les milles-et-un conseils sur mille-et-un sujets, les pauses thé, les visites à Kiara, le désormais mythique Spring Party Tour. Merci à Etienne pour les nombreux apéros organisés chez lui (et à Céline aussi !), les discussions sur les ~~vacances~~ fonctions/modèles, sur l'élevage des têtards et sur la rédaction, d'avoir supporté les bougonnements contre R. On a fait un beau bureau de taiseux (y paraît) mais c'était pour la bonne cause. Je remercie aussi Clément, toujours présent pour résoudre des problèmes de code ou de méthode et pour son esprit de contradiction. Je remercie aussi Gaëlle pour son soutien et ses conseils pour la rédaction. Merci à François pour les nombreux débats et invitations au bar.

Une pensée pour tous les colocs rencontrés pendant ces trois ans qui m'ont permis de décrocher du boulot en rentrant le soir, Fer, Philipp et CoCo mais aussi Francky (#tiramisu), Sylvain, Marianna, Angélique, Raphael, Marie, Anton, Cécile, pour les nombreuses randos, les apéros, la Sicile, les bains de St Pierre, et même le ski.

Merci à la famille malakas: Dorian pour les nombreuses soirées qui ont égaillées ma première année à Grenoble ; Mélissa pour ces nombreuses visites (et j'espère qu'il y en aura d'autres) ; Anna pour le récit annuel ces aventures en Ouzbékistan, et une pensée spéciale pour Claire et à sa petite O. qui sera bientôt parmi nous. Vivement le prochain week-end !



## Remerciements

Un grand merci à ma famille (Maman et Papa !) pour leur soutien et leurs conseils pendant ces trois années (et surtout pendant cette dernière année) et les relectures. Merci aussi à toi Arnaud (Baka !) pour les visites en Belgique et les concerts!

Ma plus grande gratitude va à Xavier, héros de l'ombre pour son soutien et ces encouragements au quotidien, pour sa confiance et sa sérénité, pour les sorties en montagnes et une initiation ratée à LoL, pour les voyages au bout du monde et les bons desserts (Miam la tarte au chocolat ! Et la fondue aussi !), pour la relecture d'une grande partie de cette thèse, et pour avoir supporté mon obsession des Dindes de Montagnes (et de R) pendant 3 ans.

Pour finir, je ne peux terminer ces remerciements sans évoquer mon Grand-père à qui j'avais assuré qu'il serait là le jour ma soutenance et qui nous a quittés trop tôt.



# CHAPTER 1: INTRODUCTION



L.Bezombes



## 1 CONTEXT

---

### 1.1 Biodiversity erosion

Since the last century, the rapid development of technologies strongly impacted the human population growth and way of life. As a consequence, dramatic effects were observed on earth ecosystems, durably altering their integrity and functionality (Vitousek et al., 1997). Those impacts are the result of a combination of anthropogenic activities at multiple intensities, time and spatial scales (Gosselin and Callois, 2018), leading to the classification of 11 167 species as threatened, including 24% of mammals and 12% of birds species (IUCN 2002). The main causes can be divided into three categories:

- **Habitat fragmentation and alteration:** a vast majority of the landscapes are nowadays shaped by the anthropogenic activities, mainly due to the growing urbanization (Antrop, 2004), the intensive agriculture and the overexploitation of natural resources (Haddad et al., 2015). Those activities are inducing a change in landscape structure and composition, making it heterogeneous and fragmented. The fragmentation is characterized by two effects, habitat loss, and isolation of the remaining habitat in patches of various areas (Andren, 1994).
- **Invasive species:** the increase of exchanges between continents leads to the transfer of species outside their distribution range. Some of those species highly compete with native species, thus deeply impacting native ecosystems (Clavero and García-Berthou, 2005). Those invasive species concern a wide variety of taxa, from protozoa (toxoplasmosis (Aguirre et al., 2007)), insects (mosquitoes (Atkinson and LaPointe, 2009)), mammals (cats, (Medina et al., 2011)), reptiles (brown tree snake (Wiles et al., 2003), to plants (Japanese knotweed (Gerber et al., 2008))).



- **Climate change:** temperature increase in the late 20<sup>th</sup> century above the natural expected climate variation has been evidenced (Crowley, 2000). The emission of greenhouse gas is mainly responsible for this climate change. This temperature rise has multiple direct and indirect impacts on earth ecosystems. It may cause the change in the occurrence of extreme climatic events and species range shift (Easterling et al., 2000). In addition, the rise of the ocean temperature and its acidification is causing the coral reef bleaching (Hoegh-Guldberg et al., 2007) thus indirectly impacting greatly the fauna. Furthermore, climate change may induce deep changes in the water cycle (precipitation and snow cover) which may produce a shift in forest composition, productivity (Lindner et al., 2010), and impact greatly species adapted to mountains ecosystems (Imperio et al., 2013).

In this context of global biodiversity erosion, efficient and sustainable management actions are needed and encouraged to preserve ecosystems. However, managing ecosystems is a challenging task that has to take into account the multiple impacts described above and at the same time maintain the use of those ecosystems by the human populations (e.g. agriculture, forestry, leisure activities). In the specific case of endangered species conservation, the management actions are crucial and urgent as the number of individuals constituting those populations is usually very low making it difficult to reverse the current trends. Those actions can be of diverse types but usually involve projects such as the creation of a protected area, the restoration of a species habitat or the translocation of individuals to establish a new population (from another wild population or from captive reproduction program). The question of the choice of spatial location and of the area extent is inherent to all such projects. Efficiency and feasibility need thus to be carefully evaluated to ensure the success of the action. In particular, due to the limited resources available for such projects, actions types and locations must be prioritized in terms of time and cost (Moilanen et al., 2011). The prioritization process thus induces that choices must be made in accordance with the objectives and constraints. Mainly, two questions are raised to ensure best the success of the action:



- What type of action should be planned?
  - Long or short terms actions
  - Special skills or material evaluation
  - Actions efficiency
- Where this action should be applied?
  - Surface choice
  - Location (s) choice

The problem of spatially prioritizing the actions when the study area is large, to address both conservation and cost efficiency, is a challenge and is in the heart of this project.

### 1.2 Local context

In the Jura massif (France), the fate of two emblematic forest birds: the Capercaillie (*Tetrao urogallus*) and the Hazel Grouse (*Bonasa bonasia*) is depending for the first on future management actions and for the second on a better knowledge of the species population's dynamics and occurrences. Those two species are threatened by multiple causes at short and long terms. The major cause among those threats is habitat loss and fragmentation. Indeed, the two species have specific requirements regarding the forest vertical and horizontal structure and composition. Yet, the forests are mainly shaped by forestry practices, which have an important role to play in the future of the two species in the Jura massif. Indeed, the adaptation of the forestry activities to the species requirement may help to increase the availability of good quality habitat. If the Capercaillie populations' occurrences are widely known in the Jura massif and thus are already the target of conservation actions, little is known on the Hazel Grouse population's presence due to a wider potential area of occurrence and to its secretive behavior.



The two species populations are studied by the Non-Governmental Organization (NGO) Groupe Tétrás Jura. The work of Groupe Tétrás Jura is focused on four different components:

- **Population monitoring** to follow up the two species dynamics in the Jura Massif together with the ONCFS (National Hunting and Wildlife Agency)
- **Habitat quality appraising** in the context of forestry activities that can affect the populations at short-term
- **Public outreach** nearby both residents and tourists
- **Leading activities aiming to protect the bird populations**

The opportunity to use two Light Detection and Ranging (LiDAR) datasets that cover a major part of the distribution of the two species in the Jura massif initiated the project to map for each species the quality of the habitat with the aim to support future forest management actions in favor of Capercaillie and future survey protocol for Hazel Grouse. Thus, this PhD project aims to propose habitat suitability predictions over the study over for each species with the hope to better plan management or survey actions.

### 1.3 Species Distribution models: a tool for management

Species Distribution Models (SDMs or often referred to Habitat Suitability Models or Ecological Niche Models in the literature) are widely used with the aim to support conservation actions (Franklin, 2009a). Indeed, such models describe the relationship between a species observations and environmental variables, allowing us to get information on the response to the environment and to spatially predict the habitat suitability under current or future conditions (Elith and Leathwick, 2009; Franklin, 2009b). The rise of SDMs as tools for species conservation is highly correlated with the emergence of remote sensing, that provides information on the landscape cover characteristics over large scales (Bradley and Fleishman, 2008). Thanks to those attributes, the used of SDMs can support decisions at multiple stages and for different types of conservation projects:



- **Identifying the environmental niche of the target species.** This can lead to decisions such as essential habitat restoration and protection (Angelieri et al., 2016; Conde et al., 2010)
- **Proposing new sites for future surveys of rare species** (Guisan et al., 2006; Marcer et al., 2013)
- **Selecting future protected area locations for one or multiple target species** (Arcos et al., 2012)
- **Identification of suitable areas for species translocation** (from wild populations or captive reproduction) (Schadt et al., 2002)
- **Identifying the risk of biological invasion and target areas for pest control** (Ficetola et al., 2007)

### 1.3.1 The problem of the transferability of the results

Thanks to those characteristics SDMs have already been successfully used to ensure the realization of different conservation actions showing the high potential of those methods (Guisan et al., 2013). The case of Madagascar is representative of a successful use of SDMs to target new areas for conservation (Kremen et al., 2008). Indeed, multi-taxa SDMs were created leading to the spatial prioritization for future protected areas. It was followed by a legal degree that outlaws mining and forestry activities in those areas. However, despite the increasing number of studies aiming to support conservation actions, only 1% of the published papers reviewed by Guisan et al. (2013) were truly addressing conservation decisions. This gap between research results and the realization of concrete conservation actions is an established problem which is not specific to the use of SDMs (Arlettaz et al., 2010; Knight et al., 2008). It is characterized by the fact that the knowledge produced by research regarding models outcomes is not turned into actions due to differences between the two communities (objectives, work daily activities, career). This incomprehension from managers was classified into three main categories by Addison et al. (2013):



- The role of models in decision making:
  - Models are not needed, expert knowledge is sufficient
  - Lack of resources to lead such projects
- Modeling practices:
  - The manager conceptual understanding of the context is not represented by models
  - Models do not capture social and economic factors
  - Models are too complicated or too simple
  - There are no sufficient or adequate data for modeling
- Model outputs:
  - Scientist communication is not well understood by managers
  - Model outputs are too uncertain to be useful

In this thesis, the need for models initially came from managers, thus the role of models in decision making was not negated.

However, to ensure the success of the project, a contribution was needed in order to obtain models that will be reliable and understood by managers. Three key points were thus identified and constructed the framework of the thesis.

- **Models should be reliable, even with the use of heterogeneous observation datasets**
- **Models should fit the ecological conceptual understanding of managers and be easy to interpret**
- **The communication of the results should be accessible to everyone**

### 1.3.2 Relevant modeling practices

Models outcomes are depending on multiple factors and the challenge is to identify the fundamental insight that will lead to uncertain and biased results. Three main issues are generally highlighted:

- The choice of environmental variables (Fourcade et al., 2018; Johnson and Gillingham, 2005)
  - If one important variable for the species is not in the models, major causes influencing its distribution may be missed
  - Ecologically irrelevant variables may still be found to be influencing the species distribution
- The characteristics of the species datasets (Cardador et al., 2017; Phillips et al., 2009)
  - Biased datasets (in space or time) may lead to poor model predictions, representing more the survey characteristics than the species distribution.
- The choice of spatial or temporal scales (Arkle et al., 2014; Graf et al., 2005)
  - Individuals are influenced by their environment at multiple spatial and temporal scales, thus missing the effect of one major variable at a particular scale may lead to poor models predictions and performances.

Those uncertainties can be measured using different models' evaluation methods, thus giving an estimate of the reliability of the results. However, this efficiency remains questionable as demonstrated by (Fourcade et al., 2018), where models based on non-biological variables (painting) were evaluated to be as good as models based on realistic environmental variables, showing that the initial choice is critical. Similarly, the choice of spatial and temporal scales should be based on previous knowledge on the species ecology in order to give relevant results. The influence of the species dataset characteristics (sample size, sampling spatial bias, presence only data versus presence-absence data) on models uncertainties has been widely demonstrated (Kramer-Schadt et al., 2013; Reddy and Dávalos, 2003). Nevertheless, a large part of the studies fails to take this into





account (Yackulic et al., 2013). Thus, the conception of reliable models is essential to support conservation actions and but requires that a careful attention should be given to those three aspects.

As a consequence, the first steps of this work were consecrated to the choice of the environmental variables and spatial scales and to the evaluation of the sampling bias. Taking the opportunity given by the specificities of the datasets the exploration of new insight was conducted. This will be described in details later (Chapter 2) due to their high dependency on the general context, the species ecology, the aims of the project and datasets characteristics.

### 1.3.3 Conceptual understanding

To fill this gap between scientists and managers, closer collaboration between researchers and decision makers should be promoted, with the aim for researchers to have a clearer vision of the managers' needs and of the specificities related to the targeted species and areas (Villero et al., 2017). Elements such as expected outcomes or future intent actions could be discussed as a first step in order to improve the transferability of the results to practitioners (Guisan et al., 2013).

The use of environmental variables that will be pertinent for managers and that will meet their conception of the problem may be a main contribution to models transferability improvement. In many studies, the use of environmental variables such as climatic condition (rain, temperature), surface cover type (forest, grassland and wetland) is not problematic, and easily understandable by everyone, but variables extracted from remote sensing datasets are not always easy to interpret. Indeed, the use of ratio or index could be obscure and often does not fit the perception of the environment by managers. Thus, for this project, the proposition of LiDAR environmental variables that will be easily interpretable was a fundamental point. Indeed, in the field of conservation biology, LiDAR extracted metrics used in SDMs were so far calculated to our knowledge using the conventional point-cloud area-based method which may not be easy to interpret and used by managers at the operational scale.



### 1.3.4 Communication

The use of a “translator” between research and managers that will clearly communicate the contribution of scientific knowledge has been proposed to improve the use of models results in conservation actions (Guisan et al., 2013). However, this transmission to an audience that often does not have the necessary background to fully understand the strengths and limits of the methods is a challenge. As the main results of this work aim to be published in scientific articles, additional reports targeting managers were produced, with the aim to improve the future use of the model results. In addition, a restitution of the results to managers from the National Forestry Office (ONF), the National Hunting and Wildlife Agency (ONCFS), National Natural reserve of the Haute Chaine du Jura (RNNHCJ), and Groupe Tétrás Jura is also programmed. This restitution aims to show a concrete example using the model results to answer to manager’s interrogations and to help them to handle the produced files.

### 1.4 Synthesis

The use of SDMs in conservation biology is still to be improved, and the main aspects that need further improvement can be divided in two categories: (1) implement relevant models regarding the datasets provided for the study, (2) improve communication and facilitate the understanding of the results by managers.

Both components are thus highly dependent on the data available and of the objectives defended by managers. As a consequence, the final objectives developed in this thesis can only be defined after the evaluation of the specificities of the datasets and of the target species characteristics. Hereafter, the use of LiDAR environmental variable in SDMs will be presented in a first part and, the two problematics surrounding the two target species will be introduced in a second part. Finally, the definition of the objectives and hypothesis development in this thesis will be presented.



## 2 REMOTE SENSING: THE RISE OF LIDAR

---

### 2.1 Remote sensing for SDMs

As previously stated, the choice of pertinent environmental variables is one of the first critical step encompassing SDMs. The necessity to obtain results that cover large spatial extent made the use of remote sensing data indispensable as field data cannot be collected at such scales. Remote sensing is defined as the process of monitoring physical characteristics at a distance from the targeted object. Those methods can be passive or active and comprise satellite or aerial imagery, RADAR and LiDAR. Satellite imagery has been used to map vegetation cover since 1970, thanks to spectral, temporal and textural differences between landscape elements. It was later used with success as an environmental variable in SDMs (Bradley and Fleishman, 2008). More recently, due to its recent development LiDAR is emerging as a powerful tool for SDMs.

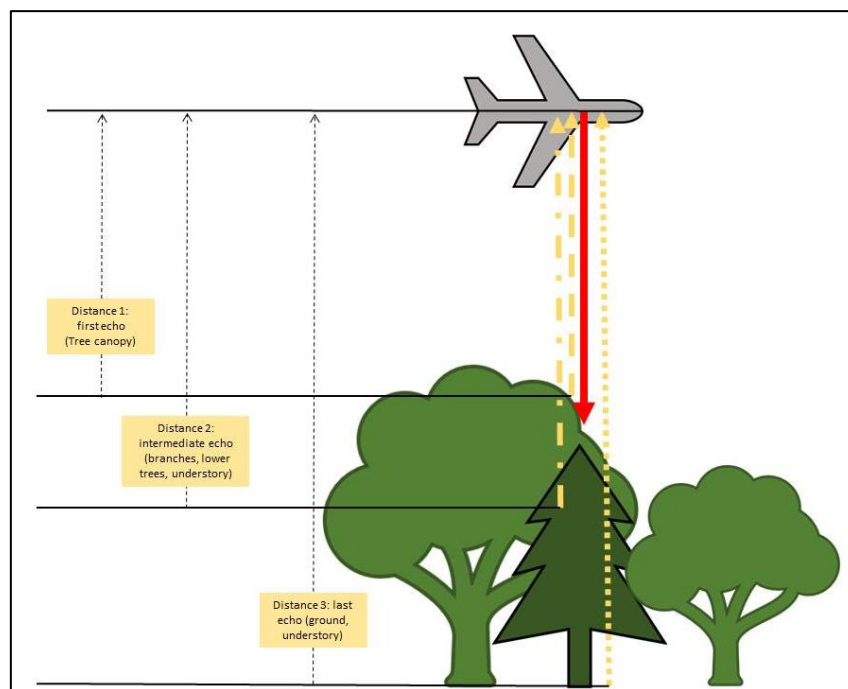
### 2.2 LiDAR: history and definitions

LiDAR is a remote sensing technique that can calculate the distance between an object and a device, by emitting a laser pulse (near-infrared 0.8 to 1.55 $\mu$ m of wavelength) and measuring the time passed between the impulsion and the return signal after being reflected by the object. The first LiDAR systems were created in 1960-1970. At the beginning only ground base system was used but soon with the main progress of GPS spatial positioning, the system was boarded in aircrafts or satellites to obtain images of the earth surface (Figure 1). It was also used for ocean floor cartography (bathymetry). LiDAR datasets can provide the coordinates and altitude of a part of an intercepted object and by a high pulse frequency a three-dimensional point cloud of this object can be obtained. The first commercial sensors in 1990 were characterized by a pulse frequency of 10kHz (Baltsavias, 1999). Today, the pulse frequency can be up to 400kHz. Thus, we can obtain datasets with a high number of points per square meters. At first, only one return or echo was recorded for each laser pulse, but as they may be multiple objects on the laser trajectory, more than one signal can be



reflected from a single emitted pulse. Multi-echoes systems were then developed to store more than one echo, usually the first and the last one (Mallet and Bretar, 2009). This characteristic is especially useful in a forest environment where a first echo will typically be reflected by the tree canopy top, while a part of the signal continues below the leaves and branches and potentially giving additional echoes from the under-canopy. Ultimately if the vegetation is not too dense (ideally during the leaf-off season) the last echo will represent the ground. In addition, laser scanning systems can now record full-waveform signals, by sampling the reflected signal with a high frequency of 1Ghz (Mallet and Bretar, 2009) giving a continuous and precise return signal.

*Figure 1 : LiDAR system measurement*





The main components of LiDAR operating systems are:

- a laser system (laser, transmitting and receiving optics)
- a position and navigation system (Global Navigation Satellite System receiver and Inertial Measurement Unit)
- a data storage device

Once the data is collected a post-processing phase is necessary. This part aims to calibrate GPS positions (using ground reference points) and to detect aberrant points. Afterwards, a classification of the echoes between categories such as vegetation, ground, and buildings can be done. At this step, a digital terrain model can be created using the points classified as ground. Generally, the precision of the point data is of 25-30cm horizontally and 10-15cm vertically.

Nowadays, LiDAR is used in multiples fields such as archeology, ecology, topography, geosciences, and meteorology.

### 2.3 LiDAR in ecology and conservation biology

Due to their high precision and to the always wider coverage availability, LiDAR data are becoming a powerful tool. It has been used in multiple ecosystems studies from tropical (Clark et al., 2004) or temperate forests (Müller et al., 2017), to urban areas (Klingberg et al., 2017) or coral reefs (Purkis, 2018). The three dimensional (3D) characteristics of LiDAR allow us to study the structure of the environment both vertically and horizontally. In addition, this can be done from a very fine scale (e.g tree individuals) to larger scales (e.g some countries like Switzerland or Denmark are entirely covered by LiDAR). LiDAR shows a high importance when studying forest ecosystems. Indeed, the 3D structure of the forest is difficult to evaluate by field measurements in particular when working on the canopy characteristics. At first, LiDAR data was mainly used for ground topography (under dense vegetation cover), overall 3D structure measurement, forests ecosystems functionalities, and prediction of forest stand attributes (Lefsky et al., 2002). Thus, applications in forestry were soon found to be common and new developments involving the extraction of objects such as trees and



associated variables (location, crown volume estimation, and species identification) using machine learning algorithm were used as forest's inventories complement, (Dassot et al., 2011). Other useful information can be calculated over a large area such as the timber volume, the density of trees, the basal area and the diameter at breast height estimations. All the variables presented can have direct application in the field of forest management (Dassot et al., 2011).

Forest ecosystems cover a large part of the landscape thus containing a large proportion of the wildlife present on earth. The characteristics of the forest ecosystems outside the abiotic conditions are mainly driven by the plants' species occurrences. The species assemblage specificities through close interactions with animals are shaping their distributions. However the plants species constituting the forest are not the only factor influencing other species distributions pattern: at the local scale, the forest structure influence was highlighted to be an important aspect as well. Indeed, the different 3D structures provide a diverse availability of micro-habitats that can have a direct or an indirect effect on individuals (Milling et al., 2018). The variation of height and vegetation density influences many other parameters such as light inflow to lower layers, local temperature or humidity. Thus, highly structured landscape supports the presence of a large diversity of niche affecting directly the ecosystem biodiversity. The 3D structure also influences animal's behavior with structures that can provide shelter from predators (or not), obstruct or improve movement or vision and cause variation in food availability. The impact of 3D structures was studied for many taxa including flying animals birds, bats, invertebrates (Fabianek et al., 2015; Froidevaux et al., 2016; Weisberg et al., 2014) and also non-flying mammals (e.g primates) (Palminteri et al., 2012). Flying animals such as birds and bats move in three dimensions, and thus many studies integrating vegetation structure were developed for those species. Indeed, the structure of the vegetation plays a large role in their ecology, from movement behavior to micro-habitat selection. Parameters such as vertical distribution diversity, canopy cover variations, vegetation heights, understory density were found to have an impact on flying species presence and global biodiversity (Flaspohler et al., 2010; Goetz et al., 2007; Müller et al., 2010; Vogeler et al., 2014). The unique characteristics of LiDAR which have demonstrated a potential high importance when studying spatial occupancy of flying species making



it a powerful tool for modeling habitat suitability of endangered avian species (Bergen et al., 2009; He et al., 2015; Tattoni et al., 2012a). As Capercaillie and Hazel Grouse are two avian species associated with forests ecosystems, LiDAR is a promising tool for describing important structural components for each species.

### 3 TWO EMBLEMATIC SPECIES FROM THE JURA MASSIF: A NEED FOR RELIABLE SPATIAL PREDICTIONS

---

#### 3.1 The Galliform order

The galliform order contains species distributed among five families: *Numididae*, *Odontophoridae*, *Phasianidae*, *Megapodiidae* and *Cracidae*. The family *Phasianidae* regroups multiple sub-families such as pheasant, quails, pavoine and grouse. The grouse sub-family, also known as *Tetraonidae*, is represented by 19 species, occurring in Europe, Asia and North America. Those birds are all adapted to cold winters with various strategies and are known to favor walking for their daily movement. Among the species, some are highly distinctive by their weight and size, as the greater sage grouse (*Centrocercus urophasianus*) or the Capercaillie. In France, four species can be observed, the Capercaillie, the Black Grouse (*Tetrao tetrix*), the Rock Ptarmigan (*Lagopus mutus*) and the Hazel Grouse.

#### 3.2 Capercaillie

##### 3.2.1 Species description

Two sub-species are documented in France *Tetrao urogallus major* and *Tetrao urogallus aquitanicus*. During the last Glacial maximum (-0.2My) the species retreated to South Asia. After this event, the population expanded again through Europe replacing and pushing the *aquitanus* clade to the south of its distribution range (Duriez et al., 2007). The Capercaillie is a highly dimorphic species where the males or cocks weight between 2.6 and 6kg and the female or hens only 1.5 to 2.2kg for the



*major* sub-species (Figure 2). The sub-species *aquitanicus* is smaller, male weight around 2.6 to 4.2kg and female 1.2 to 2kg. Their plumage is also different between sexes. Males plumage is composed in majority of black feathers, with brownish reflect on wings and greenish colors on the chest. Cocks also have a red eyebrow and a particular fan shape for their tail feathers. Females have a more cryptic plumage composed of brown with white and black barring. The coloring is more orange on their chest. Both cocks and hens have feathers on their legs to protect them from the cold in winter. The plumage for the *aquitanicus* sub-species shows some variation where males have darker feathers and female have a less orange chest than the *major* sub-species (Couturier 1980). This species is known to walk a lot in the forest, thus they have along their toes small horns tacks which provide snowshoes ability in winter.

*Figure 2 : Male and female Capercaillie*



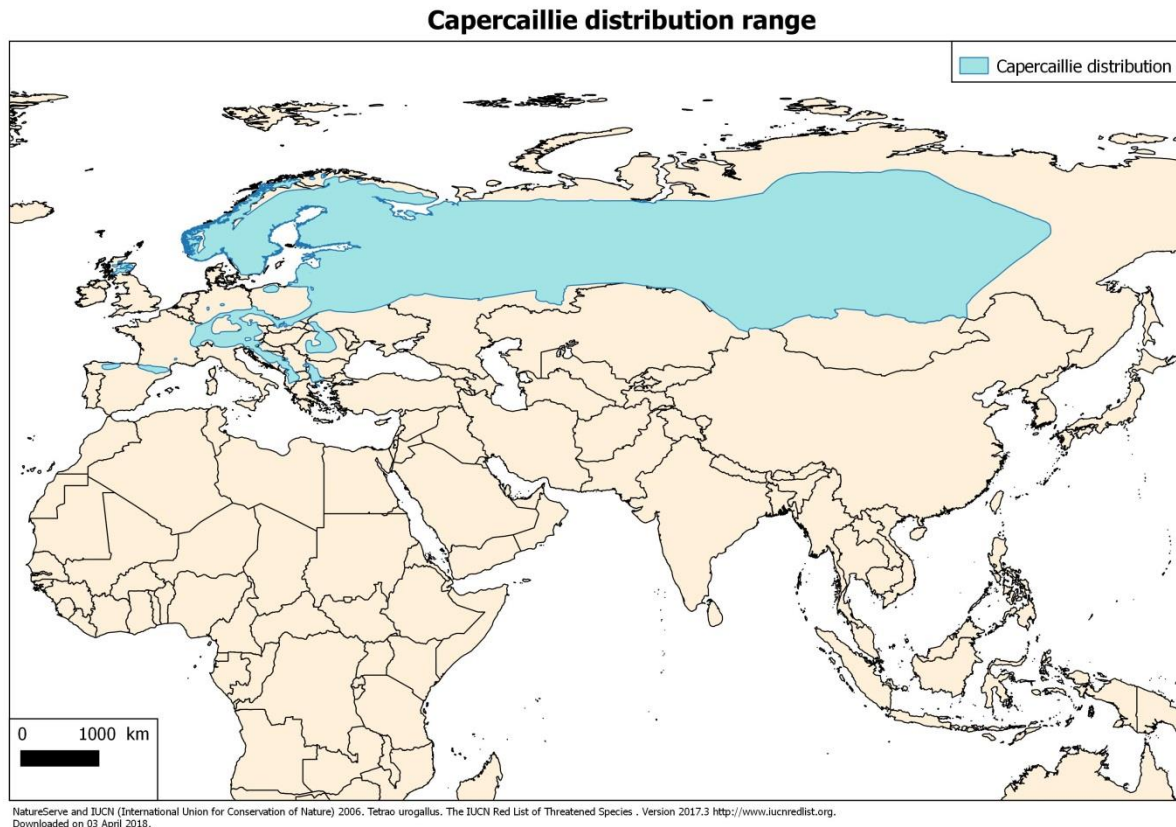




### 3.2.2 Areas of occurrences

*Tetrao urogallus major* is found from the Jura massif at its western range limit of occurrence, to the Scandinavian boreal forest, the Oural Mountains in Russia and Mongolia (Figure 3). A reintroduced population is also occurring in Scotland.

Figure 3 : Capercaillie distribution range

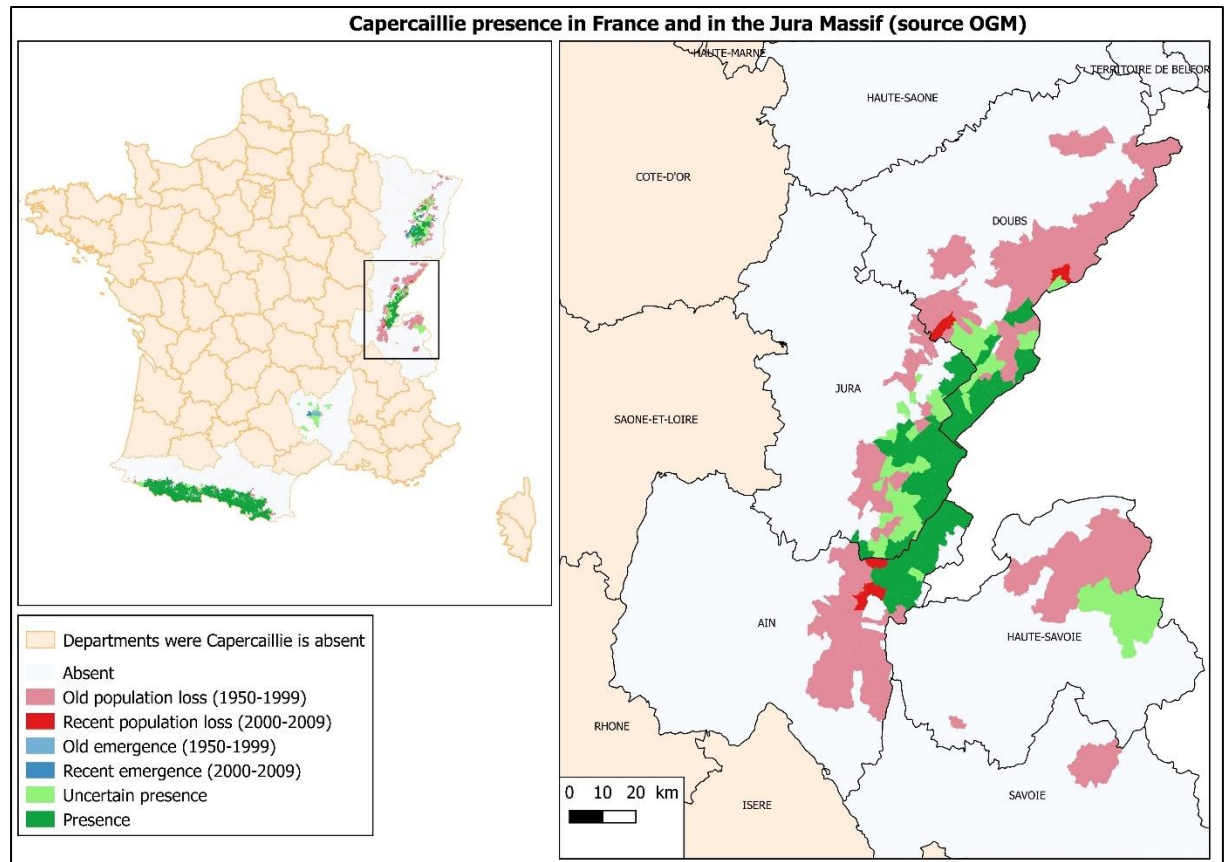


In France, four remaining population are known (Figure 4). The unique population of the sub-species *Tetrao urogallus aquitanicus* is found in the Pyrenean Massif (4000 individuals). In France, the three other populations belong to the sub-species *Tetrao urogallus major*, and are found in the Cevennes (reintroduced, 50 individuals), the Jura (500 individuals in the whole massif France - Switzerland, 250-340 in France) and the Vosges massifs (100 individuals)<sup>1</sup>.

<sup>1</sup> <http://www.observatoire-galliformes-montagne.com/Grand-Tetras.html>



Figure 4 : *Capercaillie* presence in France with a zoom in the Jura massif study area (OGM, 2009).



### 3.2.3 Reproduction

Capercaillie is a polygamous species. Social interactions are important in the reproduction process due to the distinctive mating behavior of Capercaillie. Indeed, males meet at the display place (lek) in late winter (around April in Jura Mountains), usually in the early morning and late evening before sunset. Thus, their territories are usually located within 1 km around the lek, allowing them to come back every year to the same place to display and mate (Wegge and Rolstad, 1986). Because of this social behavior, bird (both male and female) density might seem higher closer to the lek places during that period. Those places are often characterized by a low canopy cover or clear-cuts, roosting trees should be available as well. Females join the display place after the two first weeks of parades and will choose the dominant male among all the present cocks. Females are nesting on the ground and raise the chicks through late spring and summer alone.



### 3.2.4 Habitat needs

French Capercaillie populations represent the south-east border of the species range. Consequently, birds are living in forests quite different from the usual boreal forest where the major part of the worldwide population is found, showing a form of plasticity concerning their habitat selection. In the Jura Massif located in both France and Switzerland, Capercaillie is found in mixed beech-fir forests (Vittoz, 1998) and spruce-lapiaz forest between 800 and 1500 m (Figure 5). As the species overwinter their habitat requirement differs according to his seasonal needs. Capercaillie has large habitat requirement as its home range size is an average of 550 ha (Storch, 1995) but is more likely to be 150 ha in the South of Europe and in the Jura Massif Capercaillie home range is known to be even smaller: 60 ha (A. Depraz pers. comm.). This home range should provide habitat suitable for each seasonal need. The knowledge of the bird all-year-round habitat requirements is essential when it comes to species conservation and habitat management. Thus, thereafter, the knowledge from previous studies and expert knowledge for each season are presented.

*Figure 5 : Habitat of good quality for Capercaillie*



Capercaillie developed a behavioral adaptation to survive to challenging meteorological conditions in winter. This wintering behavior is essential as it attempts to minimize heat and energy loss as well as predation risk. Indeed, individuals are reducing energy expenditure by slowing their daily activities actions (2-3 hours per day) (Gjerde and Wegge, 1989a) and by avoiding unnecessary movements such as walking and flying. Thus, individuals select a few trees where they will feed (feeding trees) and perch for resting (roosting trees) during the whole winter season and at the same time minimizing the distance they have to fly or walk between



feeding and roosting trees. During this season, birds feed exclusively on abundant conifer needles in the majority from Scots pine (*Pinus sylvestris*) if available or from other conifers, like *Pinus uncinata* or European silver fir (*Abies alba*). However, roosting trees are usually deciduous trees and in the Jura Mountains, Capercaillie uses mainly beech trees (Montadert, n.d.). Furthermore, birds select trees that allow a good detection of predators and a safe way to escape by choosing large trees with one or more aisles (Thiel et al., 2007). This observation was also made by (Sachot et al., 2003a) in western Switzerland. Indeed, they observed that Capercaillie was avoiding dense canopy cover and more globally young forest stage (such as regeneration, thickets, pole stage, spruce cover, and understory beech), preferring high canopy and under-canopy fir as well as grazed forest areas in winter. We also have to keep in mind that habitat range is also influenced by habitat preference in other seasons and by social interactions (display places) (Storch, 1993a).

Spring and summer are corresponding to the nesting and rearing period. Thus, females and chicks require specific habitats with furnished herbaceous and good ground cover with high bilberry (*Vaccinium myrtillus*) density (Figure 6). This particular habitat has multiple advantages. First of all, this allows a lower risk of nests and brood predation. The ideal height of ground vegetation for capercaillie is 40 cm (Storch, 1993b), allowing them to watch for predators and still be hidden by the vegetation. Nests are usually located at the edge between old forest stand and clear-cut where the density of ground cover is particularly high (Storch, 1994). Secondly, chicks, for the first four weeks of their life, feed at 80% on insects and larvae (Wegge and Kastedalen, 2008) which are found in abundance in those habitats type. Later in the season, bilberry fruit becomes the main food component for four-weeks-old chicks and adults (Storch, 1993b).





*Figure 6 : Ideal habitat with a dense bilberry cover*



In autumn, both young and adult feed on fruits, buds and leaves. During that season bilberry are an essential part of the bird's diet and reach a peak in September (Storch, 1993b). Therefore, in autumn Capercaillie are present in high density in this habitat. Bilberry is essential because it allows the birds to create fat reserves in order to go through harsh winter. This period is also important because young birds are becoming independent and start to disperse within the forest to set up their first territory. This supposes the presence of suitable, un-occupied and reachable habitat for each young bird. Agricultural land, urban area as well as topographic barriers (such as ridge) are known to limit the dispersion of the birds. The main dispersal distance is 10 km, but some individuals were observed to disperse to longer distances (Segelbacher et al., 2003).

### 3.2.5 Protection status

The species is classified as Least Concern worldwide and in Europe<sup>2</sup> (Storch, 2007). However, the species is classified as vulnerable in France by the *Liste rouge des oiseaux nicheurs de France*

---

<sup>2</sup> <https://newredlist.iucnredlist.org/species/2267948>



*métropolitaine* (Red list of metropolitan French nesting birds<sup>3</sup>). The Capercaillie is listed in the Bird Directive annexes I, II and III which states the bird species for which conservation actions are taken or needed. The species is also listed in the Bern Convention (1979), which was approved by several countries including France. The Capercaillie is also present in the annexes I of the Habitats Directive from the European Natura 2000 network which aims to implement a special protection status on its habitat.

### 3.2.6 The causes of the population decrease in France

A continuous reproductive decline of Capercaillie has been observed over the past 80 years due to multiple causes in Europe and France (Jahren et al., 2016; Storch, 2007). French populations are small and fragmented within and in different mountains chains, making the impact of any disturbance critical.

The first threat to Capercaillie is habitat loss, fragmentation and degradation. The evolution of the forestry practices over time influences greatly the proportion of suitable habitats. Indeed, those practices determine the structure and the composition of the forest stand. Young forest stand and monospecific plantation are known to be unsuitable for the bird. Basically, forestry actions that favor a closure of the vegetation are detrimental for the populations. Habitat loss in the Jura Mountains is mainly linked to a change in forestry practices. The forest cutting evolved from a traditional uneven-age management to regular cutting and a shorter time rotation of forest stands, inducing a loss of structural diversity in the forest stands (Storch, 2007). In addition, the intensification of the forestry activities leads to a rejuvenation of the trees. As a consequence of large clear cutting, an intense beech regeneration is observed as well (Vittoz, 1998). Thus, the understory cover once suitable for bilberry growing (essential for Capercaillie) is becoming more and more closed and unfavorable. Finally, the abandonment of the forest cattle grazing is inducing the recolonization of dense herbaceous vegetation areas by fir and beech trees regeneration, losing a part of the Capercaillie habitat component.

---

<sup>3</sup> [https://inpn.mnhn.fr/docs/LR\\_FCE/UICN-LR-Oiseaux-diffusion.pdf](https://inpn.mnhn.fr/docs/LR_FCE/UICN-LR-Oiseaux-diffusion.pdf)



Another type of threat on Capercaillie populations is the human disturbances intensity. The recent boom of outdoor recreational activities such as ski (cross country and alpine), snowshoe in winter and hiking, biking in spring and summer has deleterious effects on birds individuals (Coppes et al., 2017). This is particularly true in winter when the conditions are harsh making the energy loss while escaping high. The intensity of the disturbances induces an increase of the flush distance and physiological stress response (Thiel et al., 2007, 2008). In addition, alpine ski infrastructures provoke deadly collisions with ski-lift cables which are thought to be non-negligible (Buffet and Dumont-Dayot, 2013).

The predation is also a threat to the Jura massif population. Historical Capercaillie predators are the fox (*Vulpes vulpes*), the marten (*Martes martes*) and the northern goshawk (*Accipiter gentilis*), which usually target chicks and young birds. Nowadays, wild boards are present in mountains forests where they were not observed before. It is thought that wild board predation pressure on nests and young chicks increases greatly when their density increases (Oja et al., 2015). Predation is an important factor of pressure on population dynamics as 90% of juvenile deaths were observed to be induced by predation (Wegge and Kastedalen, 2007).

Last but not least, climate change may have an impact on the Capercaillie future distribution. One point is the change of forest composition along the altitudinal gradient, which may induce the emergence of a higher proportion of beech at higher elevations. The second point is the temporal shift between temperatures increasing in spring associated with vegetation growth and the laying and chicks rearing period. The failure of the bird to be synchronized with the bilberry growth and invertebrates abundance peak may lead to lower reproductive success (Moss et al., 2001; Scridel et al., 2018). Also, further work on Capercaillie response to climate change is needed. The effect of increasing occurrence of rainfall event in spring is also believed to decrease chick's survival in their first weeks of life (Moss, 1986).



### 3.2.7 Conservation actions in the Jura Massif

The conservation actions implemented in the Jura Massif in favor of the Capercaillie are classified into two categories:

- The protection of the habitat
  - Protected areas are constituted of the National natural reserve of Haute Chaîne du Jura and four prefectural orders for biotope<sup>4</sup> protection covering 32% of the estimated regular occurrence areas
  - Habitat restoration in collaboration with the Forest National Office (ONF), involving the clearing beech regeneration stands.
- The control of anthropogenic disturbances
  - Capercaillie tranquility clause restricts forestry activities to the July-November period outside the critical Capercaillie life stage (wintering and reproduction).
  - Installation of visualization system on ski lift cables
  - Control of off-trail and off-ski activities in winter

Among all those actions, this work aims to help managers with the protection of the Capercaillie habitat, in particular, to help the decision for future habitat restoration and forestry activities management.

### 3.2.8 LiDAR variables for Capercaillie habitat model

LiDAR was already used in three Capercaillie studies mainly for conservation purpose. The identification of highly predictive LiDAR variables and of habitat quality among a study region was proposed. The choice of the variables was made using the a priori knowledge of the habitat need of the species. This can depend on the study area (boreal or temperate mountain forest), on the season or on the sex, age of interest. The extracted variables aim to capture both vertical and horizontal

---

<sup>4</sup> [http://www.parc-haut-jura.fr/fr/rubriques-editoriales/bonnes-pratiques/concilier-pratiques-hivernales-et-environnement.882-887\\_\\_1915.php](http://www.parc-haut-jura.fr/fr/rubriques-editoriales/bonnes-pratiques/concilier-pratiques-hivernales-et-environnement.882-887__1915.php)





structure variations in the environment. Graf et al. (2009a) used LiDAR data for habitat suitability models in a forest reserve of 17.7km<sup>2</sup> in Switzerland. The variables used for the models are presented in table 1. The resulting models showed moderate performance. A LiDAR only model was compared to a LiDAR and satellite extracted image variable where the latter was showing the best performance. Main contributing variables were the relative tree canopy cover (both linear and quadratic responses), indicating that horizontal variations are important for Capercaillie. A more recent study in central Europe used LiDAR as well as other predictors for Capercaillie as well. Zellweger et al. (2013a) found that vertical complexity variables were accounting the most for the model contribution (mainly the standard deviation between the 2<sup>nd</sup> and 3<sup>rd</sup> height quantiles). Models showed an overall good performance. The last study, in boreal forest, also used LiDAR variable to determine Grouse brood occurrences. They showed that brood presence was positively related to shrub and canopy cover.

LiDAR was used to improve Capercaillie habitat models with success. Those studies showed different results concerning variable contribution, sometimes showing the importance of horizontal heterogeneity and in some other study the importance of vertical heterogeneity.

*Table 1 : LiDAR variables and their description (Capercaillie models)*

Study	Ecosystem	LiDAR variables	Description
Melin <i>et al.</i> 2016	Boreal forest	Shrub cover	A shrub_cover value of 50 means that from all the echoes reflected below five meters 50% came from vegetation, not ground. The variable describe about the amount of vegetation between the heights of 0.5 and 5 m
		Canopy cover	A canopy_cover value of 70 means that 70% of all the echoes reflected above five meters. The variable describes the amount of echoes were vegetation is above 5 m of canopy cover
		Canopy height	A canopy height value of 22.5 means that 95% of the echoes are reflected below the height of 22.5 m. The value of this variable indicates the maximum canopy height, which in the example case would be 22.5 m
Graf <i>et al.</i> 2009	Alpine forest	Relative fraction of tree canopy	Calculated from aggregated height information based on Canopy height model (height >3m) by applying a moving window of 125 m *125 m
		Mean and standard deviation of tree canopy height	Calculated from aggregated height information based on Canopy height model (height >3m) by applying a moving window of 125 m *125 m
		Focal mean of relative tree canopy	Calculated with a moving window of 25 ha
		Density of tree canopy edges	A contour length per raster cell on Canopy height model at the heights of 10 m, 20 m and 30 m
Zellweger <i>et al.</i> 2013	Jura Mountains, Pre-alpine forest, Alpine forest	Standard deviation of quartile	A density of points for the 1th, 2nd, 3rd, 4th, 5th, 7th, 8th height quantiles
		Multi-storied profile	An Index representing the vertical distribution range within the 2nd and 3rd height quartile.
		Patch area	Mean patch area for understory, midstory and canopy CHM class.
		Forest gap density	Sum of forest gaps, i.e. non-forest patches <1 ha, divided by total forest area in ha.
		% of landscape	Percentage of the landscape (1 km <sup>2</sup> ) covered by patches of understory, midstory and canopy CHM class
		Clumpy	A clumpy index for patches of understory, midstory and canopy CHM class.
		Nearest neighbor	Mean nearest neighbor distances for patches of understory, midstory and canopy CHM class.
		Edge density	Density of edges (pixel transitions) between two CHM classes, i.e. non-forest/understory, non-forest/midstory, non-forest/canopy, understory/canopy.
		Canopy height heterogeneity	Total length of edges (pixel transitions) between understory, midstory and canopy CHM
		Topographic position	Mean and standard deviation of the Topographic Position Index. Measures the exposure of a site in relation to the surrounding terrain.
		Slope	Mean and standard deviation of slope

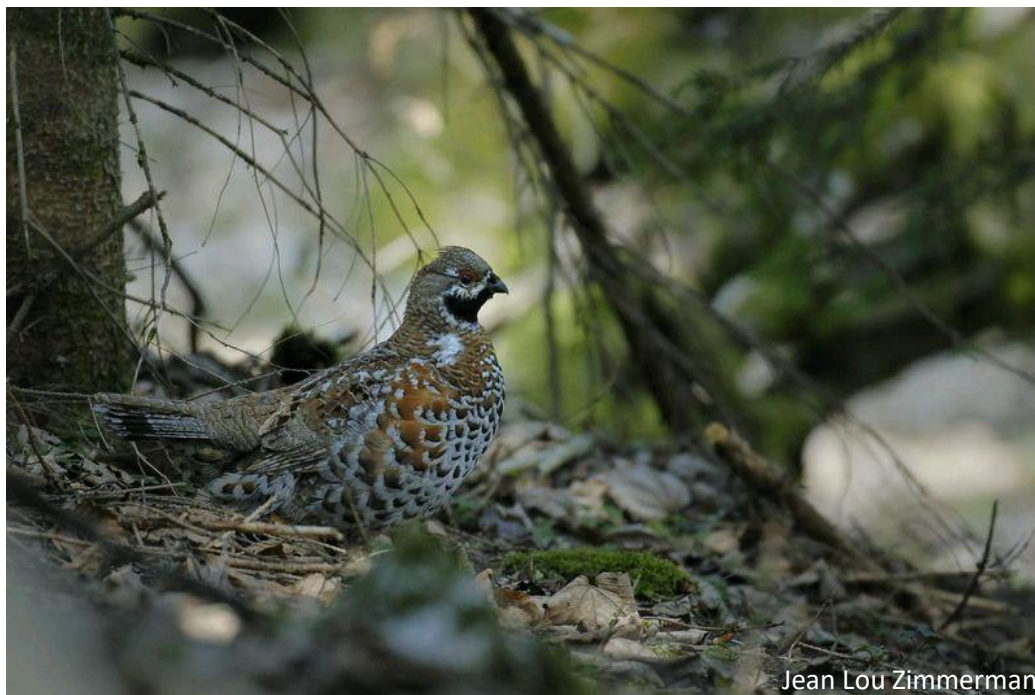


### 3.3 Hazel Grouse

#### 3.3.1 Species description

Hazel Grouse is a secretive and cryptic forest bird and the smallest representative of the Tetraonids. Adult's weight is around 400 grams. The plumage is composed of a fine pattern of brown, black and white. Wings are mostly brown while the chest is whitely flicked with black and brown. The tail is characterized by a large black strip. A sexual dimorphism is observed between the plumages of each sex. Males have a short erectile crest and a white-bordered black throat, which is not found for females (Figure 7). As Capercaillie, this species is adapted to harsh winter conditions with along their toes small horns tacks and after-feathers (hyporachis) to survive to negative temperatures. Individuals life expectancy is of 3 to 4 years also a large part of the juvenile will not survive their first year.

*Figure 7 : Male Hazel Grouse*

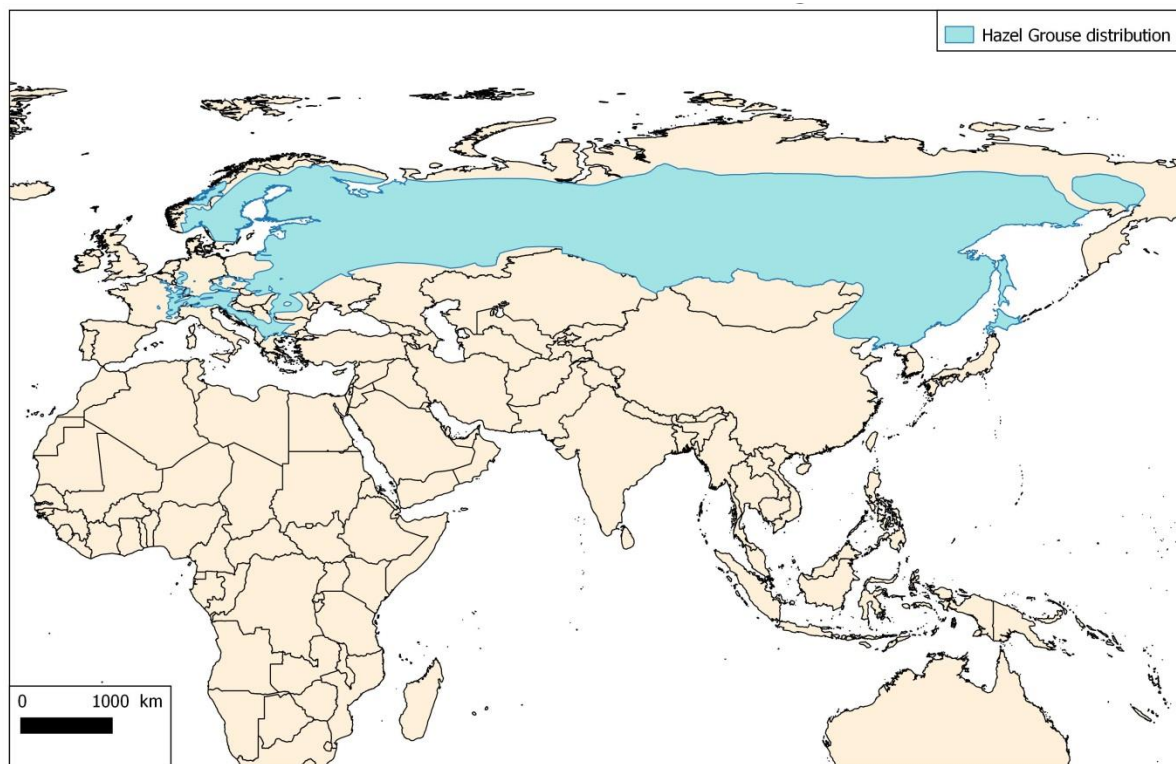




### 3.3.2 Area of occurrence

The Hazel Grouse populations are widely distributed from Europe to Asia. The populations present at the east of France are in the western border of the species range. Hazel grouse is present in Scandinavia, in the Balkan countries, and a major part of its occurring range goes through Russia (Figure 8). At the eastern border, populations are present in Hokkaido (Japan).

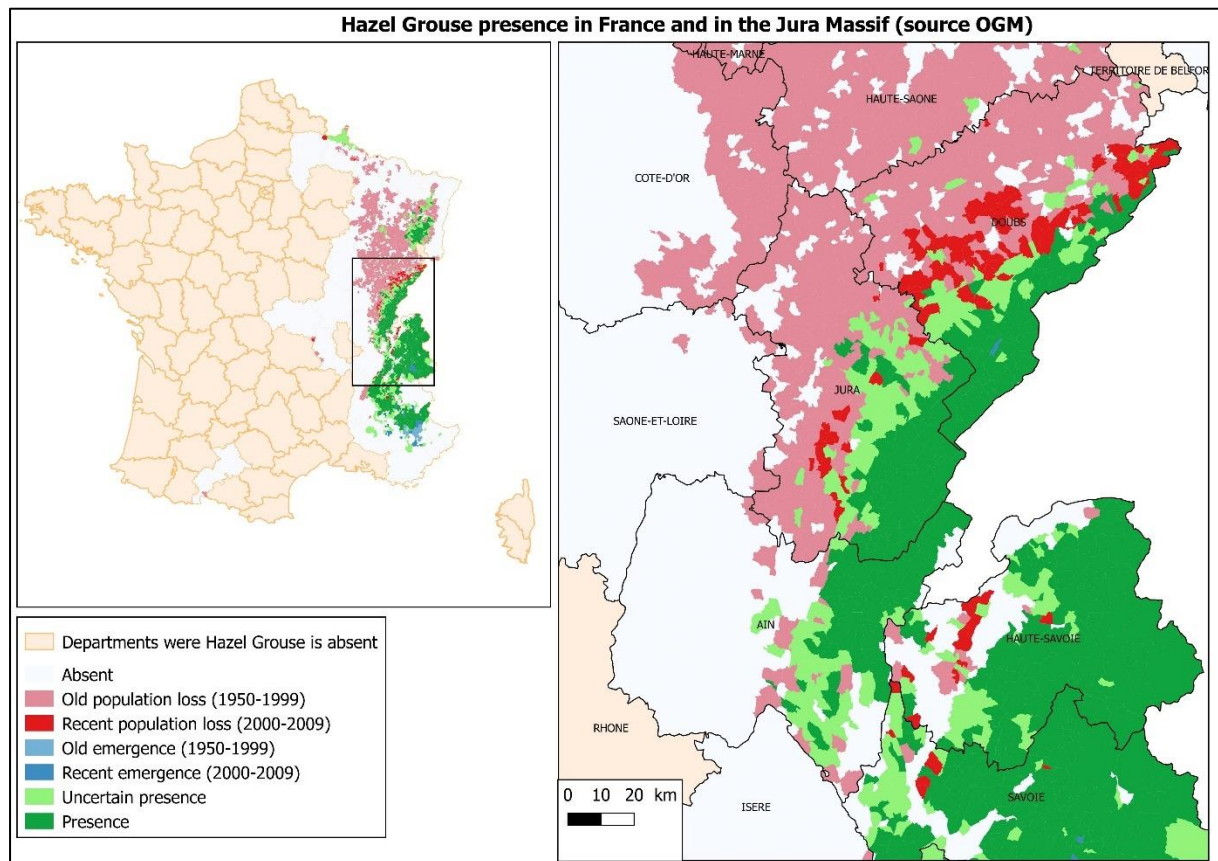
*Figure 8 : Hazel Grouse distribution range*



In France, different populations are known and are located in the Alps, Jura, Vosges and Ardennes massifs (Figure 9). If the species showed a restriction of its occurring range at the lower elevation (Jura, Vosges, Ardennes), the population present in the Alps has been expanding during the last decade (Montadert and Léonard, 2006).



Figure 9 : Hazel Grouse population repartition in France and in the Jura massif study area



### 3.3.3 Reproduction

Hazel Grouse is a monogamous species, where the couple is living on a 20-40 ha territory. Contrary to the Capercaillie, males are not concurring on a display place. Juveniles are mature as early as their first year and pairs are formed at the end of winter for the next breeding season. The reproduction period starts in March and early April through the beginning of May in the Jura Massif. Both male and female are singing in their territory and respond to each other (Rhim, 2006). The higher the density of individuals, the more time is allocated to call and song during the reproduction period. Females nest on the ground, brood seven to twelve eggs during 22 to 25 days and then raise the chicks alone during spring and summer. The chicks are strictly dependent on their mother for the first month, as females provide them heat and shelter in their first days of life. This species has a small dispersion capacity: in the French Alps the mean dispersion distance was 4 km for juvenile males and 2 km for juvenile females (Montadert and Léonard, 2006).



### 3.3.4 Habitat needs

The species habitat requirement varies worldwide, as it can be found in forest environments as different as taiga forest and temperate forest. Even within France, this bird is present in various forest environments located from 200m to 1800m a.s.l. and composed of different dominant tree species.

Nevertheless, Hazel Grouse prefers habitat with well-structured stand with at least two vegetation layers (Müller et al., 2009; Schaublin and Bollmann, 2011) (Figure 10). Thus, in the Jura mountains, this species selects habitat with high understory density of spruce (*Picea abies*) (shelters trees in winter), and rowan (*Sorbus aucuparia*) and willow (*Salix spp*) (Sachot et al., 2003b), hazel (*Corylus avellana*) and beech. This type of habitat provides in addition shelters against predators, which is particularly essential during the winter period when the cover is by definition lighter (leaf-off period). Sachot et al. (2003b), recommend a 50% cover of multilayered understory as suitable habitat for Hazel Grouse in the Swiss Jura. Contrary to Capercaillie, Hazel Grouse avoids grazed forest areas (Sachot et al., 2003b) probably because of the higher risk of predation in those areas. However, forest aisle appears to be an important habitat variable for Hazel Grouse, especially in summer when chicks forage on the ground searching for insects. Roots plate presence was also found to be an important part of the habitat as birds need bare soil for bathing and grit ingestion (Müller et al., 2009). Predator avoidance seems to be an important factor when studying this species especially in summer when hens are rearing broods (Kurki et al., 2000). Main predators in France are foxes (*Vulpes vulpes*), pine marten (*Martes martes*) and goshawks (*Accipiter gentilis*). The herbaceous layer is also an important part of the habitat in particular in summer as the vegetation will provide food for both chicks and adults.





*Figure 10: A good habitat for Hazel Grouse*



In winter, Hazel Grouse diet is composed of young buds and catkins of rowan, willow and beech depending on their presence and availability in the environment (Andreev, 1988). In spring, the Hazel Grouse diet is still composed of buds and catkins but include young leaves resources such a bilberry. In summer and autumn, the diet evolves and is now composed in majority of fruits present in the herbaceous layer such as bilberries, strawberries (*Fragaria vesca*), raspberries (*Rubus idaeus*) or currant (*Ribes ssp*). At the end of the autumn season, birds are switching again their diet to buds in particular from bilberry and rowan (Schatt, 1993).

To resume the main requirement of species, a patchwork of habitat is required, allowing the presence of both dense vegetation cover between 1m and 7m, and a furnished herbaceous layer. Heterogeneity within the vertical structure of the vegetation is also a key point.

### 3.3.5 Protection status

The Hazel Grouse is listed as least concern in the world and Europe (UICN red list 2015, 2016<sup>5</sup>). However, Hazel Grouse is listed in European Bird Directive annexes I, II where the bird species for

---

<sup>5</sup> <https://newredlist.iucnredlist.org/species/22679494/85936486>



which conservation actions are taken and needed are listed. In France, the species is listed as near threatened, but some populations at the regional scale are listed as vulnerable (Alpes) or critically endangered (Alsace, Auvergne). The species can be hunted in three departments (Isère, Savoie and Hautes-Alpes), taking an average of 50 individuals each year. In twenty-two other departments the hunting is prohibited, and two other have a hunting planning set to 0 individuals (Haute-Savoie, Jura).

### 3.3.6 The causes of the decrease in France

As a forest bird, Hazel Grouse is strongly impacted by anthropogenic activities such as forestry (Åberg et al., 2003). In France, the species is mainly disappearing from lower elevation as a result of the maturation of the deciduous forest leading to the transformation of an optimal habitat with dense coppice to unfavorable habitat. The presence of dense vegetation regeneration was historically maintained by intense harvest activities. Nowadays, forestry practices have evolved, changing deeply the habitat availability for Hazel Grouse. In the higher forests in the Jura, forestry practices with selecting cutting and the cessation of understory systematic clearing still favor good habitat for the species. Hazel Grouse is vulnerable to forest fragmentation due to its low dispersion through open land leading to the presence of isolated small populations (Åberg et al., 1995; Huhta et al., 2017). Change in predation pressure can also lead to a population decrease. Indeed, by its size, the species have many potential predators. Human disturbances such as roads were also found to have a negative effect on Hazel Grouse populations, in a lesser extent than the impact on Capercaillie (Leclercq, 1985; Rätty, 1979).

### 3.3.7 Challenges in the Jura Massif

The knowledge of the population dynamics and occurrences in the Jura Massif are still poor due to the secretive behavior of the bird and to the large potential area of presence. Nevertheless, the good knowledge of the habitat preferences of the species recommendations for its habitat conservation and restoration has been promoted nearby forest managers<sup>6</sup>. Some actions such as feeding tree plantation and the installation of a visualization system on ski lift cables were done by Groupe Tétraz Jura.

---

<sup>6</sup> <https://groupe-tetraz-jura.org/la-g%C3%A9linotte-des-bois/mesures-de-protection>





From those observations, this work aims to help to plan of future surveys through the Jura massif and improve the knowledge on population.

### 3.3.8 LiDAR variables for Hazel Grouse habitat models

More studies including LiDAR data for habitat modeling were produced on Hazel Grouse (Table 2). Two of the studies presented above for Capercaillie included Hazel Grouse models as well, thus using the same LiDAR extracted variable (Melin et al., 2016; Zellweger et al., 2013a). Three other recent studies used the LiDAR extracted variable in their models. First, Bae et al. (2014a) compared the predictive power of LiDAR, satellite image and field data survey in habitat suitability models for Hazel Grouse in the Bavarian Forest National Park (Germany). The results showed that models with LiDAR-derived variables only were giving better results than other single source-based models. However, the combination of multiple sources was giving the better overall model. Many different variables were extracted for modeling but the most contributing LiDAR variables were the number of vegetation returns (mean and standard deviation) between 5 and 10m, the penetration ratio (mean and standard deviation) between 5 and 10m, the number of vegetation returns (mean and standard deviation) between 10 and 20m, and finally the mean Shannon index for height diversity. Those results correspond to the sensibility of Hazel grouse to both horizontal and vertical composition, favoring well-structured forest. The importance of the vegetation density between 5 and 10m was highlighted as well. The same component highlighting the importance of both vertical and horizontal heterogeneity was found by (Zellweger et al., 2014). Indeed, vertical structure indicators explain 44.2% of the variance (CHH, standard deviations of quantile of height 1<sup>th</sup>, 8<sup>th</sup> clumpy index of patches (>15m)) and horizontal structure explain 65.8% of the variance. The last study recently published by (Rechsteiner et al., 2017a), created LiDAR-based models over the “Parc Regional Chasseral” (Jura Massif Switzerland). Along with the LiDAR data, forest type from Landsat images and forest boundary from topography federal office were added. The best contributing variables were average vegetation height, shrub density and canopy height variation. The model performance was evaluated to be good, thus giving relevant results.



The use of LiDAR consistently improved models for Hazel Grouse, showing the importance of the forest structure for that species. More specifically both horizontal and vertical heterogeneity are important, and the density of the canopy cover and of the middle height vegetation (2-10m) seem to be key predictors.



Table 2: LiDAR variables and their description (Hazel Grouse models)

Study	Ecosystem	LiDAR variables	Description
Melin <i>et al.</i> 2016	Boreal forest	Shrub cover	A shrub_cover value of 50 means that from all the echoes reflected below five meters 50% came from vegetation, not ground. The variable describe about the amount of vegetation between the heights of 0.5 and 5 m
		Canopy cover	A canopy_cover value of 70 means that 70% of all the echoes reflected above five meters. The variable describes the amount of echoes were vegetation is above 5 m of canopy cover
		Canopy height	A canopy height value of 22.5 means that 95% of the echoes are reflected below the height of 22.5 m. The value of this variable indicates the maximum canopy height, which in the example case would be 22.5 m
Zellweger <i>et al.</i> 2013	Jura Mountains, Pre-alpine forest, Alpine forest	Standard deviation of quartile	A density of points for the 1th, 2nd, 3rd, 4th, 5th, 7th, 8th height quantiles
		Multi-storied profile	An Index representing the vertical distribution range within the 2nd and 3rd height quartile.
		Patch area	Mean patch area for understory, midstory and canopy CHM class.
		Forest gap density	Sum of forest gaps, i.e. non-forest patches <1 ha, divided by total forest area in ha.
		% of landscape	Percentage of the landscape (1 km <sup>2</sup> ) covered by patches of understory, midstory and canopy CHM class
		Clumpy	A clumpy index for patches of understory, midstory and canopy CHM class.
		Nearest neighbor	Mean nearest neighbor distances for patches of understory, midstory and canopy CHM class.
		Edge density	Density of edges (pixel transitions) between two CHM classes, i.e. non-forest/understory, non-forest/midstory, non-forest/canopy, understory/canopy.
		Canopy height heterogeneity	Total length of edges (pixel transitions) between understory, midstory and canopy CHM
		Topographic position	Mean and standard deviation of the Topographic Position Index. Measures the exposure of a site in relation to the surrounding terrain.
Bae <i>et al.</i> 2014	Alpine forest	Slope	Mean and standard deviation of slope
		Heights	Maximum, mean, standard deviation and skewness of heights
		Heights percentile	For the 90th 75th 50th 25th 10th percentile
		Canopy cover	Vegetation returns >0.5 m/total returns
		Openness below 2m from ground	Returns <2 m/total returns
		Percentage of ground returns	Ground returns/total returns
		Percentage of vegetation return	Average and mean (0.5-10m, 10-20m, 20-30m, 30-60m heights)
		Penetration ratio	Average and mean (0.5-10m, 10-20m, 20-30m, 30-60m heights)
		Foliage height diversity with Shannon index	$ShI = -\sum p_i \ln(p_i)$ , where $p_i$ is the proportion of horizontal vegetation in the $i$ th layer
		Foliage height diversity with Simpson index	$Sml = \sum p_i^2$ , where $p_i$ is the proportion of horizontal vegetation in the $i$ th layer
Rechsteiner 2017	Jura Mountains	Average vegetation height	Average height of all vegetation return heights above 0.5 m
		Canopy height variation	Standard deviation of canopy heights based on the 90th height percentile
		Shrub density	Number of vegetation height below 5 m divided by the total number of all returns, including ground points
		Sum of small gaps	Sum of squares with a canopy height below 1.3 m, based on the 90th height percentile
Zellweger <i>et al.</i> 2014	Jura Mountains, Pre-alpine forest, Alpine forest	Shannon diversity index for CHM	Shannon's diversity index for CHM classes 2 and 3, 1 and 2 (1=matrix, 2=understory, 3=midstory, 4=canopy class)
		Shannon evenness index for CHM	Shannon's evenness index for CHM classes 2 & 3
		Area-weighted mean	Area-weighted mean of patch area for CHM classes 2, 3 and 4
		Variation of patches	Coefficient of variation of patch area for CHM classes 2, 3 and 4
		Fractional dimension index	Mean fractional dimension index of CHM classes 2, 3 and 4
		Clumpy	Clumpyness of CHM classes 2, 3 and 4
		Edge	Total edges of CHM classes 1, 2, 3 and 4
		CHH	Canopy height heterogeneity
		SD height	Standard deviation of height percentiles per layer
		SD density	Standard deviation of density percentiles per layer
		Mean height	Mean of height percentiles per layer
		Mean density	Mean of density percentiles per layer
		Penetration ratio	Penetration ratios on four height levels: 1 m, 2 m, 3 m, 4 m. Expressed as: Sum of points $\leq$ height level / sum of all points.

## 4 OBJECTIVES

---

### 4.1 General questions

The objective of the project is to propose SDMs to help future management actions in favor of Capercaillie (including Capercaillie and beech regeneration models) and to help the implementation of future Hazel Grouse survey.

As highlighted in this introduction the creation of relevant models for conservation actions can be divided into two main axes:

#### **1) Improving models performances and pertinence and understand their limits**

The main hypothesis is that the use of heterogeneous observation datasets can give relevant SDMs results if the constraints are correctly identified and corrected. The aim was thus to define the best way to analyze the data regarding their specificities which are presented in details in the different chapters that constitute the manuscript.

If this subject had already been explored in other studies (Fourcade et al., 2018; Merow et al., 2014; Phillips et al., 2009) the specificities of our datasets allowed us to explore some aspects that were poorly evaluated so far. Thus, different points will be developed:

- The specificities of the bird's observations datasets allowed us to explore the effect of spatial sampling bias on models' performance and accuracy using different sampling designs. In addition, due to the high sexual dimorphism of Capercaillie, the effect of such spatial bias in the case of a population divided in sub-groups with different behavior on models outcomes was explored.
- The effect of interaction between environmental variables on models performance and interpretation was explored using Hazel Grouse as a case study.



- The variables contribution and response were explored and compared to previous studies results in order to highlight the main strengths and limits of our models.

## 2) Improving of results transmission to managers

Beyond the willingness to provide models that are ecologically relevant, the awareness of the constraints occurring to management actions and of the target audience expectations should be taken into account during the first steps of a project, thus leading to the creation of models pertinent from both points of view. Our hypothesis is that such a framework will not decrease the performances of the models but can highly improve the transmission of the results.

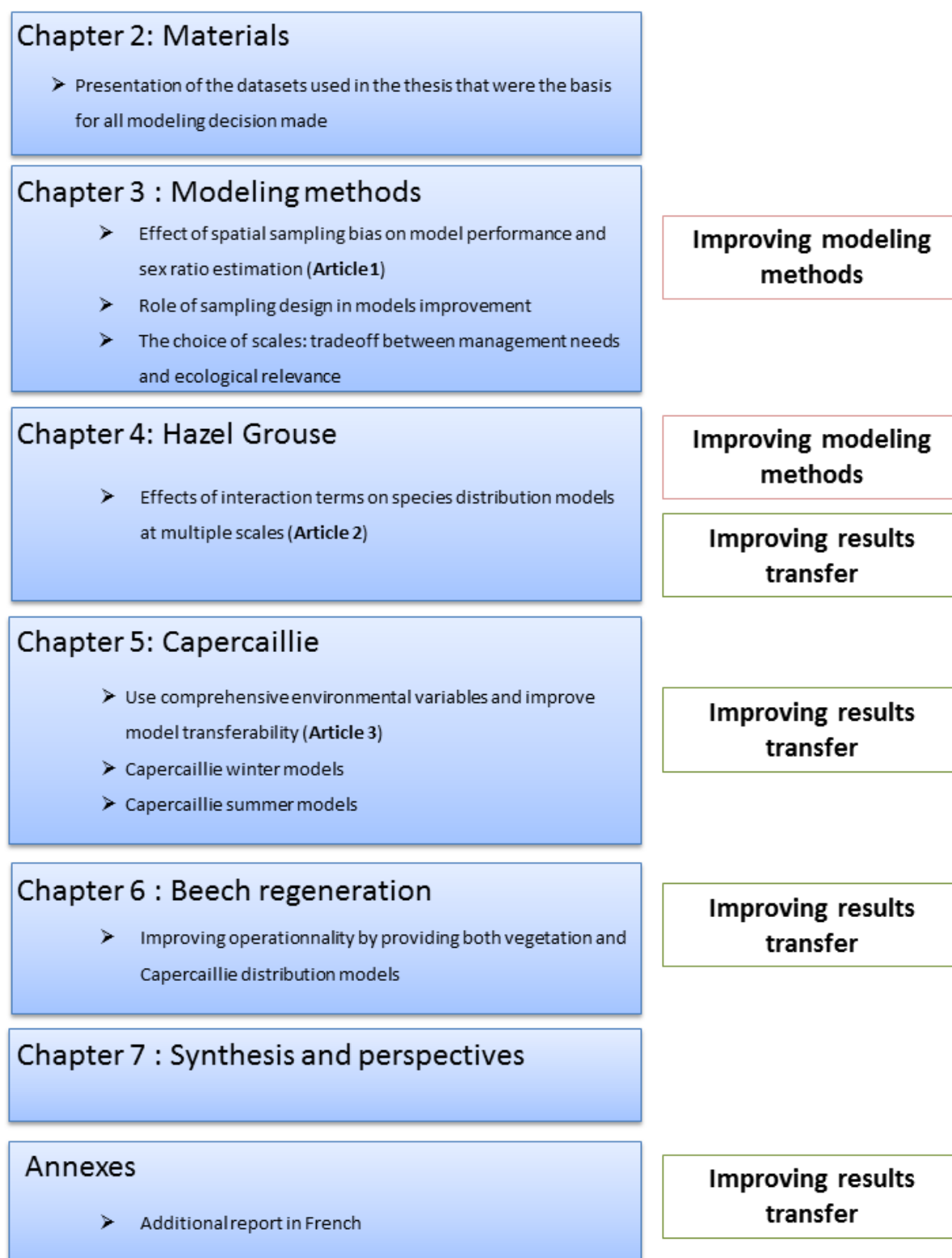
Within that purpose different points were developed:

- **Matching the manager's conceptual context:** the exploration of the benefice of using LiDAR extracted objects commonly used in forestry for endangered species distribution modeling.
- **Improving modeling practices:** because the reproduction of such models to other areas of interest is an advantage, the transferability of models results in other areas using oriented-object variables was explored using Capercaillie models.
- **Matching the manager's conceptual context:** In addition to Capercaillie distribution models, a specific management issue will be addressed by modeling the distribution of the beech regeneration.
- **Improving communication:** results communication including a report in French.

## 4.2 Manuscript organization

The thesis was thus organized in order to solve in a first part the problems encounter with models' performances (Figure 11). In a second part, the solutions found were applied to create the definitive models for both Hazel Grouse and Capercaillie. Meanwhile, models that ensure a better transfer of the results were created.

*Figure 11 : Manuscript organization*



## CHAPTER 2: MATERIALS





## 1 THE JURA MASSIF STUDY AREA

---

The Jura Massif lies in both France and Switzerland and extends over more than 340 km for a maximal width of 70 km. The study area is located in the French part of the mountains within the departments of Ain, Doubs and Jura. The climate is continental with temperature varying between - 5.9°C and 21.1°C, with maximal annual means of 11.3°C and minimal means of 2.1°C (based on data from 1981 to 2010 at the meteorological stations of La Pesse (1130 m a.s.l), Lamoura (1124 m a.s.l) and Les Rousses (1100 m a.s.l)) . The mean annual precipitation is 1874 mm (mean on the three same stations)<sup>7</sup>. Snowfall usually occurs between December and April, but the cover height showed high variations between years.

The landscape is composed of a mosaic of small urban areas, pastures, forests and fields. The massif is composed of a low plateau (elevation range from 400 m to 700 m a.s.l.) and a high plateau (elevation range from 700 m to 1620 m a.s.l.). It is dominated by the Crêt de la neige (1720 m a.s.l) and Le Reculet (1718m a.s.l) both located in the Ain department. The forest covers an area of 4800 km<sup>2</sup> in the Doubs and Jura departments whereas the Ain department comprises 1920 km<sup>2</sup> of forest (based on the IGN national forest inventory 2005-2009)<sup>8</sup>. On the lower plateau, forests are mainly deciduous-dominated composed of beech (*Fagus sylvatica*), ash (*Fraxinus excelsior*), oak (*Quercus petraea*) and spruce (*Picea abies*). Above 1 000 m forests are mainly dominated by conifers and composed of a mix of spruce, fir (*Abies alba*) and beech. At altitude higher than 1600 m, the climatic conditions limit the development of the forest, allowing the growth of alpine type grassland.

---

<sup>7</sup> <https://donneespubliques.meteofrance.fr/>

<sup>8</sup> <https://inventaire-forestier.ign.fr/>





## 2 CAPERCAILLIE AND HAZEL GROUSE DATASETS

---

Observations used for the different parts of this work originated from a large database, where the NGO “Groupe Tétrás Jura” collected data on both target species from various sources. A large part of the data was collected by Groupe Tétrás Jura, as well as ONCFS and ONF (Forest National Office) during their own fieldwork. Observations were also transmitted by volunteer ornithologists, the NGO LPO (French Bird Protection Society) and by the Natural National reserve of Haute-Chaîne du Jura. Thus, observations were collected with different protocols showing different characteristics. Older data were often unprecise, with missing date, observers and non-precise coordinates. GPS coordinates associated to observations first occur in 2001. Thus, data used for modeling were selected among those more recent and precise GPS locations. At first, the type of protocol used to collect the data was not specified in the database. Consequently, during the first part of the project, the assignation of each observation to a known type of protocol was done when it was possible. Observation data were divided into three categories: prospections, battues and unknown protocol. Older observations (before 2001) could not be assigned to any protocol. The assignation of the protocol’s type was done by crossing information on dates, observers and GPS tracks. A large part of the dataset could thus not be assigned to one specific protocol and remained unknown. Prospections occurred generally in winter and involved observers navigating in a selected area to search for presence indices. Since 2007, the observer’s tracks were also recorded and were used to assigned observations to the prospection protocol (date and distance to the observation). Prospections that were done before that date remain as unknown protocol. Since 2016, the protocol type was integrated into the database after the start of systematic parallel transects winter prospections. In summer, each year in july a “Battue” is organized. This protocol involves many volunteers, that are positioned every 10m and walks together with the aim to flush the birds, in particular, the chicks. Observations were assigned to this protocol by dates. The first observations that were assigned to this protocol were from 2008.

## 2.1 Capercaillie

Data were collected from 1995 to 2018 leading to 6713 observations. GPS locations represent 5370 observations. From those observations, 241 are from battue, and 2615 from prospections (winter and summer). Thus, 2514 observations could not be assigned to a specific protocol. Within the data assigned to a protocol, 583 are from summer, 147 from autumn, 2838 from winter and 1802 from spring. The number of observed individuals is largely differing between protocols (Table 3) with a large majority coming from Prospection or unknown protocols. A sex ratio biased toward male was observed in those two cases but was not observed for the battue protocol (Table 3). Here, it should be highlighted that for 2018 winter prospection data, the sex was not systematically provided because the sex will be ultimately determined by genetic analysis. Unfortunately, those analyses were not done before the end of the project, leading to a higher occurrence of unknown sex in 2018 data. Finally, different types of indices were collected, with a high occurrence of two major types: dropping and roosting trees. Feathers, traces, sight are present as well in the database. Some of the observations signs types are also unknown.

*Table 3 : Number of observation by protocols and sex for Capercaillie*

Protocol	Sex	N observations
Prospection	Male	1454
	Female	904
	Unknown	251
Battue	Male	109
	Female	103
	Unknown	23
Unknown	Male	1301
	Female	610
	Unknown	615

## 2.2 Hazel Grouse

Like Capercaillie, data were collected from 1995 to 2017 leading to 6854 observations. GPS locations represent 5541 observations (since 2002). From those observations, 282 are from battue 2898 from prospections and 2361 observations could not be assigned to a specific protocol. Within those data, 501 are from summer, 147 from autumn, 3559 from winter and 1334 from spring. The number of



observed individuals is differing between protocols with a large majority coming from prospection or unknown protocols. In addition, the sex of the bird was determined in only rare cases because of the small sexual dimorphism observed for this species (Table 4). The major types of indices collected are droppings, traces and roosting trees. The presence of igloos and the sight of individuals are also often reported. Other types of observation are present as well in the database, as feathers or dead birds. A part of the dataset is composed of non-assigned types of observation.

*Table 4 : Number of observations by protocol and sex for Hazel Grouse*

Protocol	Sex	N observations
Prospection	Male	19
	Female	2
	Unknown	2870
Battue	Male	4
	Female	0
	Unknown	265
Unknown	Male	35
	Female	23
	Unknown	2312

### 3 LIDAR DATASETS

---

#### 3.1 LiDAR campaigns

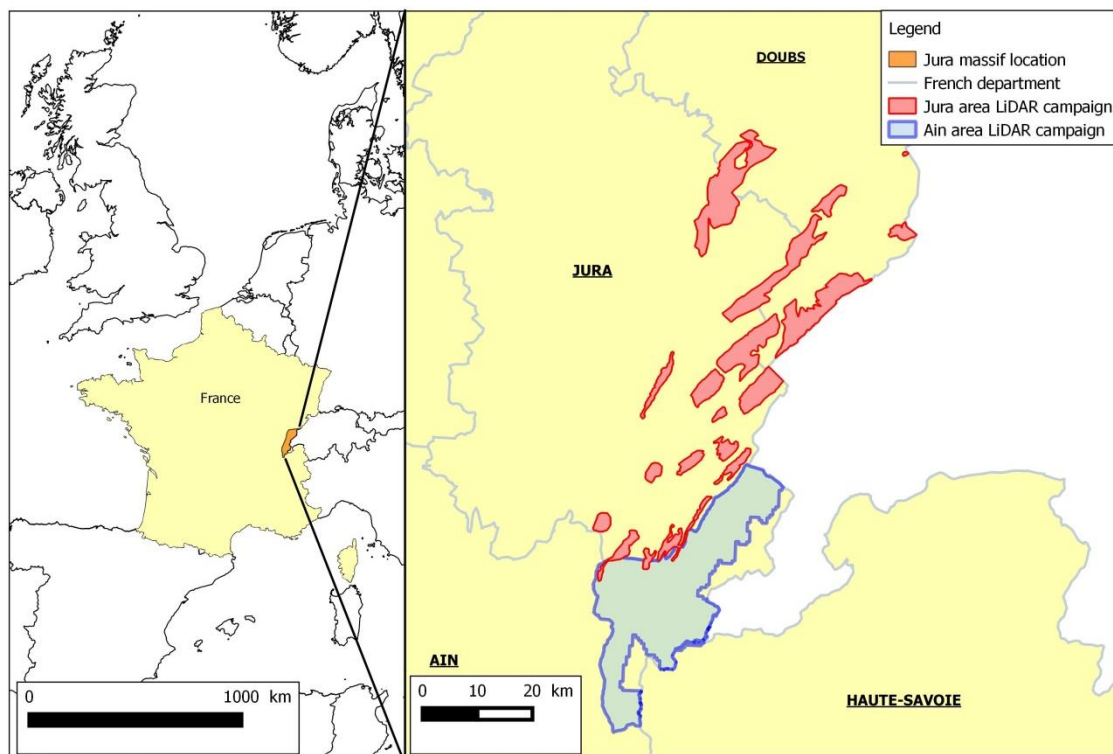
The study area was covered by two different airborne LiDAR surveys (Figure 12). Both surveys were conducted using a Riegl LMS Q680i system (RIEGL Inc., Horn, Austria). The system was operated with a maximum half-scan angle of 60°, with all return points being recorded.

In the first area referred hereafter as Ain study area, the LiDAR campaign was conducted in autumn 2014 and covered 626 km<sup>2</sup>. The flight means height was 600m above ground. We obtained an average point density of 21.3 points/m<sup>2</sup> with a mean vertical error below 0.5 cm. During this campaign, the whole landscape was covered with LiDAR.



The second Lidar survey was conducted in summer 2016 and covered a surface of 431km<sup>2</sup>. The flight means height was 600m above ground. We obtained an average point density of 18 points/m<sup>2</sup> with a mean vertical error below 0.5 cm. Contrary to the first campaign, the LiDAR datasets area is composed only of forest patches that are potentially occupied by Hazel Grouse and Capercaillie, which leads to having fragmented and incomplete description of the landscape.

*Figure 12 : Coverage of the two LiDAR campaigns in the French Jura Massif constituting two study areas.*



### 3.2 LiDAR extracted variables: descriptions

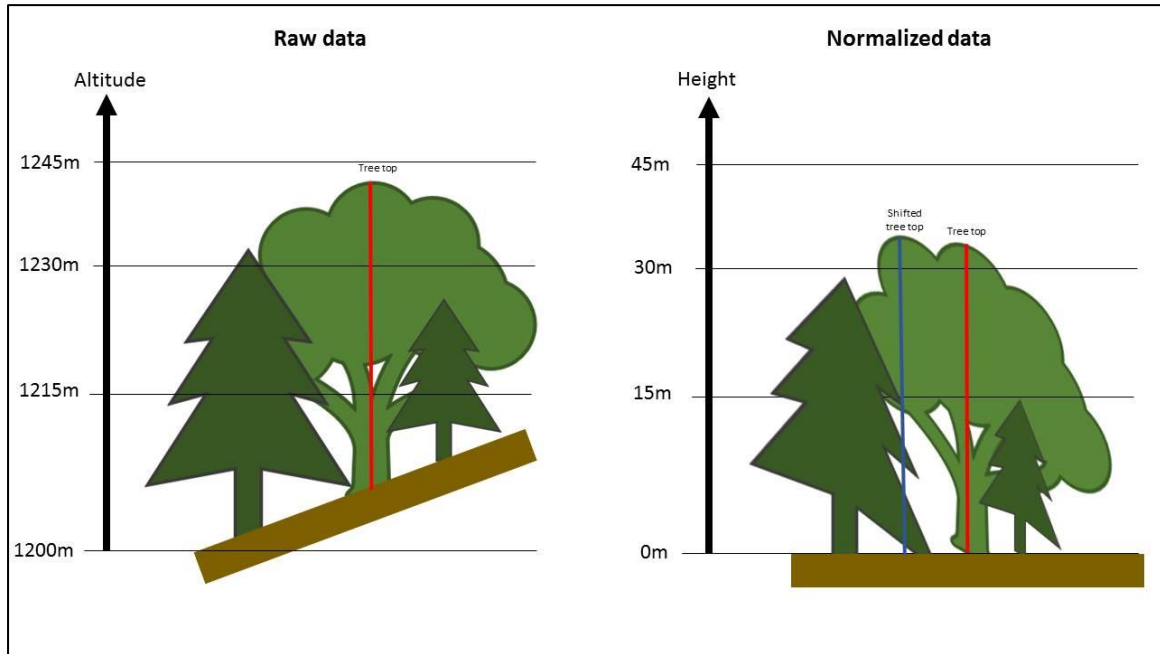
LiDAR metrics were calculated using two R packages: lidR (Roussel and Auty, 2016) and LidaRtRee developed by M. Monnet. We divided the variables extracted in three types: point-cloud (PC) area-based metrics, object-oriented (OO) metrics and LiDAR interpolated metrics (the code used is presented in Annexe 2 and the package LidaRtRee is available on gitlab<sup>9</sup>). The comparison between point-clouds area-based and object-oriented metrics type on models' performance and results interpretability is presented in Chapter 5. This section aims to describe each calculated variable in

<sup>9</sup> <https://gitlab.irstea.fr/jean-matthieu.monnet/lidaRtRee/wikis/Forest-habitat-metrics>



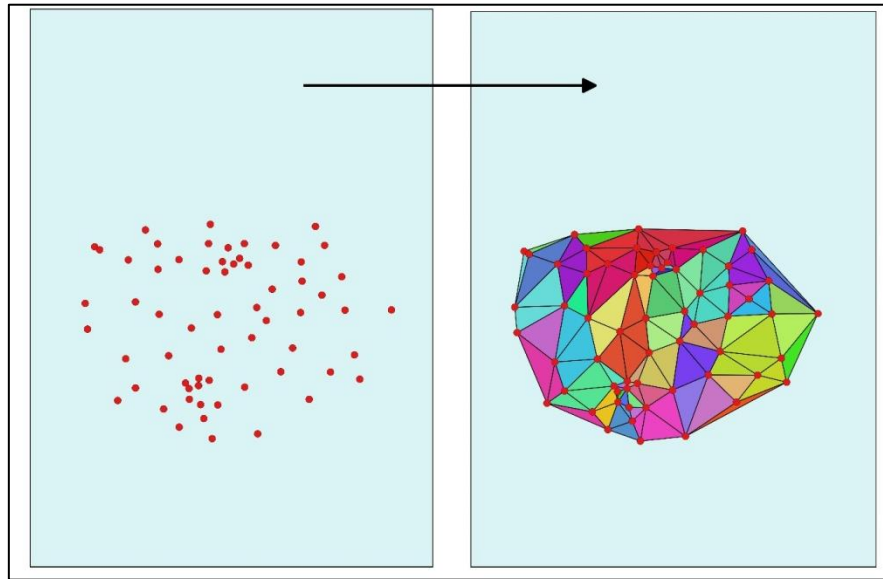
detail. Before calculating any metrics, point-clouds were normalized using the function `lasnormalize` from the `LidR` package (Figure 13).

*Figure 13 : The normalization of the LiDAR point cloud generate the deformation of the tree crown and a shift of the tree top location.*



This function subtracts a digital terrain model (DTM) to the z coordinates values of each point to set the ground at a height of 0 m. We used a Delaunay triangulation, which makes a linear interpolation within triangles to create the DTM from ground classified points (Figure 14). Triangles are drawn between each closest point of the ground point clouds. The normalization is done by subtracting the altitude of the projection of each point on the Delaunay triangulation to the point altitude (Roussel and Auty, 2016).

*Figure 14 : example of a Delaunay triangulation applied on a point cloud*



### 3.2.1 Point-cloud area-based metrics

Point-cloud area-based metrics are defined as indices describing the LiDAR point clouds over a given and fixed surface (pixels). Multiple metrics were calculated to represent different characteristics of the point-cloud such as heights, point densities and penetration ratios (Table 5). The height metrics included the maximum height, the mean height and the height of the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup> and 90<sup>th</sup> quantiles. The point densities and the penetration ratios were calculated by height portions: 0.5-1m, 1-2m, 2-5m, 5-10m, 10-20m and 20-30m. The point density metric is represented by the number of points within a height portion divided by the total number of points (soil and vegetation). The penetration ratio is calculated as the number of points in the lower portions minus the cumulative sum of this portion and the next highest one.



Table 5 : Point-cloud area-based metrics

Type of metrics	Metric Name	Usual name (in text)	Description
Points cloud metrics	H.max	MaxVeg	Maximum height
	H.mean	MeanVeg	Mean height
	H.nb0_0.5relative_density	Canopy005	Point density between 0 and 0,5m
	H.nb0.5_1relative_density	Canopy0501	Point density between 0.5 and 1m
	H.nb1_2relative_density	Canopy0102	Point density between 1 and 2m
	H.nb2_5relative_density	Canopy0205	Point density between 2 and 5m
	H.nb5_10relative_density	Canopy0510	Point density between 5 and 10m
	H.nb10_20relative_density	Canopy1020	Point density between 10 and 20m
	H.nb20_30relative_density	Canopy2030	Point density between 20 and 30m
	H.nb30_60relative_density	Canopy3060	Point density between 30 and 60m
	H.nb0_0.5ratio	penetrationratio005	Penetration ratio between 0 and 0,5m
	H.nb0.5_1ratio	penetrationratio0501	Penetration ratio between 0,5 and 1m
	H.nb1_2ratio	penetrationratio0102	Penetration ratio between 1 and 2m
	H.nb2_5ratio	penetrationratio0205	Penetration ratio between 2 and 5m
	H.nb5_10ratio	penetrationratio0510	Penetration ratio between 5 and 10m
	H.nb10_20ratio	penetrationratio1020	Penetration ratio between 10 and 20m
	H.nb20_30ratio	penetrationratio2030	Penetration ratio between 20 and 30m
	H.nb30_60ratio	penetrationratio3060	Penetration ratio between 30 and 60m
	H.nb0_0.5	veglayer005	Number of Vegetation and unclassified point between 0 and 0,5m
	H.nb0.5_1	veglayer0501	Number of Vegetation and unclassified points between 0.5 and 1m
	H.nb1_2	veglayer0102	Number of Vegetation and unclassified points between 1 and 2m
	H.nb2_5	veglayer0205	Number of Vegetation and unclassified points between 2 and 5m
	H.nb5_10	veglayer0510	Number of Vegetation and unclassified points between 5 and 10m
	H.nb10_20	veglayer1020	Number of Vegetation and unclassified points between 10 and 20m
	H.nb20_30	veglayer2030	Number of Vegetation and unclassified points between 20 and 30m
	H.nb30_60	veglayer3060	Number of Vegetation and unclassified points between 30 and 60m
	H.p10	Q10	Vegetation and unclassified points height quantile 10
	H.p25	Q25	Vegetation and unclassified points height quantile 25
	H.p50	Q50	Vegetation and unclassified points height quantile 50
	H.p75	Q75	Vegetation and unclassified points height quantile 75
	H.p90	Q90	Vegetation and unclassified points height quantile 90
	H.sd	SdVeg	Height standard deviation
	H.simpson	simpson	Simpson index of height
	I.mean.1stpulse	I.mean.1stpulse	First pulse mean intensity
	nbSol	GroundR	Number of ground points
	nbVeg	nbVeg	Number of vegetation and unclassified points
	CHM0.sd	CHM0.sd	Canopy height Model
	CHM1.sd	CHM1.sd	Canopy height Model standard deviation smoothed with a 1 pixel kernel
	CHM2.sd	CHM2.sd	Canopy height Model standard deviations smoothed with a 2 pixels kernel
	CHM4.sd	CHM4.sd	Canopy height Model standard deviations smoothed with a 4 pixels kernel
	CHM8.sd	CHM8.sd	Canopy height Model standard deviation smoothed with a 8 pixels kernel
	CHM16.sd	CHM16.sd	Canopy height Model smoothed with a 16 pixels kernel
	CHM.mean	CHM.mean	Canopy Mean height Model
	CHM.PercInf0.5	CHM.PercInf0.5	Proportion of Canopy height Model < 0.5m within the 25*25 pixel in % (based on 0.5*0.5 surface)
	CHM.PercInf1	CHM.PercInf1	Proportion of Canopy height Model < 1m within the pixel in % (based on 0.5*0.5 surface)
	CHM.Perc2_5	CHM.Perc2_5	Proportion of Canopy height Model < 5m within the pixel in % (based on 0.5*0.5 surface)
	CHM.Perc1_5	CHM.Perc1_5	Proportion of Canopy height Model with a maximum height between 1m and 5m within the pixel in % (based on 0.5*0.5 surface)
	CHM.PercSup5	CHM.PercSup5	Proportion of Canopy height Model > 5m within the pixel in % (based on 0.5*0.5 surface)
	CHM.PercSup10	CHM.PercSup10	Proportion of Canopy height Model > 10m within the pixel in % (based on 0.5*0.5 surface)
	CHM.PercSup20	CHM.PercSup20	Proportion of Canopy height Model > 20m within the pixel in % (based on 0.5*0.5 surface)

### 3.2.2 Object-oriented metrics

Object-oriented metrics are defined as elements constituted of pre-specified characteristics that will be common to all individual objects. Those elements are called objects as they aim to represent real individual components of the landscape. Two object elements were defined and then extracted from the LiDAR point clouds (Table 6). First, individual trees objects were extracted using features that

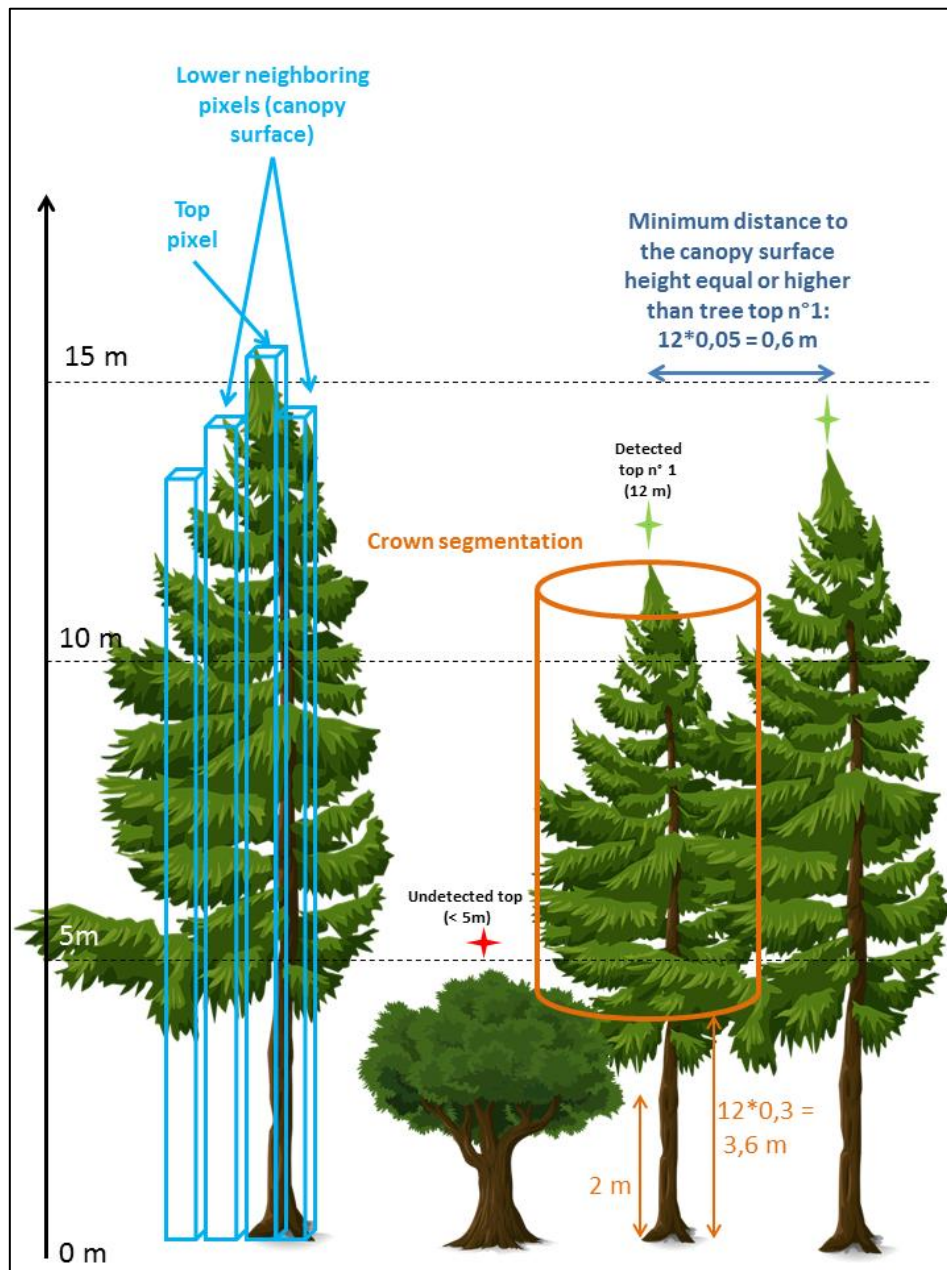


constitute the shape of a tree. The main point here is that the top of tree will be detected in the LiDAR point cloud as a local maximum, which can be filtered. This method has some limitation has smaller trees localized under and near a higher one can thus be hidden. False tree tops can also be detected because of the presence of large branches or forks. The window size for detecting those local maxima is thus an important parameter. A trade-off between detecting only higher overall tree tops and detecting all top at a very fine resolution with the risk to obtain false tree detections needs to be found. Once local maxima are detected, trees are then segmented using a watershed algorithm. In our case, we choose the following parameters for tree top detection (adapted from method 1 in Eysn et al 2014), (Figure 15):

- A distance minimum from the detected tree top to the next higher pixel of 0 (thus compare the pixel value to its neighboring pixels)
- A distance minimum from the detected tree top has a proportion of the height to the next higher pixel of 0.05
- The minimum height for tree top detection of 5m
- A minimum crown height of 2 meters for segmentation
- A number of the minimum crown height as a proportion of the tree top height of 0.3



Figure 15 : Tree top detection and crown segmentation



Concerning the gap detection, we first made a distinction between structural gaps define as physical low vegetation area from functional gaps defined as an entity that can provide specific habitat requirements. In our case, gaps should provide enough light for young trees regeneration, bilberries and herbaceous growth (Hinsley et al., 2008). Thus, the vegetation height around this surface must not completely hide the sun. A tradeoff was chosen to take this effect into account for gap detection. Therefore, a gap was defined as a minimum surface of  $25\text{m}^2$  constituted of vegetation below 1m height, where the width is equal or larger than half of the height of the surrounding trees (Figure 16).

Figure 16 : The method used to extract gap objects from LiDAR datasets

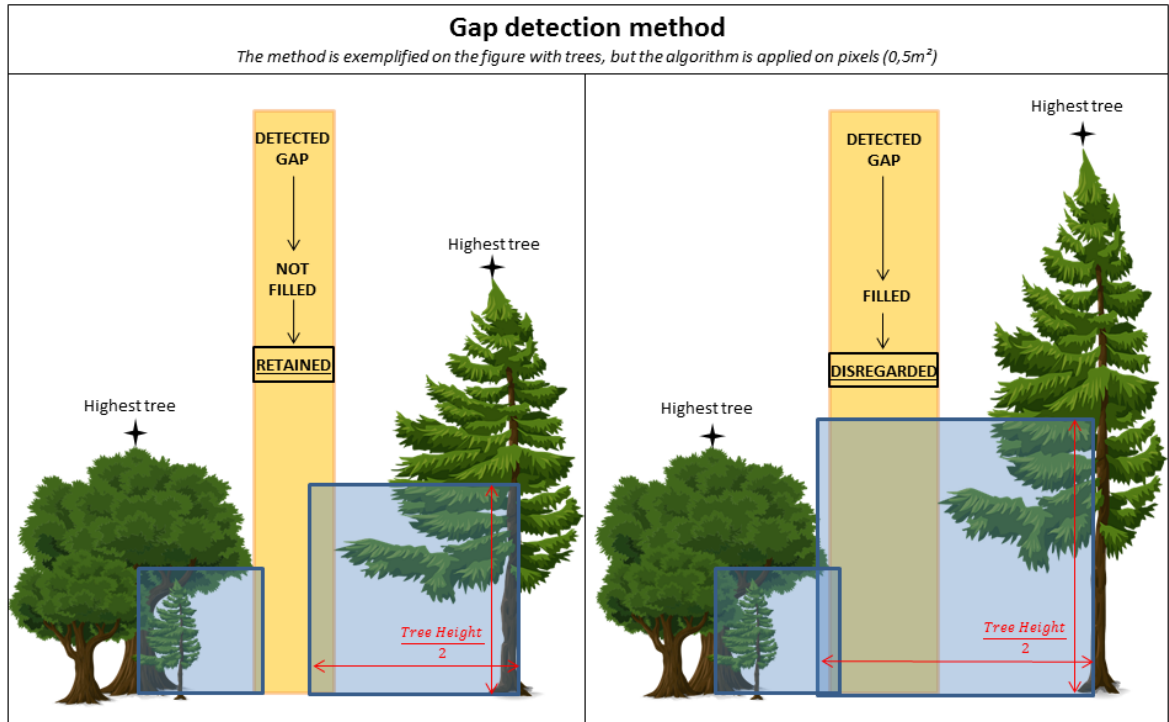


Table 6 : Object-oriented metrics

Type of metrics	Metric Name	Usual name (in text)	Description
Object oriented metrics	Tree.density	Tree_density	Tree density ( number/ha)
	Treeinf10.density	Tree_densityinf10	Tree lower than 10 meter but higher than 5m density (number/ha)
	Tree10.density	Tree_density10	Tree higher than 10 meter density (number/ha)
	Tree20.density	Tree_density20	Tree higher than 20 meter density (number/ha)
	Tree30.density	Tree_density30	Tree higher than 30 meter density (number/ha)
	Tree.giniH	Tree_gini	Tree height gini index
	Tree.meanH	Tree.meanH	Tree mean height
	Tree.meanS	Tree.meanS	Tree mean surface
	Tree.sdH	Tree.sdH	Tree height standard deviation
	Tree.sumS	Tree.sumS	Total surface covered by trees
	Tree.sumV	Tree.sumV	Total volume of trees
	Gap_Prop_1000_inf	Prop_open_area	Proportion in the pixel that contains an object grassland (surface < 1000m <sup>2</sup> and vegetation height < 1m )
	Gap_Prop200_1000	Middle_gap	Proportion in the pixel that contains an object large gap from 200m <sup>2</sup> to 1000m <sup>2</sup> (Height is <1m and half height of surrounding trees is less than half of the gap width)
	Gap_Prop25_200	small_gap	Proportion in the pixel that contains an object large gap from 25m <sup>2</sup> to 200m <sup>2</sup> (Height is <1m and half height of surrounding trees is less than half of the gap width)

### 3.2.3 Modeled forest metrics

Modeled forest metrics were also calculated using LiDAR and vegetation data field measurements (Munoz et al., 2015). The National Forestry Office (ONF), completed surveys on 377 permanent located under the LiDAR Ain coverage<sup>10</sup>. Among those 274 plots were taken with a more precise GPS to be used to model LiDAR metrics. During the last survey (2015), trees with a diameter at breast

<sup>10</sup> <https://www.rnn-hautechainedujura.fr/composants/uploads/2016/10/rapport-observatoire-2015-v150118.pdf>



larger than 17.5 cm were localized. Distance to the trunk, azimuth, species and diameter, were measured for each tree. A GPS was used to localize the center of the 15 m radius plot. From the inventory dataset five metrics were calculated (Table 7):

- The basal area for all trees, which is the sum of the trunk section surfaces measured at 1.30m then divided by the plot surface. It is expressed in m<sup>2</sup>/ha.
- The basal area for coniferous trees only
- The s density, which is the number of trees with a diameter at 1.30 m larger than 17.5 cm and expressed as a number of stem per ha
- The stem density for coniferous trees only
- The mean trunk diameter for each plot

At the same time, basic point-cloud metrics were extracted from the LiDAR dataset among plots (heights, intensity, return number and point classification). Multiple regression models were then applied with the parameters calculated from the field survey as variables and LiDAR metrics as predictors. The best model which gave the best predictions regarding the measured parameters was selected. Predictions over the whole Ain study area were then applied.

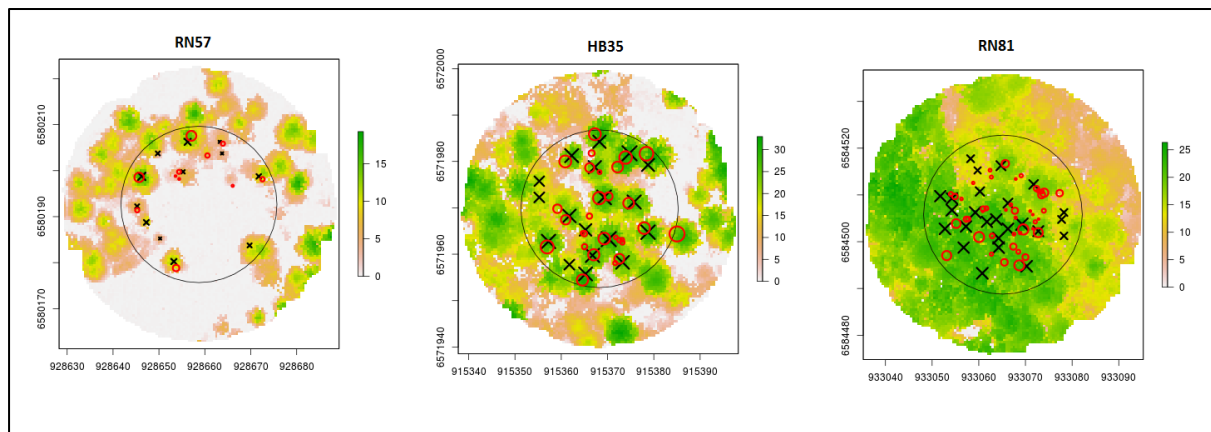
*Table 7 : LiDAR predicted metrics*

Type of metrics	Metric Name	Usual name (in text)	Description
LiDAR and Field vegetation datasets	Gha	Gha	Basal area
	Dmean	Dmean	Mean trunk diameter
	prop_conif	prop_res	Proportion of Coniferous trees (among tree diameter > 17,5, ONF data)
	Nha	Nha	Stem density
	prop_conifG	prop_resGha	Proportion of coniferous trees in basal area (among tree diameter > 17,5, ONF data)

### 3.2.4 Tree detection accuracy

Using the 377 permanent plots presented above, the accuracy of the tree top detection was examined over the Ain study area. In those plots, 14 438 trees were localized and measured. From the first campaign LiDAR dataset 8 005 trees were extracted. This gives a rate of 52.45 % of trees detected.

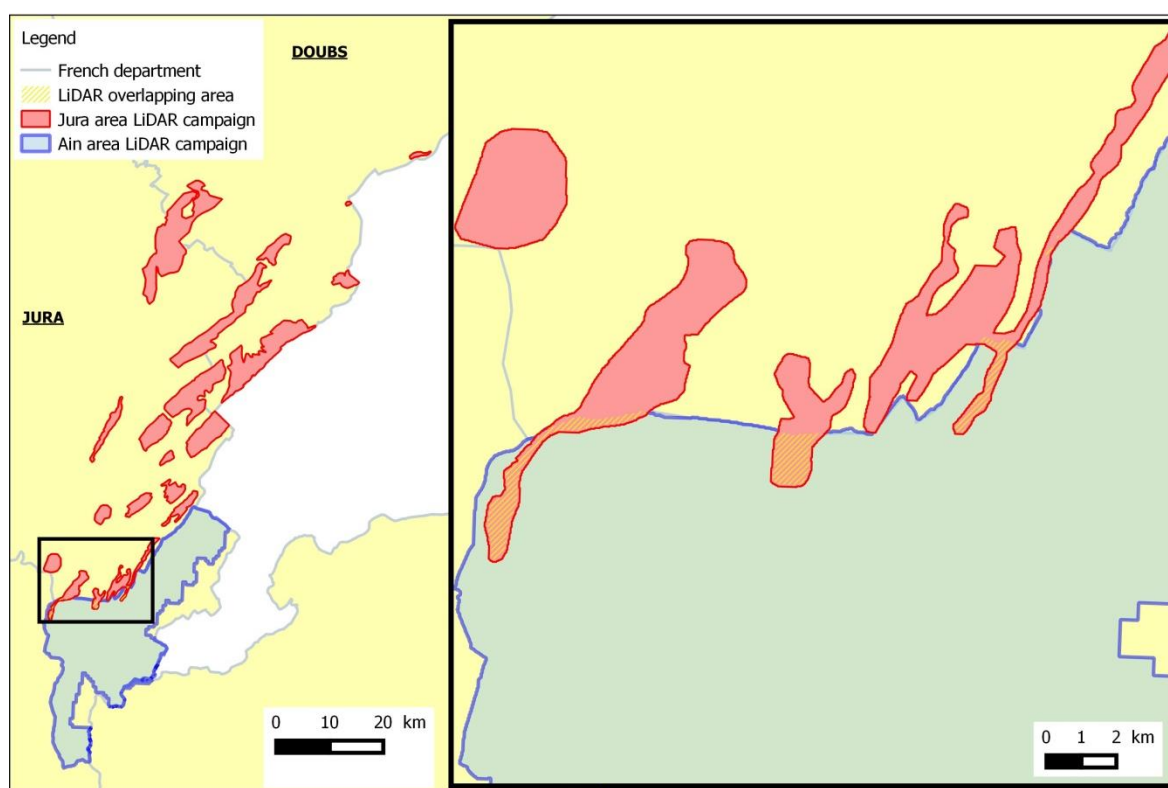
*Figure 17 : Trees localization in the field (red circles) and LiDAR detected trees (black crosses) on three field plots, larger symbols are indicating higher trees.*



### 3.2.5 Comparison of LiDAR metrics between the two campaigns

A small area (4.4km<sup>2</sup>) was covered by the two LiDAR campaigns. This gives the opportunity to compare the transferability of the LiDAR metrics among LiDAR campaigns.

*Figure 18 : Location of the two campaigns overlapping areas.*



For 59 metrics, the correlation between the distributions of values between the two campaigns was calculated (Table 8). We use a Pearson correlation from the function `layerStats` from the package



raster (Hijmans and van Etten, 2014). We first remark that 50 metrics among 59 are highly correlated ( $>0.7$ ). The metrics that represent upper layer metrics showed a better conservation between LiDAR campaigns. This result was expected as metrics calculated on lower height layers are more sensitive to external variation such as the frequency of the laser pulse or canopy cover seasonal variations (Hill and Broughton, 2009). Metrics that count the number of return (number of vegetation or ground return), are depending a lot on the LiDAR dataset, therefore the use of relative metrics should be preferred.

*Table 8 : Correlation between the overlapping areas for each metrics*

Metrics	Correlation (Pearson)	Type	Metrics	Correlation (Pearson)	Type
Canopy3060	0.98	Point-cloud	Canopy1020	0.91	Point-cloud
penetrationratio3060	0.98	Point-cloud	CHM.Perc2_5	0.91	Point-cloud
Q90	0.98	Point-cloud	Tree_density10	0.91	Object-oriented
CHM.PercSup20	0.98	Point-cloud	penetrationratio1020	0.9	Point-cloud
MaxVeg	0.97	Point-cloud	Simpson	0.9	Point-cloud
Canopy2030	0.97	Point-cloud	CHM.PercInf0.5	0.9	Point-cloud
penetrationratio2030	0.97	Point-cloud	Tree_sumS	0.9	Object-oriented
Q75	0.97	Point-cloud	Tree_density	0.89	Object-oriented
SdVeg	0.96	Point-cloud	penetrationratio0102	0.88	Point-cloud
CHM0.sd	0.96	Point-cloud	penetrationratio0205	0.88	Point-cloud
CHM1.sd	0.96	Point-cloud	Canopy0102	0.87	Point-cloud
CHM2.sd	0.96	Point-cloud	Tree_densityinf10	0.85	Object-oriented
CHM4.sd	0.96	Point-cloud	Tree_gini	0.85	Object-oriented
Tree_density30	0.96	Object-oriented	Tree_sdH	0.85	Object-oriented
MeanVeg	0.95	Point-cloud	veglayer3060	0.84	Point-cloud
CHM8.sd	0.95	Point-cloud	penetrationratio0510	0.83	Point-cloud
CHM.mean	0.95	Point-cloud	Canopy0501	0.81	Point-cloud
CHM.PercSup10	0.95	Point-cloud	Tree_meanS	0.8	Object-oriented
Prop_open_area	0.95	Object-oriented	penetrationratio0501	0.78	Point-cloud
Tree_density20	0.94	Object-oriented	middle_gap	0.75	Object-oriented
Tree_meanH	0.94	Object-oriented	small_gap	0.71	Object-oriented
Canopy0510	0.93	Point-cloud	veglayer2030	0.66	Point-cloud
CHM16.sd	0.93	Point-cloud	Q25	0.66	Point-cloud
CHM.PercSup5	0.93	Point-cloud	veglayer0102	0.52	Point-cloud
Tree_sumV	0.93	Object-oriented	veglayer0510	0.52	Point-cloud
Canopy0205	0.92	Point-cloud	veglayer0205	0.51	Point-cloud
Q50	0.92	Point-cloud	veglayer1020	0.42	Point-cloud
CHM.PercInf1	0.92	Point-cloud	GroundR	0.26	Point-cloud
CHM.Perc1_5	0.92	Point-cloud	Q10	0.14	Point-cloud
			nbVeg	0	Point-cloud

### 3.2.6 Create metrics at different scales

All metrics were calculated at different scales. This task was done with a moving window method by using the function `r.neighbors` from GRASS GIS 7.4 (© 2003-2017 GRASS Development Team) via the R package `rgrass7`. The mean and the standard deviation were calculated for each selected scale. To calculate values located at the border the hypothesis that the landscape over the border was similar to the landscape inside the study area was made. Thus the empty half surface was completed by duplicating the data outside the study area with the window width.

### 3.2.7 Non-vegetation LiDAR metrics

Using the digital terrain model, slope, exposition and curvature were calculated for 25\*25m pixels. The slope was calculated in degrees and the exposition in degrees as well where 0° and 360° are representing north. Those variables meant to be used in beech regeneration distribution models.

## 3.3 Synthesis

Numerous variables from LiDAR point-clouds were calculated with success. Among all those metrics, only a few were later selected as variables for Hazel Grouse and Capercaillie and beech regeneration models. This selection process is described in each corresponding chapter 4, 5 and 6.

## 4 BEECH REGENERATION DATASETS

---

To model the beech regeneration spatial distributions over the two study areas, two different vegetation survey datasets were used. The first one was collected by the Groupe Tétrás Jura over the Jura study area. The second one is constituted of vegetation survey plots collected by the ONF in the Ain department.

### 4.1 Groupe Tétrás Jura dataset

This dataset is constituted of multiple surveys (2014 to 2017) which first aimed to collect data on the Capercaillie habitat suitability (Figure 19). Plots were systematically localized using a 100 m width

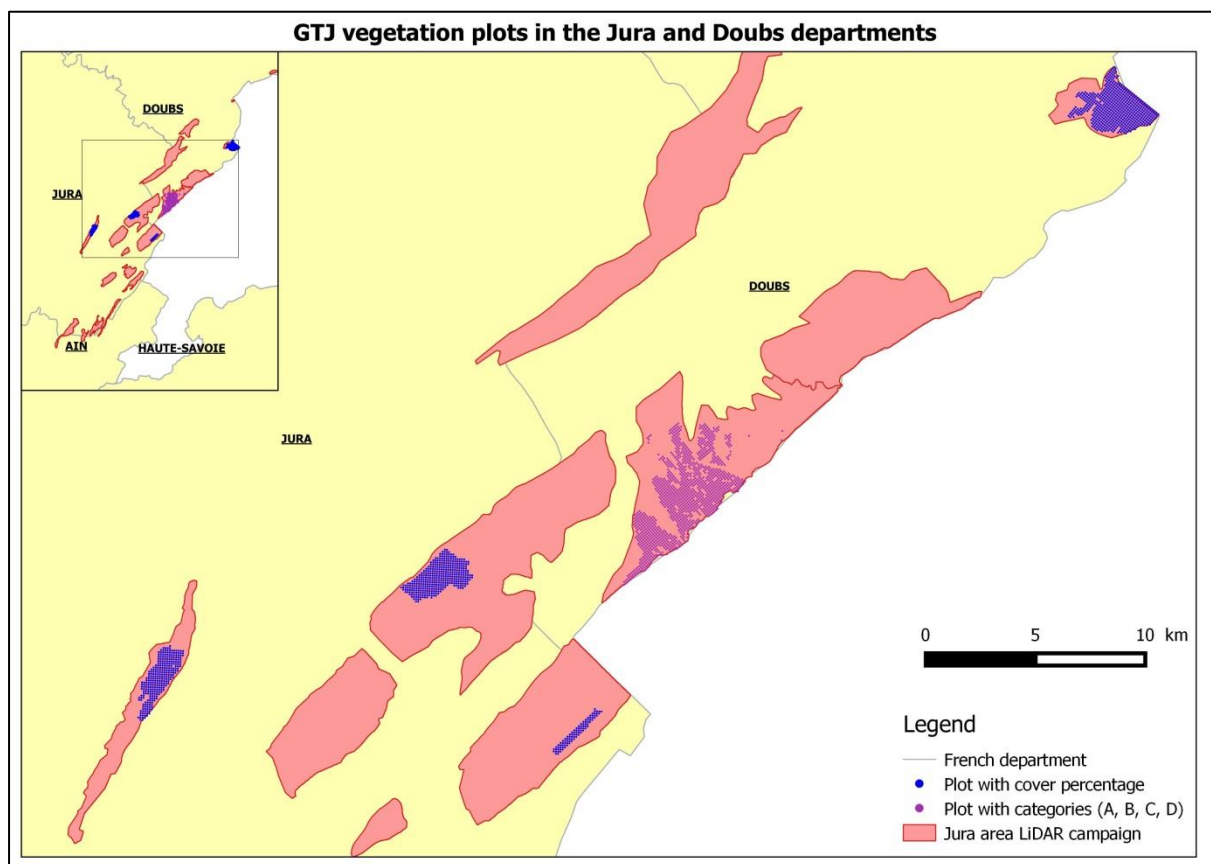


square grid. Two different types of information were collected resulting on differences between surveys. In some surveys the percentage of cover were measured for different species and strata whereas for other four categories indicating the suitability of the different strata for the Capercaillie from poor to good was measured. In the first case 1380 plots were surveyed in four different forest units (Echallon, Chalet neuf, Chaux du Dombief and Mont Noir). In the second case, 1276 plots were surveyed in the Risol forest. In both datasets, the beech regeneration between 50 cm and 3 m was measured. In order to increase the number of plots, percentage of cover were converted to categories A to D to be merged with the second dataset. Indeed, the categories were defined as follow:

- **A:** Good habitat, percentage of beech regeneration cover is lower than 20%
- **B:** Fair habitat, percentage of beech regeneration cover ranges between 20% and 30%
- **C:** Poor habitat, percentage of beech regeneration cover ranges between 30% and 50%
- **D:** Degraded habitat, percentage of beech regeneration cover is higher than 50%

The categories were transformed into a binary variable where the categories D, C and B were assumed to be presence (regeneration cover > 20%) and the category A to be absence.

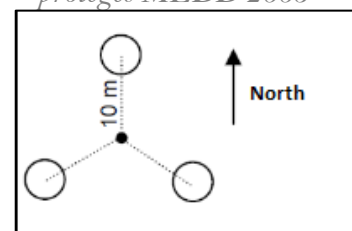
Figure 19 : GTJ vegetation plots



## 4.2 ONF dataset

The ONF constituted a large database where 377 permanent plots were surveyed once in 2008 and a second time in 2015. Those plots were located to cover most of the Ain forests following a systematic sampling (600m width square grid), leading to a large cover. During those surveys, multiple indicators were measured such as species inventory (herbaceous to trees), large tree measurement (diameter, height, species) and regeneration dynamic. To collect data on regeneration, three sub-plots of 1.5 m diameter were surveyed at 10m from the center of the survey plot location (Figure 20). Here the number of individuals was counted by species, and this in three categories: stem from 50 cm to 1.5 m height, stem lower than 1.5 m with a diameter of 2.5 cm and stem with a diameter between 2.5 cm and 7.5 cm. In order to fit better with the data collected by Groupe Tétrás Jura, the three classes were summed together to produce a variable number of beech stem. Afterward, all plots where

Figure 20 : Sub-plots location, Protocole suivi d'espaces naturels protégés MEDD 2005

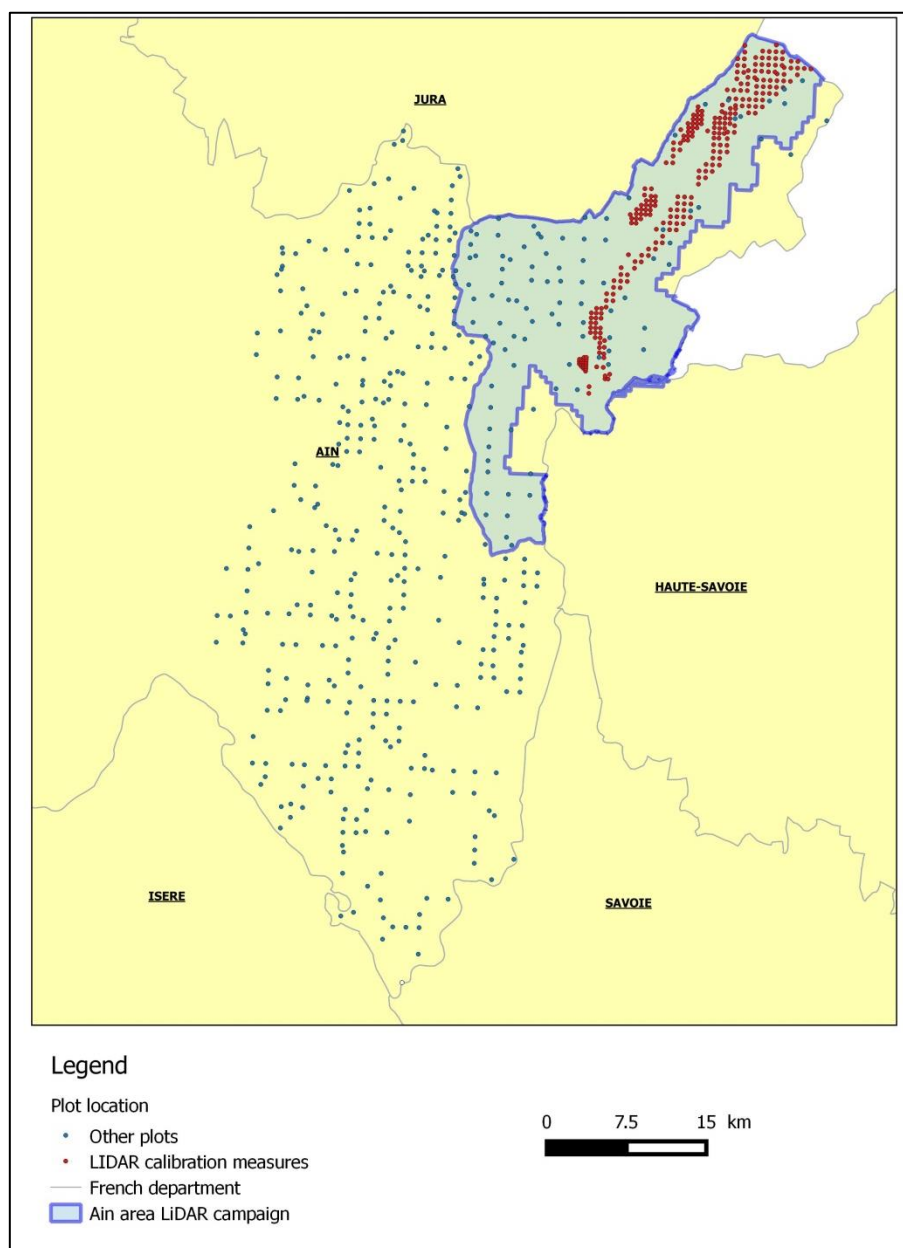






beech regeneration was not observed (thus not present in the database) were selected and were extrapolated to be absence of beech regeneration (0 individuals). Those data were transformed into binary response presence-absence. When the number of individuals was below 5, the value 0 (absence) was attributed and all other values were considered as presence. Furthermore, among those plots 219 were measured with a more precise GNSS receiver in order to calibrate models using LiDAR variables (Figure 21). To conserve the precision of the measure, only those plots were used for beech regeneration models.

*Figure 21 : ONF permanent vegetation plots*



## 5 ADDITIONAL DATASETS

---

### 5.1 Satellite images

For both study area, the Normalize Difference Vegetation Index (NDVI) was extracted from satellite images. Satellite imagery data from Copernicus Sentinel-2 images taken on a clear day were used (17/07/2016 Copernicus Sentinel-2 data 2016 processed at level 2A by CNES for the THEIA Land data center)<sup>11</sup>. Vegetation index are widely used to describe the vegetation characteristics. For this, variations at wavelength red-edge (700 nm) are favored as it is characteristic of green vegetation. The Difference Vegetation Index (DVI) is represented by the difference between the reflectance in near infrared and red. From this, the NDVI is obtained by dividing the difference index by the sum of the near infrared and red reflectance.

### 5.2 Anthropogenic disturbances

In order to account for human disturbances, three different variables were created: distance to roads, distance to ski runs and buildings surface area. The roads were extracted from Route 500® database provided by the French national institute for geographical and forest information (IGN). The selected roads are local, regional, departmental and national asphalted roads which are connected to urban areas or to the principal transport infrastructures. Forest tracks and trails are not a part of this dataset. The distance to the closest road was calculated afterwards for each pixel of the study area. Similarly, from ski runs locations (Groupe Tétrás Jura dataset) minimum distance for each pixel was calculated.

The building locations and surfaces were extracted from the IGN BDTOPO®. This surface was calculated for each pixel. The sum of those surfaces were calculated at different scales and divided by the surface of the selected scale to create the building area proportion variable.

---

<sup>11</sup> <http://www.theia-land.fr/fr/node/816>

# CHAPTER 3: MODELING METHODS



L. Bezombes



## 1 FOREWORD

---

As a first step in the process of using SDMs, the algorithm choice and the global modeling framework should be defined regarding the characteristics of the datasets and the final use of the model results (Elith and Leathwick, 2009). In this chapter, the different modeling methods will be presented and the choices made for both birds and beech regeneration models will be documented. In a first part, a detailed overview of SDMs principles and assumptions will be presented. In a second step, the issues regarding the datasets characteristics will be investigated in a research article submitted to the Journal of Applied Ecology. To provide additional leads the role of sampling design for SDMs improvement will be investigated as well. Finally, the issue concerning the choice of scales will be exposed. For an easier comprehension, the problems related to the choice of environmental variables will not be addressed in this chapter because it is highly dependent on the species habitat needs (Chapter 4, 5 and 6).



## 2 SDMs: PRINCIPLES AND ASSUMPTIONS

---

### 2.1 Ecological niche and habitat definitions

Before going into the principle of SDMs, it is needed to best understand what is driving species distributions. The presence of a species in an area is depending on three pre-requisites (Guisan et al., 2017; Pulliam, 2000; Soberón, 2007):

- **Suitable abiotic conditions**
  - The abiotic conditions represent a combination of all the physical characteristics at a specific place. When those entire characteristics meet the species requirement for their survival and reproduction, it is thus representing the fundamental niche of the species.
- **Suitable biotic conditions**
  - The biotic conditions represent the interactions with the other organisms such as predation, competition or mutualism and commensalism. Those interactions can greatly constrain the communities present at the location.
- **Accessibility to the location**
  - Finally, for an area to be occupied by a species, the location must be reachable: being within the dispersal capacity of the species and surrounded by suitable biotic and abiotic conditions.

The realized niche is by definition the niche occupied by the species given the three constraints presented above. The complexity of those interactions is making it difficult to map the realized distribution without having a complete view of the interactions between the entire biotic and abiotic components that constitute an ecosystem.



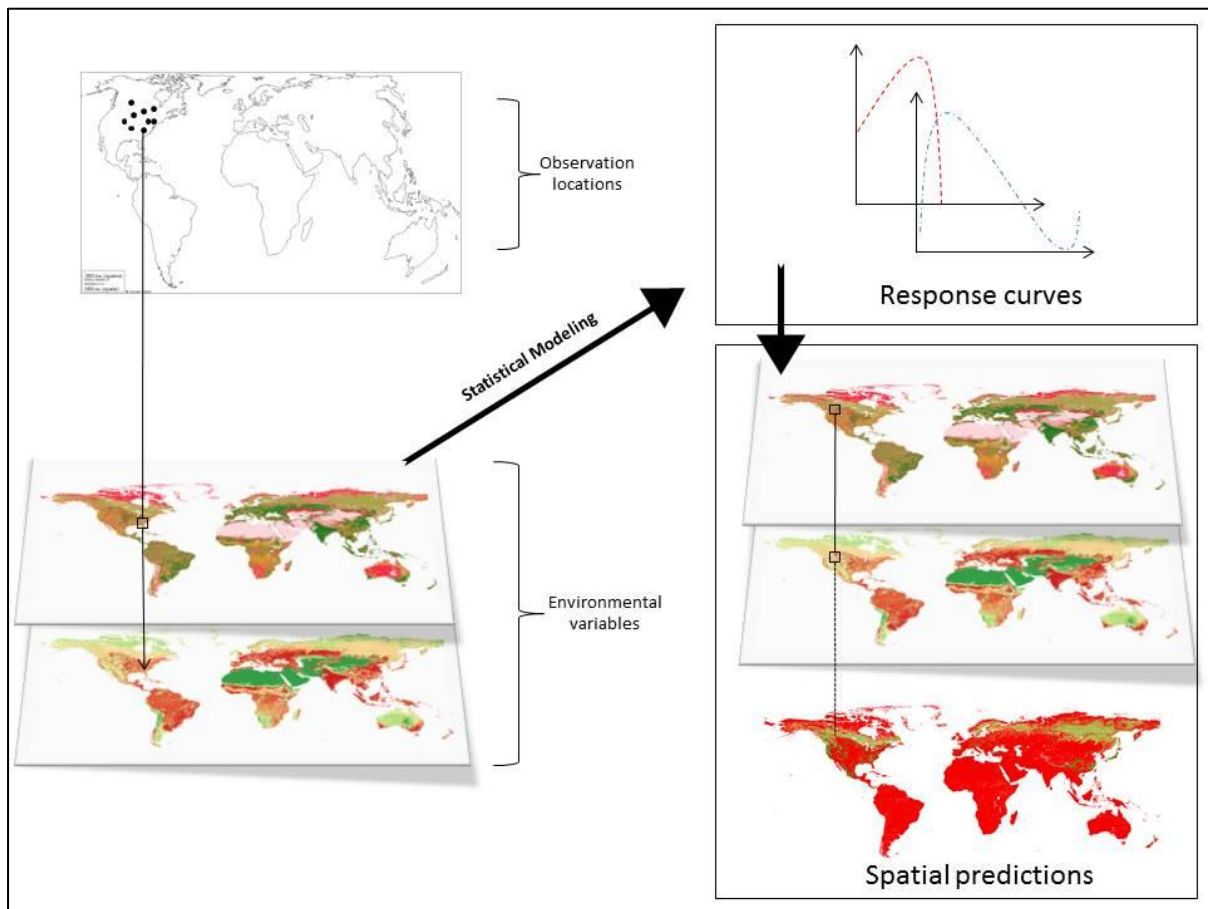
The description of the environmental biotic and abiotic condition at a particular location where a species can or may survive is called habitat (Kearney, 2006). Thus, the habitat is defined by the physical attributes at a locality and includes the interactions with vegetation as a part of the habitat description (e.g. grassland, tropical forest). However, elements such as the interactions with other animal species or the location accessibility are not part of the habitat description.

### 2.2 Concept

The concept of SDMs is to link the environmental characteristics to the presence or abundance of a target species (Franklin, 2009b; Guisan and Zimmermann, 2000). Those models can be used either to explain the causes of the species distribution or to predict this distribution in time or space (Elith and Leathwick, 2009). The species relationship with the environmental variables is done using statistical modeling techniques that fit a response curve to each of those variables (Figure 22). The response curves allowed the quantification of the suitable condition range for the species and consequently are used to explain better the distribution of a species (Elith and Leathwick, 2009; Mac Nally, 2000).



Figure 22 : SDMs concept illustrated with two environmental variables.



The response curves are an essential element that will shape the insight on species preferences and prediction results. Those responses can be of different complexities from unimodal response (suitable, non-suitable) to continual responses (linear, quadratic or more complex responses) (Merow et al., 2014). The choice of the shape of the response is depending on the algorithm and must be selected according to the objectives of the study (explanation versus prediction).

Ultimately the combination of those responses associated with the spatially explicit environmental variables allows the prediction of the species distribution in a geographical space. Those predictions are divided in two categories: models used to predict the distribution to un-sampled sites within a same region (which is the case in this project) or prediction to other geographic areas and/or to future and past climate (Elith and Leathwick, 2009). The latter use is challenging as by definition it does not fulfill the assumptions behind SDMs (Elith et al., 2010), but as we do not aim to produce models projection in other areas or within time, those specificities will not be presented here.

## 2.3 Assumptions

The SDMs are based on different assumptions that must be kept in mind through the modeling process (Guisan et al., 2017):

- Equilibrium
  - A major postulate is to assume that the targeted species is at equilibrium within its environment (Franklin, 2009b; Guisan and Zimmermann, 2000), meaning that the relationship between the species and environmental variables does not change in space or time. The typical case where this assumption is not fulfilled is for invasive species, for which the distribution is under expansion.
- Statistical methods
  - The choice of an appropriate statistical method is essential as different algorithm types lead to different types of response, thus influencing the results (Guisan and Zimmermann, 2000). This choice must also take into account the characteristics of the datasets. The choice of an appropriate algorithm will be presented hereafter (2.4 the choice of the statistical method).
- Target species observations datasets
  - The target species dataset is suitable to answer the research questions: the population dynamics do not influence the individual repartition in space or time. Indeed, the presence of populations outside the fundamental niche of the species, also defined as sink populations (death rate is higher than the reproduction rate) can have important issues on models' relevance. However, such sink populations are difficult to detect within observations data.
  - There is no spatial or temporal bias occurring in the dataset. In order to correctly quantify the habitat suitability using models, the entire range of environmental





combinations available for the species should be sampled. This problem will be addressed in detail in the sections 3 and 4 of this chapter.

- The observations are independent (e.g. not subject to spatial autocorrelation). Spatial autocorrelation is defined when processes and patterns that influence the spatial distribution of a species are themselves spatially structured, or when a species is subject to intrinsic population dynamics (territoriality, dispersal, social interactions). Non independent observations may lead to inaccurate models if not taken into account. However, it is difficult to evaluate whether data are independent or not. The case of Capercaillie, with data collected during the reproduction period when males and females meet at the display place is a good example of auto-correlated observations.

- Environmental predictors

- It is assumed that all environmental predictors impacting the species distribution are available for modeling at the relevant ecological scale. The problem of the choice of environmental variables will be addressed in the next chapters (4, 5 and 6) but the issue with scale selection will be addressed in the section 5 of this chapter.
- Environmental variables do not contain errors. The errors resulting from LiDAR metrics calculation were presented in the chapter 2 section 2.

## 2.4 Link between the niche concept and SDMs

A consensus on the link between the niche theory and models outcomes has not been found yet and the clarification of the niche concept was identified as one major challenge for SDMs by Araújo and Guisan (2006). Indeed, the description of the realized niche faces many difficulties and the use of only environmental predictors only gives the potential habitats for the species as other factors such as biotic interactions and population dynamics are not part of the modeling process (Guisan and



Zimmermann, 2000). Concretely, different consequences on the species spatial distribution can be observed:

- **False positive**, the habitat is suitable but the species is not present because of the influence of one or multiple biotic factors that were not present in the model.
- **False negative**, the habitat is not optimal but the species is nevertheless observed due to fluctuation in population dynamics or due to other biotic factors that were not present in the model.

## 2.5 The choice of the statistical method

### 2.5.1 Objectives of the study

The models for Hazel Grouse and Capercaillie aim to spatially predict their distribution but the interpretation of the response is nevertheless important, even if the ecological requirements for the two species are well known. Indeed, those responses may support the results nearby managers by confirming the importance of certain variables for each species. On the contrary, concerning the beech regeneration models, knowing the effect of each variable was not the priority.

### 2.5.2 Datasets specificities

As stated here above, the different algorithms must also be chosen regarding the characteristics of the datasets (Elith and Leathwick, 2009). Indeed, observations datasets can be divided in three categories:

- **Presence only data**: generally originated from non-predefined sampling surveys such as citizen science. Here only the observation of the targeted species is reported. The Capercaillie and Hazel Grouse datasets are presence only data.
- **Abundance data**: originated from either pre-defined or non-pre-defined sampling, the observer notes here the abundance of the species at each plot. The raw beech regeneration dataset is an abundance dataset (percentage of cover or individual counts).



- **Presence-absence data:** generally collected from pre-defined sampling framework such as random or stratified plot surveys. Here, the observer notes for each plot the presence or the absence of the targeted species. The reclassification of beech regeneration dataset in two categories is thus a presence-absence dataset.

Algorithms handling presence only datasets operate a comparison between the presence observations and the background points (also called pseudo-absence). Such methods include algorithms such as ecological niche factor analysis (ENFA), bioclim, domain, Maxent and logistic regression (Brotons et al., 2004).

Presence-absence and abundance data are analyzed using algorithm such as generalized linear models (GLM), generalized additive models (GAM), classification and regression trees (CART, Random forest, or boosted regression trees (BRT)) (Brotons et al., 2004; Guisan and Zimmermann, 2000).

From those informations, the selection of the algorithm built on the equivalence of Maxent with an infinitely weighted logistic regression for the case of Hazel Grouse and Capercaillie was made (Phillips et al., 2017). Indeed, Maxent is arguably the most frequently used SDM method designed to handle presence only data. It allows the selection of different features defining the response curve shape (linear, quadratic, product, threshold and hinge) (Elith et al., 2011; P Anderson et al., 2006; Phillips et al., 2006).

For the beech regeneration models, three methods were chosen:

- The Random Forest algorithm (Breiman, 2001). It is known to give high classification accuracy and is able to model complex interactions between predictors (Cutler et al., 2007). Random Forest use random subsets of both the environmental predictor variables and species occurrence data and build multiple trees. From those a final tree can be constructed based on an average of all trees (Breiman, 2001).



- The logistic regression (Austin, 1980). It can analyze continuous and discrete environmental variables to predict qualitative response. The response curve represents the probability of the species being present.
- The GLM with Poisson distribution (Vincent and Haworth, 1983). It uses abundance data as input variable, the linear predictor is related to the mean of the response variable by a link function thus constituting a flexible family of regression models (Guisan and Zimmermann, 2000).



### 3 ARTICLE 1: DISTRIBUTION MODELS OF DIMORPHIC SPECIES: CORRECTING SPATIAL SAMPLING BIAS IMPROVES ESTIMATES OF HABITAT ASSOCIATION AND SEX RATIO

---

Anouk Glad<sup>1,2</sup> <https://orcid.org/0000-0001-5810-189X>

Björn Reineking<sup>1</sup> <https://orcid.org/0000-0001-5277-9181>

Jörn Pagel<sup>3</sup>

Jean-Matthieu Monnet<sup>1</sup> <https://orcid.org/0000-0002-9948-9891>

<sup>1</sup> Univ. Grenoble Alpes, Irstea, UR LESSEM, BP 76, 38402 St-Martin-d'Hères, France

<sup>2</sup>Groupe Tétras Jura

<sup>3</sup>Institute of Landscape and Plant Ecology, University of Hohenheim, Stuttgart, Germany

Corresponding author: Anouk Glad, [gladanouk@gmail.com](mailto:gladanouk@gmail.com)

**It was submitted to the Journal Diversity and distribution on September 2<sup>nd</sup> 2018.**

**Supporting information is available in Annex 4.**

#### 3.1 Abstract

Species distribution models (SDMs) are widely used in ecology and conservation. A key challenge for SDMs are presence records from surveys with spatially inhomogeneous sampling intensity, which can strongly impact inference and model accuracy. Several methods have been proposed to correct sampling bias, and accordingly corrected models are generally assumed to accurately represent the species distribution over the study area. However, this assumption may be wrong for species in which subgroups (e.g. by sex or age) differ in their habitat association.

We explored the effect of spatial sampling bias on estimated habitat preferences and sex ratios in dimorphic species, using a virtual species and a case study of a sexually dimorphic forest bird, the Capercaillie (*Tetrao urogallus*). Specifically, we evaluated three sampling designs (uniform random, systematic, and spatially biased) and two correction methods, first on two spatially segregated virtual species sub-groups, and secondly on two Capercaillie presence only datasets from a systematic and a spatially biased sampling design. In each case, occurrence data included information on subgroup identity.

Sampling design had an important effect on the accuracy of parameter estimation for the virtual species: Non-corrected spatially biased sampling resulted in biased estimates of habitat association and sex ratios, but both biases could successfully be corrected with established methods and did not occur for uniform random and systematic sampling. For the Capercaillie datasets, the estimated sex ratio was closer to 50% for observations from the systematic design. However, bias correction methods did not alter the values of estimated sex ratios.

We conclude that current methods for correcting spatially biased sampling can improve estimates of both habitat association and sex ratios for dimorphic species, and we suggest that information on subgroups with distinct habitat associations should be included in SDMs analyses whenever possible.

**Keywords:** sexual dimorphism, species distribution model, spatial sampling bias, sampling design, sex ratio, Capercaillie, virtual species

### 3.2 Introduction

The impact of human activities on the environment at different scales, from local (urbanization, agriculture), to global (climate change) enforces the need to better understand the relationships between species and their environment, and to better predict their distribution. Species distribution models (SDMs) are an important tool in ecology and conservation biology (Elith and Leathwick, 2009; Franklin, 2009a, 2009b; Johnson and Gillingham, 2005), as they quantify the link of organism occurrences to spatial environment characteristics (Elith and Leathwick, 2009). Additionally, SDMs can predict spatial distributions that are used to plan conservation actions, wildlife management and



monitoring strategies (e.g. new sampling designs). Modeling methods need to be chosen according to the scientific question, the specificities of the target species life cycle and, crucially, according to the dataset characteristics (Franklin, 2009a; Guisan and Zimmermann, 2000; Merow et al., 2013).

Occurrence data for distribution modeling comes from many sources, and often do not originate from controlled sampling survey designs such as systematic transects or random plots, but from preferential spatial sampling or opportunistic observations (Geldmann et al., 2016). Data collected without a sampling framework are subject to sampling bias (Guisan et al., 2017), which can be of different sources: uneven record intensity in space or time, uneven sampling effort, and variations in detection efficiency among observers (Geldmann et al., 2016). Here, we focus on the effects of uneven record intensity in space, since spatially biased sampling is a major cause of poor model predictions (Araújo and Guisan, 2006; Guisan and Zimmermann, 2000; Renner et al., 2015), and has been shown to have a strong influence on the accuracy results (Elith et al., 2011, 2006; Phillips et al., 2006). Spatial sampling bias often occurs in the absence of a pre-defined sampling scheme, because observers tend to survey areas depending on their personal preferences (accessibility, higher potential of observations, good knowledge of the study area) (Isaac and Pocock, 2015). As the sampling bias characteristics are specific to each dataset and their influence unknown most of the time, it must be approximated by the modeler. Consequently, the effect of sampling bias on prediction may often be underestimated, as many published results are not corrected for sampling bias before analysis (Yackulic et al., 2013).

Multiple methods have already been tested in order to correct sampling bias and generate better models and predictions. Phillips et al. (2009) proposed the targeted background method, that distributes background points in space with the same bias as the observer process. It has been used in numerous studies (Kramer-Schadt et al., 2013), and was shown to give better results than randomly distributed background points (Warton et al., 2013). However, when the targeted background points are generated over a too restricted area, it can also reduce model accuracy and result in lower prediction performances (Fourcade et al., 2014; Thuiller et al., 2004). (Cardador et al.,



2017) recently proposed to include the potential causes of bias as predictor variables in the models, associated with a random background sampling.

Models successfully corrected for spatial sampling bias and models created from spatially non-biased datasets are thought to represent the species distribution over the study area with the underlying hypothesis that all individuals are sampled with the same probability, thus covering the panel of different habitat needs within the population. This assumption can be wrong in the case of spatially biased sampling involving spatial segregation within species subgroups (e.g. sex, age). Those differences in habitat selection are well known for a variety of species. However, despite this knowledge, the effect of spatial sampling bias on the reliability of predictions when working on spatially segregated sub-groups has been little studied. Conde et al. (2010) investigated the case of the jaguar in Central America and concluded that the use of non-gender model underestimated the effect of fragmentation on female habitat use, which is an important parameter when planning conservation actions. Different seasonal habitat uses were observed for bats (Hayes et al., 2015). Multiple studies showed that differences in habitat preference within sex in avian species exist and depend on multiple parameters like dispersal capacities, food needs, mating system, morphology (Cody, 1985).

A known example for different habitat preferences between males and females is the Capercaillie (*Tetrao urogallus*). Given their size, weight, and color, the male's strategy to avoid predation is known as "detect predator and escape" or "detect predator and self-defense", whereas females prefer to hide in dense vegetation in boreal forests (Gjerde and Wegge, 1989; Rolstad, 1988; Rolstad et al., 1988). The same tendency where males use more open areas than females was observed in the Jura massif and in the Alps (Storch, 1993; Thiel et al., 2007). Management guidelines tend to highlight only the male's preferred habitats (Montadert. M. pers. comm.). In addition, the fact that presence signs are easier to find in more open areas associated with an underestimation of the differences in habitat use between sexes may have lead observers to intentionally concentrate the sampling effort on those habitats preferred by males, as e.g. exemplified by a study on Capercaillie genetics by Mollet et al., 2015 where observers were encouraged to focus on roosting and feeding





trees, on hiding sites, on internal forest edges, and on root plates and on tree stumps. Indeed, assuming that the true sex ratio is 0.5, in that study the sex ratio was biased toward males, reflecting the low proportion of female signs found through this sampling protocol (Mollet et al., 2015). A second study, on the Cantabrian Capercaillie (*Tetrao urgallus cantabricus*) sub-species found the same results (Morán-Luis et al., 2014).

In this study, we evaluate the effect of uneven spatial sampling, using different survey design and correction methods in the case of spatially segregated sub-groups using both a virtual species, and Capercaillie as a highly sexually dimorphic case study species (Catry et al., 2004; Ruckstuhl and Neuhaus, 2000; Wolf et al., 2005).

We hypothesize that using spatially biased sampling leads to (H1) biased estimates of habitat association and (H2) to a biased subgroup ratio within the virtual population, but that these biases can be corrected by the two methods introduced above (H3). In addition, from our knowledge of previous studies on Capercaillie habitat use, we hypothesize that, in the absence of a systematic survey protocol, (H4) the datasets will be biased toward males, but that (H5) sampling bias correction methods and the implementation of a systematic sampling scheme can correct the sex ratio estimation. The general hypotheses H1-H3 were assessed using a virtual species approach, whereas Capercaillie-specific hypotheses H4-H5 were tested with empirical data.

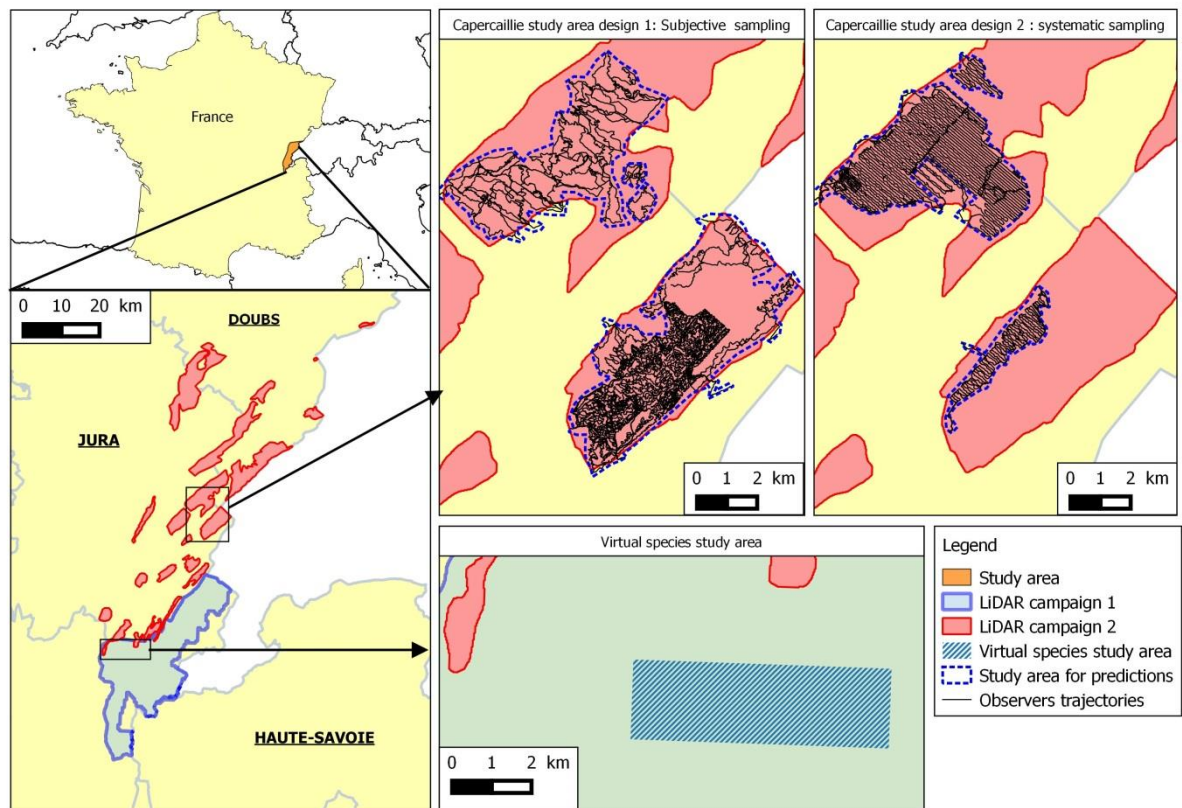
### 3.3 Material and methods

#### 3.3.1 Environmental variables

The environmental variables used in the study for the both virtual species and Capercaillie models were derived from LiDAR datasets stemming from two acquisition campaigns over two distinct areas (Fig. 23). In the first area, the LiDAR campaign was conducted in fall 2014 and covered 626 km<sup>2</sup> over the Ain department (France). The second LiDAR campaign was conducted in summer 2016 and covered a surface area of 431 km<sup>2</sup> over two other French departments, the Jura and Doubs. LiDAR point-cloud metrics were calculated with R package *lidaRtRee* (<https://gitlab.irstea.fr/jean-matthieu.monnet/lidaRtRee>). First, LiDAR raw point clouds were normalized with the Delaunay

interpolation method from the LidR package (Roussel and Auty, 2016). Point-cloud summary metrics such as mean point density between 10-20 m and between 20-30 m, maximum height, Simpson height index and slope were calculated within pixels of size 25 m \* 25 m.

*Figure 23 : Study areas with LiDAR surveys, Capercaillie sampling designs and localizations for virtual species case study.*



### 3.3.2 Capercaillie case study

#### Study area

The study area is located in the French Jura Massif, within the departments of Doubs and Jura (Fig. 1). The landscape is composed of a mosaic of small urban areas, pastures, forests and fields. The massif is composed of a low plateau (elevation range from 400 m to 700 m a.s.l.) and a high plateau (elevation range from 700 m to 1620 m a.s.l.). The forest covers a surface of 480 000 ha in the Doubs and Jura departments. Here, we will focus on two forests (forêt du Risoux and forêt du Mont Noir) that were surveyed using both subjective and systematic surveys (Fig. 1).



### Datasets description

- Design 1: Subjective sampling

Long-term Capercaillie winter surveys were organized without a strongly prescriptive framework between 2007 and 2015 by the “Groupe Tétras Jura” non-governmental organization (NGO). Observers surveyed forests known as favorable for Capercaillie, by navigating through the focus area according to observers’ preferences. Using a Global Navigation Satellite System (GNSS) receiver, they recorded their survey trajectories and Capercaillie sign locations (mostly faeces, prints and feathers). Whenever possible, sex was assigned to each observation. This can be done with a fairly high accuracy due to animals’ high sexual dimorphism; according to preliminary results from a genetic analysis signs such as droppings can be visually assigned to sexes with a 90% accuracy (A. Depraz pers. Comm.). All non-sex assigned observations were removed, and the 318 remaining observations were used for the analysis (211 males and 107 females).

- Design 2: Systematic sampling

The implementation of a new survey protocol in the winters of 2016 and 2017 for a large scale Capercaillie genetic survey involved the collection of droppings following a standardized pre-defined path. Transect trajectories were separated by 80 m, and observers were requested to stay within 20 m of their assigned transect. As in design 1, real observers’ trajectories as well as sign locations were recorded using a GNSS receiver. 29 observations where the sex was determined were collected (17 males and 12 females).

#### 3.3.3 Virtual species case study

Here, a virtual species showing sexual differences in habitat preferences distribution was simulated. 100 simulations of presence were created from this distribution and a virtual data collection according to three sampling schemes was applied. Then SDMs were parametrize using three different methods.



### Species distribution intensity

Virtual species presence only datasets were created following an inhomogeneous Poisson process model using three environmental variables with low collinearity. All variables were normalized by subtracting mean pixel value and dividing by the standard deviation before modeling. The intensity  $\lambda(s)$  represents the expected number of individuals per unit (pixel) at position  $s$ , and it is computed as:

$$\ln \lambda(s) = \beta_0 + \sum_{p=1}^p x_p(s) \times \beta_p$$

where  $x_p(s)$  represents the environmental covariate  $p$  at location  $s$ , and  $\beta_p$  are the model coefficients associated with each environmental variable, and  $\beta_0$  denotes the intercept term (Renner et al., 2015). The coefficients  $\beta_0$  and  $\beta_p$  were chosen in order to obtain intensity values per pixel between 0 and 2 (Table 1). From the intensity map, the number of individual presences per pixel was independently drawn from a Poisson distribution with mean  $\lambda(s)$ .

As we study the influence of species subgroups (such as the sexual dimorphism effect), two occurrence distributions were created, one for each subgroup with different  $\beta_p$  coefficients. Those groups will be hereafter referred to as male and female. The intercept  $\beta_0$  was set for male in order to have a known sex ratio of 0.5 in the study area ( $(\text{Nobs females})/(\text{Nobs females} + \text{Nobs males})$ ) (Table 9).

The three environmental variables selected to create those distributions were: the Simpson index of vegetation heights, the canopy density between 10-20 m, and slope. Together, these variables capture different aspects of the landscape (topography, vegetation density, and vegetation height heterogeneity). The correlation between those variables was low: 0.18 (Slope/Simpson), 0.12 (Simpson/Canopy density) and 0.21 (Slope/Canopy density). Those variables were used on a study area of 13 km<sup>2</sup> (20 800 pixels), located in the French Ain department (Fig. 23).

*Table 9 : Parameter values for the virtual species sub-groups 1 (female) and 2 (male)*

	Slope	Canopy density 10-20m	Simpson index	Intercept
Sub-group 1	-0,5	-0.3	-1	-3.8
Sub-group 2	0.01	-0.02	0.08	-1,0

### Modeling

Maxent it is arguably the most frequently used SDM method designed to handle presence only data (Elith et al., 2011; P Anderson et al., 2006; Phillips et al., 2006). Here, we use a recent reimplementaion of the algorithm building on the equivalence of Maxent with an infinitely weighted logistic regression (Maxnet package; Phillips et al., 2017). All analyses were conducted using R (R Core Team, 2016) The same environmental variables used to create the virtual species distribution were used as predictors for modeling.

### Virtual sampling

In order to obtain different virtual species sampling datasets, three basic sampling designs (random, systematic, and non-systematic) were applied for each species subgroup (Fig. 24). The random sampling was done using homogenous sampling within the entire area, using 2000 random points. The systematic sampling was created using parallel transects separated by 200 m generated with the DSSim R package for a total length of 43,9km. Concerning the non-systematic sampling design, subjective trajectories of real Capercaillie surveys over the selected area were used (total length 31.2km). The comparison of the density curves among the sampling design showed the presence of a spatial sampling bias with non-systematic design.

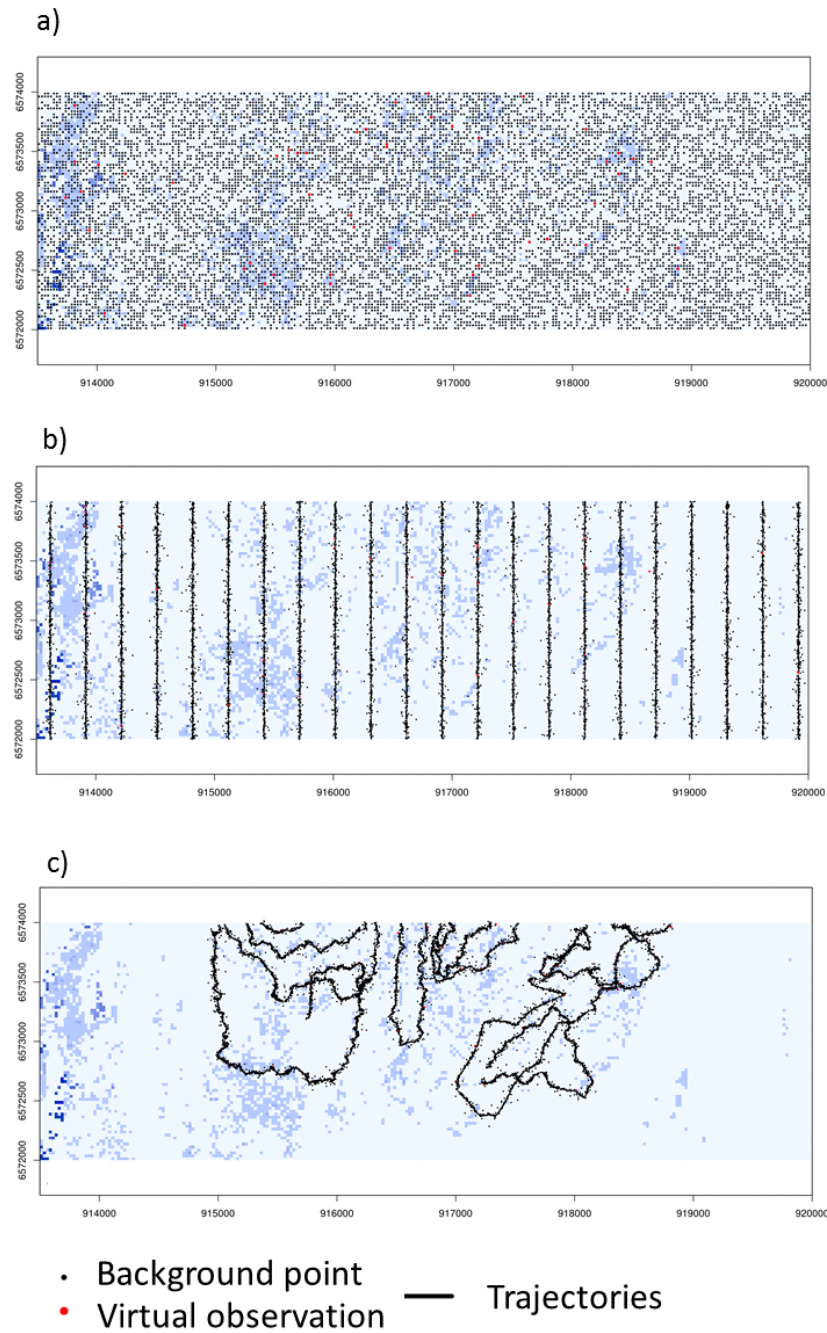
In a first step, for all cases with known trajectories (i.e. the two non-random designs), the observation dataset was created from the realized presences (or signs) distribution by adding a distance-dependent detection probability – the detection probability was maximal on the tracks and the diminished with increasing distance from the tracks. The detection probability followed a logspline distribution fitted to empirical sign location distances to their corresponding trajectory. In a second step, for all cases, a random binomial draw was applied, with the virtual species count taken



as the number of trials per pixel and distance-dependent detection probability as probability of success. The resulting count data was converted into presence only records, by removing the locations with no observations. Locations with more than one observation were replicated according to the number of signs found at the location.



Figure 24 : Overview of the three-sampling designs for the virtual species case study over the intensity distribution map. a) represents the random sampling design and random background point, b) represents the transects sampling design and targeted background point sampling, c) represent the subjective sampling design and targeted background points.



#### Bias correction methods

Two different bias corrections methods were applied to systematic and non-systematic sampling designs: targeted background points and a distance-to-trajectory variable. Both were compared with



a non-corrected model with uniform randomly localized background points. For the targeted background point method, 10 000 background points were randomly sampled along transects, where the frequency of distances between each background point and transect followed the same logspline distribution as the one used to generate the detection probability. The second bias correction method was applied by using the distance to trajectories as a predictor variable in the model, combined with 10 000 uniform randomly distributed background points. The location of the background points were the same across models replicates.

### Parameter estimation

To control the effect of two bias correction methods on model accuracy, the estimations of  $\hat{\beta}_p$  from the fitted models were compared to the true values of  $\beta_p$  used to generate the intensity map for one virtual sub-group. The mean estimation was compared to the true parameter value with a Wilcoxon rank sum test.

### Species subgroups ratio estimation

The species subgroups ratio, here interpreted as sex ratio, was calculated with two approaches: first, from the number of observations sampled from the virtual sampling and second from the sum of model predicted relative intensities  $\hat{\lambda}(s)$ . For the random background point and targeted background point models, predictions were made for the study area using the three environmental variables. For the distance to trajectories method, distance values were set to 0 in order to predict the potential range of the species. The sex ratio estimates were compared with the true sex ratio calculated from the 100 distributions of signs generated over the landscape using a Wilcoxon rank sum test.

#### 3.3.4 Capercaillie case study

##### Modeling

The same modeling method as the one applied for the virtual species study case was chosen to build Capercaillie models (section on modeling and bias correction). Three environmental variables were selected, describing the characteristics of the canopy cover and the vertical structure known to be





essential for Capercaillie (Graf et al., 2009): the density of the canopy between 10-20 m and 20-30 m as well as the Simpson index for vegetation heights. The two bias correction methods presented here above were used and compared with random background point models. The sex ratio was calculated as described in the virtual species case. The differences in ratio of female observations between the two sampling designs was tested using an Exact Rate Ratio Test, assuming Poisson counts.

## 3.4 Results

### 3.4.1 Virtual species case study

#### Parameter estimation

Parameters were estimated for both virtual species sub-groups, for all three environmental variables. Only the results of the first sub-group are presented here (for the second sub-group see Annexes III Supporting information Fig. S1; Table S1). Non-corrected models and corrected models mean parameter estimates were compared with the true parameter value (Fig. 25). The estimated parameters differed significantly from the true value in most cases, there was considerable variations in the size of this discrepancies across scenarios. Parameter estimation for the random design and random background point showed a mean of estimated parameters close to the real parameter value, for the three variables in the sub-group 1 case but not different for slope only. When using the transect design, the estimated values of Simpson were not non-significantly different in the case of targeted background points (Table 10). However, all other estimated values were significantly different from the true parameter value. Nevertheless, minor variations are observed between the three correction cases, and estimated values are still relatively close to the true parameter value. For the subjective design, when no correction is applied (random background points), the deviations of estimated parameters from the true value were much larger. This discrepancy was reduced for the three variables by the two correction methods, with no significant differences observed for the slope variable anymore. Finally, an increase of the variance when estimating the parameters in the case of sampling bias correction was observed.

Figure 25 : Parameter estimations (100 models) for sub-group 1 (female)

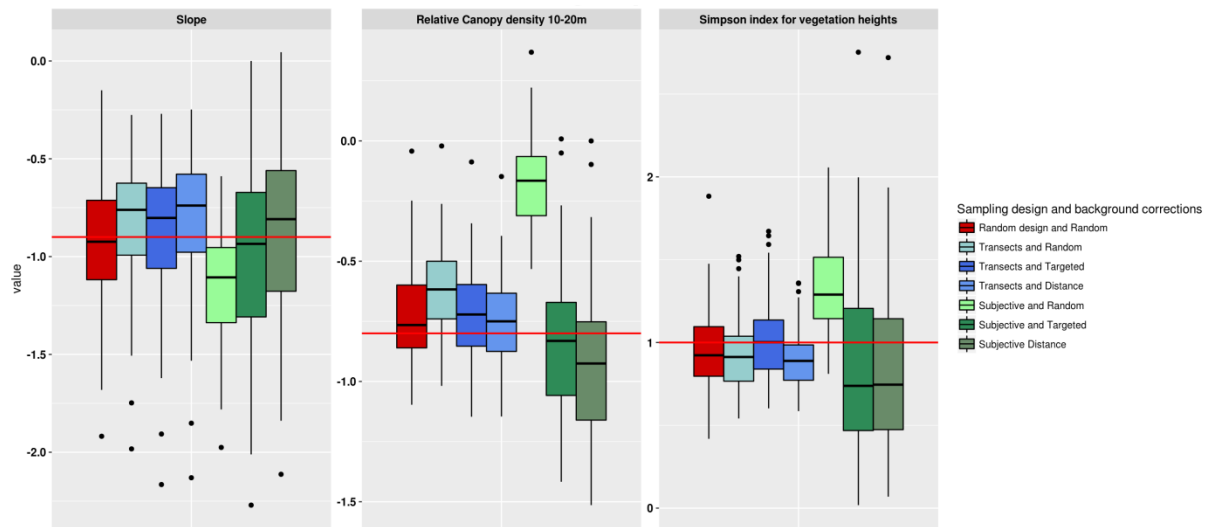


Table 10 : Wilcoxon test p-values on parameter estimation mean values

Case	Slope	Canopy	Simpson
<b>Sub-group 1</b>			
Random and Random	0.55	0.0035	8e-04
Transects and Random	0.00069	1.5e-14	0.00011
Transects and Targeted	0.054	6.2e-05	0.47
Transects and Distance	8.2e-05	0.0034	8.9e-08
Subjective and Random	2e-13	2.6e-17	6.4e-17
Subjective and Targeted	0.14	0.046	0.006
Subjective and Distance	0.24	3.1e-06	0.0021

#### Species subgroups sex ratio estimation

For the systematic sampling design, the estimated sex-ratios, based on either raw observations uncorrected SDMS were already very close to the true value, but the use of any of the two corrections method still improved this estimation (Fig. 26. Table 11). However, for the subjective design, the sex-ratio is highly biased in favor of one subgroup when the sex ratio is estimated from the raw number of observations or from intensity maps predicted from a non-corrected model. When a correction method is applied, the estimated sex ratio is better estimated in the case of targeted background, but not significantly different (p-value of 0.47), whereas the mean value observed with the distance

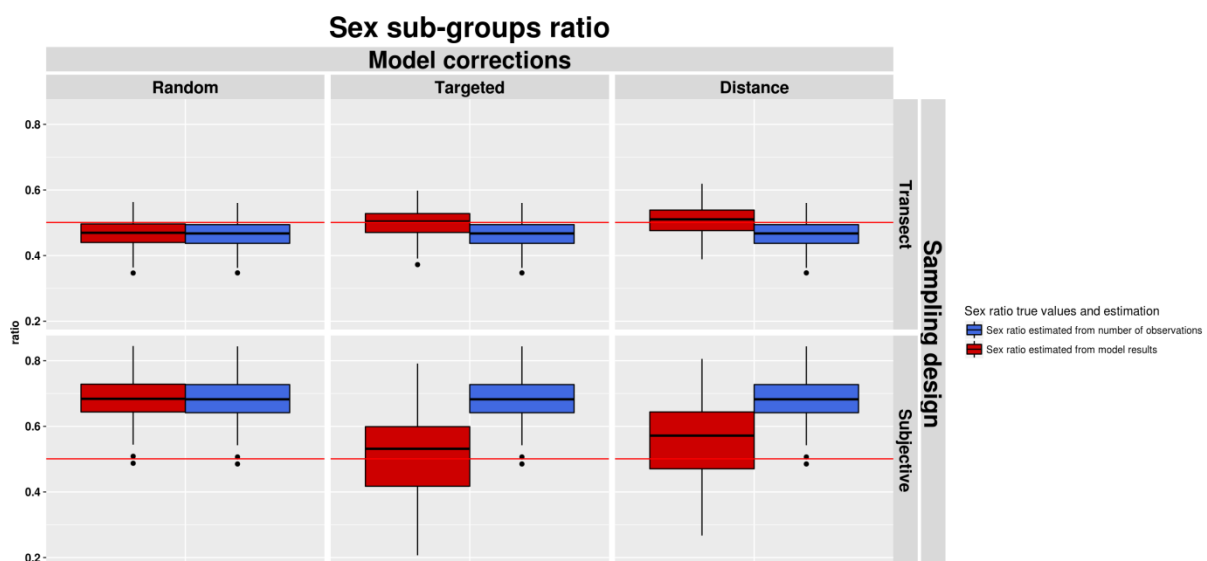


to trajectories method is closer to the true ratio value and significantly different (p-value 0.0012). As for the estimation of the variable parameters, the variance observed within the distribution of the ratio estimation increases when a correction method is applied.

Table 11 : P-values from the Wilcoxon test on the ratio estimation distribution.

Case	Estimated ratio
Transects and Random	1.8e-10
Transects and Targeted	1
Transects and Distance	1
Subjective and Random	1.7e-31
Subjective and Targeted	0.47
Subjective and Distance	0.0012

Figure 26 : Sex ratio estimated in the virtual case study from the number of observations and from predictions of SDMs that use different sampling designs and correction methods. The red line represents the mean sex ratio calculated from all simulated presences.

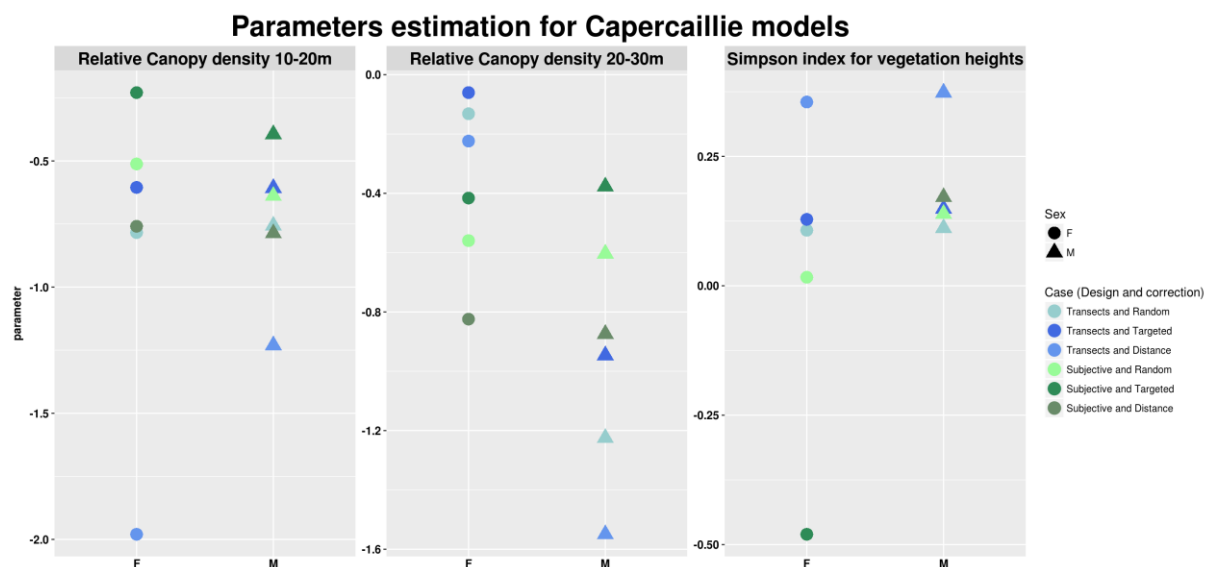


### 3.4.2 Capercaillie case study

Parameter estimation was compared between the modeling cases and the sexes. Variations among the different sampling designs and corrections were observed for both sexes, but no clear pattern regarding sampling design or the bias correction was found. For the relative canopy density (20-30m)

variable, a difference between males and females was observed in the case of systematic sampling: this pattern was not found for the values estimated from subjective sampling (Fig. 27).

*Figure 27 : The parameters estimation of Capercaillie for each sex in function of the sampling design and applied corrections.*



The sex ratio was closer to the value of 0.5 in the case of systematic sampling, in all cases (Table 12). This ratio was lower in the case of subjective sampling in all cases. Nevertheless, the difference in ratio of female observations between the two sampling designs was not significant ( $p$ -value = 0.58). No effect of spatial-bias correction methods on the sex ratio estimate was found, i.e. corrected models predicted a similar sex-ratio than non-corrected models.

*Table 12 : Sex ratio estimates for Capercaillie*

	Number of observations	Non-corrected	Targeted background	Distance to trajectories
Systematic	0,41	0,41	0,42	0,41
Subjective	0,34	0,34	0,35	0,34

### 3.5 Discussion

The aim of this study was to evaluate the effect of different sampling designs on species sub-group ratio estimations. Two bias correction methods were tested to correct bias sampling schemes on two sampling designs, i.e. a transect and a subjective sampling design.



The sampling design had an important effect on the accuracy of the model results in our study, in accordance with previous research (Thibaud et al., 2014). As expected, the systematic sampling design allowed a better estimation of the parameters with or without corrections for potential spatial bias, as it covers a larger part of the landscape variations constraining the observer to also visit areas with expected lower habitat quality such as denser vegetation cover (canopy or understory), thus reducing possible spatial bias (Cardador et al., 2017). The use of real observer trajectories to sample a virtual species confirmed that a non-negligible sampling bias can occur in the absence of a pre-defined sampling design. The Knowledge of observer trajectories during their survey associated with one method of bias correction allowed a better estimation of each variable parameter in all cases. These results are highlighting the importance of having a good knowledge of the sampling effort will help to produce more accurate models (Kramer-Schadt et al., 2013; Ranc et al., 2017). However, those better estimations are made at the cost of an increase of the variance of the parameter estimations. The well-known targeted background point method performed equally well compared to the alternative method using “distance-to-trajectory” as predictor variable combined with random background points. However, differences between these methods might be more pronounced in different settings, as the overall effect of bias correction also depends on the modeling technique (Thibaud et al., 2014), and the occurrences of the species (wide or narrow range) (Ranc et al., 2017), and the sample size (Thibaud et al., 2014) .

The estimation of the ratio between species sub-groups using a virtual species showed that a major bias can occur when two subgroup habitat preferences differ. The use of systematic sampling allowed a correct estimation of the ratio, which was even better when correction methods were applied. However, when the observations were sampled using a subjective design, the ratio was highly biased toward one subgroup. The application of the two correction methods allowed a more reliable but not fully corrected estimation of this ratio as Syfert et al. 2013 observed in their study. This shows that if the implementation of a systematic design is not possible, the application of a correction method such as the two tested in this study is still a good option. As expected, our results indicate that if the difference between sub-groups is not taken into account when the dataset is

spatially biased, it is likely that the full range of the potential habitat will not be covered. This can greatly impact future conservation actions (creation of protected areas, habitat restoration) by lowering the importance of one type of habitat or the impact of anthropogenic activities on a part of a population (Conde et al., 2010; Jiménez et al., 2017). Still, if models are developed for each subgroup separately with a correction for spatial sampling bias, those differences in habitats may appear clearly by comparing the response curves and the predicted maps.

In our study case on Capercaillie, the sex ratio was leaning toward males for both sampling designs. This ratio was closer to the value for a balanced sex ratio for the observations collected with the systematic design. Proportionally more female signs were collected with this protocol than with the subjective one. However, the number of observations available for the systematic design was very low (29), thus more data are needed to confirm this pattern. Nevertheless, the ratio toward males observed with systematic sampling (0.59) was slightly lower than the one reported by previous studies (0.63 and 0.625) that used non-systematic sampling (Mollet et al., 2015; Morán-Luis et al., 2014). Neither of the two correction methods had an effect on the estimated sex ratio. This may be due to multiple factors, such as the limited number of observations, the pertinence of the environmental variables used in the model (Johnson and Gillingham, 2005) which can influence the overall model accuracy. The productions of overall inaccurate models may limit the effects of the corrections methods. We can also suppose that a lower than expected difference in habitat use between sexes is occurring in Capercaillie populations. In addition, though the same forest areas were surveyed, some parts of the forest were surveyed only with systematic or only with subjective design (Fig. 23), limiting the direct comparability of the sex ratios.

Even if the differences in habitat preferences between sexes for Capercaillie were small in this study, many other species show more differentiated habitat use that could affect model accuracy significantly. This was observed in mark-recapture studies (McKnight and Ligon, 2017) and telemetry studies (Kolts and McRae, 2017). While systematic protocols are more robust to differences between subgroups, their use is unlikely to solve the entire problem of biased detectability. Despite the spatial sampling bias, difference in detection probability can still introduce bias towards one sub-group in



models (Guillera-Arroita, 2016; Lahoz-Monfort et al., 2014). This limitation is well known in studies that aim to estimate population size (McKnight and Ligon, 2017). In the specific case of Capercaillie, differences in behavior related to the higher risks of predation on females, due to their smaller size. Their presence in denser vegetation cover may induce a lower detection rate in such environments, combined with a tendency to avoid long movements, and thus inducing a more restricted spatial distribution of droppings, further reducing the chance of observing signs (Mollet et al., 2015). Furthermore, it has been observed that females in winter have a gregarious behavior, leading to the presence of one to three individuals in a single roosting tree which cannot be reliably detected by the observation of signs only such as dropping (A. Depraz, pers. comm.).

### 3.6 Conclusion

The use of non-systematic sampling protocols can be a problem when working with species where sub-groups have different habitat use behavior. When using a non-systematic sampling design, the habitat preferred by one sub-group can be underrepresented in the dataset, rendering the predictions from SDMs biased. The two spatial bias correction methods tested here were able to correct this effect as evidenced in a virtual species study case, but not in the case of a dimorphic forest bird, Capercaillie. We argue that spatial correction methods can be useful when developing SDMs for sub-groups with distinct behaviors. But their effectiveness cannot be taken for granted and other potential sources of bias, such as sex-differences in detectability, require additional attention.

### 3.7 Authors' Contributions

A.G. co-designed the methodology, analyzed the data and wrote the manuscript. J.-M. M. J.P, and B.R. designed the methodology, and contributed to model development and data analyses. J.-M. M. developed the original code for LiDAR variable extraction. All authors contributed to and gave approval of the final manuscript.

### 3.8 Acknowledgements

The first LiDAR survey was financed the by European Territorial Cooperation program INTERREG IV France-Switzerland 2007-2013 and the Ain department. The second LiDAR campaign was financed by the Bourgogne Franche-Comté regional council. The project was financed by the departmental council of Ain. We would like to thank Alexandra Deparz and Marc Montadert for constructive discussions, collaboration and support.

### 3.9 Data accessibility

Data and R code are available at the repository below. Observer names were removed from Capercaillie observations and track datasets due to privacy legislation. The reviewers may access it by using temporarily a guest account (please fill the IDs in the Standard tab):

[https://gitlab.irstea.fr/users/sign\\_in](https://gitlab.irstea.fr/users/sign_in)

account: grouse.project

password: grouse.project

### 3.10 Bibliography

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>

Cardador, L., Diaz-luque, J.A., Hiraldo, F., Gilardy, J.D., Tella, J.L., 2017. The effects of spatial survey bias and habitat suitability on predicting the distribution of threatened species living in remote areas. *Bird Conserv. Int.* 1–12.

Catry, P., Campos, A., Almada, V., Cresswell, W., 2004. Winter segregation of migrant European robins *Erithacus rubecula* in relation to sex, age and size. *J. Avian Biol.* 35, 204–209. <https://doi.org/10.1111/j.0908-8857.2004.03266.x>

Cody, M.L., 1985. *Habitat selection in birds*. Academic Press.





Conde, D.A., Colchero, F., Zarza, H., Christensen Jr., N.L., Sexton, J.O., Manterola, C., Chávez, C., Rivera, A., Azuara, D., Ceballos, G., 2010. Sex matters: Modeling male and female habitat differences for jaguar conservation. *Biol. Conserv.* 143, 1980–1988. <https://doi.org/10.1016/j.biocon.2010.04.049>

Elith, J., H. Graham\*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* 40, 677.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57.

Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS One* 9, e97122.

Franklin, J., 2009a. Mapping species distributions: spatial inference and prediction. Cambridge University Press.

Franklin, J., 2009b. Moving beyond static species distribution models in support of conservation biogeography. *Divers. Distrib.* 16, 321–330.

Geldmann, J., Heilmann-Clausen Jacob, Holm Thomas E., Levinsky Irina, Markussen Bo, Olsen Kent, Rahbek Carsten, Tøttrup Anders P., Leung Brian, 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers. Distrib.* 22, 1139–1149. <https://doi.org/10.1111/ddi.12477>



Gjerde, I., Wegge, P., 1989. Spacing pattern, habitat use and survival of capercaillie in a fragmented winter habitat. *Ornis Scand.* 219–225.

Graf, R.F., Mathys, L., Bollmann, K., 2009. Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps. *For. Ecol. Manag.* 257, 160–167.

Guillera-Arroita, G., 2016. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40, 281–295. <https://doi.org/10.1111/ecog.02445>

Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat suitability and distribution models: with applications in R.* Cambridge University Press.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.

Hayes, M.A., Cryan, P.M., Wunder, M.B., 2015. Seasonally-dynamic presence-only species distribution models for a cryptic migratory bat impacted by wind energy development. *PloS One* 10, e0132599.

Isaac, N.J., Pocock, M.J., 2015. Bias and information in biological records. *Biol. J. Linn. Soc.* 115, 522–531.

Jiménez, S., Domingo, A., Brazeiro, A., Defeo, O., Abreu, M., Forselledo, R., Phillips, R.A., 2017. Sexual size dimorphism, spatial segregation and sex-biased bycatch of southern and northern royal albatrosses in pelagic longline fisheries. *Antarct. Sci.* 29, 147–154.

Johnson, C.J., Gillingham, M.P., 2005. An evaluation of mapped species distribution models used for conservation planning. *Environ. Conserv.* 32, 117–128.

Kolts, J.R., McRae, S.B., 2017. Seasonal home range dynamics and sex differences in habitat use in a threatened, coastal marsh bird. *Ecol. Evol.* 7, 1101–1111.



Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366–1379.

Lahoz-Monfort, J.J., Guillera-Aroita, G., Wintle, B.A., 2014. Imperfect detection impacts the performance of species distribution models. *Glob. Ecol. Biogeogr.* 23, 504–515.  
<https://doi.org/10.1111/geb.12138>

McKnight, D.T., Ligon, D.B., 2017. Correcting for unequal catchability in sex ratio and population size estimates. *PloS One* 12, e0184101.

Merow, C., Smith, M.J., Silander, J.A., 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36, 1058–1069.  
<https://doi.org/10.1111/j.1600-0587.2013.07872.x>

Mollet, P., Kéry, M., Gardner, B., Pasinelli, G., Royle, J.A., 2015. Estimating population size for Capercaillie (*Tetrao urogallus* L.) with spatial capture-recapture models based on genotypes from one field sample. *PloS One* 10, e0129020.

Morán-Luis, M., Fameli, A., Blanco-Fontao, B., Fernández-Gil, A., Rodríguez-Muñoz, R., Quevedo, M., Mirol, P., Bañuelos, M.-J., 2014. Demographic status and genetic tagging of endangered capercaillie in NW Spain. *PloS One* 9, e99799.

P Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., R Leathwick, J., Lehmann, A., Li, J., G Lohmann, L., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197.



R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ranc, N., Santini, L., Rondinini, C., Boitani, L., Poitevin, F., Angerbjörn, A., Maiorano, L., 2017. Performance tradeoffs in target-group bias correction for species distribution models. *Ecography* 40, 1076–1087.

Renner, I.W., Elith, J., Baddeley, A., Fithian, W., Hastie, T., Phillips, S.J., Popovic, G., Warton, D.I., 2015. Point process models for presence-only analysis. *Methods Ecol. Evol.* 6, 366–379.

Rolstad, J., 1988. Autumn habitat of capercaillie in southeastern Norway. *J. Wildl. Manag.* 747–753.

Rolstad, J., Wegge, P., Larsen, B.B., 1988. Spacing and habitat use of capercaillie during summer. *Can. J. Zool.* 66, 670–679.

Roussel, J.-R., Auty, D., 2016. lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 1.0.0. <https://github.com/Jean-Romain/lidR>.

Ruckstuhl, K.E., Neuhaus, P., 2000. Sexual segregation in ungulates: a new approach. *Behaviour* 137, 361–377.

Storch, I., 1993. Patterns and strategies of winter habitat selection in alpine capercaillie. *Ecography* 16, 351–359.

Syfert, M.M., Smith, M.J., Coomes, D.A., 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PloS One* 8, e55158.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A.C., Guisan, A., 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Methods Ecol. Evol.* 5, 947–955. <https://doi.org/10.1111/2041-210X.12203>

Thiel, D., Unger, C., Kéry, M., Jenni, L., 2007. Selection of night roosts in winter by capercaillie *Tetrao urogallus* in Central Europe. *Wildl. Biol.* 13, 73–86.



Thuiller, W., Brotons, L., Araújo, M.B., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27, 165–172.

Warton, D.I., Renner, I.W., Ramp, D., 2013. Model-based control of observer bias for the analysis of presence-only data in ecology. *PloS One* 8, e79168.

Wolf, J.B., Kauermann, G., Trillmich, F., 2005. Males in the shade: habitat use and sexual segregation in the Galapagos sea lion (*Zalophus californianus wollebaeki*). *Behav. Ecol. Sociobiol.* 59, 293–302.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243. <https://doi.org/10.1111/2041-210x.12004>

## 4 THE ROLE OF SAMPLING DESIGN: CAN WE REDUCE SAMPLING BIAS EFFECT?

---

### 4.1 Introduction

A dataset is defined as spatially biased when the selected study area is unevenly sampled. Here, two situations can occur: (1) all types of habitats are sampled, but some types are more sampled in proportion than others, (2) some habitats types in the landscape are not sampled at all. In the first case (1), if the sampling bias is not corrected, the resulting distribution will likely represent the sampling survey effort rather than the species real distribution, which can lead to a misinterpretation of SDMs results (Reddy and Dávalos, 2003) (e.g the target species favors habitats close to road and trails). In the second case (2), as only part of the landscape is sampled, it can make predictions within non-sampled habitat types unreliable.

As we saw in the precedent section, sampling bias correction methods are a powerful tool to obtain better parameter estimates than with non-corrected models when working with such biased datasets. Nevertheless, those correction methods have their limits, as the two methods tested earlier were not able to correct for the sex ratio estimation with the Capercaillie datasets. Thus, it is not surprising that the improvement of sampling design for SDMs is one of the five challenges for future SDMs highlighted by (Araújo and Guisan, 2006). Indeed, it is acknowledged that random or systematic sampling lead in general to more accurate models. This was confirmed by the results presented in the first part of this chapter where better parameter estimations were observed in the case of random or systematic sampling. Unfortunately, the use of such designs is time and cost consuming as there is often a large area to be covered. In addition, environments that are rough for the observers such as dense forest, or landscapes with the presence of cliffs or rivers, as well as the presence of private propriety in the study area are often an obstacle to promote such unbiased sampling schemes. Then, knowing those constraints, how can the data sampling design ameliorate models' performance and reliability without using a fully random or systematic sampling? In this



section, the effect of sampling design will be further investigated, with the aim to propose alternative sampling strategies which may ameliorate the results of SDMs. The hypothesis that the combination of datasets originated from systematic sampling design with spatially biased datasets may improve models performance and accuracy was thus formulated.

### 4.2 Methods

Following the methods presented in the submitted paper in the precedent section, two new sampling designs were tested on the virtual sub-group 2 (male) case, in addition to the random, transect and subjective sampling design initially presented.

The first design is a combination of the subjective sampling with an additional systematic transect localized in a part of the landscape adding new information on the environment. This design was created to observe the effect of a better coverage of the study area on the parameters estimation. Thus, the transect location was an essential aspect, as it is likely to influence the parameter estimation performance. This location was chosen among five transects, and the proportion of new information added for the three environmental variables regarding the landscape and the subjective design was evaluated using density curves.

In a second part, a sampling design combining all systematic transects and subjective sampling trajectories was created to investigate the effect of the over-sampling a part of the landscape in the case of a full coverage of the study area.

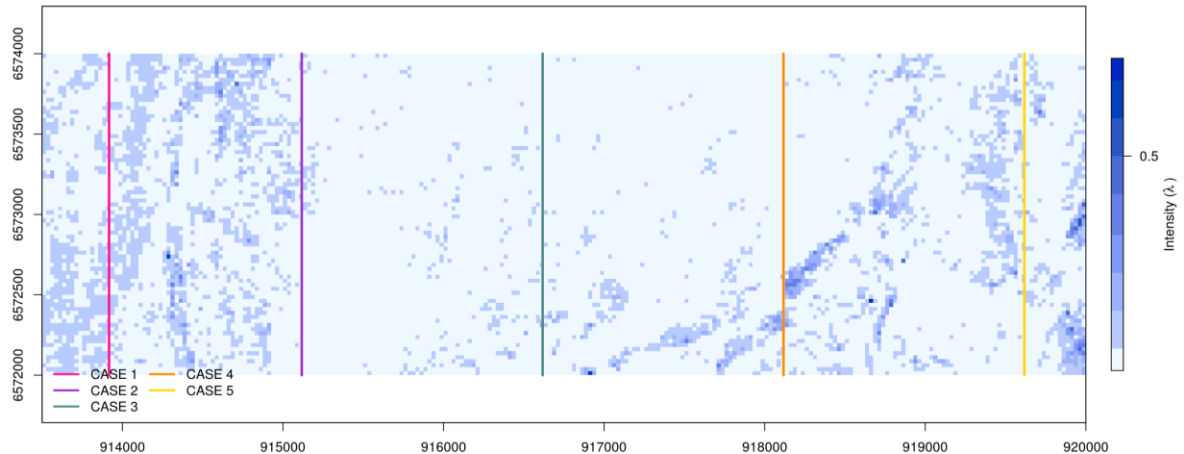
To control the effect of two bias correction methods on model accuracy, the estimations of  $\hat{\beta}_p$  from the fitted models were compared to the true values of  $\beta_p$  used to generate the intensity map for one virtual sub-group. The mean estimation was compared to the true parameter value with a Wilcoxon rank sum test.

## 4.3 Results

### 4.3.1 Choice of the single transect location

The location of the selected transect was chosen among five locations (Figure 28).

*Figure 28 : Location of five candidate transects*

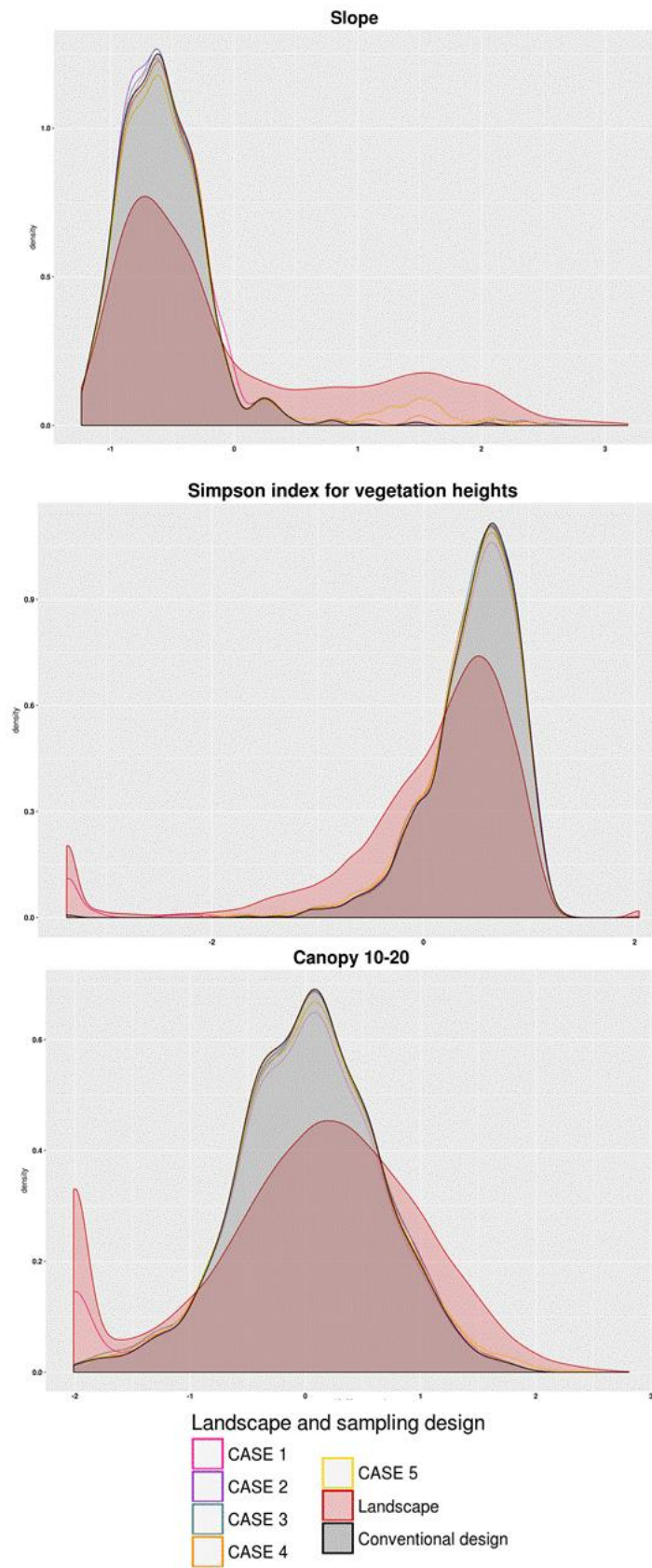


Density curves show that two transects are adding more information on the landscape than the others regarding the subjective sampling (Figure 29). Indeed, the transect case 1 is sampling low values of Simpson index and Canopy 10-20m, that are not sampled with the subjective sampling. The case 5 transect samples higher values of slope that are not sampled with the subjective design. Finally, the transect case 5 was arbitrarily chosen, and added to the subjective trajectories as a new sampling design.





Figure 29 : Density curves for the three environmental variables along subjective trajectories, candidate transects and within the landscape.





#### 4.3.2 Parameter estimation results

The results of the parameter estimation for the sub-group 2 are presented in the figure 30.

First of all, contrary to the estimation observed with the sub-group 1 in the precedent section, no strong effect of the bias was observed for the canopy density estimation parameter, leading to a good estimation in all cases. Those estimations were non-significantly different only for all subjective cases, subjective and one transects, subjective and all transects with no correction (table 13). Yet, an increase of the variance was observed when correction methods were applied.

The estimation of the slope parameter was better with the random and transects sampling design. In the case of subjective sampling, the estimation was wrongly estimated when non-corrected for sampling bias. When a transect was added, the estimation of the parameter was closer to the true value, when no corrections were applied, but was not fully corrected. The addition of one of the correction methods solved this problem. In the case of all subjective trajectories and transects combined, the estimation of the slope with no correction was close to the true value, but was still not better than results obtained with the transects design with no correction. The application of the corrections to this case gave relevant results with a lower variance increase than observed with the subjective only and subjective design combined with one transect cases. However, despite those observations, the estimations were still non-significantly different for the random design and for the subjective and one transect design corrected with targeted background point cases (table 13).

The same pattern was observed for the Simpson index variable, but a clear improvement of the parameter estimation with the addition of one transect (with no-correction) was not observed. The estimation was non-significantly different from the true value in only four cases: transects, subjective, subjective and one transect, subjective and all transects corrected with the targeted background point method (table 13).



Figure 30 : Parameters estimation of 100 models for the virtual species (sub-group 2) in different sampling cases.

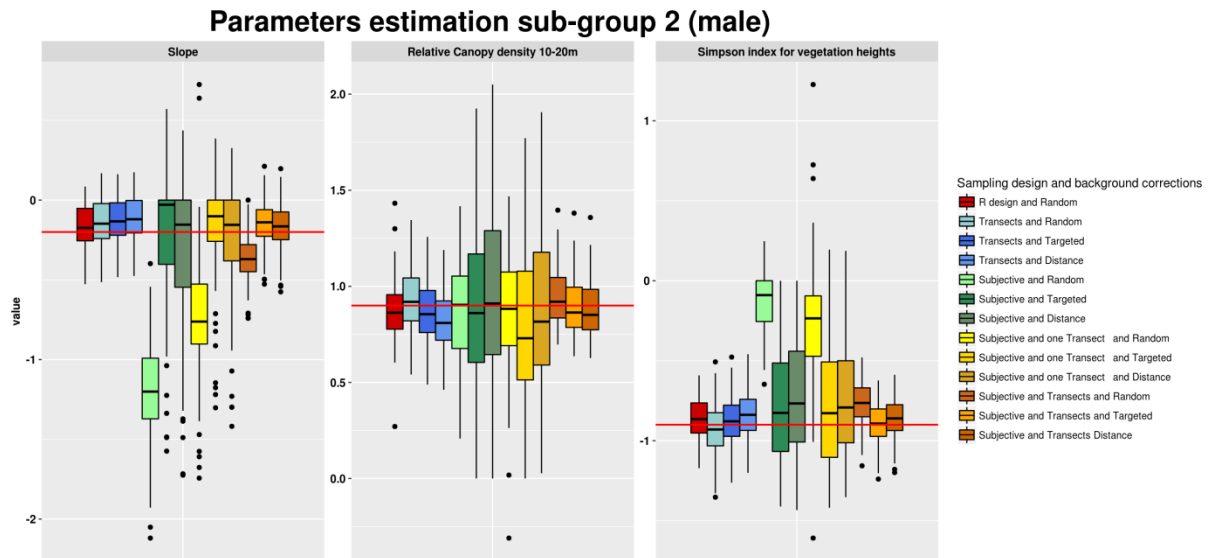


Table 13 :  $p$ -values from Wilcoxon rank test

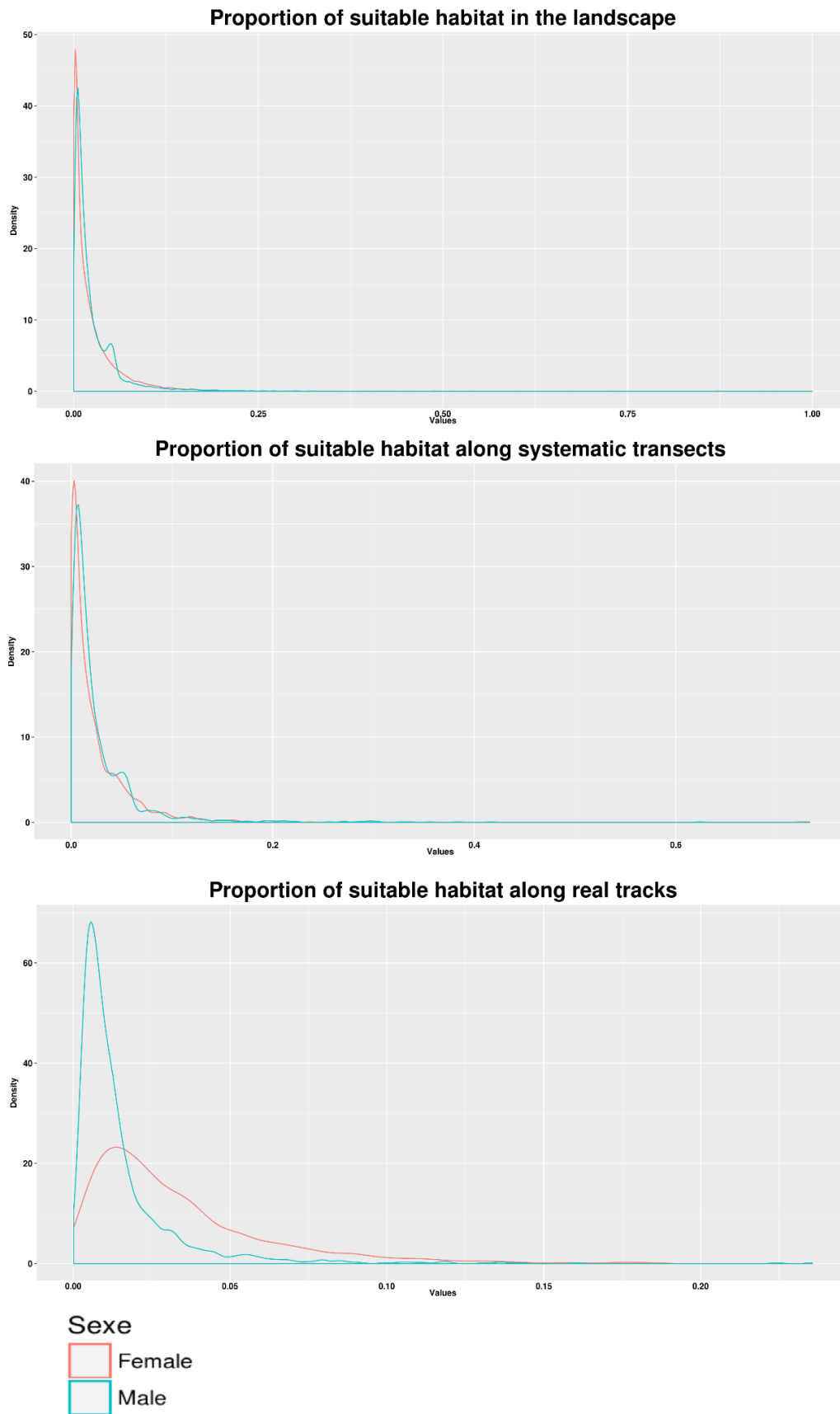
Case	Slope	Canopy	Simpson
<b>Sub-group 2</b>			
R and Random	0.41	0.047	0.012
Transects and Random	0.011	0.058	0.046
Transects and Targeted	0.00018	0.054	0.11
Transects and Distance	1e-04	5e-06	1.3e-05
Subjective and Random	4e-18	0.3	7.8e-13
Subjective and Targeted	0.0094	0.49	0.42
Subjective and Distance	1.6e-05	0.45	0.034
Subjective and one Transects and Random	8.2e-16	0.77	3e-17
Subjective and one Transects and Targeted	0.72	0.018	0.32
Subjective and one Transects and Distance	0.0089	0.66	0.022
Subjective and Transects and Random	1.1e-16	0.025	5.8e-12
Subjective and Transects and Targeted	0.001	0.16	0.88
Subjective and Transects and Distance	0.031	0.025	0.0091

#### 4.4 Discussion

Interestingly with the case of this virtual species second sub-group (male), the effect of spatial sampling bias was not observed on the variable canopy 10-20m contrary to the effect observed with the sub-group 1 (female, see precedent section). This might be due to a better sampling of the sub-group 2 occurrence range in proportion to its availability in the landscape compared to the sub-group 1 in the case of subjective sampling (Figure 31), leading to a good representation of the preferred habitat type for this variable. Consequently, the sampling bias effect was higher in the case of the female virtual sub-species and thus the effect of the corrections was stronger as well.



Figure 31 : Proportion of suitable habitat in the landscape and along systematic transect and subjective trajectories, blue is male and red is female.



The addition of systematic transects to subjective sampling designs can participate to the amelioration of the parameters estimation. However, in the case of one additional transect, this contribution is highly dependent on its location, as a stronger positive effect was observed for the variable slope only. This result is concordant with the observation of the density curves, where the transect location was bringing additional information on slope variations in particular. When all trajectories were combined (transects and subjective), the bias for two variables was lower when not corrected, showing that if the sampling bias is unknown, the implementation of a systematic survey can greatly ameliorate the models without making assumptions on the bias occurring in the previous dataset. The combination of those two designs showed that when the correction is applied, the parameter estimation is even better, with a lower variance increase than with the other cases. This observation is joining the results obtained by Koshkina et al. (2017) that a combined use of systematic and biased datasets can ameliorate models accuracy. Thus, if the sampling bias cannot be fully estimated (which is the case in most studies) the implementation of systematic or random additional sampling may ameliorate models' accuracy when added to the initial biased dataset. When it is impossible to pursue a full sampling in the study area with a random or systematic design, several systematic transects can be added, but should be preferably located in non-prospected areas in order to obtain additional information on the species occurrence (Cardador et al., 2017). Furthermore, nowadays multiple resources are available to construct sampling schemes and multiple other combinations than those tested here can be generated such as probabilistic sampling, stratified sampling, and mixed random and systematic sampling design (Pavlacky Jr et al., 2017). Nevertheless, those sampling designs need to be adapted to the aim of the study. This not only includes the objectives, but also the landscape and species characteristics and the socio-economical factors to be applied in the field (Lengyel et al., 2018) and thus ultimately improve future models.



## 5 THE CHOICE OF SPATIAL SCALE

---

The way individuals perceive and respond to their environment is a key point which needs to be addressed when implementing SDMs. The individual's perception is influenced by multiple factors at different hierarchical levels in space or time (Wiens, 1989). Mayor et al. (2009) in their review described six different types of levels used to define habitat selection at multiple scales:

- Causal levels
- Organizational levels (species, population, individual)
- Environmental or geographic levels in spaces (forest patch, tree species, trees individuals)
- Environmental or geographic levels in time (disturbances cycle, season, daily light)
- Behavioral levels in space (species range, population range, home range, territory)
- Behavioral levels in time (generation time, annual breeding, migration, feeding bout)

This implies that one individual will respond to different environmental components at different levels at the same time. For example, an individual Capercaillie may choose a particular tree because its shape will provide shelter, but also because it is near other tree species which will provide food and all of them are present within a large suitable forest patch. If one of these elements is not achieved, for example, the trees are not a part of a reasonable forest patch; this individual will not choose that particular tree at the first place. The imbrication complexity of those multiple levels is challenging to take into account in species distribution modeling but their understanding is nevertheless essential. Therefore, the development of “multi-scale” modeling methods has been promoted since the first definition of the concept by Stommel (1963) and multiple publications have highlighted the importance of scales (Wiens, 1989).

To achieve this objective, a five level hierarchical framework linked with their potential adapted from (Johnson, 1980) was proposed by Meyer and Thuiller (2006):



- **Level 0:** Geographical ranges within world or parts of worlds (the objective could be to assess biodiversity or climate change)
- **Level 1:** Regions containing populations (the objective could be reserve design, meta-population management, land-use planning)
- **Level 2:** Individual home range within region where a population is present (the objective could be habitat management, conservation of used area)
- **Level 3:** Individual life requirement at patch scale within the home range (the objective could be habitat management or mitigation of impacts)
- **Level 4:** Micro-habitat requirement, habitat characteristics that can be related to individual life requirement (the objective could be protection of key life history attributes)

Thus, in our case, the identification of which level should be analyzed considering our objectives was made. For Hazel Grouse the main objectives are to propose habitat suitability maps in order to plan future surveys in the Jura. Concerning Capercaillie, the identification of unsuitable habitats in order to help future management actions is the main purpose of the study. In both cases, models which include variables at intermediate level are necessary to prioritize management actions at coarse resolution, for example how many local patches should be we plan to restored in order to increase the global suitability of the territory or forest stand area. The two lowest levels are also important as it may facilitate the interpretation of the results and help the decision of forest managers for the choice of the local management action itself.

Regarding those observations, we choose three levels of interest for our study:

- Home range level (n°2)
- Patch level (n°3)
- Micro-habitat level (n°4)





### 5.1 From multiple levels to multi-scale selections

The major challenge in multi-scale habitat modeling is to define the best scales at which the species respond to each of the different variables used as predictors (Mayor et al., 2009).

In order to achieve this objective, different strategies of multi-scale modeling have been conducted. McGarigal et al. (2016) in their recent review classified those strategies in 6 different categories (table 14).

*Table 14 : Multi-scale model strategies from (McGarigal et al., 2016)*

Strategy	Description
A priori single scale	One scale for each variable within each level
A priori separate scale	A single but different scale for each variable (not empiric), variables are not combined
A priori multiple scales	A single but different scale for each variable (not empiric), variables are combined into a single model
Pseudo-optimized single scale	Evaluation of all variables simultaneously across a range of scales, and select one best scale model
Pseudo-optimized multiple scales	Evaluation of each variable at multiple scale separately, and are then combine together at their best scale to produce a final model
Pseudo-optimized multiple scales alternative	All variables are evaluated together at multiple scales, scales were not pre-defined but selected by an optimization algorithm

From the use of those different strategies, a few recommendations were proposed by (McGarigal et al., 2016) in order to improve multi-scale selection modeling.

- (1) clearly define levels of habitat selection under investigation following the categories introduced by (Meyer and Thuiller, 2006)
- (2) promote multi-scale approach among levels
- (3) for each level of habitat selection, identify and justify the choice of scales
- (5) true optimization approach should be used when possible
- (6) compare multi-scale analysis within a level to the best single-scale models



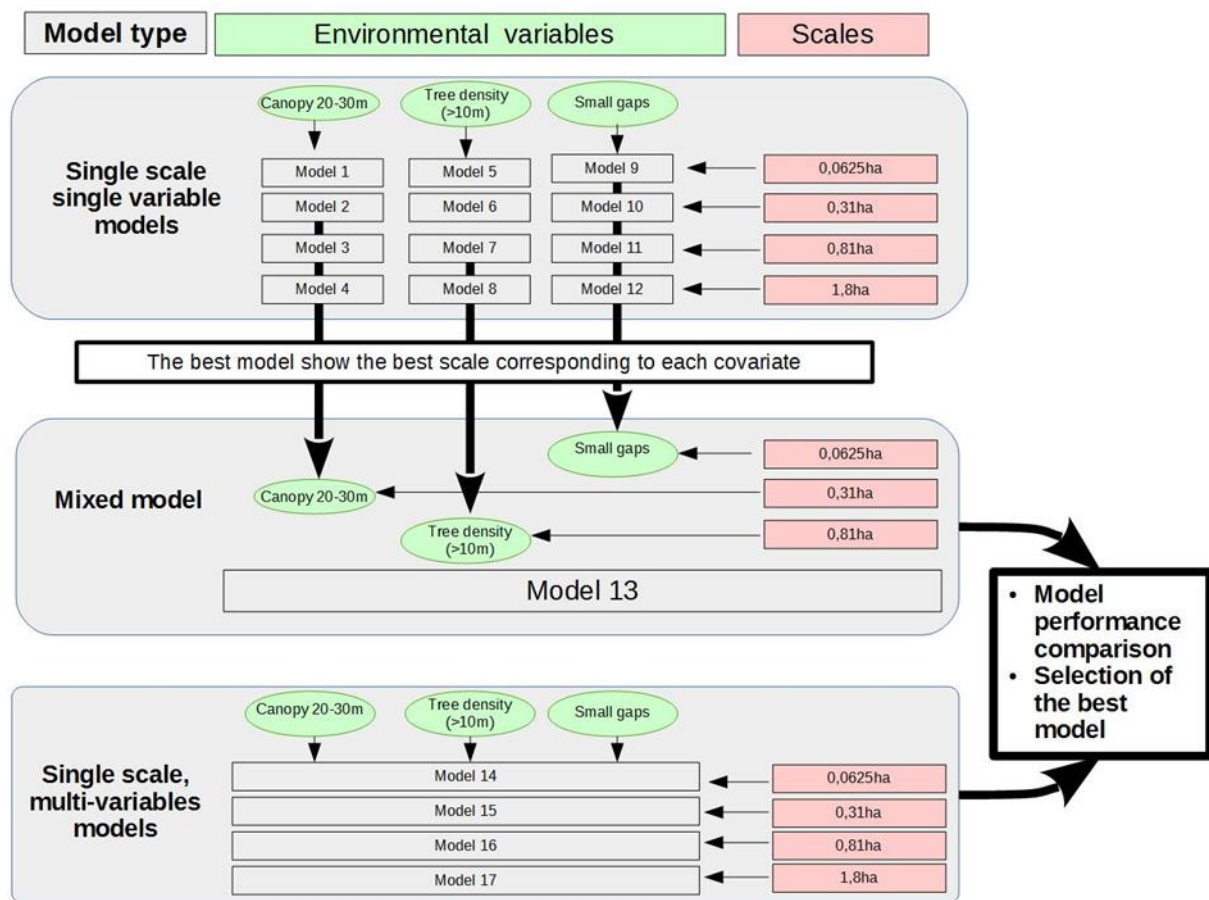
## 5.2 Selection of scales and methods

We already defined three levels of interest for our models. Following the recommendation (2) above, we will use multiple scales within each level. This choice is also supported by the fact that multi-scale models were found to better estimate the relationship between habitat characteristics and Capercaillie presence (Graf et al., 2005). As the scales are defined within levels, it is depending on each species life requirement; those values were defined in consequence.

Concerning Capercaillie, the scales were defined with the help of experts as large differences in home range surface were observed according to the study area. For the home range level, three scales were selected, 15.8ha, 27.5 ha and 56 ha, thus representing different potential surface areas. The two patch level scales (0.81 and 1.8ha) are representing the area that an individual can cover in one day. Two scales were selected to describe the micro-habitat, 0.31 ha and 0.0625 ha, thus representing the immediate environmental component to birds location.

Hazel Grouse has a smaller home range than Capercaillie with couples living on a 20-40 ha territory. As a consequence, the scales 15.8 ha and 27.5 ha were chosen. At the patch level, three scales were selected: 7 ha, 1.8 ha and 0.81 ha. The micro-habitat scales were similar to those presented for Capercaillie (0.31 and 0.0625 ha).

Using the different selected scales, two methods that use optimization were chosen: pseudo-optimized single scale and pseudo-optimized multi-scale methods. The results will be compared to identify the best models and the best scale for each variable. This framework is presented in the figure 32 below.

*Figure 32 : Framework used to produce and compare models using different scales*

This framework was used to produce both Hazel Grouse and Capercaillie models (chapter 4 and 5).

## 6 CONCLUSION

In this chapter, the principles and assumptions behind SDMs were reviewed as a first step on the way of producing reliable and pertinent models. The importance of the choice of the modeling technics according to the study objectives and datasets characteristics was highlighted. Thus, the choice of the algorithms that were used for both bird and beech regeneration was documented according to our objectives.

The role of spatial sampling bias on SDMs performance and relevance was examined more closely. Indeed, the specific case of Capercaillie as a dimorphic species and the opportunity to work with data



from two different sampling designs was taken to study the effects and reliability of two spatial bias corrections methods. In a second part, the role of sampling design as a tool to ameliorate SDMs outputs was examined. From those two studies, we conclude that current methods for correcting spatially biased sampling can improve estimates of both habitat association and sex ratios for dimorphic species. More importantly, we believe that information on subgroups with distinct habitat associations should be included in SDMs analyses whenever possible. Finally, the addition of data from systematic sampling designs can substantially ameliorate the parameters estimation when used with appropriate correction methods.

In a last part, the importance of spatial scales was presented as well as the framework created for Hazel Grouse and Capercaillie modeling.

In the next chapters, recommendations reviewed here were applied in order to produce SDMs.

# CHAPTER 4: HAZEL GROUSE

## DISTRIBUTION MODELS



Groupe Tétraz Jura



## 1 INTRODUCTION

---

The species population showed a dramatic regression at lower altitudes between 1950 and 1990 in France. In the Jura Massif, the population continued to disappear from the lower elevation areas during the last decade as well (2000-2009) (Bellon 2011). In this context, we seek to improve our knowledge of the species actual distribution across the Jura Massif with the conception of a new survey protocol to collect future data on the species. Indeed, so far Hazel Grouse surveys conducted by the NGO “Groupe Tétrás Jura” were paired with Capercaillie surveys, restricting the actual data to upper elevation environment, where Capercaillie is still present. However, Hazel Grouse potential distribution is much larger, thus obtaining field data for that particular species over its full potential presence range is time-consuming. The improvement of the sampling survey design must fulfill two objectives:

- 1) Optimize the protocol to give us new information on the species distribution
- 2) The protocol should be cost-effective, regarding the funds available for the project.

For this purpose, we elaborate, in this chapter, multi-scale species distribution models for Hazel Grouse, which are meant to select areas for future sampling. In addition to this objective, the datasets were used to explore the effect of interaction features on models performances and on the interpretation of the response curves. This will be presented in an article (n°2) submitted to the journal wildlife biology. In a second step, the univariate responses curves and predicted maps obtained with those models will be presented.



## 2 ARTICLE 2: EFFECT OF INTERACTION TERMS ON SPECIES DISTRIBUTION MODELS PREDICTIONS INTERPRETATION AT MULTIPLE SCALES.

---

Effect of interaction terms on species distribution models predictions interpretation at multiple scales.

Anouk Glad<sup>1,3</sup> <https://orcid.org/0000-0001-5810-189X>

Björn Reineking<sup>1</sup> <https://orcid.org/0000-0001-5277-9181>

Marc Montadert<sup>2</sup>

Alexandra Depraz<sup>3</sup>

Jean-Matthieu Monnet<sup>1</sup> <https://orcid.org/0000-0002-9948-9891>

<sup>1</sup> Univ. Grenoble Alpes, Irstea, UR LESSEM, BP 76, 38402 St-Martin-d'Hères, France

<sup>2</sup> ONCFS, 90 impasse des daude, 74320 Sévrier, France

<sup>3</sup> Groupe Tétràs Jura, 9 impasse du tacon, 39370 les Bouchoux, France

Corresponding author: Anouk Glad, [gladanouk@gmail.com](mailto:gladanouk@gmail.com)

**The manuscript was submitted to the journal Wildlife Biology on August 1<sup>st</sup> 2018.**

**Supplementary material is available in Annex 5.**

### 2.1 Abstract

The Hazel Grouse is a locally endangered forest bird occurring in France. Due to its secretive behavior, the estimation of its distribution range in the Jura Massif is challenging. Species Distribution Models are widely used for conservation and wildlife management purposes. In the case of Hazel Grouse, such models could be used to plan future surveys. To produce a relevant model, the

choice of environmental variable response features is important. The product feature is particularly important as it can take into account the effect of an interaction between two variables. Thus, the impact of the addition of an interaction term on the model performance, response curve interpretation using a three dimensional curve has been evaluated. Yet, the addition of an interaction term between two selected variables did not improve model performance for Hazel. The choice of scales for the selected environmental variables is also known to be an important aspect of SDMs modeling. However, in past studies on Hazel Grouse, very few used multi-scale analysis for SDMs modeling. Thus, using seven different scales, model's performance and variable contribution were evaluated using both single-scale, multi-variable models and multi-scale, multi-variable models (mixed model) over two study areas. The expectation that, the best scales observed, will be similar between the two study areas was also not validated by our results. Nevertheless, four main contributing common variables were found to have a high impact in both study area models and at all scales: the standard deviation of the canopy density, the proportion of open areas, the penetration between 2-5m and the tree density >10m.

Keywords: Species distribution modeling, interaction feature, response curve, Hazel Grouse

## 2.2 Introduction

Hazel Grouse (*Bonasa bonasia*) is a locally endangered forest bird occurring in France. Indeed, the species population showed a dramatic regression at low altitudes between 1950 and 1990 in the country mainly due to habitat loss and fragmentation (Huhta et al. 2017). In particular, in the Jura Massif, the population continued to disappear from the lower elevation areas during the past decades as well (2000-2009) (Bellon 2011). The species is a discrete bird, difficult to observe making the estimation of its distribution challenging. In this context, improving the knowledge of the species' potential distribution across the Jura Massif is needed.

Species Distribution Models (SDMs), also called habitat suitability models, are widely used in conservation biology (Franklin 2009a, 2009b; Guisan and Zimmermann 2000). SDMs aim to relate multiple presence locations to relevant environmental parameters (Guisan, Thuiller, and





Zimmermann 2017). Thus, those models can provide researchers and practitioners accurate and useful information on the habitat preferences of a species over local or large landscapes (Elith et al. 2011). Those results are often used as a management or conservation tool (Guisan et al. 2013). However, obtaining pertinent and valuable predictive models requires that implications and limits be clearly understood (Merow et al. 2014).

Maxent it is arguably the most frequently used SDM method (Elith et al., 2011; P Anderson et al., 2006; Phillips et al., 2006). When using this method, a first element is the choice of the features and interactions. Features represent different possible transformations of the variables, which aims to get closer to the complex response of the species to each particular variable. The simplest feature is a linear response, but it is recommended to include other features such as quadratic, hinge, product or threshold in models (Elith et al. 2011). Indeed, those choices will constrain the response of the species to each variable in different ways, leading to different models in terms of complexity and interpretations (Merow et al. 2014). As an example, quadratic features can indicate the tolerance of the species to the variation from the best conditions which cannot be observed with linear features (Elith et al. 2011). The product feature is particularly important as it can take into account the effect of an interaction between two variables. In this case, the effect of one variable varies according to the values of a second variable. Thus, a documented choice of the features is necessary as it can improve both model performance and the interpretation of the environmental variables regarding the biology of the species (Elith et al. 2011). The interpretation of the results is an important aspect, as it can highlight the limitation of models. In particular, species response to the environmental variables interpretation, regarding their biological plausibility for the species is highly valuable. This is usually done by calculating the univariate response for each variable. However, the spatial predictions over the full landscape are the results of the variable variations combination which is difficult to visualize with such a type of univariate response (Zurell, Elith, and Schröder 2012). The interaction terms can capture the covariation between two variables, as well as a threshold within a gradient (is a homogenous environment more suitable with low or dense vegetation?) (Merow et al.



2014). Thus, the first aim of this study is to evaluate the effect of the addition of a product feature on the interpretability and the accuracy of model prediction.

The choice of scales for the selected environmental variables is also known to be an important step for SDMs. Indeed, the perception of their environment by individuals is influenced by multiple factors at different hierarchical levels in space (eg. landscape, home range, patch levels) or time (e.g. hour, day, season) (Wiens 1989). This implies that one individual will respond to different environmental components at different levels at the same time. Therefore, the development of “multi-scale” modeling methods has been promoted since the first definition of the concept by Stommel (1963), and multiple publications have since then emphasized the importance of scales. To address this, we can distinguish two types of approach; 1) define a scale the environmental variable a priori using expert knowledge of the species, 2) use empirical methods and optimization to select the best scales among a range of values using models own results. However, in past studies on Hazel Grouse habitat suitability, very few used multi-scale analysis. Indeed, to our knowledge, only five studies used between two and four scales (Zellweger et al. 2014; Schäublin and Bollmann 2011; Aberg et al. 2000; Swenson 1993; Sitzia et al. 2014). Thus, the second aim of this study is to address this issue. For this purpose, models will be created at different selected scales using a pseudo-optimized framework and subsequently compared (McGarigal et al. 2016).

Finally, the choice of environmental variables that will be meaningful from the species’ point of view and thus will capture the influence of the environment on individuals is an essential and well known pre-requisite to create relevant models (Johnson and Gillingham 2005). The Hazel Grouse is a cryptic bird, found in a large variety of habitats due to its large distribution from Hokkaido (Japan), through Russia and Scandinavia, to the Alps and Carpathian massifs. Thus, this species can be found in diverse habitats from the young successional stage to the old-growth deciduous stand, and from mixed to coniferous dominated forests. Hazel Grouse is known to occur in well-structured stand with at least two vegetation layers associate with a dense understory and an horizontally heterogeneous habitat (Müller, Schröder, and Müller 2009; Schäublin and Bollmann 2011). As a consequence, the use of LiDAR extracted variables, which can capture the three dimensional structure of the vegetation, in



SDMs has recently been promoted for this species (Melin et al. 2016; Bae et al. 2014; Zellweger et al. 2014, 2013; Rechsteiner et al. 2017). However, so far the variety of variable and of forest environment (boreal, alpine and Jura forest types) used, could not lead to the emergence of a consensus on which LiDAR variables best predict the distribution of the Hazel Grouse. Because we had the opportunity to compare models from two separate LiDAR datasets, which originate from two different LiDAR campaigns the third objective of this study will be to investigate models transferability and limits regarding the choice of the variables. In addition, the fact that only one among five studies using LiDAR variables (Zellweger et al. 2014) was conducted to our knowledge with a multi-scale analysis, can be highlight, and ensure again the need for the promotion of multi-scales models frameworks.

In sum, using the study case of an endangered species for which the production of an appropriate and reliable species distribution model is needed, we propose three objectives:

1. To evaluate the effect a product feature on model performance, response interpretability and predictions
2. To compare the performance and the contribution of variables among scales
3. Choice and transferability of LiDAR variable for Hazel Grouse

We expect that the addition of an interaction term will improve models performance and that the response curve interpretation will be facilitated by the use of a three dimensional curve. We also hypothesize that multi-scales, multi-variable models will give better results than single scale models as it was showed in multiple studies (Timm et al. 2016; Rostro-García et al. 2016; Vergara et al. 2016). In addition, we expect that, the best scales observed, will be similar between the two study areas. We hypothesize, that the same important predicting variables will be observed at both study sites and at each single scale and mixed models, thus representing important environmental component for the species.



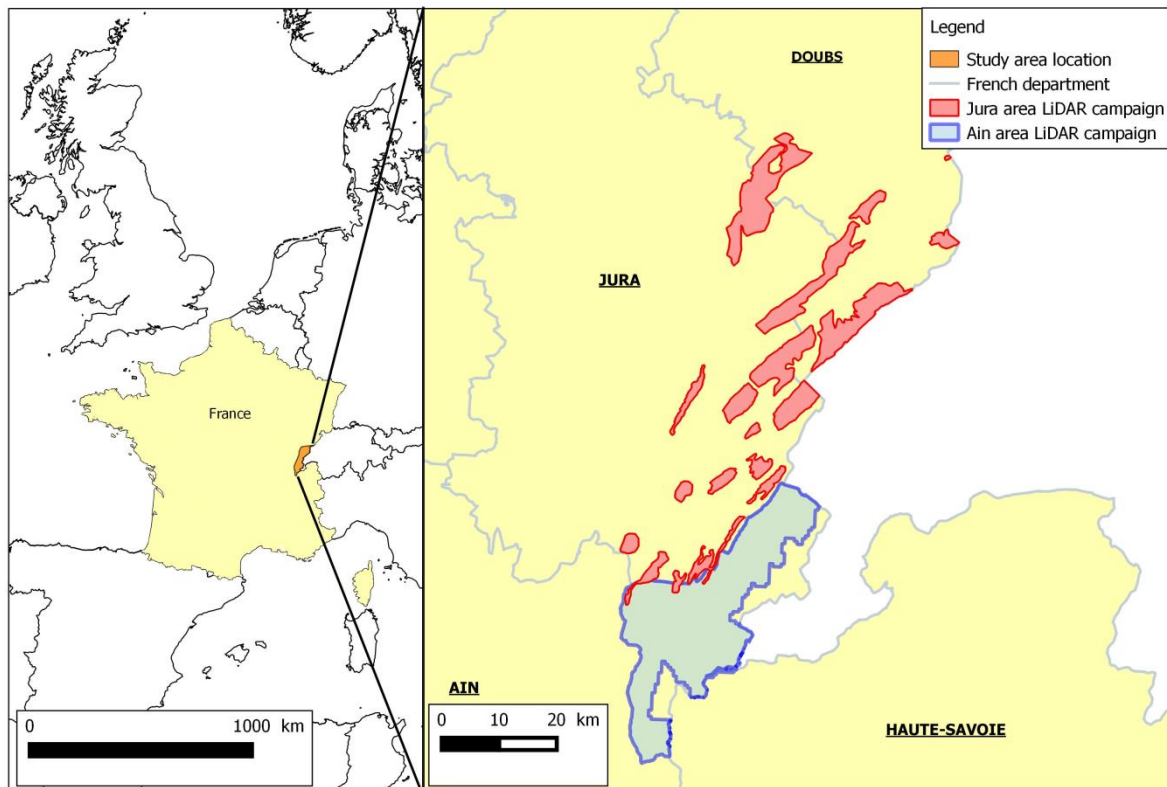
## 2.3 Material and methods

### 2.3.1 Study areas

The study area lies in the east of France, in the Jura Massif, located within the departments of Ain, Doubs and Jura (Fig. 33). The landscape is composed of a mosaic of small urban areas, pastures, forests and fields. The massif is composed of a low plateau (elevation range from 400 m to 700 m a.s.l.) and a high plateau (elevation range from 700 m to 1620 m a.s.l.). The forest covers an area of 4800 km<sup>2</sup> in the Doubs and Jura departments. The Ain department comprises 1920 km<sup>2</sup> of forest (based on the IGN national forest inventory 2005-2009, <https://inventaire-forestier.ign.fr/> ). On the lower plateau, forests are mainly deciduous-dominated composed of beech (*Fagus sylvatica*), ash (*Fraxinus excelsior*), oak (*Quercus petraea*) and spruce (*Picea abies*). Above 1 000 m forests are mainly dominated by conifers, and composed of a mix of spruce, fir (*Abies alba*) and beech. The Ain department is characterized by an unforested subalpine level at higher altitudes. The climate is continental with temperature varying between -5.9°C and 21.1°C and a mean annual precipitation of 1874 mm (<https://donneespubliques.meteofrance.fr/>).



*Figure 33 : Location of the study area within three French departments, (Ain, Jura, and Doubs).  
Two LiDAR campaigns were conducted in two different parts of the study area.*



### 2.3.2 LiDAR datasets

The study area was covered by two different airborne LiDAR surveys (Fig. 33). Both surveys were conducted using a Riegl LMS Q680i system (RIEGL Inc., Horn, Austria). The system was operated with a maximum half-scan angle of 30°, full wave-form digitization.

In the first area, referred to hereafter as “Ain study area”, the LiDAR campaign was conducted in fall 2014 and covered 626 km<sup>2</sup>. An average point density of 21.3 points/m<sup>2</sup> with a vertical error of 10cm was obtained. During this campaign, the whole landscape was covered with LiDAR.

The second LiDAR campaign hereafter referred as “Jura study area”, was conducted in summer 2016, and covered a surface of 431km<sup>2</sup>. An average point density of 18 points/m<sup>2</sup> and vertical error of 10cm was obtained. Contrary to the first campaign, only forest areas that are potentially occupied by Hazel Grouse were surveyed, which led to a fragmented partial description of the landscape.



Due to differences within LiDAR campaigns (e.g. the season of flight survey), the environmental variables extracted from LiDAR were calculated and used separately in models.

### 2.3.3 Others environmental datasets

Non-LiDAR datasets were also used to calculate environmental metrics. First, satellite imagery data from Copernicus Sentinel-2 images taken on a clear day were used (17/07/2016 Copernicus Sentinel-2 data 2016 processed at level 2A by CNES for the THEIA Land data center, <http://www.theia-land.fr/fr/node/816> ). In addition, forest parameters collected from forest vegetation surveys acquired by the French National Forest Office (ONF) in the Ain department were also used (“Protocole de suivi dendrometrique des réserves forestières”, including permanent plot data and tree location sampling for LiDAR individual tree calibration). Datasets on anthropogenic infrastructures were also added i.e. roads from the French national institute for geographical and forest information (IGN) Route 500® and building locations and surface areas from IGN BDTOPO®.

### 2.3.4 Hazel Grouse surveys

Long-term Hazel Grouse surveys were organized between 2007 and 2016 by the “Groupe Tétraz Jura” a non-governmental organization (NGO) in the Jura Massif. Surveys were conducted in winter, from January to April, in the presence of a snow cover. Observers surveyed the forest outside trails, using a Global Navigation Satellite System (GNSS) receiver to record their survey tracks and Hazel Grouse sign locations (mostly sign such as feces, prints and feathers). A total of 961 sign observations were collected within the Jura study area, and a total of 139 sign observations were collected in the Ain study area.

### 2.3.5 Extraction of LiDAR metrics

Different variables were selected for Hazel Grouse habitat suitability models. From LiDAR datasets, two types of metrics were extracted, point-cloud metrics and object-oriented metrics. LiDAR raw point clouds were normalized with the Delaunay interpolation method from the LidR package (Roussel and Auty 2016). Both point-cloud and object-oriented metrics were calculated using LidR and a new R package, lidaRtRee ( <https://gitlab.irstea.fr/jean-matthieu.monnet/lidaRtRee>). Point-



cloud metrics were calculated within pixels of 25 m \* 25 m. Objects were segmented from the canopy height model at a 0.5 m resolution. Summary statistics such as mean, ratio, proportions or percentages were then calculated within pixels of 25 m \* 25 m. Two types of objects were extracted: trees and gaps. Concerning trees, a tree top detection method was used (method #1 in (Eysn et al. 2015)). Gaps were defined as areas with a canopy height lower than 1 m, surrounded by trees whose height was less than half of the gap width.

The selected metrics for Hazel Grouse distribution modeling are presented in table 15. Those metrics were chosen with the aim to capture habitat elements that are important for Hazel Grouse. Hazel Grouse needs a moderate canopy cover (Mathys et al. 2006; Hofstetter et al. 2015), with the presence of canopy gaps that favor the presence of denser understory composed of young trees such as spruce, rowan (*Sorbus aucuparia*), willow (*Salix* sp.), thus providing good feeding sites (Zellweger et al. 2014; Sachot, Perrin, and Neet 2003). Those aspects can be explored with two tree density per hectare variables, `Tree_density10` (number of detected trees with height higher than 10 m), `Tree_densityinf10` (number of detected trees with height higher than 5 m and lower than 10 m), and the `middle_gap` variable. The `middle_gap` variable encompasses the proportion of area cover by gaps of surface between 200 and 1000m<sup>2</sup>.

In addition, Hazel Grouse is generally found in horizontally heterogeneous habitat (Mathys et al. 2006; Müller, Schröder, and Müller 2009), especially within the shrub layer which must favor both dense coniferous habitats offering shelters and more open areas offering food resources (Sachot, Perrin, and Neet 2003; Schaublin and Bollmann 2011; Zellweger et al. 2014). Thus, the standard deviation of canopy density 20-30 m (number of points between 20 and 30 m divided by the total number of points) was added as well as the standard deviation of both penetration ratio 0.5-1 m and 2-5 m, in order to have information on different vertical levels. The penetration ratio was here used as a proxy for the shrub layer density and represents the number of LIDAR points between 0.5-1 m or 2-5 m divided by the total number of points below those layers.



Vertical structure heterogeneity was represented in our models by the Gini index of detected tree heights, and Hazel Grouse was expected to favor vertically heterogeneous habitat. As Hazel Grouse does not favor large open areas, the variable proportion of open areas, which is the proportion of surface covered by gaps larger than 1000m<sup>2</sup>, was also included. In mountains areas, this variable is likely to represent alpine grassland or pastures, but at lower altitude, it will comprise areas such as fields and urban areas.

It is well known that Hazel Grouse favors conifer dominated forests, which provide both shelters and food resources (Sachot, Perrin, and Neet 2003). Within the Ain study area modelled forest metrics were calculated allowing us to create a variable proportion of coniferous trees weighted by basal area. This variable was calculated using a modeling of forest parameters from both LiDAR point-cloud metrics and forests vegetation survey. The classification of tree types was done using random-forest (Breiman 2001). As no permanent plots were present in the Jura study area, NDVI (Normalized Difference Vegetation Index) derived from sentinel images was used as a proxy for the proportion of coniferous species. The high NDVI values were expected to be preferred, as a higher NDVI value indicates a highly productive forest. In both cases, the standard deviation was also added, with the aim of taking into account the horizontal heterogeneity of each variable.

Furthermore, the Ain study area is composed for a large part of non-forest areas (including urban areas). Thus, with the attention for accurate models, two variables representing human disturbances were added to the Ain models: the distance to main roads and buildings area proportion. The selected roads are local, regional, departmental and national asphalted roads which are connected to urban areas or to the principal transport infrastructures. Forest tracks and trails are not a part of this dataset. Indeed, high traffic was found to influence Hazel Grouse presence along the road, but no effects were found on the disturbance by ski trails and forest tracks which can be even a good habitat by allowing the growth of the shrub layer (Räty 1979; Leclercq 1985). Buildings with a surface above 20m<sup>2</sup> indicating the presence of an urban area, such as housing, schools, and farms (category BATI\_INDIFFERENCIE, IGN BDTOPO®) were used to calculate the building area proportion variable.



*Table 15 : Environmental metrics selected to predict Hazel Grouse suitable habitat*

Type of variable	Variable Name	Description	Area
LiDAR	Canopy2030sd	Standard deviation of point density between 20 and 30m	Ain and Jura
	Ratio0501sd	Standard deviation of penetration ratio between 0,5 and 1m	Ain and Jura
	Ratio0205	Mean Penetration ratio between 2 and 5m	Ain and Jura
	Ratio0205sd	Standard deviation of Penetration ratio between 2 and 5m	Ain and Jura
	Tree_densityinf10	Mean tree lower than 10 meter but higher than 5m density (number/ha)	Ain and Jura
	Tree_density10	Mean tree higher than 10 meter density (number/ha)	Ain and Jura
	Tree_gini	Mean tree height gini index	Ain and Jura
	Middle_gap	Proportion of surface covered by gaps of size 200 to 1000 m <sup>2</sup>	Ain and Jura
	Grassland	Proportion of surface covered by open areas (objects with vegetation height<1m and area > 1000m <sup>2</sup> )	Ain and Jura
	prop_resGsd	Standart deviation of proportion of coniferous trees in basal area (among tree diameter > 17,5, ONF data)	Ain
	prop_resG	Mean proportion of coniferous trees in basal area (among tree diameter > 17,5cm)	Ain
IGN BD TOPO®	Buildings	Total surface of buildings within a 56ha area	Ain
IGN ROUTE 500® Version 2.1	Road	Distance to the closest main road (meters)	Ain
Sentinel-2 © Copernicus data / ESA	ndvisd	Standard deviation of NDVI	Jura
	ndvi	Mean NDVI	Jura

### 2.3.6 Spatial scale choice

At least two different scales within each level define hereafter were chosen to assess habitat suitability (Table 16). The values of the largest scales surface areas were chosen using the estimation of Hazel Grouse home range. By contrast, the smallest scale represents the perception of the local environment by individuals; therefore three scales at the patch level and two scales for the micro-habitat level were proposed. For six of the seven scales (0.31 ha to 27.5 ha) the mean and the standard deviation were calculated among the neighboring pixels using a moving window. This task was done with the *rgrass7* R package (Bivand et al. 2016).

At the smallest scale presented here (0.0625 ha) the raw value calculated in the pixel 25\*25m was used. The standard deviation was not calculated and thus not used for modeling. Therefore, models at the smallest scale are different from all other models.

*Table 16 : Habitat levels and corresponding scales for the models*

Hazel Grouse			
Level	Definition	Scales	Objectives
1	Home range level	27.5 ha	Capture the different habitats present at a large scale
		15.8 ha	
2	Patch level	7 ha	Identify the suitability of the local environment
		1.8 ha	
		0.81 ha	
3	Micro-habitat level	0.31 ha	
		0.0625 ha	

### 2.3.7 Modeling framework

For all models, we use a recent reimplementation of the algorithm building on the equivalence of Maxent with an infinitely weighted logistic regression (Maxnet package; Phillips et al. 2017). 10 000



background points were randomly located along observer sampling tracks, where distances between each background point and survey tracks followed a logspline distribution fitted to Hazel Grouse observations data. The correction of sampling bias was done using a targeted background point method (Phillips et al. 2009). All variables were normalized by subtracting the pixel value from the mean and by dividing the result by the standard deviation before the modeling process. In addition, the correlation between variables was evaluated prior to each analysis using a Pearson test. When the value was higher than 0.7, one of the variables was dropped (Dormann et al. 2013) (except for the proportion of open areas, which was nevertheless necessary to identify non-forest areas with accuracy). For all models, both linear and quadratic terms were selected. In addition, the product between canopy2030sd and proportion of open areas was added to all multi-variable models, with the intended purpose to link the landscape heterogeneity to the availability of forest and non-forest areas. In order to calculate the Area Under Curve (AUC) value which is a method to evaluate model performance, a 5 k-fold cross-validation was performed. The estimation of the contribution of each variable to the model was evaluated by comparing the beta values. All beta values associated with the same variable (whether linear quadratic or product) are added to obtain the final contribution value. The beta product value was divided by two before being added to the linear and quadratic values of each variable. In addition, the variables associated with a contribution value below 0.05 were dropped for result analysis keeping only the most essential variables.

### 2.3.8 Evaluation of the effect of an interaction between two variables

To evaluate the effect of interaction on a single-scale, multi-variables models were all declined in three different categories. For this, the standard deviation penetration ratio 2-5m variable was chosen, as it is known that Hazel Grouse is sensitive to understory characteristics. This variable was combined with the mean proportion of open areas in order to make the distinction between forest and non-forest areas. Thus, for six of the seven scales (0.0625 ha scale was excluded as the standard deviation was not calculated), three different cases were created. First, models with no added interaction terms as described in the precedent section. For the second case, an interaction term between the mean proportion of open areas and the standard deviation of the penetration ratio 2-



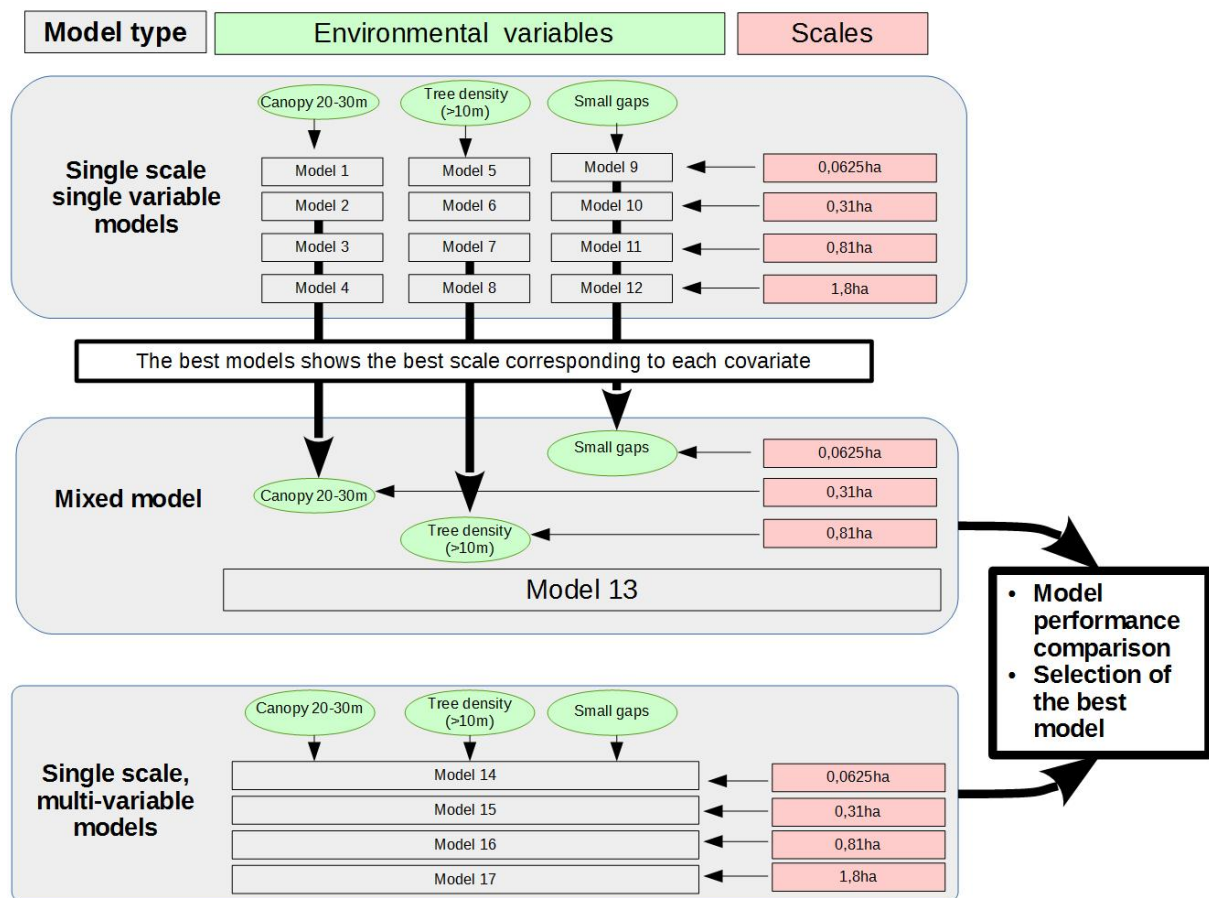
5m was added (interaction 1 case). Finally, a second interaction term was added to the interaction 1 model formula: the product between the mean and the squared standard deviation of the same variables (interaction 2 case). Model performances were then compared. For each combination of variables, a three dimensional response curve was created. One hundred values were selected across the range of each variable, thereby giving 10 000 unique combinations of values. The predictions were calculated for each combination, when all other variables were set to their mean value.

### 2.3.9 Framework for best model selection

In order to choose the best model, models are compared using two methods recently described by McGarigal et al. (2016). The first one is a pseudo-optimized multiple scale method, where each variable is evaluated separately at each scale (single-variable, single scale models), and after, is combined at their best scale. Those models will be referred to as “Mixed models” (Fig. 2). In a second stage a pseudo-optimized single scale method, where the different selected variables are evaluated simultaneously across a range of scales was applied. Those models will be referred to as “Single scale, multi-variables models” (Fig. 34). Models described as no interaction cases were used to evaluate the effect of scales.



Figure 34 : Schematic representation of model selection framework. Only three variables and four scales are represented, as an example.



## 2.4 Results

### 2.4.1 Evaluation of the effect of an interaction between two variables on response curve's interpretation and predictions

The results pertaining to single-scale, multi-variable, model's performance for the three interaction cases are presented in Table 17. Model performance did not vary between the three cases. Only the model at 27.5ha with the interaction 1 case gave a better result than the two other treatments at the same scale, for the Ain study area. In addition, the interaction feature at some scales does not participate in the model (Betas estimate =0) (Supplementary material Appendix 1 Table A1). Consequently, those models were similar to models without any interaction or with a simple interaction. In the Jura study area, at scale 0.81 ha no differences between the three case models were observed.



*Table 17 : AUC results for all single-scale, multi-variable models and the three interaction cases: no interaction, interaction between mean proportion of open area and standard deviation of penetration ratio 2–5m (interaction 1 case), interaction between mean proportion of open area, standard deviation of penetration ratio 2–5m and standard deviation of penetration ratio 2–5m (interaction 2 case).*

Scale	Study area	No interaction	Interaction 1 case	Interaction 2 case
0.0625ha	JURA	0.59	NA	NA
0.31ha		0.61	0.61	0.61
0.81ha		0.62	0.62	0.62
1.8ha		0.62	0.62	0.62
7ha		0.61	0.61	0.61
15.8ha		0.61	0.61	0.61
27.5ha		0.61	0.61	0.61
0.0625ha	AIN	0.67	NA	NA
0.31ha		0.68	0.68	0.68
0.81ha		0.7	0.7	0.7
1.8ha		0.7	0.7	0.7
7ha		0.74	0.74	0.74
15.8ha		0.75	0.75	0.75
27.5ha		0.72	0.73	0.72

When the interaction terms contributed to the models, in some cases, the response of the two variables showed a highly different pattern (Fig. 3). This response pattern was different in each study area. At lower scales, the differences in curve shape were very low or even nonexistent in each study area.

Here, the results for two scales are presented in detail: 7 ha and 27.5 ha (Fig. 35). The other responses are presented in Supplementary material Appendix 1, Fig. A1.1 to A1.4. For the Ain models at the scale of 7 ha, the response was highly different for each case. The model with no interaction



term showed that a high ( $>0.3$ ) or a very low standard deviation were predicted as favorable for the whole range of mean values. In the interaction 1 case, a high proportion of mean open areas is associated with a low suitability value. The same pattern was observed for moderate standard deviations of penetration ratio as well. In the last case, only values combining a low mean and a low standard deviation, or a high mean and a high standard deviation were predicted as favorable. At the 27.5 ha scale, the shape of the curve in the case of no-interaction-term added, was similar to the one described at the 7 ha scale (Fig. 36). The response curve observed for the two other cases was highly similar to the response observed at the 7 ha in the interaction 1 case.

In the Jura study area, the response curves pattern was different. At the 7 ha scale, higher suitable values were predicted for intermediate values of mean proportion of open areas with a low effect of the standard deviation. Nevertheless, the predictions were observed to be lower than those observed in the two first cases for high standard deviation values in the interaction 2 case. At the largest scale (27.5ha), in the no interaction case, higher suitability were predicted for a combination of moderate mean of proportion of open area and penetration ratio 2-5 standard deviation. The addition of an interaction term slightly changed the shape of the response by predicting high standard deviation values as non-suitable when the mean proportion of open area is low or reaches 100 %. The same pattern was observed in the interaction 2 case.



Figure 35 : Three dimensional response curves at 7 ha scale: suitability predictions as function of the mean proportion of open areas and of the standard deviation of 2–5 m penetration ratio, depending on the inclusion of interaction terms in the models (top row : Ain study area, and bottom row : Jura study area)

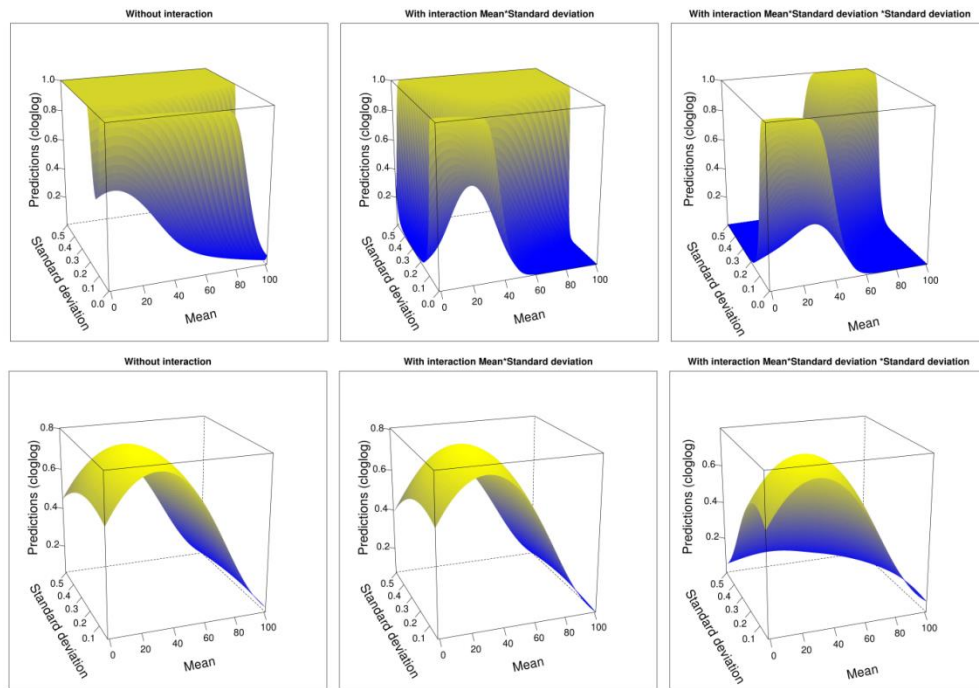
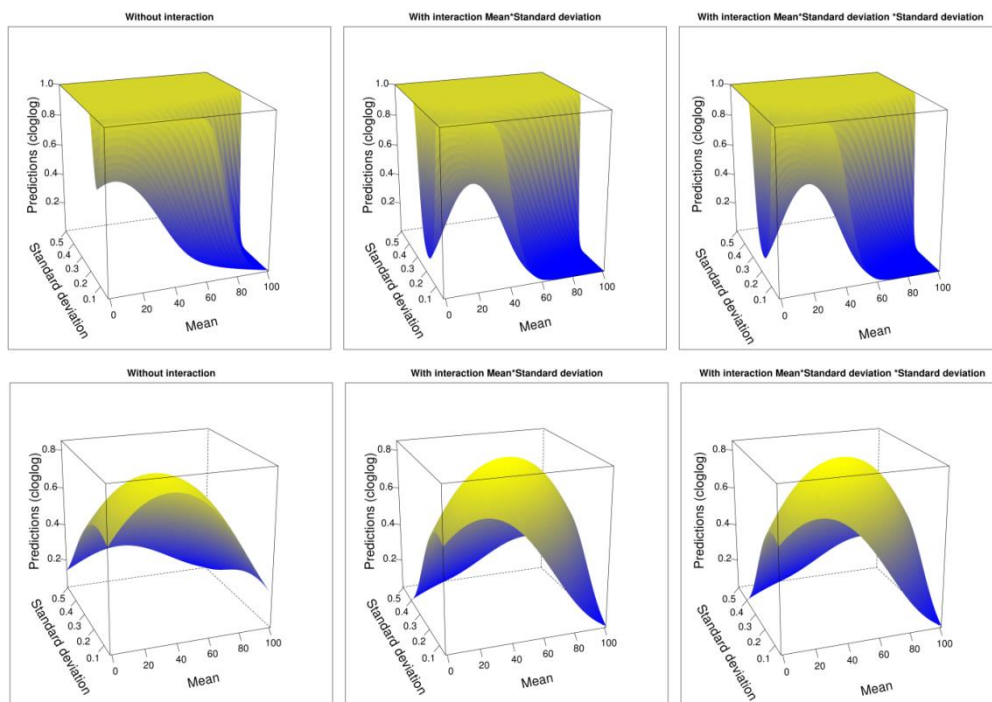


Figure 36 : Three dimensional response curves of each of the three cases at 27.5 ha scale (top row: Ain study area, and bottom row: Jura study area)







### 2.4.2 Influence of scales

#### Single-scale models

Single variable, single scale models were produced in each area and the results can be consulted in Supplementary material Appendix 2, Table. A2.1 and A2.2. Differences in best scale selection were observed between the two study areas. The middle\_gap and Tree\_gini variables were selected at the largest scales in the Ain study area but at smaller scales in the Jura models. The opposite result was observed for the proportion of open areas, Tree\_density10 and Ratio0205 variables, which were selected at small scale in the Ain models but at larger scales in the Jura models. For multiple variables, more than one scale gave the same model performances. The case of Road can be highlighted, as none of the seven scales gave a better result.

#### Multi-variables, multi-scale models

From the results of the single scale, single variable models, two mixed models were elaborated. When the same best value of AUC was observed for multiple scales, the smallest best scale was kept, as the aim is to produce a predicted map precise enough at the local scale (Table 18). The AUC value for the Ain and the Jura mixed models were respectively 0.7 and 0.62.

*Table 18 : Scales choice for the mixed models*

Variable	Ain	JURA
Grassland	0,31ha	7ha
NDVI	NA	0,31ha
NDVIsd	NA	0,31ha
Middle_gap	27,5ha	0,0625ha
Tree_gini	7ha	0,31ha
Tree_density10	0,31ha	27,5ha
Tree_densityinf10	0,31ha	0,81ha
Ratio0205sd	0,31ha	7ha
Ratio0205	1,8ha	0,81ha
Ratio0501sd	27,5ha	15,8ha
Canopy2030sd	0,81ha	0,31ha
Prop_resG	0,31ha	NA
Prop_resGsd	0,31ha	NA
Buildings	27,5ha	NA
Road	0,0625ha	NA

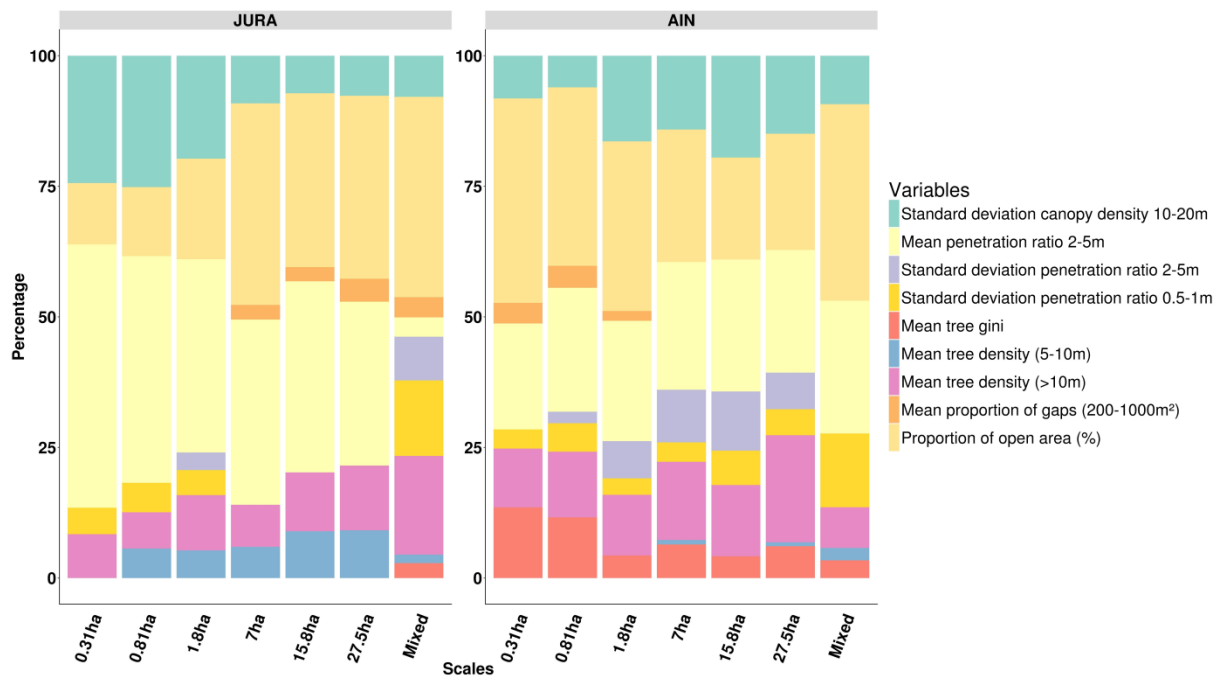
### 2.4.3 Model performance and variable contributions across scales (with no added interaction terms)

Mixed models did not perform better than all single scale, multi-variable models in the Ain study area. Among the Jura models, the performance observed was similar with the best single scales, multi-variable models and the mixed model (Table 17).

The Ain models gave better performance, than the Jura models. The best models were observed at the scales of 0.81 ha and 1.8 ha for Jura (AUC 0.62), while a scale of 15.8 ha gave the best results for the Ain models (AUC 0.75) (Table 3).

The variable contributions were calculated for multi-variables, single-scale models from 0.31ha to 27.5ha. Variables common to both areas are presented in Fig. 37, all variable contributions are presented in supplementary material Appendix 3, Fig. A3.

Four variables among ones common to models contribute the most in all models from the two study areas: the standard deviation of canopy 10-20m, proportion of open areas, mean penetration ratio 2-5 m and mean tree density (>10m) metrics. However, the contribution of Mean tree density 5-10m was high only in the Jura models; whereas the mean tree gini variable shows a non-negligible contribution in all models from Ain only. Differences between areas concerning the mean proportion of gap (200-100 m<sup>2</sup>) contribution were also observed. Indeed, in the Jura models, the contribution was higher at larger scales, unlike in the Ain models where it contributed the most at small scales. At the largest scale and in the mixed model, the building surface ratio contributes for 11% and 24% respectively but does not contribute to smaller scale models. In addition, the road variable used in the Ain models substantially impacts models from 53.8% (0.625ha) to 10.5% (15.8h).

*Figure 37 : Contributions of common variables for both study areas.*

## 2.5 Discussion

The aims of this study were (1) to measure the effect of the interaction feature on response curve interpretation and models performance, (2) to evaluate the effect of scales on model performances (3) to evaluate the transferability of LiDAR variables among study sites, for an endangered avian species in the French Jura Massif: the Hazel Grouse.

The hypothesis that the addition of product features facilitates the interpretation of species responses of specific selected variables was evaluated. Our results showed that the introduction of an interaction term between two selected variables (standard deviation of penetration ratio 2-5m and mean proportion of open area) did not improve models performances in our case. The contribution of the interaction term was in most cases very low for small scales, showing that the combined variation of the two variables was greater at large scales for Hazel Grouse. This result is not surprising, as the landscape horizontal heterogeneity described by the standard deviation is an essential data on Hazel Grouse habitat at broader scales (Mathys et al. 2006). The representation of two variables in a three dimensional surface graph can improve the understanding of species response, as it shows which is the best combination of values for the species (Zurell, Elith, and



Schröder 2012). This representation makes it possible to see easily that for example a same standard deviation value, the predicted value is highly different regarding the value of the mean of proportion of open area (or inversely), making the understanding of the species requirements more robust. When the added interaction terms contributed to the model, differences were observed in the responses curves shapes, but this influence depends on both the scale and study area.

However, the analysis of those responses showed that a prediction for extreme values may be wrongly extrapolated by the model, resulting by components being predicted as suitable when what is known of the species contradicts this observation. This illustrates the limits of feature choice that constrains the response to only two possible shapes. This limitation may be solved by adding other features like hinge or threshold, at the cost of increased model complexity (Merow et al. 2014). The use of other methods such as Random forest (Breiman 2001) or Boosted regression trees (Elith, Leathwick, and Hastie 2008) that allowed the construction of more complex responses may also be promoted (Merow et al. 2014).

Finally, models with no interaction terms added were retained as a gain in performance obtained through more complex models was not evidenced. The overall AUC values showed that our models had a moderate performance, on both sites, with better results for the Ain area. The overall moderate performance of our models may be the result of multiple causes. This moderate performance, especially observed in the Jura study area can be due to a high proportion of good habitats in the landscape, where it can be difficult to hierarchize predictions among overall good habitat pixels. The difference in performance between the two study areas may also be explained by the fact that the same variables were not used in each area, like the NDVI (Jura only), the proportion of coniferous trees (Ain only), the distance to roads (Ain only) and the building area ratio (Ain only).

Finally, the hypothesis that mixed models give better results than single scale models as it was showed in multiple studies was not confirmed in our study (Timm et al. 2016; Rostro-García et al. 2016; Vergara et al. 2016). The expectation that, the best scales observed, will be similar between the two study areas was also not validated by our results. This may indicate that results are affected



by the Hazel Grouse datasets (e.g number of observation) (Thibaud et al. 2014; Jiménez-Valverde, Lobo, and Hortal 2009) or by the differences between the LiDAR campaigns. In addition, the selection of a scale may be dependent on other factors, such as species population density, interactions with other species or time (seasons, migrations) (Prokopenko, Boyce, and Avgar 2017; Oatway and Morris 2007; Mateo-Sánchez et al. 2016). Therefore there is no overall best scale for a specific variable, and accordingly different multi-scale studies can lead to different conclusions (Mayor et al. 2009). This might be improved by using fully optimized scale selection with no a priori scale choice (McGarigal et al. 2016).

Nevertheless, similar contributions of LiDAR between the two study areas were expected. Four main contributing common variables were found to have a high impact in both study area models: the standard deviation of the canopy density, the proportion of open areas, the penetration between 2-5m and the tree density >10m. This result indicates that those four variables are describing important environmental component for the species. Those results are consistent with the literature as understory characteristics, canopy cover and height as well as the horizontal heterogeneity are known to be important components for Hazel Grouse (Rechsteiner et al. 2017; Bae et al. 2014; Zellweger et al. 2014).

### 2.6 Conclusion

Species distribution models for Hazel Grouse over two study areas over the French Jura Massif were created. The addition of an interaction between a mean and a standard deviation variable when it participated in the model slightly changed the pattern of the predictions and the predicted maps, but not model performance. A representation in a three dimensional graph helps to better visualize and understand the effect of the interaction between two variables. Interestingly, the mixed model did not perform better than single-scale, multi-models. In addition, the best selected scales were different between study areas, indicating more work is needed on this species concerning multi-scales models. Nonetheless, differences observed between study areas (best scale model), four same



main contributing common variables were evidenced on both sites at all scales, indicating the importance of those parameters for Hazel Grouse.

### 2.7 Acknowledgement

The first LiDAR survey was financed the by European Territorial Cooperation program INTERREG IV France-Switzerland 2007-2013 and the Ain department. The second LiDAR campaign was financed by Bourgogne Franche-Comté regional council. The project was financed by the departmental council of Ain.

### 2.8 Bibliography

Aberg, J, Gunnar Jansson, Jon E Swenson, and Grzegorz Mikusinski. 2000. "Difficulties in Detecting Habitat Selection by Animals in Generally Suitable Areas." *Wildlife Biology* 6 (2): 89–100.

Bae, Soyeon, Bjoern Reineking, Michael Ewald, and Joerg Mueller. 2014. "Comparison of Airborne Lidar, Aerial Photography, and Field Surveys to Model the Habitat Suitability of a Cryptic Forest Species—the Hazel Grouse." *International Journal of Remote Sensing* 35 (17): 6469–89.

Bellon, B. 2011. "Évolution de La Répartition Communale Du Petit Gibier de Montagne En France Décennie 2000-2009." *Supplément Faune Sauvage*, no. 290.

Bivand, R, R Krug, M Neteler, and S Jeworutzki. 2016. "rgrass7: Interface between GRASS 7 Geographical Information System and R." *R Software Package Google Scholar*.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Dormann, Carsten F, Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R García Marquéz, Bernd Gruber, Bruno Lafourcade, and Pedro J Leitão. 2013. "Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance." *Ecography* 36 (1): 27–46.

Elith, Jane, John R Leathwick, and Trevor Hastie. 2008. "A Working Guide to Boosted Regression Trees." *Journal of Animal Ecology* 77 (4): 802–13.



Elith, Jane, Steven J Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J Yates. 2011. “A Statistical Explanation of MaxEnt for Ecologists.” *Diversity and Distributions* 17 (1): 43–57.

Eysn, Lothar, Markus Hollaus, Eva Lindberg, Frédéric Berger, Jean-Matthieu Monnet, Michele Dalponte, Milan Kobal, Marco Pellegrini, Emanuele Lingua, and Domen Mongus. 2015. “A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space.” *Forests* 6 (5): 1721–47.

Franklin, Janet. 2009a. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press.

———. 2009b. “Moving beyond Static Species Distribution Models in Support of Conservation Biogeography.” *Diversity and Distributions* 16 (3): 321–30.

Guisan, Antoine, Wilfried Thuiller, and Niklaus E Zimmermann. 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press.

Guisan, Antoine, Reid Tingley, John B. Baumgartner, Ilona Naujokaitis-Lewis, Patricia R. Sutcliffe, Ayesha I. T. Tulloch, Tracey J. Regan, et al. 2013. “Predicting Species Distributions for Conservation Decisions.” *Ecology Letters* 16 (12): 1424–35. <https://doi.org/10.1111/ele.12189>.

Guisan, Antoine, and Niklaus E Zimmermann. 2000. “Predictive Habitat Distribution Models in Ecology.” *Ecological Modelling* 135 (2): 147–86.

Hofstetter, Lea, Raphaël Arlettaz, Kurt Bollmann, and Veronika Braunisch. 2015. “Interchangeable Sets of Complementary Habitat Variables Allow for Flexible, Site-Adapted Wildlife Habitat Management in Forest Ecosystems.” *Basic and Applied Ecology* 16 (5): 420–33.

Huhta, Esa, Pekka Helle, Vesa Nivala, and Ari Nikula. 2017. “The Effect of Human-modified Landscape Structure on Forest Grouse Broods in Two Landscape Types.” *Ecosphere* 8 (9).

Jiménez-Valverde, A, J Lobo, and J Hortal. 2009. “The Effect of Prevalence and Its Interaction with Sample Size on the Reliability of Species Distribution Models.” *Community Ecology* 10 (2): 196–205.



Johnson, Chris J, and Michael P Gillingham. 2005. "An Evaluation of Mapped Species Distribution Models Used for Conservation Planning." *Environmental Conservation* 32 (2): 117–28.

Leclercq, B. 1985. "Influence Des Routes et Voies de Pénétration Humaine Sur Les Comportements de Grands Tétràs et de Gélínottes Dans Le Haut-Jura Français." Actes du colloque "Routes et faune sauvage", Strasbourg.

Mateo-Sánchez, María C, Aitor Gastón, Juan I García-Viñas, Jorge Cuevas, César López-Leiva, Alfredo Fernández-Landa, Nur Algeet-Abarquero, Miguel Marchamalo, Marie-Josée Fortin, and Santiago Saura. 2016. "Seasonal and Temporal Changes in Species Use of the Landscape: How Do They Impact the Inferences from Multi-Scale Habitat Modeling?" *Landscape Ecology* 31 (6): 1261–76.

Mathys, Lukas, Niklaus E Zimmermann, Niklaus Zbinden, and Werner Suter. 2006. "Identifying Habitat Suitability for Hazel Grouse *Bonasa Bonasia* at the Landscape Scale." *Wildlife Biology* 12 (4): 357–66.

Mayor, Stephen J, David C Schneider, James A Schaefer, and Shane P Mahoney. 2009. "Habitat Selection at Multiple Scales." *Ecoscience* 16 (2): 238–47.

McGarigal, Kevin, Ho Yi Wan, Kathy A Zeller, Brad C Timm, and Samuel A Cushman. 2016. "Multi-Scale Habitat Selection Modeling: A Review and Outlook." *Landscape Ecology* 31 (6): 1161–75.

Melin, M., L. Mehtätalo, J. Miettinen, S. Tossavainen, and P. Packalen. 2016. "Forest Structure as a Determinant of Grouse Brood Occurrence – An Analysis Linking LiDAR Data with Presence/absence Field Data." *Special Section: Drought and US Forests: Impacts and Potential Management Responses* 380 (November): 202–11. <https://doi.org/10.1016/j.foreco.2016.09.007>.

Merow, Cory, Mathew J Smith, Thomas C Edwards, Antoine Guisan, Sean M McMahon, Signe Normand, Wilfried Thuiller, Rafael O Wüest, Niklaus E Zimmermann, and Jane Elith. 2014. "What Do We Gain from Simplicity versus Complexity in Species Distribution Models?" *Ecography* 37 (12): 1267–81.





Müller, Daniel, Boris Schröder, and Jörg Müller. 2009. "Modelling Habitat Selection of the Cryptic Hazel Grouse *Bonasa Bonasia* in a Montane Forest." *Journal of Ornithology* 150 (4): 717–32.

Oatway, ML, and DW Morris. 2007. "Do Animals Select Habitat at Small or Large Scales? An Experiment with Meadow Voles (*Microtus Pennsylvanicus*)." *Canadian Journal of Zoology* 85 (4): 479–87.

Phillips, Steven J., Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, and Mary E. Blair. 2017. "Opening the Black Box: An Open-Source Release of Maxent." *Ecography*, May, n/a-n/a. <https://doi.org/10.1111/ecog.03049>.

Phillips, Steven J, Miroslav Dudík, Jane Elith, Catherine H Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data." *Ecological Applications* 19 (1): 181–97.

Prokopenko, Christina M, Mark S Boyce, and Tal Avgar. 2017. "Extent-Dependent Habitat Selection in a Migratory Large Herbivore: Road Avoidance across Scales." *Landscape Ecology* 32 (2): 313–25.

Räty, M. 1979. "Effect of Highway Traffic on Tetraonid Densities." *Ornis Fennica* 56 (4): 169–70.

Rechsteiner, Christian, Florian Zellweger, Anatole Gerber, Frank T. Breiner, and Kurt Bollmann. 2017. "Remotely Sensed Forest Habitat Structures Improve Regional Species Conservation." *Remote Sensing in Ecology and Conservation*, n/a-n/a. <https://doi.org/10.1002/rse2.46>.

Rostro-García, Susana, Lhendup Tharchen, Leandro Abade, Christos Astaras, Samuel A. Cushman, and David W. Macdonald. 2016. "Scale Dependence of Felid Predation Risk: Identifying Predictors of Livestock Kills by Tiger and Leopard in Bhutan." *Landscape Ecology* 31 (6): 1277–98. <https://doi.org/10.1007/s10980-015-0335-9>.

Roussel, Jean-Romain, and David Auty. 2016. *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. R Package Version 1.0.0. <https://github.com/Jean-Romain/lidR>.



Sachot, Sebastien, Nicolas Perrin, and Cornelis Neet. 2003. "Winter Habitat Selection by Two Sympatric Forest Grouse in Western Switzerland: Implications for Conservation." *Biological Conservation* 112 (3): 373–82.

Schäublin, Sabrina, and Kurt Bollmann. 2011. "Winter Habitat Selection and Conservation of Hazel Grouse (*Bonasa Bonasia*) in Mountain Forests." *Journal of Ornithology* 152 (1): 179–92.

Sitzia, Tommaso, Matteo Dainese, Thomas Clementi, and Silvano Mattedi. 2014. "Capturing Cross-Scalar Variation of Habitat Selection with Grid Sampling: An Example with Hazel Grouse (*Tetrastes Bonasia* L.)." *European Journal of Wildlife Research* 60 (2): 177–86.

Stommel, Henry. 1963. "Varieties of Oceanographic Experience." *Science* 139 (3555): 572–76.

Swenson, Jon E. 1993. "The Importance of Alder to Hazel Grouse in Fennoscandian Boreal Forest: Evidence from Four Levels of Scale." *Ecography* 16 (1): 37–46.

Thibaud, Emeric, Blaise Petitpierre, Olivier Broennimann, Anthony C. Davison, and Antoine Guisan. 2014. "Measuring the Relative Effect of Factors Affecting Species Distribution Model Predictions." *Methods in Ecology and Evolution* 5 (9): 947–55. <https://doi.org/10.1111/2041-210X.12203>.

Timm, Brad C., Kevin McGarigal, Samuel A. Cushman, and Joseph L. Ganey. 2016. "Multi-Scale Mexican Spotted Owl (*Strix Occidentalis Lucida*) Nest/roost Habitat Selection in Arizona and a Comparison with Single-Scale Modeling Results." *Landscape Ecology* 31 (6): 1209–25. <https://doi.org/10.1007/s10980-016-0371-0>.

Vergara, Maria, Samuel A. Cushman, Fermín Urra, and Aritz Ruiz-González. 2016. "Shaken but Not Stirred: Multiscale Habitat Suitability Modeling of Sympatric Marten Species (*Martes Martes* and *Martes Foina*) in the Northern Iberian Peninsula." *Landscape Ecology* 31 (6): 1241–60. <https://doi.org/10.1007/s10980-015-0307-0>.

Wiens, John A. 1989. "Spatial Scaling in Ecology." *Functional Ecology* 3 (4): 385–97.



Zellweger, Florian, Veronika Braunisch, Andri Baltensweiler, and Kurt Bollmann. 2013. "Remotely Sensed Forest Structural Complexity Predicts Multi Species Occurrence at the Landscape Scale." *Forest Ecology and Management* 307 (November): 303–12. <https://doi.org/10.1016/j.foreco.2013.07.023>.

Zellweger, Florian, Felix Morsdorf, Ross S Purves, Veronika Braunisch, and Kurt Bollmann. 2014. "Improved Methods for Measuring Forest Landscape Structure: LiDAR Complements Field-Based Habitat Assessment." *Biodiversity and Conservation* 23 (2): 289–307.

Zurell, Damaris, Jane Elith, and Boris Schröder. 2012. "Predicting to New Environments: Tools for Visualizing Model Behaviour and Impacts on Mapped Distributions." *Diversity and Distributions* 18 (6): 628–34.



### 3 UNIVARIATE RESPONSE CURVES ANALYSIS

---

The analysis of the univariate response curve helps to understand the species preferences and the model predictions. Here, the description and interpretation of the response for Hazel Grouse single-scale, multi-variable models will be presented.

#### 3.1 Methods

##### 3.1.1 Response curves calculation

The response curves were calculated using normalized variables. In order to obtain meaningful and interpretable results, the corresponding raw variables values were calculated afterwards. In addition, the density curves of the variables through the study areas were calculated with the aim to facilitate the interpretation of the responses.

##### 3.1.2 Expected response

As Hazel Grouse occurs in forests with moderate canopy cover, a response of the bird favoring moderate tree (>10m) density but higher densities for smaller trees (5-10m height) was expected. A moderate proportion of middle size gaps was expected to be optimal at larger scales.

Due to the importance of a horizontally heterogeneous habitat for Hazel Grouse, a response that favors moderate standard deviation values for the metrics penetration ratio 0.5-1 m, penetration ratio 2-5 m and canopy density 20-30 m was expected. However, higher values were expected to be unfavorable as it represents an extremely fragmented habitat with the presence of forest and non-forest areas (Bae et al., 2014b; Rechsteiner et al., 2017b). This effect should be particularly important over the Ain study area, which comprises a large part of non-forest areas.

Hazel Grouse favors vertically heterogeneous habitat thus higher values of tree gini index were expected to be more suitable. As Hazel Grouse does not favor large open areas, large proportions of open areas should be predicted unsuitable for the species at all scales.



The positive response of Hazel Grouse to a high proportion of coniferous trees and its negative response to a complete absence of either tree types were expected as the bird favors conifer-dominated forests (Sachot et al., 2003a). Similarly, high NDVI values were expected to be preferred, as a higher values indicate a highly productive forest. Concerning the response to the standard deviation of those two variables, the preference of Hazel Grouse for low values, indicating small variations of the proportion of coniferous trees in the landscape, was expected.

Finally, a negative response to human disturbances represented in our models by the distance to main roads and the buildings surface area ratio is expected.

### 3.2 Results

Univariate response curves of the four most contributing variables in both study areas presented in the precedent section (the penetration ratio 2-5 m, the standard deviation of the canopy density 20-30 m, the proportion of open areas, the density of tree (> 10 m)) will be described in detail. In addition, the densities of each variable across the landscape were plotted in order to facilitate the interpretation of the responses curves. All response curves are presented in Figures 39 to 41.

In both study areas, the response of the species to the mean penetration ratio showed the same quadratic pattern, with an optimal value at 0.3.

The response curves were different in each study area for the variable standard deviation of canopy density 20-30m. In the Ain study area, low values were preferable at small scales (0.31 ha, 0.81 ha and 1.8 ha) with linear responses. At the scales, 7 ha and 15.8 ha quadratic shape with an optimum for moderated values were observed. At the largest scale, higher values of standard deviation were predicted as the more favorable. In the Jura study area, the highest suitability was predicted for a standard deviation of 0 at all scales. Within the Ain landscape, the density curve showed that values between 0 and 0.05 are the most represented and that the maximum value is 0.1, whereas within the Jura landscape a second peak is observed for the values between 0.05 and 0.1.



Concerning the variable proportion of open areas, the response is also different between study sites. Indeed, in the Ain study area, a low proportion of open areas (5% to 25%) was more suitable at large scales. At small scale, a proportion of 0% was more favorable. Within the Jura study area, a proportion between 20 and 50% was predicted as more suitable. Interestingly, at small scales, a maximum proportion of open areas was showed to be more favorable, due to the linear shape of the curve. In the landscape, most of the values were close to an open area proportion of 0%. A second peak was observed for a proportion >90% in the Ain area but not in the Jura study area.

In the Ain study area, the density of trees (>10m) responses showed different patterns among scales. At the 0.31 ha scale, a density of 200 trees per hectares was more favorable. However, at larger scales, a density of tree between 0 and 100 trees was more suitable for the species. In the Jura study area, at small scales, low densities of trees were favored (linear response), but at large scales values around 200 trees were showed as not favorable contrary to lower and higher densities. Within the landscape, in the Ain area, two main well represented ranges of values were observed: a density of 0 trees for one peak, and a density of 200 trees for the second peak. Within the Jura study area, a high proportion of 200 trees was observed as well.

Figure 38 : Responses curves in the Ain study areas (part 1)

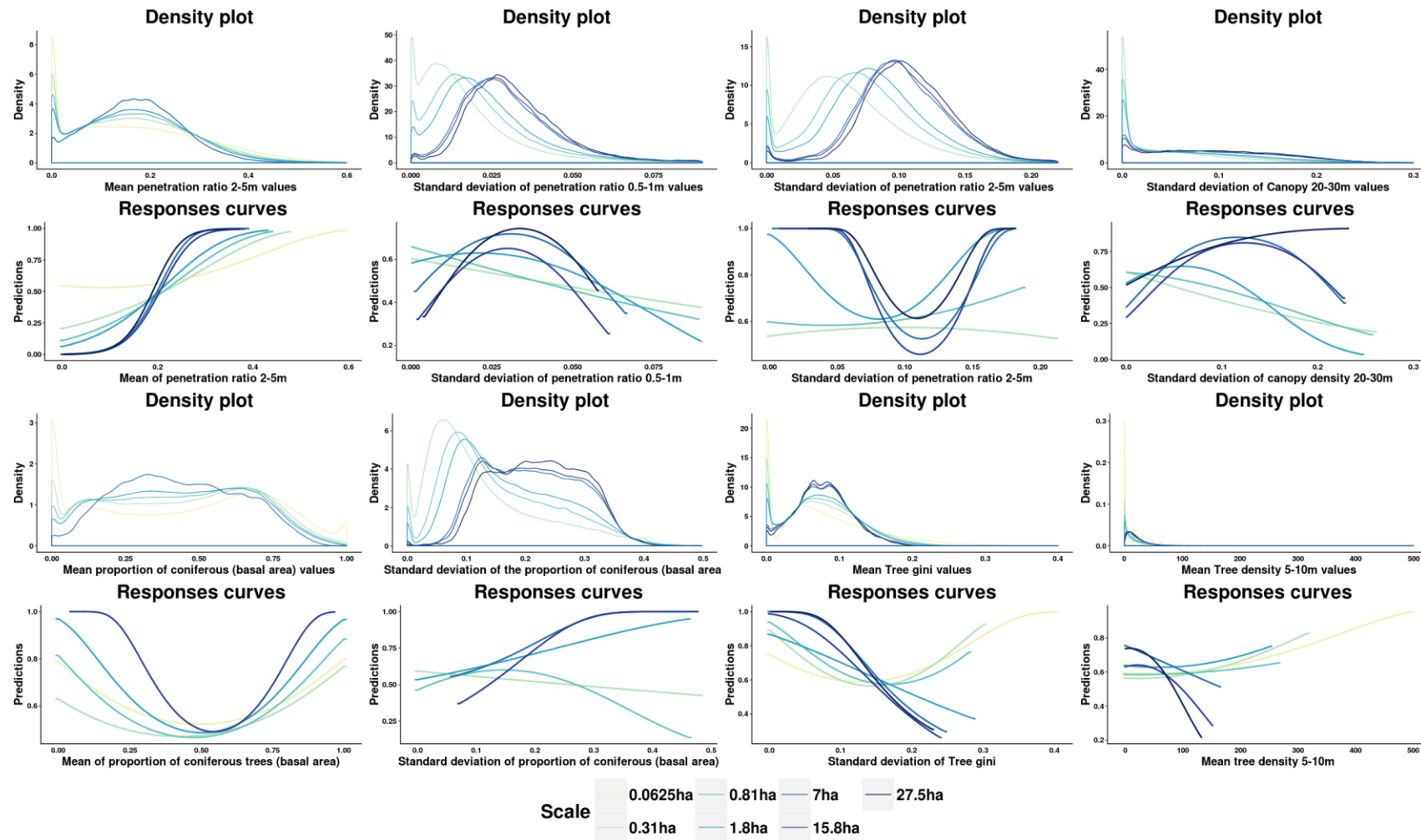




Figure 39 : Responses curves in the Ain study areas (part 2)

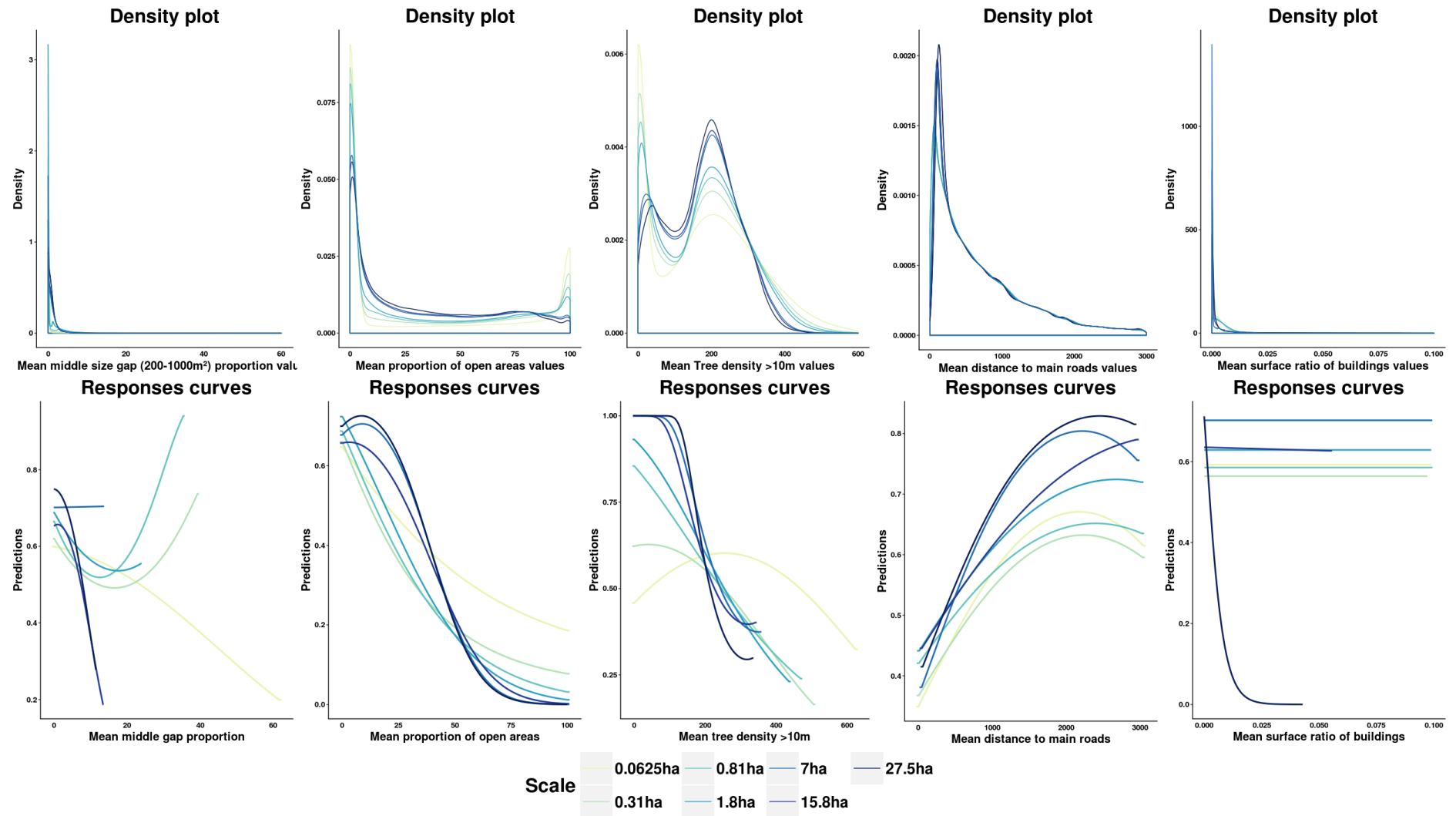






Figure 40 : Responses curves in the Jura study areas (part 1)

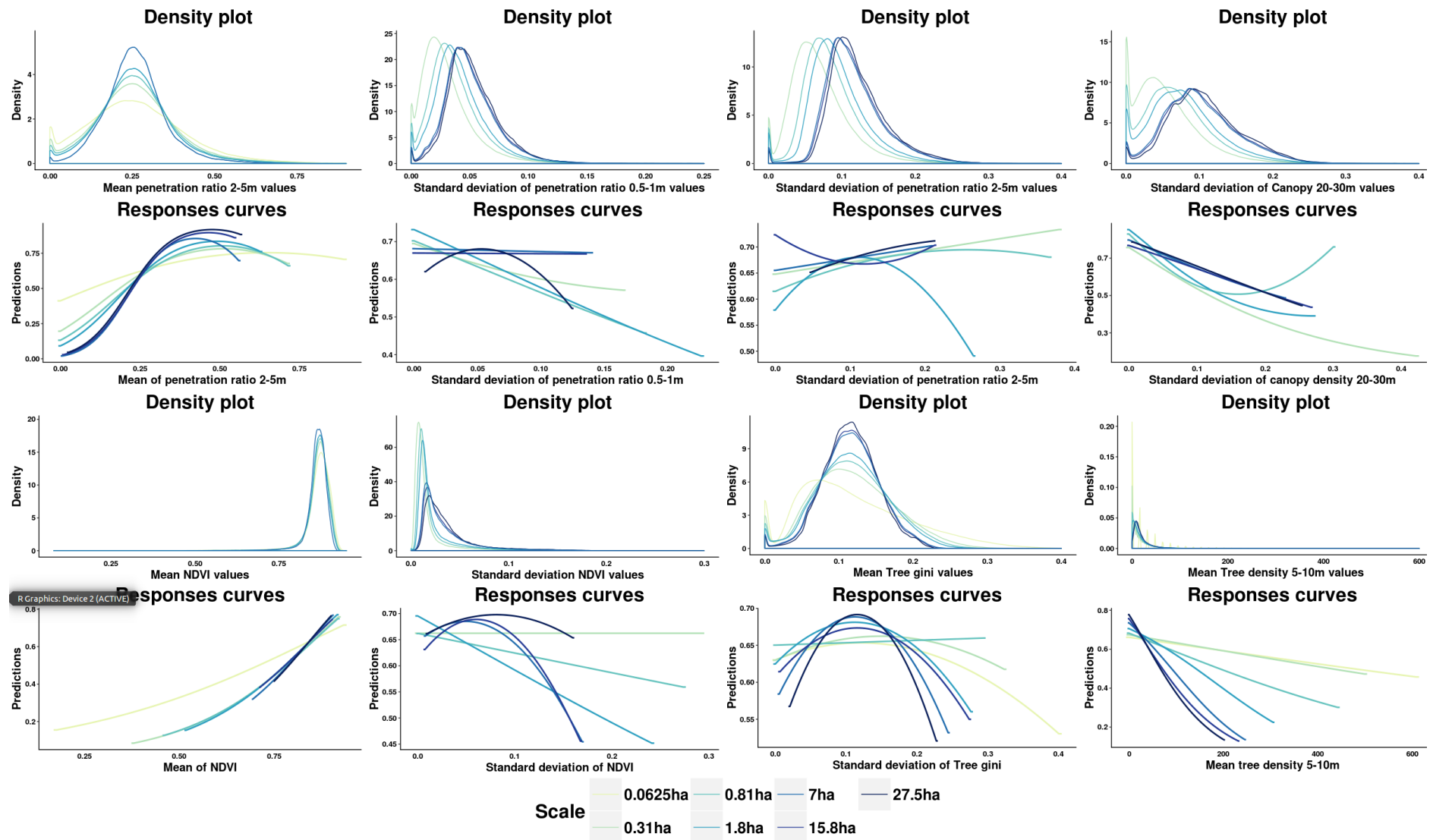
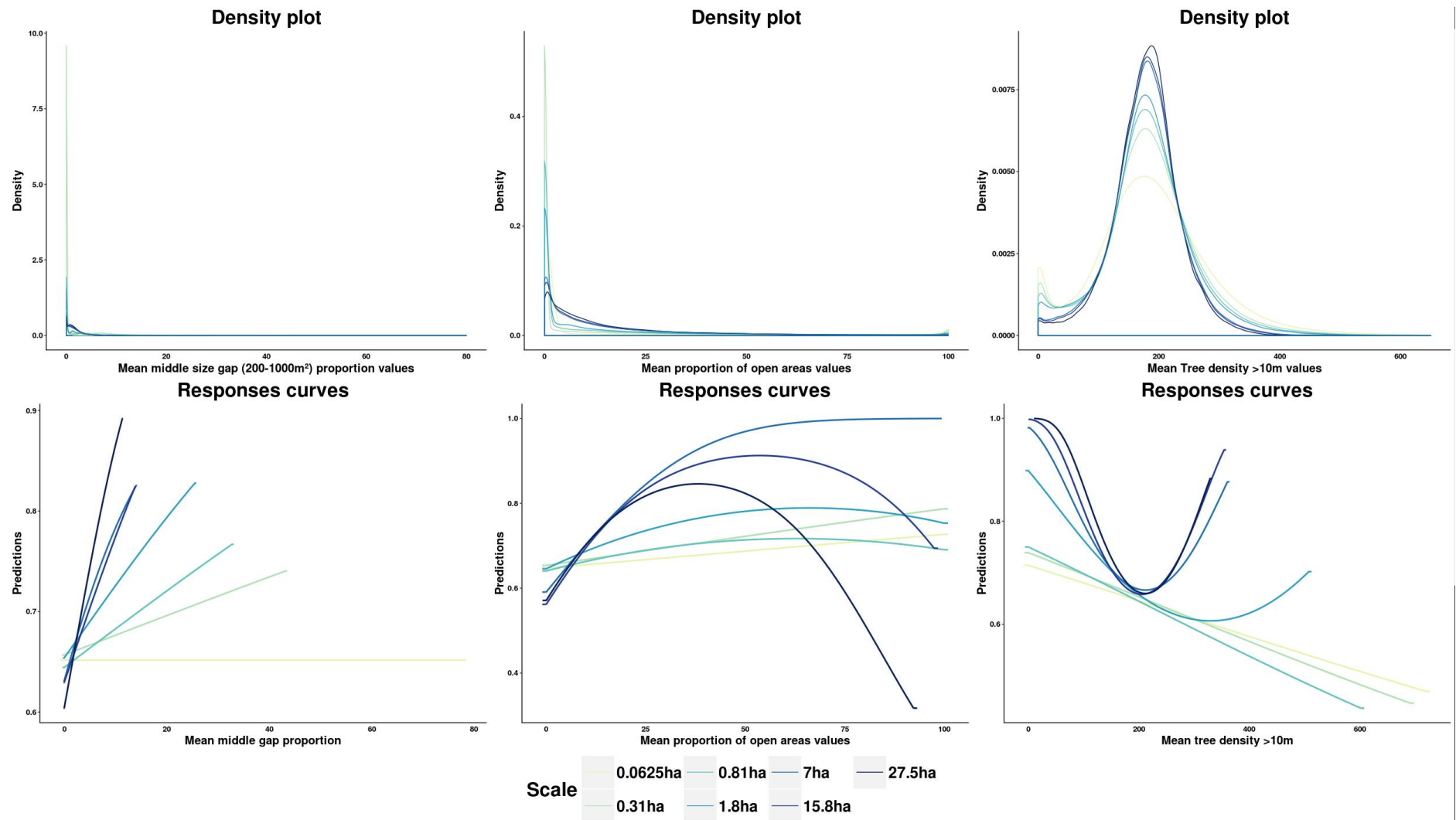




Figure 41 : Responses curves in the Jura study areas (part 2)





### 3.3 Discussion

The response to the mean penetration ratio 2-5m variable is consistent between the study areas, confirming that the characteristics of the middle-high vegetation layer are a key driver for Hazel Grouse distribution (Bae et al., 2014b). In addition, a moderate tree density was predicted as favorable in both areas joining the observations of (Sachot et al., 2003a) that recommend a canopy cover of 50% for the species.

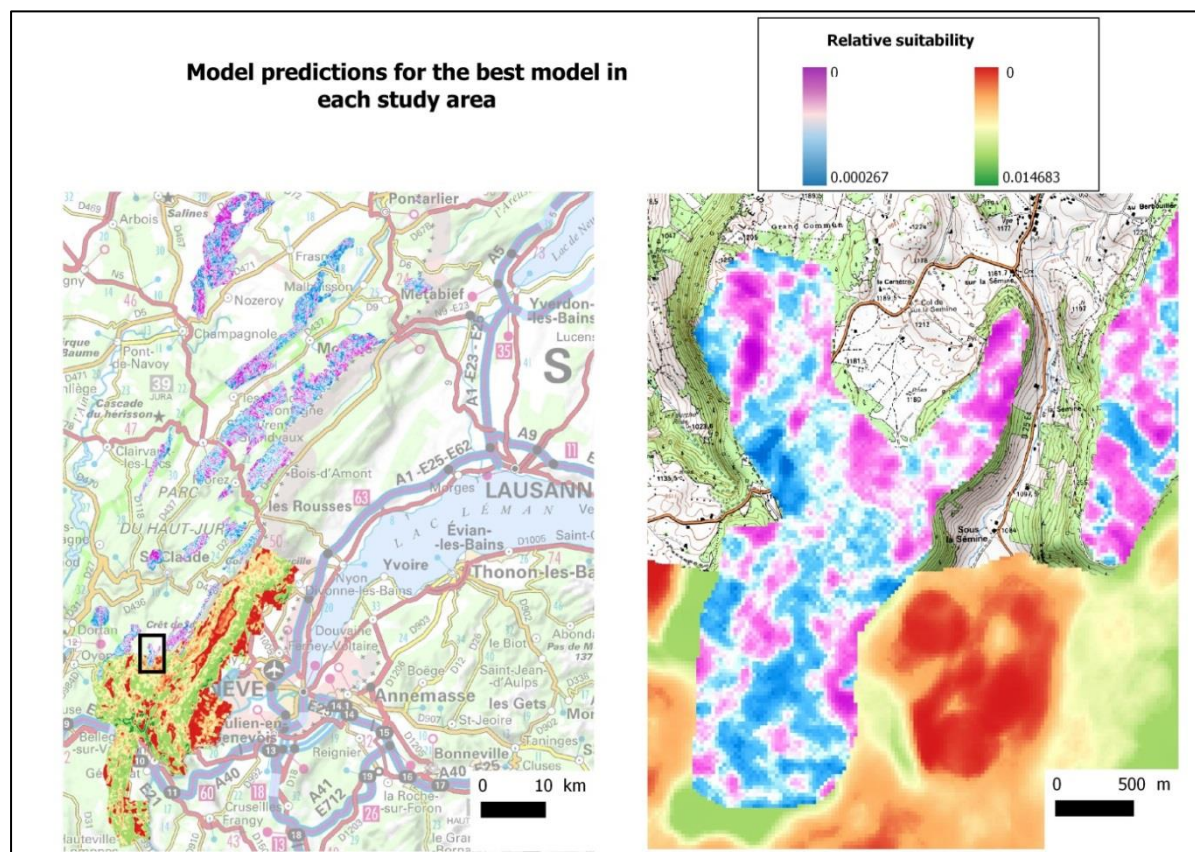
However, for the other variables differences were observed in the responses curve shapes between both scales and study areas. As the two study areas are located in the same mountain massif, high differences in habitats used by the different Hazel Grouse populations are not likely to occur. Thus, the differences observed between models are mainly due to the study area characteristics. The density curves from the two study areas confirmed the existence of those differences in terms of distributions shape and values, which may have influenced the models' performance and responses. This overall result indicates that the models are sensitive to the choice of scale and that this effect is depending on the study area location and extent. This might be particularly important at large scales, where a strong border effect can occur for the forest patches close to the border of the study area causing high variation in standard deviation and mean estimations. This might explain why the preference for high values of variable standard deviation of canopy density 20-30m was not observed in the Jura study area which is fragmented in forest patches. Similarly, in the Jura area, the models failed to predict the open areas as non-favorable, which might be due to the fact that all observations and tracks outside the forest areas were removed as there was no LiDAR available for modeling. The predictions accuracy may be ameliorated by the acquisition of an extended LiDAR coverage allowing the reducing of the border effect observed in the Jura study area.

## 4 RELATIVE SUITABILITY MAP PREDICTIONS

### 4.1 Results

Habitat suitability predictions were calculated for each model and study area. The exponential (also called raw) output was created, giving relative values which indicate the suitability of the pixel relative to the other pixels in the study area. Such maps are thus indicating the best habitat regarding the habitat quality in the overall study area only. In the Jura model, which is displayed for one of the best models at the scale 0.81 ha (Figure 42) the fragmentation of the potential favorable habitat in multiple patches was observed. At larger scales, in the Ain study area, the results are smoothed, showing large no favorable areas corresponding to open areas and the presence of cities. Nevertheless, in more favorable areas, the fragmentation between moderate to high suitability patches was still observed showing that the best habitats are fragmented.

*Figure 42 : Predicted maps of two best models within the two study areas.*





## 4.2 Discussion

The predicted maps showed in both areas that the Hazel Grouse habitat is fragmented at large scale because of the discontinuities between the forest patches. Within those forest patches, the habitat is highly fragmented. The causes of this fragmentation are probably multiple, but differences in forest management practices between municipalities and also between private owners are probably participating to this spatial structure. We can also notice that the models predicted suitable habitat patches in the state forest of La Fresse located at low altitudes and where the occurrence of the species is yet unknown. The predicted map can be used as a basis for future sampling design by optimizing the surveys between suitable and non-suitable patches in under-sampled areas.

## 5 CONCLUSION

---

SDMs for Hazel Grouse over two study areas over the French Jura Massif were created and showed that the suitable habitat for the species is fragmented. Four same main contributing common variables describing the characteristics of the understory and of the canopy were evidenced on both sites at all scales, indicating the importance of those parameters for Hazel Grouse. The negative impact of open areas was also observed. A representation in a three-dimensional graph can help to better visualize and understand the effect of the interaction, but in our study case this was not an essential variable impacting the model accuracy. The prediction of potentially suitable habitat at low elevation in the Jura massif enforces the need for wider surveys of Hazel Grouse in the future.



# CHAPTER 5: CAPERCAILLIE

## DISTRIBUTION MODELS



Bernard Bellon



## 1 INTRODUCTION

---

The aim of this chapter is to set out Capercaillie distribution models that will be used to plan future forest management actions in favor of the species. To achieve this objective, models using meaningful LiDAR variables for both the target species and managers that will use the models results will be created. Indeed, if LiDAR point clouds area based variables are commonly used; those variables are difficult to interpret as those are generally composed of indexes (ratios, strata point's densities) that do not fit the conceptual view of managers. Yet object-oriented metrics that may be easier to interpret due to the correspondence with real world objects (e.g trees) are not used in SDMs outside of forestry applications.

In a first part, the comparison between LiDAR point-cloud area-based metrics to potentially more interpretable object-oriented metrics in models will be explored over the Ain study area. Variations in performances, response curves, variable contributions and predictions will be investigated using single-scale, multi-variable models. This study lead to the submission of a third article in the Remote Sensing in Ecology and Conservation journal.

In a second part, models for Capercaillie will be set out for both Ain and Jura study areas (corresponding to the two LiDAR campaigns) using the object-oriented metrics tested in the Ain study area (Article 3) with an additional LiDAR modelled based variable. The availability of a vegetation plots dataset in the Ain study area, where numerous components were measured such as the basal area, allowed us to estimate the basal area using the LiDAR point-cloud (see Chapter 2 for more details). The basal area is an index widely used in forestry by practitioners, thus the use of such a variable combined with object-oriented variables in Capercaillie distribution models might be useful for improving the transfer of the results to managers. Those models are meant to be used and distributed by Groupe Tétràs Jura, consequently the results obtained in this section for winter and summer are also presented in a report in French (Annexe 8). Capercaillie models were created using the same variables and methods for both the winter and summer seasons except for the variable



distance to ski runs (winter only). Indeed, winter data represented most of the available data (See Chapter 2) and is also a critical season for Capercaillie survival (Gjerde and Wegge, 1989b). Summer habitat use has different specificities as it is the period when female raise their chicks (e.g. presence of grassland and shelters). The presence of bilberry patches is becoming increasingly essential as birds need to prepare for the next winter (Storch, 1993c). Unfortunately, a specific model focusing on brood's habitat could not be intended due to a lack of observations (n=51).





## 2 ARTICLE 3: MAPPING AND MANAGING BIRD HABITAT WITH LiDAR-BASED MULTI-SCALE OBJECT-ORIENTED INDICATORS

---

Anouk Glad<sup>1,3</sup> <https://orcid.org/0000-0001-5810-189X>

Björn Reineking<sup>1</sup> <https://orcid.org/0000-0001-5277-9181>

Marc Montadert<sup>2</sup>

Alexandra Depraz<sup>3</sup>

Jean-Matthieu Monnet<sup>1</sup> <https://orcid.org/0000-0002-9948-9891>

<sup>1</sup> Univ. Grenoble Alpes, Irstea, UR LESSEM, BP 76, 38402 St-Martin-d'Hères, France

<sup>2</sup> ONCFS, 90 impasse des daude, 74320 Sévrier, France

<sup>3</sup> Groupe Tétrás Jura, 9 impasse du tacon, 39370 les Bouchoux, France

Corresponding author: Anouk Glad, [gladanouk@gmail.com](mailto:gladanouk@gmail.com)

**Submitted to the journal “Remote sensing in ecology and conservation” on August 8<sup>th</sup> 2018.**

**Supporting information is available in Annex 6.**

### 2.1 Abstract

Species Distribution models (SDMs) are widely used to plan actions for species of conservation interest. Models that will be turned into conservation actions need predictors that are both ecologically pertinent and fitting managers' conceptual view of ecosystems. Remote sensing technologies such as Light Detection and Ranging (LiDAR) can describe landscapes at high resolution over large spatial areas and have already given promising results for modeling forest species distributions. However, the point-cloud (PC) area-based LiDAR variables generally used for this purpose may be challenging to interpret and thus use by managers at the operational scale. Here, we

start with the assumption that object-oriented (OO) metrics in SDMs will be better interpreted and used by managers. We tested two hypotheses: (1) the use of OO variables in SDMs will give similar performance as PC area-based models; and (2) OO variables will improve model's transferability to other areas. Using the case of an endangered forest bird, the Capercaillie (*Tetrao urogallus*), model's performances and predictions were compared with the two variable types.

Models using OO variables showed slightly lower discriminatory performance than PC-based models (average  $\Delta$ AUC = -0.025 and -0.013 for females and males, respectively), with differences most pronounced at larger spatial scales. OO-based models were as transferable (absolute difference in correlation of predictions  $\leq$  0.01) or more transferable than PC-based models, with the exception of models for females at large spatial scales.

In sum, LiDAR derived object-oriented metrics showed promising results for use in species distribution models for conservation, and their potential in improving the scientist-stakeholder interface through their better interpretability should be explored.

Keywords: species distribution models, LiDAR, object-oriented metrics, conservation

## 2.2 Introduction

Species Distribution Models (SDMs), also known as habitat-suitability models or environmental niche models, are statistical models that relate species occurrence to environmental factors (biotic and/or abiotic). SDMs quantify the species' response to the environmental variables and provide predicted spatial distribution maps. Both results can be used in order to plan conservation actions in favor of a target species (Franklin, 2009, 2009; Guisan and Thuiller, 2005; Johnson and Gillingham, 2005): the response to the environmental variables can help to characterize the optimal habitat and the distribution maps can be used to target areas for prospective conservation projects (Franklin, 2009). Regardless the choice of modeling method, the use of environmental predictors that are ecologically relevant is essential to obtain reliable results and to best explain the species distribution in the landscape (Fourcade et al., 2018; Johnson and Gillingham, 2005).



Advances in remote sensing technologies are making it a powerful tool for improving SDMs (He et al., 2015). One of these methods is Light Detection and Ranging (LiDAR), which provides highly precise data on the three dimensional structure of the environment at high resolution over large areas (Vierling et al., 2008). LiDAR has been used in SDMs for a wide variety of species including forest birds, which are sensitive to variations in vegetation structure (Bergen et al., 2007, 2009). The use of LiDAR has given promising results by significantly improving model's performance either alone or when combined with other predictors (Bae et al., 2014; Farrell et al., 2013; Graf et al., 2009).

However, the production of accurate models is not enough to obtain better conservation action results, and pitfalls between scientists' research and effective conservation decisions by managers are still to be reduced (Guisan et al., 2013; Villero et al., 2017). A better impact of modeling results depends on communication, mediation, and close collaboration with decision makers from the beginning of the process (Guisan et al., 2013). Concerning this aspect, the LiDAR variables proposed so far in different studies to improve SDMs may not be ideal, as they appear to be difficult to interpret and use by managers at the operational scale: to our knowledge, LiDAR extracted metrics for use in conservation biology were so far overwhelmingly calculated using the point-cloud (PC) area-based method, which summarizes the characteristics of the point-cloud within a given a region of interest (pixel). As a concrete example, variables such as the standard-deviation of 1th quantile (Zellweger et al., 2013), the proportion of echos above five meters (Melin et al., 2016) or height skewness of the returns located between 1.5 m and 5 m (Kortmann et al., 2018) were used to assess the local canopy or shrub heterogeneity and density. However, it seems yet difficult to link an optimal value of each of those variables to an actual canopy or shrub cover assessment as perceived by managers in the field. As a consequence, those results may not easily be converted into effective management solutions and actions in the field because of a lack of understanding between the two communities.

Advances in image processing promoted the use of object-oriented (OO) methods with the objective to overcome PC area-based limitations (Benz et al., 2004). Contrary to the area-based metrics, object-related metrics are defined by known common characteristics (structure, shape, texture),



describing a real-world object such as buildings, roads, or trees. Thus, with LiDAR data, object types can be defined by their three dimensional characteristics and be extracted from the raw point-clouds. This type of approach has been used already as a forest management tool, exploiting single tree detection methods (Munoz et al., 2015; Othmani et al., 2013). The promotion of this type of approach is thus particularly interesting for forest species that need specific habitat management actions.

The Capercaillie (*Tetrao urogallus*) is a species of conservation concern at the regional scale, occurring in the Jura massif (France and Switzerland). Capercaillie is threatened in Western Europe, showing a contraction of its occurring range due mainly to habitat loss and alteration (Mikoláš et al., 2017; Storch, 2007). More particularly, the isolated population located in the Jura massif is critically endangered (Sachot, 2002; Storch, 1994). In France, Capercaillie favors old mixed forests constituted of a mosaic of habitats (Sachot et al., 2003). The species is strictly dependent on a high proportion of moderate canopy density favoring the growth of a key food resource: the bilberry (*Vaccinium myrtillus*) (Storch, 1993). In addition, according to Capercaillie seasonal needs, its home range needs to contain mixed patches of young forest stages (predator shelter), clear-cuts (broods rearing), and patches of clear canopy density with fir (*Abies alba*) (Rolstad, 1988; Rolstad et al., 1988; Sachot, 2002; Storch, 1994). In this context, obtaining a potential Capercaillie distribution at the regional scale holds great potential to optimize and plan conservation actions. Numerous studies on this species have already been conducted in other countries to produce such species distribution maps (Braunisch and Suchant, 2007; Graf et al., 2007; Sachot et al., 2006; Teuscher et al., 2013). More recently, in order to improve models, LiDAR point-cloud area based derived predictors were used as well (Graf et al., 2009; Kortmann et al., 2018; Melin et al., 2016; Zellweger et al., 2013).

We hope that the use of OO LiDAR metrics as environmental variables, when the object types are decided on in close collaboration with practitioners, may improve the use of SDMs results for conservation management projects.



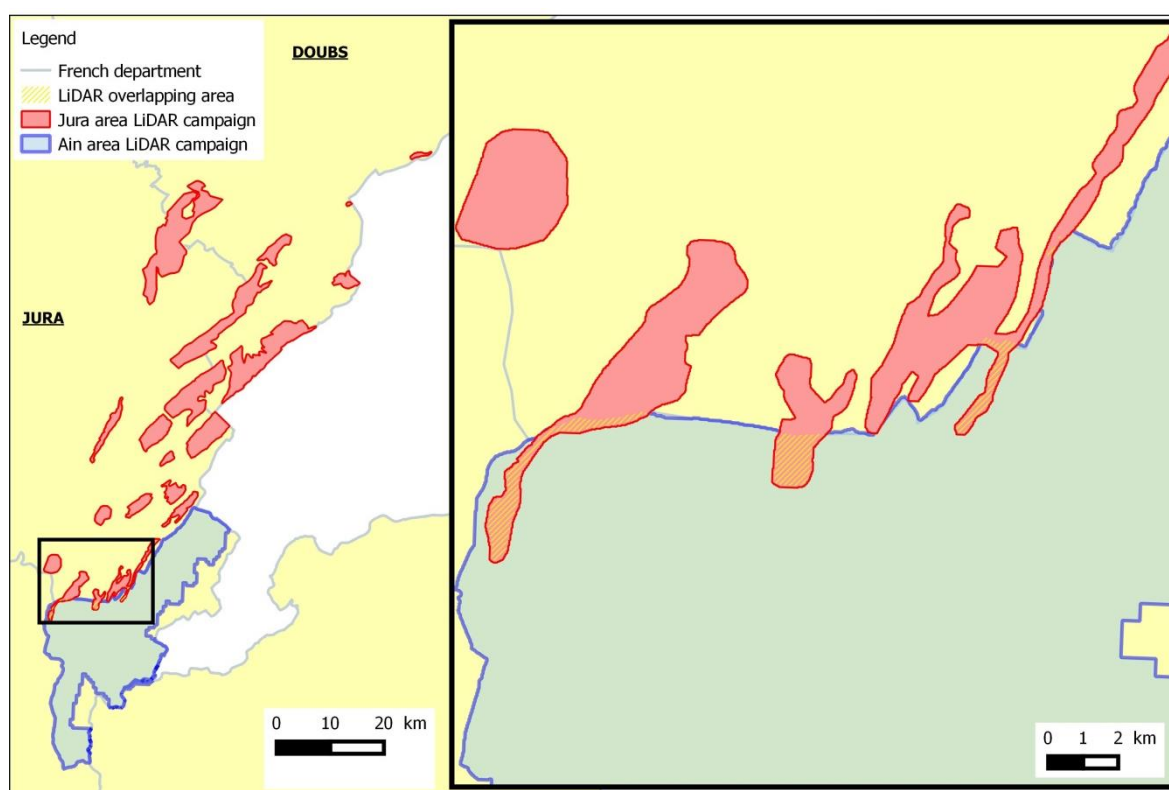
To explore the potential of OO LiDAR metrics as environmental variables, we evaluated two hypotheses on their performance relative to PC area based models: (H1) OO Lidar metrics show a similar SDMs performances as PC area-based models, and (H2) OO variables improve model's transferability to a different LiDAR dataset, as the structure of an object is less dependent on the acquisition characteristics such as point density or flight season (leaf-on/leaf-off) (Gaulton and Malthus, 2010; Tiede et al., 2007). We compared models' performances between the two types of variables for Capercaillie winter observations in the French Jura, and the similarity between the predicted maps was assessed. In a second step, the transferability efficiency was evaluated for each variable type using a second independent LiDAR dataset.

### 2.3 Material

#### 2.3.1 Study area

The study area is situated in the east of France, in the Jura massif, located within the departments of Ain, Jura and Doubs (Fig. 43). The landscape is composed of a mosaic of small urban areas, pastures, forests, and fields. The massif is composed of a low plateau (elevation range from 400 m to 700m a.s.l.) and a high plateau (elevation range from 700 m to 1620 m a.s.l.). On the lower plateau, forests are mainly deciduous dominated forests composed of beech (*Fagus sylvatica*), ash (*Fraxinus excelsior*), oak (*Quercus petraea*) and spruce (*Picea abies*). The high plateau is characterized mainly by coniferous dominated forest, composed of mixed spruce, fir and beech above 1 000 m and by the presence of a grassland subalpine landscape at the highest altitudes (IGN national forest inventory 2005-2009). The climate is continental with temperature varying between -5.9 °C and 21.1 °C and a mean annual precipitation of 1187 mm (<https://donneespubliques.meteofrance.fr/>).

*Figure 43 : The study area and the two LiDAR campaigns.*



### 2.3.2 LiDAR datasets

The study area was covered by two airborne LiDAR surveys (Fig. 1) using a Riegl LMS Q680i system (RIEGL Inc., Horn, Austria). The system was operated with a maximum half-scan angle of 30°, and full wave-form digitization.

The first LiDAR campaign, hereafter referred as “Ain study area”, was conducted in autumn 2014 and covered 626 km<sup>2</sup>. An average point density of 21.3 points/m<sup>2</sup> with a vertical error of 10 cm was obtained.

The second LiDAR campaign, hereafter referred as “Jura study area”, was conducted in summer 2016, and covered a surface of 431 km<sup>2</sup>. An average point density of 18 points/m<sup>2</sup> with a vertical error of 10 cm was obtained.

Those two campaigns overlap over an area of 4.4 km<sup>2</sup>, over the Ain department.



### 2.3.3 Capercaillie observations dataset

Long-term Capercaillie winter surveys were organized between 2007 and 2018 by the “Groupe Tétrás Jura”, a non-governmental organization. Using a Global Navigation Satellite System (GNSS) receiver, observers recorded their survey tracks and Capercaillie signs (feces, prints and feathers) locations. Sex was assigned to each observation when possible. All non-sex assigned observations were removed and the 379 remaining observations were used for the analysis (207 males and 172 females) in order to create separate models for each sex due to their differences in habitat use (Thiel et al., 2007).

## 2.4 Methods

### 2.4.1 Environmental predictors

#### Spatial Scales choice

Spatial scale is known to be an essential aspect for understanding the habitat use of individuals and populations (McGarigal et al., 2016). We thus created models at multiple spatial scales. Six scales were chosen among three hierarchical orders: home range (15.8 ha, 27.5 ha, and 56 ha), patch (0.81 ha and 1.8 ha) and micro-habitat (0.31 ha) (Meyer and Thuiller, 2006). For each scale, the mean and the standard deviation of LiDAR metrics (presented in detail in the next section) were calculated among the neighboring pixels using a moving window. This task was done with the R package *rgrass7* (Bivand et al., 2016).

#### LiDAR derived predictors

Raw point-clouds were normalized with the Delaunay interpolation method implemented in the *LidR* package (Roussel and Auty, 2016). Both PC area-based and object-oriented metrics were calculated using R packages *LidR* and *lidaRtRee* (<https://gitlab.irstea.fr/jean-matthieu.monnet/lidaRtRee>) using the normalized point clouds (Fig. 44). The R code used to calculate all the metrics is presented in Supporting Information S1.



- Points-clouds (PC) area based

Point-cloud metrics were calculated for pixels of 25 m \* 25 m. Six variables were selected with the aim to capture habitat components that are important for Capercaillie.

First, the relative density of the canopy between 10 m and 20 m was chosen as Capercaillie prefer moderate canopy cover allowing the bilberry to cover the ground layer (Storch, 1993). To take into account the horizontal heterogeneity, which is a key aspect for Capercaillie, the variables standard deviation of the relative canopy density 20-30 m and the standard deviation of the penetration ratio between 2-5 m were selected. The preference of Capercaillie for heterogeneous cover for both canopy and understory layers was expected (Graf et al., 2009). The encroachment of the understory layer is not favorable for the species, leading to a poor proportion of the bilberry in the habitat (Sachot et al., 2003). To model this aspect, the mean penetration ratio between 2 m and 5 m as well as the height quantile 25th was selected. The latter variable also indicates unfavorable low vegetation area (non-forest). Finally, the vertical heterogeneity of the vegetation was included with the simpson index as predictor variable. The Capercaillie is expected to favor multi-layered forest stands (Graf et al., 2007).

- Object-oriented

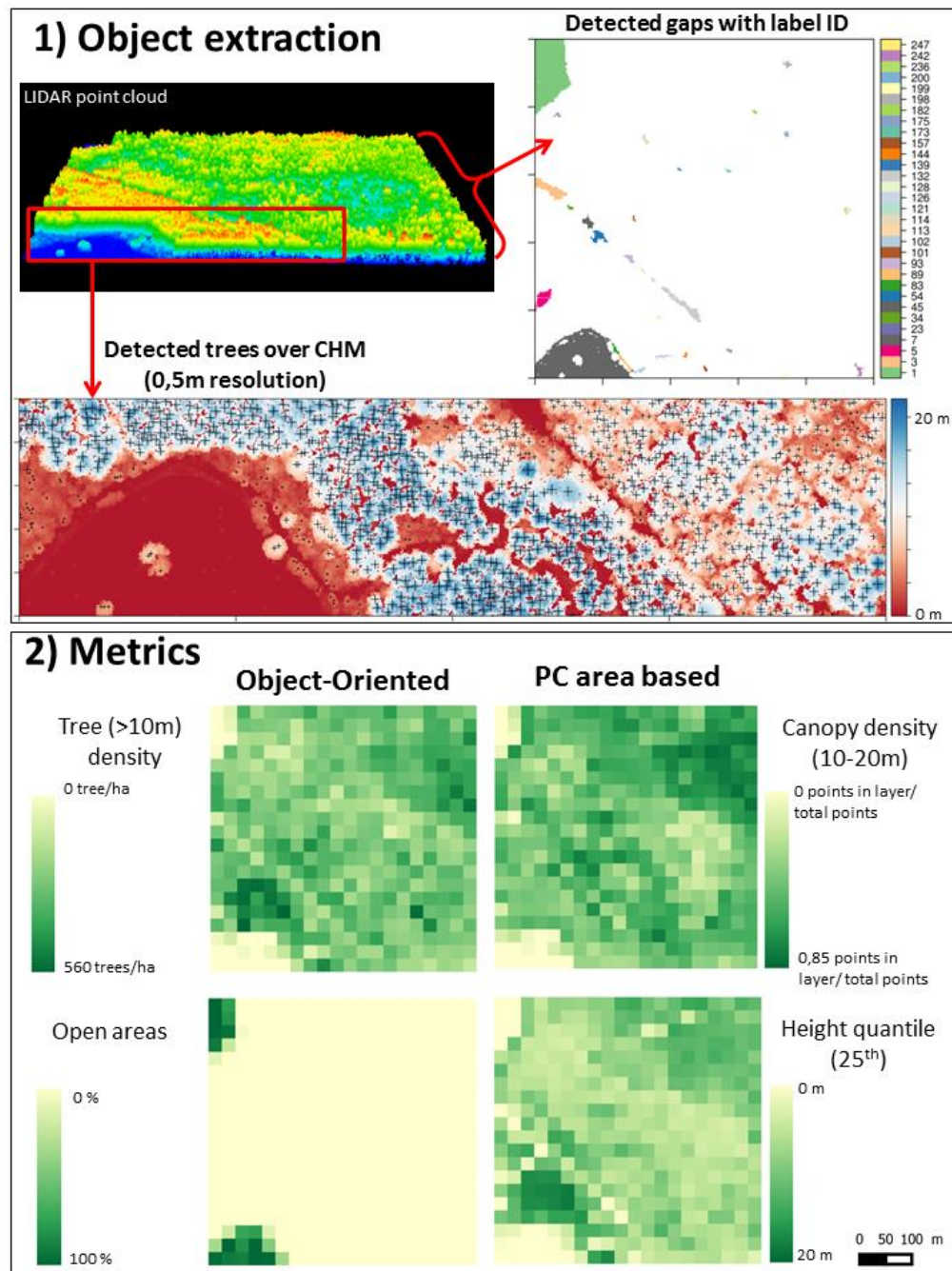
Objects were segmented on the canopy height model at 0.5 m resolution. Summary statistics on objects such as means, standard-deviations, and percentages were then calculated within pixels. Two types of objects were extracted: trees and gaps. Concerning trees, we used a tree top detection method (method #1 in (Eysn et al., 2015)); metrics such as density of trees were derived afterwards over pixels. Gaps were selected among areas characterized by a height lower than 1 m surrounded by trees whose height is less than twice the gap width. From this, habitat metrics were derived such as percentage of gaps with different surface categories. The metrics were chosen in order to represent different habitat components as presented with the point-clouds metrics. The density of the canopy is represented by the density of trees higher than 10 m. The horizontal heterogeneity was represented by two variables: the standard deviations of the density of trees higher than 20 m and of





the proportion of small gaps (25-200 m<sup>2</sup>). Density of the understory is represented by the density of trees between 5 m and 10 m. The presence of gaps in the canopy favors the presence of the herbaceous layer needed by the chicks and that are used by the adults to escape predators (Thiel et al., 2007). The presence of gaps in the canopy was represented by the proportion of middle size gaps (200-1000 m<sup>2</sup>) and the presence of non-forest area by the proportion of open areas (gaps larger than 1000 m<sup>2</sup>).

Figure 44 : Visualization of calculated LiDAR metrics



- Other metrics

To take into account the potential effect of disturbances on Capercaillie (Thiel et al., 2008), the distance to the closest ski runs (Groupe Tétrás Jura dataset) were used to quantify this effect in all models.



### 2.4.2 Species distribution modeling methods

Models were created for each sex and scale separately. Here, we used a recent reimplementation of the algorithm building on the equivalence of Maxent with an infinitely-weighted logistic regression using the R package maxnet (Phillips et al., 2017). The correction of sampling bias was done using a targeted background point method (Phillips et al., 2009). 10 000 background points were randomly located along observer's sampling trajectories, where distances between each background point and trajectories were following a logspline distribution fitted to Capercaillie observations data. All variables were standardized to zero mean and standard deviation of one before modeling. In addition, the Pearson correlation between variables was evaluated prior to each analysis. When the absolute correlation value was higher than 0.7, one of the variable was dropped (Dormann et al., 2013), except for variables chosen to represent the interaction between landscape heterogeneity and landscape large scale component. For all models, both linear and quadratic terms were selected. In addition, the product between two variables was added with the intended purpose to link the landscape heterogeneity to the availability of forest and non-forest areas. For the PC area-based models, the interaction between standard deviation of the Canopy density 20-30 m and the height quantile 25th was selected. In the case of object-oriented models, the interaction between the tree density (>20 m) and proportion of open area was selected.

### 2.4.3 Comparison of model results

The results obtained from the two types of models at the different scales will be compared using different methods.

#### Model performance

First, the performance of each model was evaluated using both the Area Under Curve (AUC), which corresponds to the probability that a randomly drawn occurrence point has a higher predicted occurrence probability than a randomly drawn background point, and the continuous Boyce index which measures how much model predictions differ from random distribution of the observed presences across the prediction gradients (Hirzel et al., 2006). The AUC was evaluated on a 5-fold

cross-validation and the Boyce index with the presence data. The Boyce index was calculated using the function `boyce.ecospat` from the R package `ecospat` (Di Cola et al., 2017).

### Variables contribution and responses

The estimation of the contribution of each variable to models was evaluated by comparing the model coefficients. All absolute coefficient values associated with the same variable (linear, quadratic or product) were added to obtain the final contribution value. The model coefficient of product variables was divided by two before being added to linear and quadratic values of each variable. In addition, the variables associated with a contribution value below 0.05 were dropped for the result analysis, keeping only the most important variables. The response curves were calculated for the three most contributing LiDAR based variables.

### Predictions

Finally, the predictions were made for each model over the Ain study area and also extrapolated to the Jura study area. Afterwards, the correlation between PC area-based and OO predicted maps were calculated for each scale over the Ain study area. In addition, the correlation between predictions by type PC area-based or OO were examined over the overlapping surface of the two LiDAR campaigns. Those predicted maps were generated using the exponential (also called raw) output.

## 2.5 Results

### 2.5.1 Model performance

Model performance across variable types, spatial scales, and sexes was evaluated using two indices: AUC and Boyce (Table 1). Models showed moderate to good performance according to AUC (0.71 - 0.8); performances according to the Boyce index showed higher spread (-0.39-0.94). The correlation between the two performance indices was low.

OO-based models had overall a lower performance than PC-based models; on average, OO model AUC was lower by -0.025 for females, and -0.013 for males. As measured by AUC, the largest



difference between OO and PC variables were found at the coarsest spatial scale of 56 ha: While PC-based models at this scale were among the best models, OO-based models performed worse than at finer scales. The Boyce index showed a contrasting pattern: OO-based models outperformed PC-based models at the same spatial scale in four and five (out of six scales) cases for females and males, respectively. For both males and females, the highest Boyce indices were from OO-based models, and the lowest indices were from PC-based models (Table 19, Supporting Information S2 A1).

Spatial scale had a substantial effect on model performance. However, there were no consistent best spatial scales for males or females across the two variable types or the two performance metrics (Table 1).

Models for females were less performant than models for males in all cases according to AUC (maximum AUC difference for the same spatial scale: 0.04 for PC-based, and 0.05 for OO-based models). In contrast, according to the Boyce index, the best models for both PC- and OO-based models were for females.

*Table 19 : Model performance (AUC and Boyce index)*

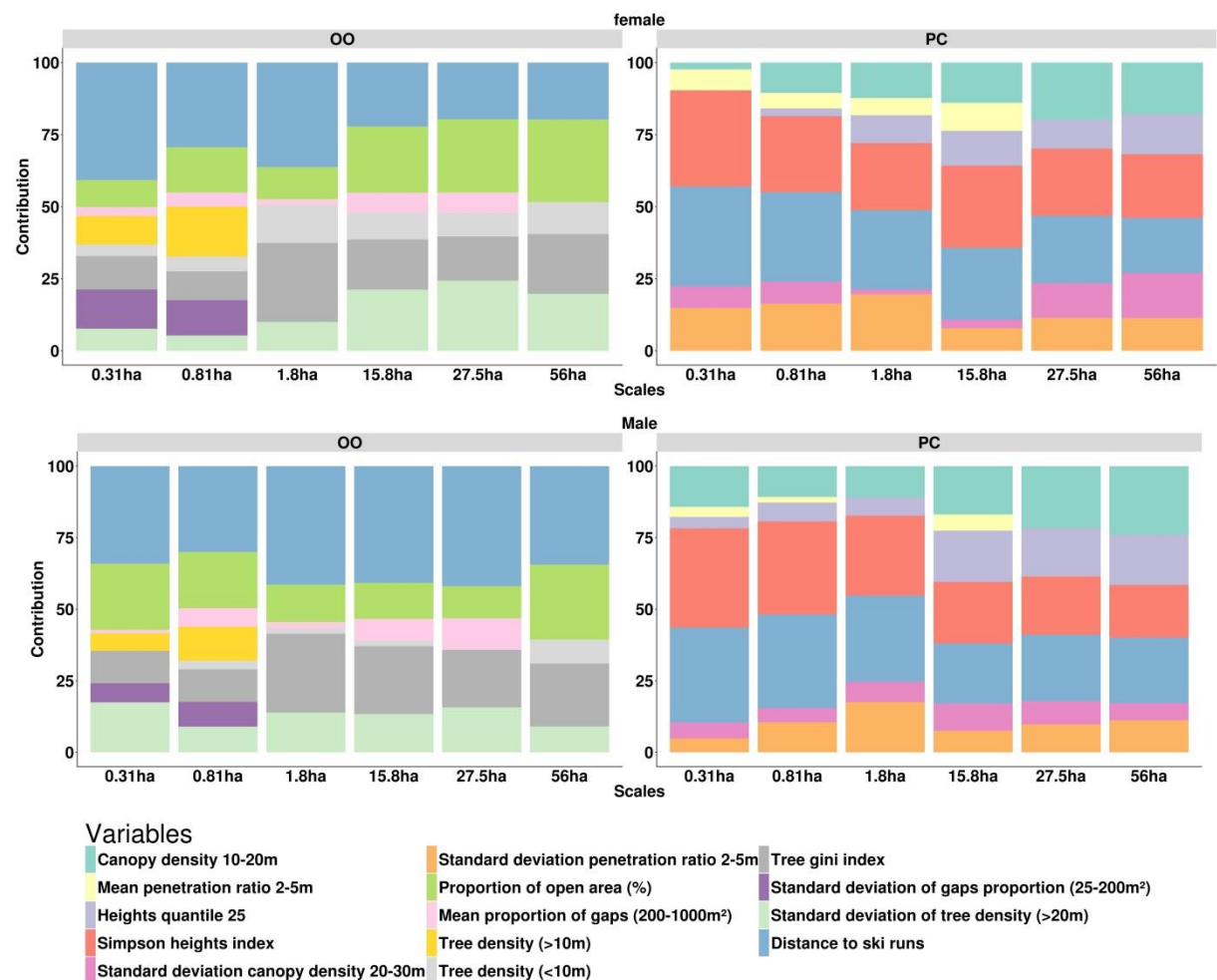
Scale	Female			
	AUC		Boyce	
	PC	OO	PC	OO
0,31ha	0,74	0,72	0,55	0,94
0,81ha	0,75	0,72	0,85	0,9
1,8ha	0,76	0,73	0,73	0,47
15,8ha	0,73	0,72	0,09	0,41
27,5ha	0,74	0,73	-0,39	0,5
56ha	0,76	0,71	0,92	-0,16
Scale	Male			
	AUC		Boyce	
	PC	OO	PC	OO
0,31ha	0,75	0,75	0,52	0,38
0,81ha	0,76	0,78	-0,28	0,78
1,8ha	0,78	0,77	0,05	0,74
15,8ha	0,79	0,78	0,51	0,82
27,5ha	0,8	0,78	0,6	0,79
56ha	0,8	0,74	0,09	0,5



## 2.5.2 Variable contributions and response

In both PC area-based and OO models, and at all scales for both sexes, the variable “distance to ski runs” was highly contributing (Fig. 45). The response curve showed that an optimal distance of 1km for both sexes (Supporting Information S2 A2). For OO models, the tree gini index, the standard deviation of tree density (>20 m), and the proportion of open areas were contributing the most at each scale and for both sexes. The tree density (>10 m) and the standard deviation of gaps proportion (25-200 m<sup>2</sup>) were contributing only at the scales 0.31 ha and 0.81 ha for both sexes. Regarding the PC area-based models, the Simpson height index, the canopy density 10-20m, the height quantile 25th and the standard deviation of penetration ratio 2-5m were contributing the most for all scales and both sexes.

Figure 45 : Contribution of the variables to each model





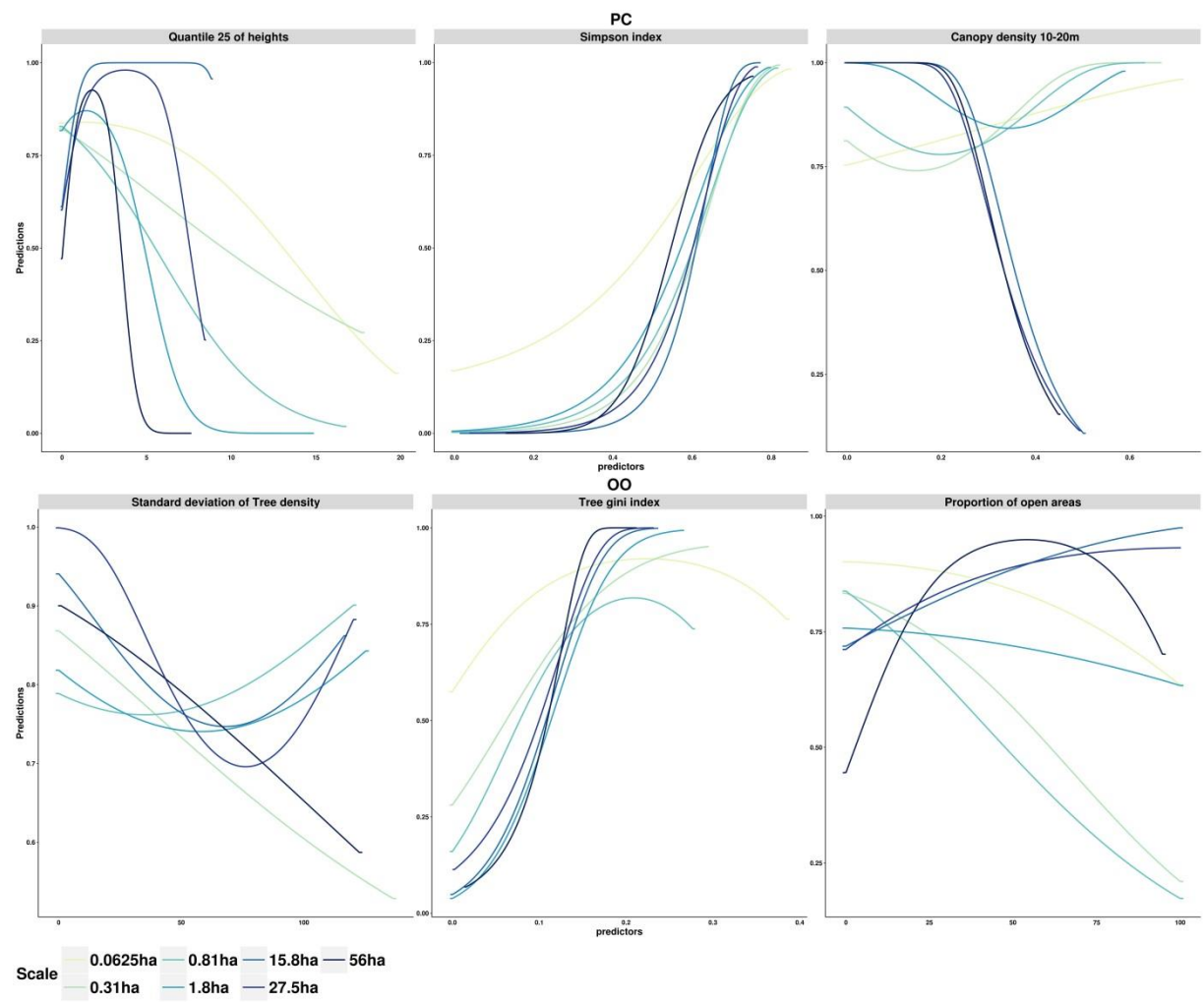
We calculated response curves of the three most contributing LiDAR variables for each model type (Fig. 46 and 47). The response curves of the tree gini index and the simpson heights index showed a monotonously increasing suitability across scales, with the exception of models for males at fine resolution (0.31 ha and 0.81 ha), where models indicated an optimum at intermediate values for the tree gini index.

The response of the variable quantile 25th of heights showed similar quadratic responses between sexes, except for two fine scales (0.31 ha and 0.81 ha), where a negative linear response was observed for females. A height from 2 to 5 meters of the understory height was more favorable. The response to the canopy density 10-20 m for females showed that an optimum of 0.2 points/pixels was preferred. At large spatial scales, values of 0 points/pixels also showed to be suitable. The responses were highly different for males; showing at the smallest scales a more suitable habitat for either low or high canopy density, when an optimum at 0.2 points/pixels was observed at large scales, with a similar response at low density values as females.

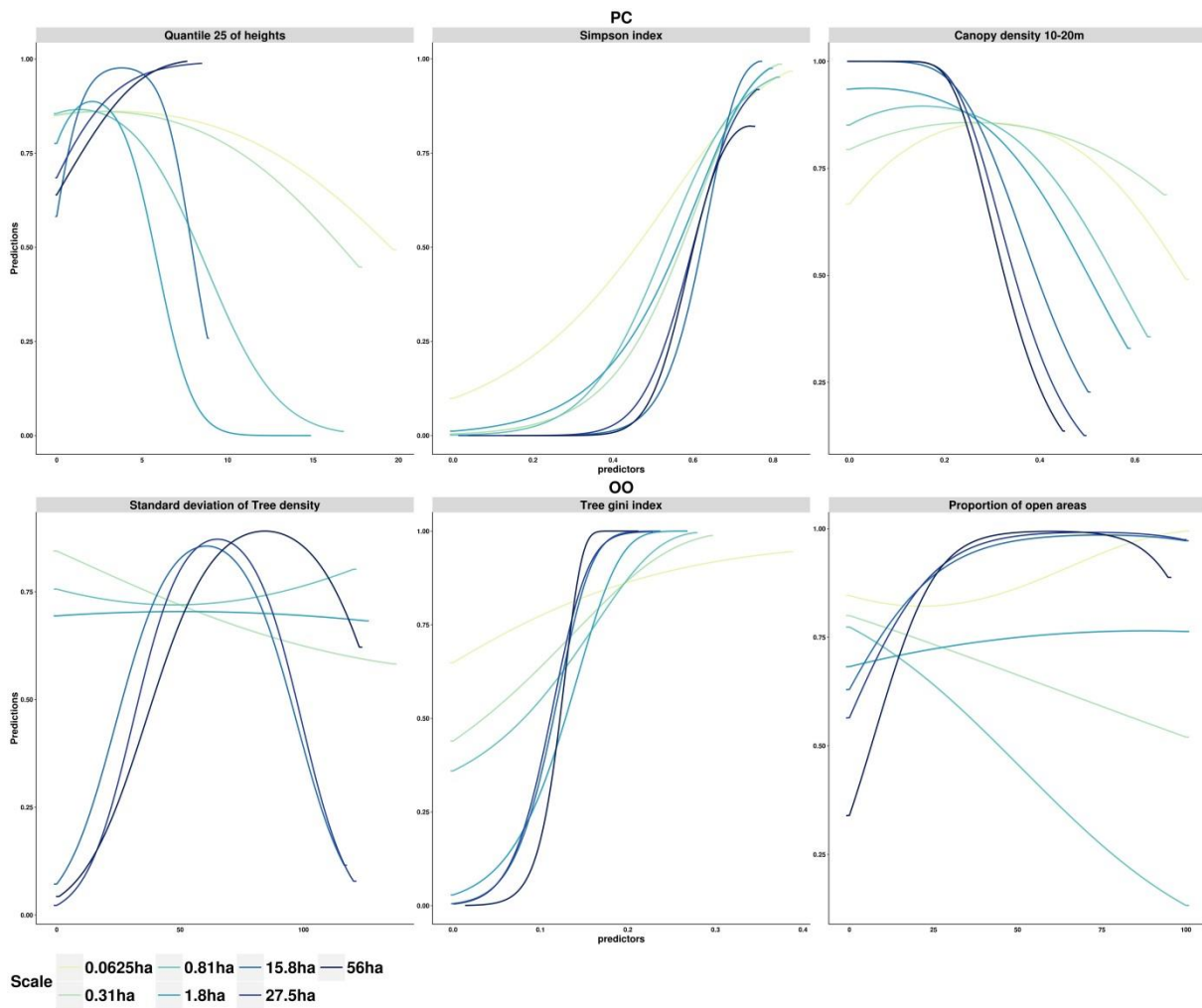
Regarding the standard deviation of tree density (< 20m) from OO type models, male and female had different responses. For males, a negative linear response was observed for two scales (0.31 ha and 56 ha). The other responses indicated values between 50 trees/ha and 70 trees/ha as unsuitable for the species, whereas extreme lower and higher values were more suitable. For females, the exact opposite response was observed at larger scales with more suitable habitat predicted for the same values. At smallest scales, the observed response showed very slight variations, being overall favorable. For the proportion of open areas variable, a higher percentage of open area was unsuitable for males at small scales. At 15.8 ha and 27.5 ha, a positive response was found whereas at the large 56 ha scales, an optimum at 50% of open area was found. For females, a linear negative response was observed for two scales (0.31 ha and 0.81 ha), but a positive response was observed for the other scales. For the largest scale 56 ha, an optimum at 50% was observed.



Figure 46 : Male response curves





*Figure 47 : Female response curves*

### 2.5.3 Predictions

Pearson correlations between predicted maps were calculated between models of each variable type for the same scale, to evaluate overall similarity in model predictions between OO- and PC-based models (Table 20). In a second step, predictions for the second study area (Jura) were compared over a surface common to the two LiDAR campaigns, to assess the transferability of the models across areas and LiDAR campaigns.

The correlation between PC area-based and OO predictions were ranging from 0.32 to 0.74. The highest correlations between predictions, for both female and male models, were found at the two smallest scales (0.31 ha and 0.81 ha), with correlations >0.7 for males and females. At larger scales, correlations decreased, with lowest correlations at the largest scale (56 ha; 0.32 and 0.56 for females and males, respectively).

*Table 20 : Pearson correlation between PC area-based and OO predictions over the Ain study area.*

Scale	Female	Male
0,31ha	0,79	0,81
0,81ha	0,71	0,74
1,8ha	0,39	0,64
15,8ha	0,58	0,66
27,5ha	0,58	0,7
56ha	0,32	0,53

Regarding the transferability of the model to other areas, the correlation ranged between 0.5 and 0.82 for females and 0.61 and 0.88 for males. Comparing to PC area-based models, OO models had a higher heterogeneity between scales, and also showed the most extreme correlation values (low and high) (Table 21). In sum, with the exception of models for females at large scales ( $\geq 27.5$  ha), OO models were as transferable (absolute difference in correlation  $\leq 0.01$ ), or more transferable than PC-based models.

*Table 21 : Pearson correlation between predictions of each model type, calculated with two LiDAR datasets.*

Scale	Female		Male	
	PC	OO	PC	OO
0,31ha	0,74	0,79	0,7	0,7
0,81ha	0,67	0,75	0,67	0,66
1,8ha	0,67	0,82	0,61	0,8
15,8ha	0,67	0,67	0,62	0,78
27,5ha	0,73	0,5	0,72	0,76
56ha	0,74	0,65	0,68	0,88

## 2.6 Discussion

The aim of this study was to evaluate the use of object-oriented LiDAR extracted variables to predict the distribution of an endangered forest bird, the Capercaillie, in the Jura massif. To achieve this objective, we compared model performance and transferability using two types of LiDAR extracted variables: commonly used point-cloud (PC) area-based and object-oriented metrics.

The hypothesis that models with OO variables show a better or similar performance than PC-based models for the distribution of Capercaillie was not confirmed by our observations. Indeed, OO models overall performed slightly less well (according AUC) than PC area-based models and this for



both sexes, particularly at larger spatial scales. However, the influence of scale was highlighted by both model types, showing high variations in model performance.

A high contribution and a similar response shape was observed for each variable describing the heterogeneity of the vertical structure (the PC area based metric Simpson index for canopy height and the OO metric Gini index for tree height), highlighting the importance of the forest stand heterogeneity for the species. Furthermore, this result confirms the equivalence between those two variable types in our models. However, concerning the other variables, a general pattern was not observed. Indeed, in PC area-based models, the canopy density and height were contributing the most, indicating the importance of the cover and of the vertical structure, whereas in OO models the proportion of open areas and the standard deviation of the Tree density (>20 m) were the most contributing, indicating here the importance of the horizontal heterogeneity. The response curves observed the less contributing variables were also different between scales and sexes for both PC area-based and OO, making difficult to extract general conclusions from these results.

The consistency of the predictions between model types is moderate, with correlation between PC area-based and OO maps being in some case low, in particular at the 56ha scale (0.32, 0.53) and for the 1.8ha scale (0.39 for female's model). This result indicates that our models have in some cases important divergence in spatial prediction, leading to an uncertainty regarding model accuracy. This can be due to the fact that the models created from PC area-based and OO may capture different characteristics of the environment, leading to differences in the prediction. In particular, the extraction of understory objects is still a challenge due to the difficulty to find the vertical separation between small and tall trees, and the extraction of such objects was thus not addressed in this study. Information on the understory is provided by two PC area-based variables in our models: the penetration ratio 2-5m (mean and standard deviation) and the height quantile 25. On the contrary, in OO models, the minimum height for tree top detection was 5 m. Yet, understory is known to constitute an essential component of the capercaillie habitat, which may explain why OO models were performing less than PC area-based models. Nevertheless, recent progress in LiDAR processing may soon fill this gap and allow the extraction of object bellow the canopy (Hamraz et al., 2017).



Also, the role of feature types (here quadratic and linear) in models may greatly influence the predictions, in particular for extreme values. Indeed, the response curves presented for the three main contributing variables were in some cases highly different between sexes and scales. In some cases, this might be due to the limited shapes of response curves that can be represented by quadratic features, and the use of more flexible, but adequately constrained, response curves by using new features (e.g. hinge, splines) could reduce this effect (Ranc et al., 2017).

There was some evidence that models using OO variables had higher transferability than PC-based models particularly at finer spatial scales. Overall, the transferability to another LiDAR dataset was moderate, with 10 cases with values below 0.7 out of 18 models. Those observations are indicating that the use of different LiDAR campaigns may require separate models to be more accurate (Chasmer et al., 2006; Hopkinson, 2007).

Nevertheless, the advantages of using object-oriented variables when diffusing the results to managers are still important. The use of those variables may allow a better comprehension of the results as the model is more closely related to variables measured in the field by forest managers and used as indicators to plan forestry management actions. Indeed, objectives are usually quantified as stems densities or as number/percentages of gaps. In addition, if variables are easier to interpret, the results can be more readily used to raise awareness of conservation of a species or a habitat with other stakeholders such as private owners.

### 2.7 Conclusion

The use of LiDAR extracted object-oriented variables had a limited impact on model performance, accuracy and transferability. Indeed, more variation in terms of performance was observed among scales and sex, indicating that the type of variable used is not the most important component influencing model accuracy. Nevertheless, the use of those LiDAR object-oriented variables may still improve communication between researchers and managers, and should be therefore further explored in the field of conservation biology.



## 2.8 Acknowledgements

The first LiDAR campaign was financed by European Territorial Cooperation program INTERREG IV France-Switzerland 2007-2013 and the Ain department. The second LiDAR campaign was financed by “Conseil régional Bourgogne Franche-Comté”. The project was financed by the departmental council of Ain.

## 2.9 Data accessibility

Data will be hosted on the IRSTEA servers.

## 2.10 Bibliography

- Bae, S., Reineking, B., Ewald, M., and Mueller, J. (2014). Comparison of airborne lidar, aerial photography, and field surveys to model the habitat suitability of a cryptic forest species—the hazel grouse. *Int. J. Remote Sens.* 35, 6469–6489.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., and Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *Integr. Geodata Imag. Autom. Refinement Update Spat. Databases* 58, 239–258.
- Bergen, K.M., Gilboy, A.M., and Brown, D.G. (2007). Multi-dimensional vegetation structure in modeling avian habitat. *Ecol. Inform.* 2, 9–22.
- Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., and Radeloff, V.C. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *J. Geophys. Res.-Biogeosciences* 114, G00E06-G00E06.
- Bivand, R., Krug, R., Neteler, M., and Jeworutzki, S. (2016). *rgrass7: interface between GRASS 7 geographical information system and R*. R Softw. Package Google Sch.
- Braunisch, V., and Suchant, R. (2007). A model for evaluating the ‘habitat potential’ of a landscape for capercaillie *Tetrao urogallus*: a tool for conservation planning. *Wildl. Biol.* 13, 21–33.
- Chasmer, L., Hopkinson, C., Smith, B., and Treitz, P. (2006). Examining the influence of changing laser pulse repetition frequencies on conifer forest canopy returns. *Photogramm. Eng. Remote Sens.* 72, 1359–1367.
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F.T., D’amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., and Dubuis, A. (2017). *ecospat: an R package to support spatial analyses and modeling of species niches and distributions*. *Ecography* 40, 774–787.



Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., and Leitão, P.J. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46.

Eysn, L., Hollaus, M., Lindberg, E., Berger, F., Monnet, J.-M., Dalponte, M., Kobal, M., Pellegrini, M., Lingua, E., and Mongus, D. (2015). A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space. *Forests* 6, 1721–1747.

Farrell, S.L., Collier, B.A., Skow, K.L., Long, A.M., Campomizzi, A.J., Morrison, M.L., Hays, K.B., and Wilkins, R.N. (2013). Using LiDAR-derived vegetation metrics for high-resolution, species distribution models for conservation planning. *Ecosphere* 4, UNSP-UNSP 42.

Fourcade, Y., Besnard, A.G., and Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* 27, 245–256.

Franklin, J. (2009). Moving beyond static species distribution models in support of conservation biogeography. *Divers. Distrib.* 16, 321–330.

Gaulton, R., and Malthus, T.J. (2010). LiDAR mapping of canopy gaps in continuous cover forests: A comparison of canopy height model and point cloud based techniques. *Int. J. Remote Sens.* 31, 1193–1211.

Graf, R.F., Bollmann, K., Bugmann, H., and Suter, W. (2007). Forest and landscape structure as predictors of capercaillie occurrence. *J. Wildl. Manag.* 71, 356–365.

Graf, R.F., Mathys, L., and Bollmann, K. (2009). Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps. *For. Ecol. Manag.* 257, 160–167.

Guisan, A., and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., et al. (2013). Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435.

Hamraz, H., Contreras, M.A., and Zhang, J. (2017). Vertical stratification of forest canopy for segmentation of understory trees within small-footprint airborne LiDAR point clouds. *ISPRS J. Photogramm. Remote Sens.* 130, 385–392.

He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., and Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.* 1, 4–18.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–152.



- Hopkinson, C. (2007). The influence of flying altitude, beam divergence, and pulse repetition frequency on laser pulse return intensity and canopy frequency distribution. *Can. J. Remote Sens.* 33, 312–324.
- Johnson, C.J., and Gillingham, M.P. (2005). An evaluation of mapped species distribution models used for conservation planning. *Environ. Conserv.* 32, 117–128.
- Kortmann, M., Heurich, M., Latifi, H., Rösner, S., Seidl, R., Müller, J., and Thorn, S. (2018). Forest structure following natural disturbances and early succession provides habitat for two avian flagship species, capercaillie (*Tetrao urogallus*) and hazel grouse (*Tetrastes bonasia*). *Biol. Conserv.* 226, 81–91.
- McGarigal, K., Wan, H.Y., Zeller, K.A., Timm, B.C., and Cushman, S.A. (2016). Multi-scale habitat selection modeling: a review and outlook. *Landsc. Ecol.* 31, 1161–1175.
- Melin, M., Mehtätalo, L., Miettinen, J., Tossavainen, S., and Packalen, P. (2016). Forest structure as a determinant of grouse brood occurrence – An analysis linking LiDAR data with presence/absence field data. *Spec. Sect. Drought US For. Impacts Potential Manag. Responses* 380, 202–211.
- Meyer, C.B., and Thuiller, W. (2006). Accuracy of resource selection functions across spatial scales. *Divers. Distrib.* 12, 288–297.
- Mikoláš, M., Tejkal, M., Kuemmerle, T., Griffiths, P., Svoboda, M., Hlásny, T., Leitão, P.J., and Morrissey, R.C. (2017). Forest management impacts on capercaillie (*Tetrao urogallus*) habitat distribution and connectivity in the Carpathians. *Landsc. Ecol.* 32, 163–179.
- Munoz, A., Bock, J., Monnet, J., Renaud, J., Jolly, A., and Riond, C. (2015). Évaluation par validation indépendante des prédictions des paramètres forestiers réalisées à partir de données LiDAR aéroporté. *Rev. Fr. Photogrammétrie Télédétection* 211, 81–92.
- Othmani, A., Voon, L.F.L.Y., Stolz, C., and Piboule, A. (2013). Single tree species classification from Terrestrial Laser Scanning data for forest inventory. *Pattern Recognit. Lett.* 34, 2144–2150.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., and Blair, M.E. (2017). Opening the black box: an open-source release of Maxent. *Ecography* n/a-n/a.
- Ranc, N., Santini, L., Rondinini, C., Boitani, L., Poitevin, F., Angerbjörn, A., and Maiorano, L. (2017). Performance tradeoffs in target-group bias correction for species distribution models. *Ecography* 40, 1076–1087.
- Rolstad, J. (1988). Autumn habitat of capercaillie in southeastern Norway. *J. Wildl. Manag.* 747–753.



- Rolstad, J., Wegge, P., and Larsen, B.B. (1988). Spacing and habitat use of capercaillie during summer. *Can. J. Zool.* 66, 670–679.
- Roussel, J.-R., and Auty, D. (2016). *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. R package version 1.0.0. <https://github.com/Jean-Romain/lidR>.
- Sachot, S. (2002). Viability and management of an endangered capercaillie (*Tetrao urogallus*) metapopulation.
- Sachot, Perrin, N., and Neet, C. (2003). Winter habitat selection by two sympatric forest grouse in western Switzerland: implications for conservation. *Biol. Conserv.* 112, 373–382.
- Sachot, Perrin, N., and Neet, C. (2006). Viability and management of an endangered Capercaillie (*Tetrao urogallus*) metapopulation in the Jura Mountains, Western Switzerland. *Biodivers. Conserv.* 15, 2017–2032.
- Storch, I. (1993). Habitat selection by capercaillie in summer and autumn: Is bilberry important? *Oecologia* 95, 257–265.
- Storch, I. (1994). Habitat and survival of capercaillie *Tetrao urogallus* nests and broods in the Bavarian Alps. *Biol. Conserv.* 70, 237–243.
- Storch, I. (2007). Grouse, Status Survey and Conservation Action Plan 2006-2010.
- Teuscher, M., Brandl, R., Foerster, B., Hothorn, T., Roesner, S., and Mueller, J. (2013). Forest inventories are a valuable data source for habitat modelling of forest species: an alternative to remote-sensing data. *Forestry* 86, 241–253.
- Thiel, D., Unger, C., Kéry, M., and Jenni, L. (2007). Selection of night roosts in winter by capercaillie *Tetrao urogallus* in Central Europe. *Wildl. Biol.* 13, 73–86.
- Thiel, D., Jenni-Eiermann, S., Braunisch, V., Palme, R., and Jenni, L. (2008). Ski tourism affects habitat use and evokes a physiological stress response in capercaillie *Tetrao urogallus*: a new methodological approach. *J. Appl. Ecol.* 45, 845–853.
- Tiede, D., Lang, S., and Maier, B. (2007). Transferability of a tree-crown delineation approach using region-specific segmentation. pp. 21–26.
- Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., and Clawges, R.M. (2008). Lidar: shedding new light on habitat characterization and modeling. *Front. Ecol. Environ.* 6, 90–98.
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., and Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodivers. Conserv.* 1–21.
- Zellweger, F., Braunisch, V., Baltensweiler, A., and Bollmann, K. (2013). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *For. Ecol. Manag.* 307, 303–312.





## 3 CAPERCAILLIE DISTRIBUTION MODELS

---

### 3.1 Material

#### 3.1.1 Study areas

Models and predictions for Capercaillie were made over the “species management area”, defined by Groupe Tétraz Jura. This area represents the surface where active management activities for the Capercaillie can be planned across the Jura Massif: areas characterized by a known population presence or by a favorable habitat, which can be eventually recolonized, according to expert knowledge.

#### 3.1.2 Capercaillie datasets

For the winter season, we used the dataset coming from the long-term Capercaillie winter “prospections” survey collected between 2007 and 2018 by Groupe Tétraz Jura. The observer’s survey tracks along with the presence signs’ locations were used for modeling. Thus, 379 observations were used in the analysis (207 males and 172 females) in the Ain study area and 1105 observations (691 males and 414 females) in the Jura study area (unknown sex observations were removed).

For the summer season, very few “prospections” surveys were conducted, resulting in a low number of individuals observations associated with observer’s tracks. Therefore, those data collected between 2007 and 2017 were pooled with other opportunistic data collected with non-defined protocols. Only signs of presence localized using a GNSS receiver were used to avoid unprecise localization. Observations collected during the annual summer count “Battue” protocol were excluded, as it is composed mostly of locations of individuals that were flushed away by the observers, leading to a probably biased habitat association (Montadert pers. com.). In addition, few observations had information on the individual sex that is why separated models could not be created (contrary to winter models). This decision was supported by the hypothesis that the sex ratio

bias is estimated to be less important in summer (Montadert pers. com). The summer dataset was thus composed of 134 observations in the Jura study area and 161 observations in the Ain study area (both sexes).

### 3.1.3 LiDAR datasets and variables

The LiDAR datasets from both campaigns were used in this section. For more details on the LiDAR datasets please refer to the corresponding section in the Chapter materials (Chapter 2). The selected variables were the same as those listed in the object-oriented case (see this Chapter section 2), only the variable tree density (>10m) was replaced by the basal area (Gha) in the Ain study area (Table 22). The standard deviation variable was not used at the smallest scale (0.0625 ha) but the raw metrics were used instead.

*Table 22 : LiDAR variables used for Capercaillie models.*

Ain study area		Jura study area	
Season	LiDAR metric	Season	LiDAR metric
Winter and summer	Basal area (m <sup>2</sup> /ha)	Winter and summer	Tree density (> 10m)
Winter and summer	Standard deviation of tree density (>20 m)	Winter and summer	Standard deviation of tree density (>20 m)
Winter and summer	Standard deviation of the proportion of gaps (25-100 m <sup>2</sup> )	Winter and summer	Standard deviation of the proportion of gaps (25-100 m <sup>2</sup> )
Winter and summer	Tree density (< 10m)	Winter and summer	Tree density (< 10m)
Winter and summer	Proportion of gaps 200-100 m <sup>2</sup>	Winter and summer	Proportion of gaps 200-100 m <sup>2</sup>
Winter and summer	Proportion of open area (> 1000 m <sup>2</sup> )	Winter and summer	Proportion of open area (> 1000 m <sup>2</sup> )
Winter and summer	Tree gini index	Winter and summer	Tree gini index
Winter	Distance to ski runs	Winter	Distance to ski runs

### 3.1.4 Other datasets

To take into account the potential effect of disturbances on Capercaillie (Thiel et al. 2008), the distance to the closest ski runs (Groupe Tétrás Jura dataset) was used to quantify this effect in all models. The intensity of the disturbances induces an increase of the flush distance and physiological stress response (Thiel et al., 2007, 2008). In addition, alpine ski infrastructures provoke deadly



collisions with ski-lift cables which are thought to be non-negligible (Buffet and Dumont-Dayot, 2013).

### 3.2 Methods

As with previous models, the package maxnet was used (Phillips et al., 2017). Variables that were highly correlated ( $>0.7$ ) were removed from the analysis (Ain models: standard deviation gaps proportion (25-200 m<sup>2</sup>) for the scales 1.8 ha to 56 ha. The model performance was evaluated by the AUC using a 5 fold dataset partition. The contribution and response curves for each variable were evaluated. Two different sampling bias correction methods were applied, one for winter models and one for summer models due to differences in the Capercaillie dataset.

#### 3.2.1 Winter spatial sampling bias correction

The spatial sampling bias correction method used to create winter models was the targeted background point method presented in the Chapters 3 and 4 and in the precedent section.

#### 3.2.2 Summer spatial sampling bias correction

As survey observers tracks were available for very few data in summer, the correction method used for the winter models could not be applied to the summer models. Instead, the sampling bias impact was approximated by calculating the number of visits (using the date and the observer name) extracted from both the Capercaillie and Hazel Grouse observation datasets (Fithian et al., 2015a). By using this method, if one person observed multiple signs in the same day, this was still counted as one visit. The number of visits was calculated within pixels of 4km<sup>2</sup> and disaggregated to a 25\*25 m pixel resolution afterwards. This variable was then used as an environmental variable, like one of the corrections method already tested in the Chapter 3 (Method distance to transect). This was then associated with the generation of 10 000 points randomly distributed over each study area.



### 3.2.3 Scales optimization

The models were created by applying two methods recently described by (McGarigal et al., 2016):

- A pseudo-optimized single scale method, where the different selected variables are evaluated simultaneously across a range of scales (see article 2), referred as “Single scale, multi-variables models” (Figure 32).
- A pseudo-optimized multiple scales method, where each variable is evaluated alone at each scale (single-variable, single scale models), and after, are combine together at their best scale, referred as “Mixed models” (Figure 32).

## 3.3 Winter models results

### 3.3.1 Mixed model creation

Single-variable, single-scale models results are presented in table 23. Differences in best scale selection were observed between both study areas and sexes. For the Jura study area, the variables proportion of open area and tree density (>10m) were selected at the patch level for both sexes, whereas all the other variables showed best performances at the home range level. For Ain female models, all variables were selected at the patch level, except for the proportion of gaps (200-1000m<sup>2</sup>) which was selected at the home range level (56ha). On the contrary, only the basal area and the tree density (<10m) were selected at the patch level for males. The other variables showed best performance at the home range level.



Table 23 : AUC values calculated for single-scale, single-variable winter models

Sex	Area	Level	Scale	Open area	SD Tree density (> 20m)	gaps proportion (200-1000m²)	Tree gini index	Tree density (>10 m)	Tree_density (>10 m )	SD gaps porportion (25-200 m²)	Distance to ski runs
Female	JURA	Micro-Habitat	0.0625ha	0.54	NA	0.53	0.56	0.57	0.51	NA	0.49
			0.31ha	0.56	0.56	0.55	0.59	0.59	0.51	0.59	0.49
		Patch	0.81ha	0.57	0.61	0.56	0.6	0.6	0.53	0.62	0.49
			1.8ha	0.58	0.63	0.57	0.61	0.61	0.55	0.67	0.49
		Home range	15_8ha	0.56	0.65	0.62	0.62	0.58	0.59	0.69	0.49
			27_5ha	0.52	0.65	0.63	0.63	0.58	0.59	0.68	0.49
			56ha	0.49	0.64	0.63	0.63	0.58	0.59	0.68	0.49
Male		Micro-Habitat	0.0625ha	0.56	NA	0.53	0.61	0.57	0.56	NA	0.58
			0.31ha	0.58	0.59	0.55	0.64	0.6	0.6	0.62	0.58
		Patch	0.81ha	0.6	0.62	0.56	0.66	0.6	0.62	0.63	0.58
			1.8ha	0.6	0.65	0.58	0.67	0.61	0.63	0.66	0.58
		Home range	15.8ha	0.59	0.66	0.63	0.68	0.56	0.65	0.7	0.58
			27.5ha	0.58	0.67	0.65	0.69	0.56	0.65	0.7	0.58
			56ha	0.55	0.66	0.66	0.69	0.56	0.65	0.7	0.58

Sex	Area	Level	Sex	Open area	SD Tree density (> 20m)	gaps proportion (200-1000m²)	Tree gini index	Basal area	Tree_density (>10 m )	SD gaps porportion (25-200 m²)	Distance to ski runs
Female	AIN	Micro-Habitat	0.0625ha	0.51	NA	0.53	0.59	0.5	0.55	NA	0.7
			0.31ha	0.52	0.54	0.57	0.64	0.51	0.59	0.62	0.7
		Patch	0.81ha	0.53	0.56	0.58	0.66	0.54	0.59	0.63	0.7
			1.8ha	0.51	0.56	0.6	0.68	0.53	0.6	0.68	0.7
		Home range	15_8ha	0.5	0.54	0.66	0.65	0.5	0.57	0.65	0.69
			27_5ha	0.49	0.55	0.67	0.64	0.5	0.56	0.63	0.69
			56ha	0.45	0.54	0.68	0.61	0.54	0.51	0.61	0.69
Male		Micro-Habitat	0.0625ha	0.48	NA	0.53	0.56	0.50	0.57	NA	0.71
			0.31ha	0.53	0.57	0.52	0.63	0.53	0.64	0.61	0.71
		Patch	0.81ha	0.53	0.58	0.48	0.64	0.51	0.65	0.65	0.71
			1.8ha	0.51	0.61	0.53	0.65	0.51	0.67	0.66	0.71
		Home range	15.8ha	0.52	0.64	0.64	0.65	0.51	0.68	0.67	0.7
			27.5ha	0.53	0.66	0.66	0.64	0.51	0.67	0.66	0.7
			56ha	0.56	0.63	0.63	0.58	0.53	0.62	0.61	0.69



From the results of the single-scale, single-variable models, two mixed models were elaborated. When the same best value of AUC was observed for multiple scales, the smallest best scale was kept, as the aim is to produce a predicted map precise enough at the local scale.

### 3.3.2 Models performances

Models' performances were differing between scales from poor to fair (0.63 to 0.79) (Araújo et al., 2005; Guisan et al., 2017). Male models gave better performance than female models. The best performance was observed for mixed models except for female Jura models where the highest AUC value was observed at the 15.8 ha scale (Table 24).

*Table 24 : AUC values for winter models*

Scale	AIN		JURA	
	Female	Male	Female	Male
0,0625ha	0,69	0,74	0,63	0,68
0,31ha	0,72	0,76	0,66	0,71
0,81ha	0,72	0,78	0,69	0,73
1,8ha	0,73	0,78	0,72	0,74
15,8ha	0,72	0,78	0,73	0,75
27,5ha	0,73	0,78	0,73	0,75
56ha	0,73	0,75	0,72	0,74
Mixed	0,75	0,79	0,74	0,77

### 3.3.3 Variables contribution

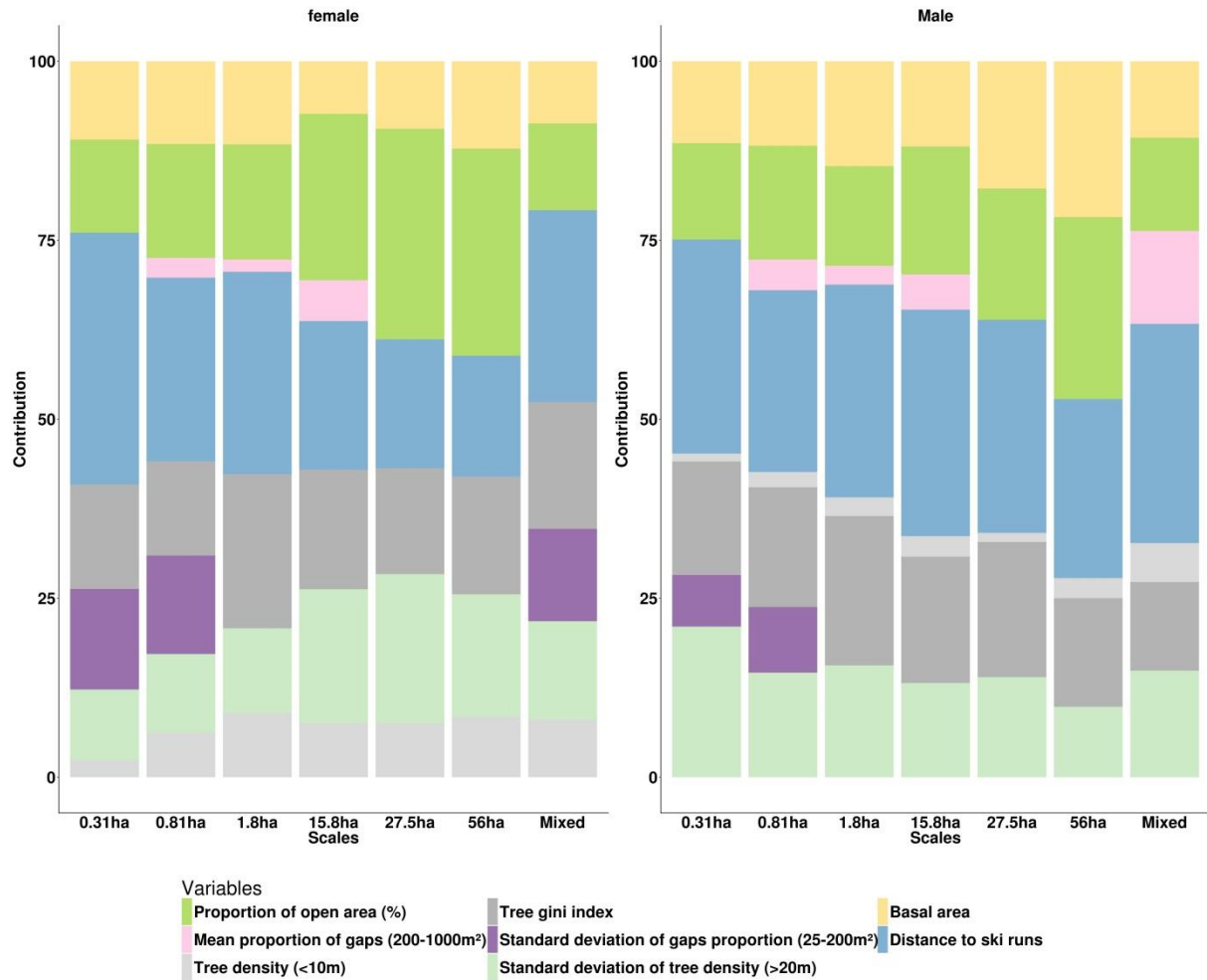
In both areas and for both sexes, the variable distance to the ski runs was highly contributing to the models.

In the Ain study area, four variables contributed the most at all scales for both male and female: tree gini, basal area, proportion of open area and standard deviation of the tree density (>20m) (Figure 48). The standard deviation of small gap proportion contributed at small scale for both sexes. The tree density (<10m) contributed to female models but showed a low contribution to male models. The mean proportion of gaps (200-1000m<sup>2</sup>) contributed only to the 0.81 ha, 1.8 ha and 15.8 ha scales for both sexes. In the mixed models, the contribution of the different variables was more homogenous. Differences between male and female models were observed: the variable standard deviation of the gaps proportion (25-200m<sup>2</sup>) contribute to female mixed models but not in male



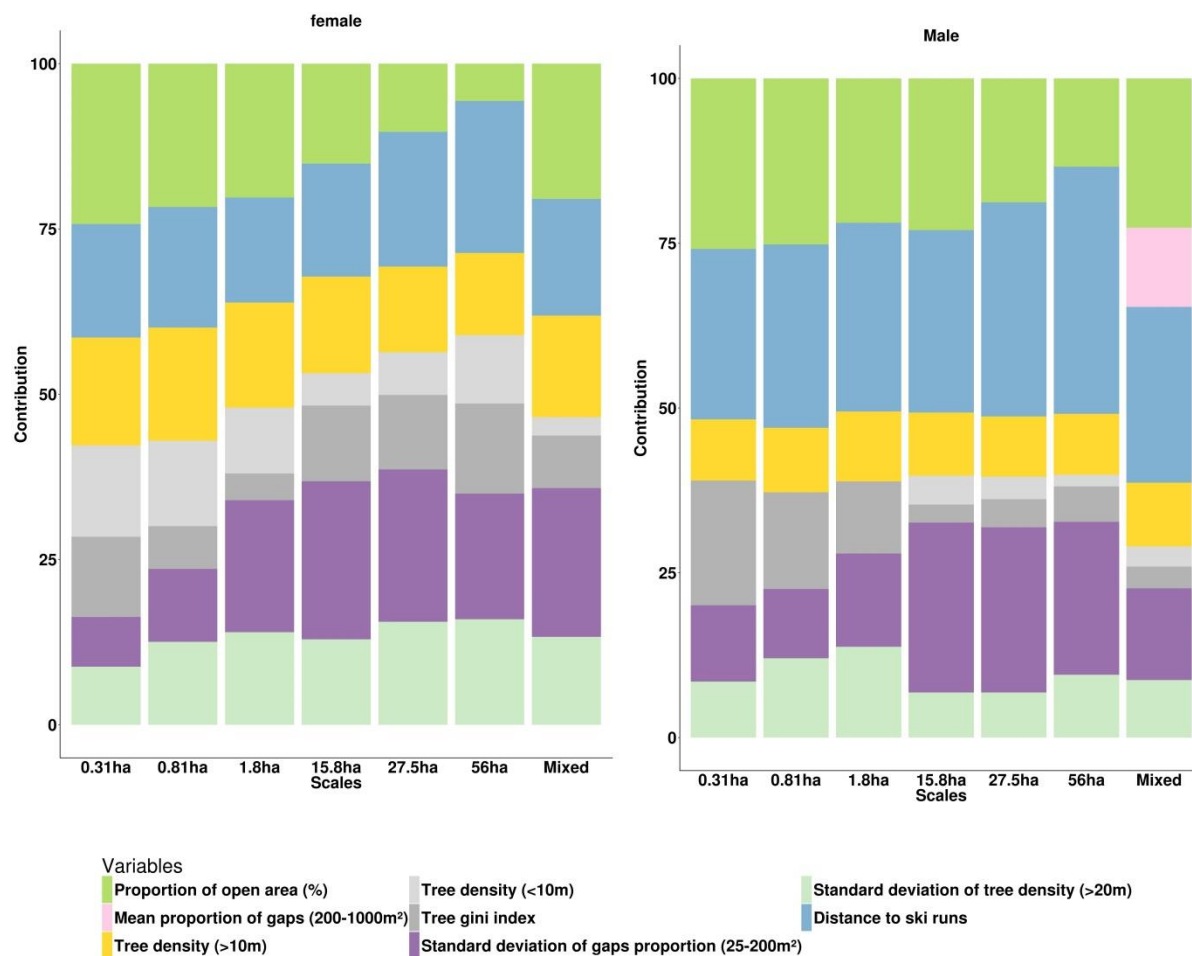
mixed models. Similarly, the variable mean proportion of gaps (200-1000m<sup>2</sup>) contributed in male mixed models but not in female mixed models.

Figure 48 : Contribution of the variables in the Ain study area for both sexes.



In the Jura study area, the proportion of open area, the tree density (>10m) and the standard deviation of gaps proportion (25-200m<sup>2</sup>) contributed the most for both sexes outside of the variable distance to ski runs (Figure 49). The tree gini index contributed more for female models but is an important variable for the 0.31ha scale male models. The mean proportion of gaps (200-1000m<sup>2</sup>) was contributing once again only for the male mixed model.

*Figure 49 : Contribution of the variables in the Jura study area for both sexes.*

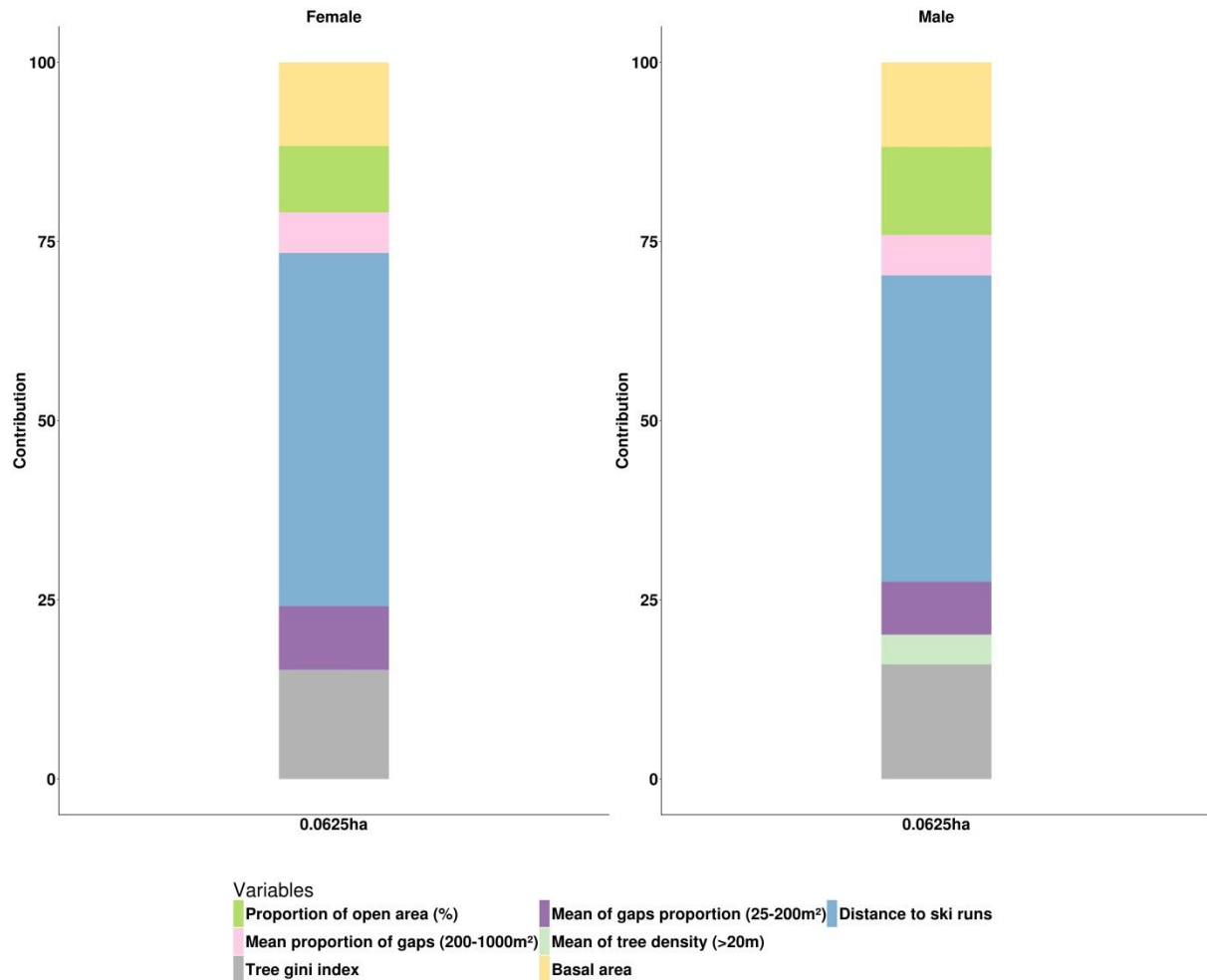




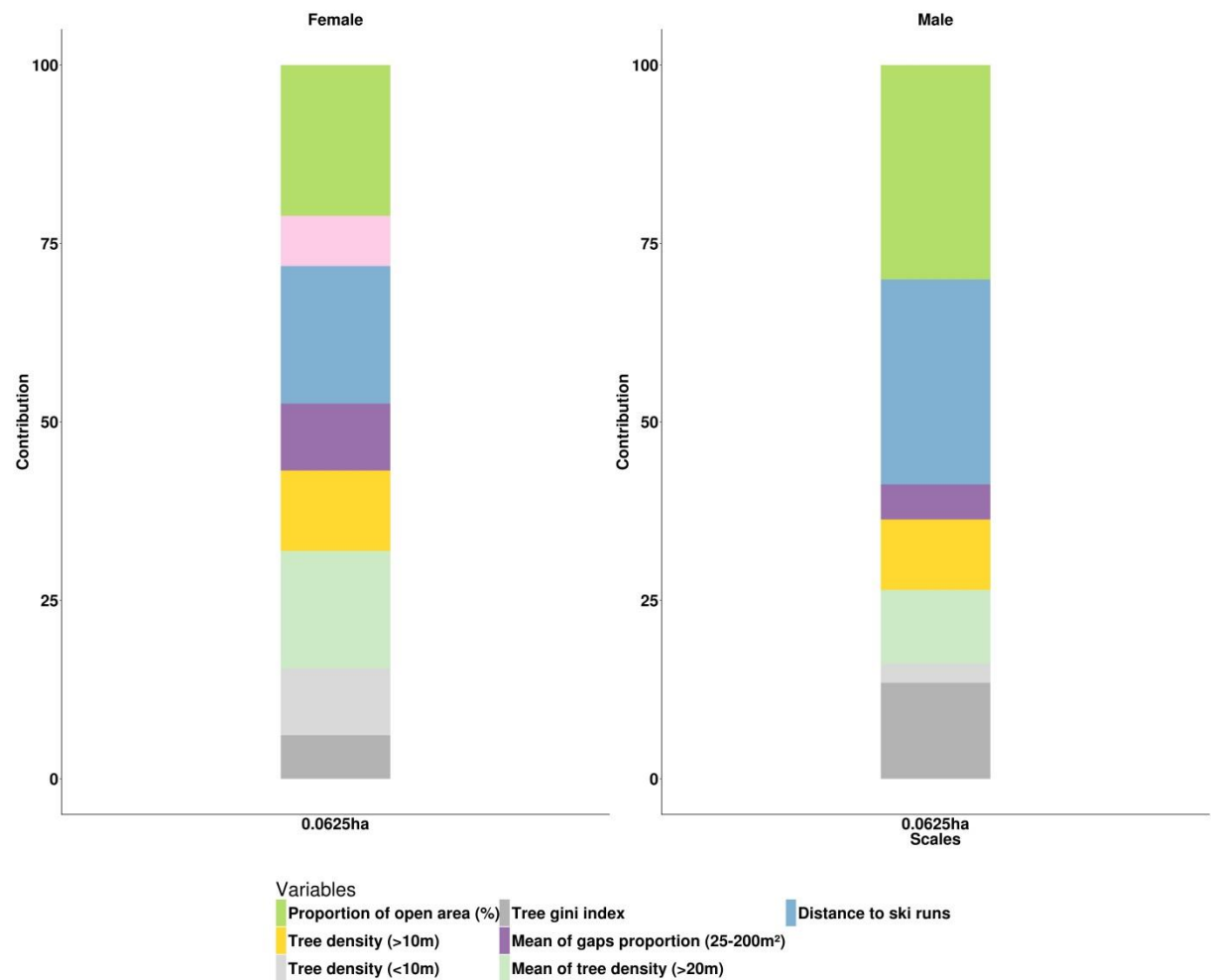


At the smallest scale (0.0625ha) in the Ain study area, variables contribution was similar between sexes (Figure 50). The variable distance to ski runs contributed the most, followed by the tree gini index, the basal area and the proportion of open area.

*Figure 50 : Contribution for the scale 0.0625ha, in the Ain study area*



In the Jura study area, the most contributing variables were the distance to ski runs, the proportion of open area and the tree density (>20m) (Figure 51). In the male model, the tree gini index and the tree density (>10m) also highly contributed, whereas the tree density (<10m), the mean gap proportion (25-200m²) and the tree density (>10m) contributed to the female model.

*Figure 51 : Contribution for the scale 0.0625ha, in the Jura study area*

### 3.3.4 Response curves

#### Ain study area

Response curves for males and females area were highly similar, except for two variables: the standard deviation of the tree density (>20m) and the standard deviation of the gaps proportion (25-200m²) (Figures 52 and 53). Here, the response curves of only the four main contributing variables will be described.

The response curves to the variable distance to ski runs were similar between scales, showing an optimum for distances equal to 1km.

At the three largest scales and for the mixed model, the tree gini index value of 0.15 and higher were preferred. At lower scales, an optimum was observed for the value 0.2.



The response to the basal area showed that intermediate values are more suitable (between 30 and 50 m<sup>2</sup>/ha). At the 0.0625ha scale, the response reaches a plateau at 50m<sup>2</sup>/ha.

The response curve of the proportion of open area showed that a proportion of 25% is preferred at large scales. However, the curve shape reaches a plateau, which gives high suitability for higher proportion values as well. At smallest scales, the response showed better suitability of the habitat for high proportion of open areas.

Figure 52 : Response curves for female models in the Ain study area.

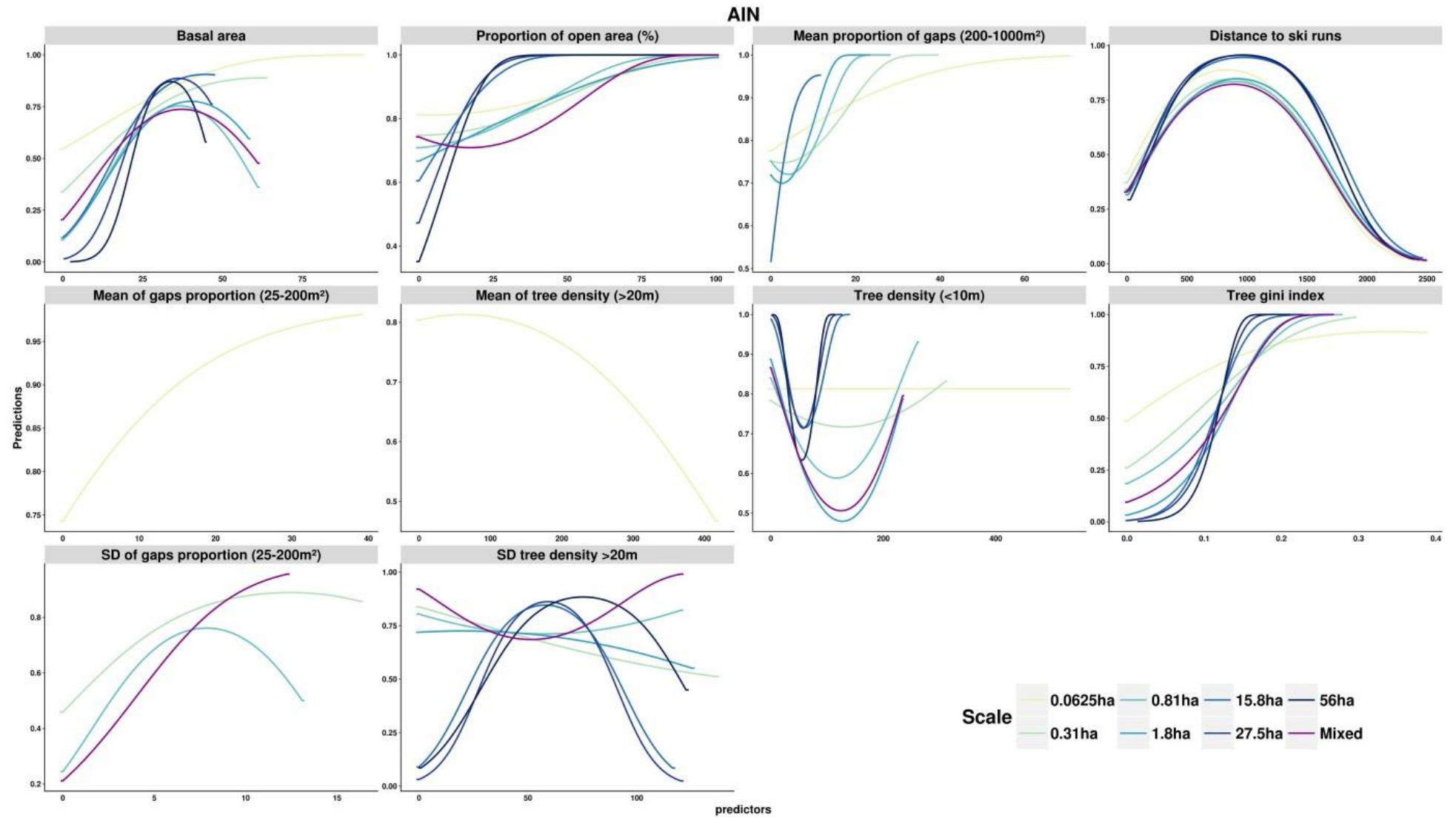
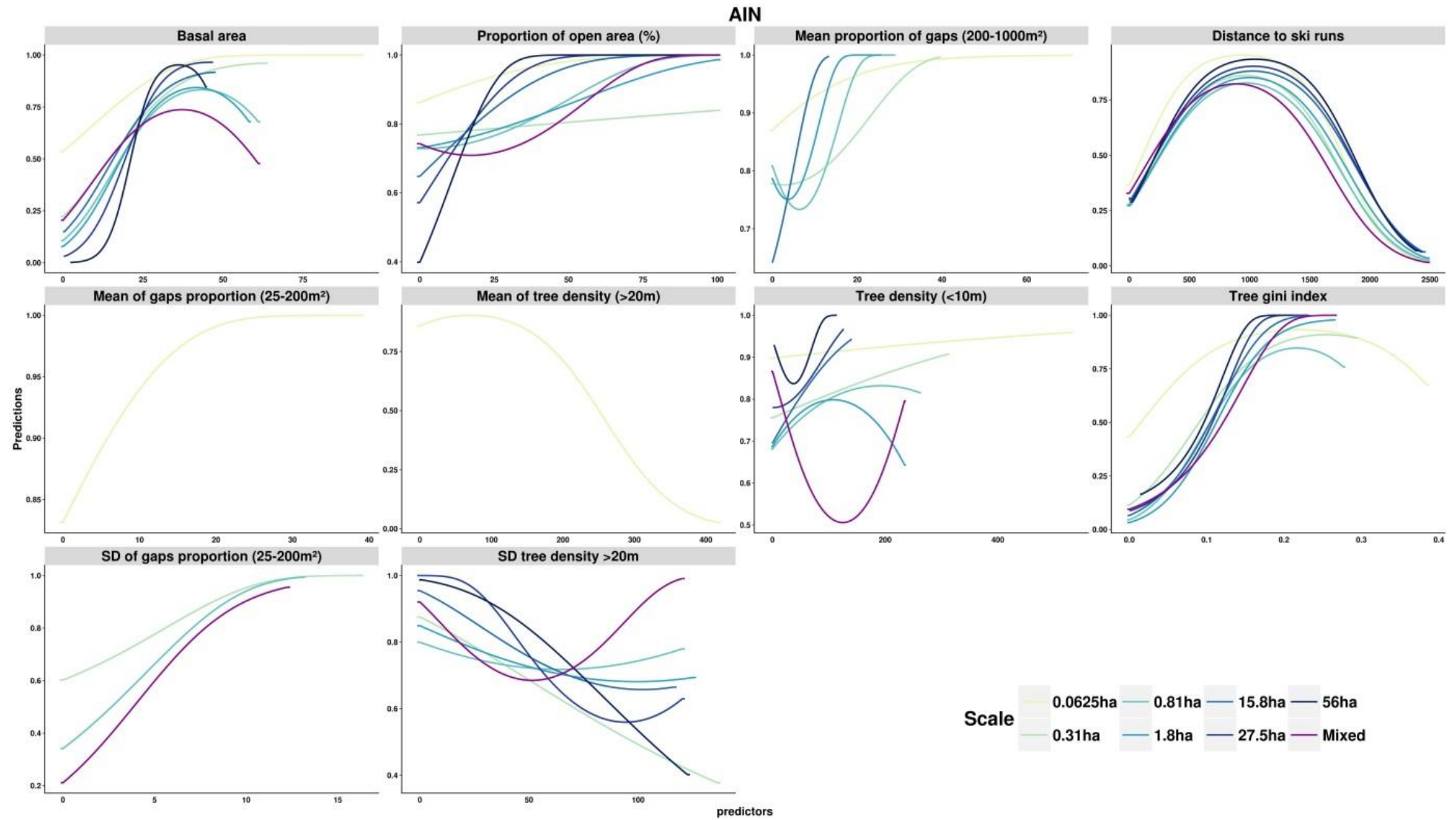




Figure 53 : Response curves for male models in the Ain study area.





### Jura study area

The responses were different between sexes for the proportion of open area, tree density ( $>10\text{m}$ ), and tree gini index (Figure 54 and 55).

The response to the distance to ski runs was similar to the response observed in the Ain area, showing an optimum at 1km.

For female models, high proportions of open area were unfavorable at all scales, whereas the same pattern is observed only at small scales for males. Indeed, at the largest scales, intermediate proportion values were predicted as unfavorable (25 to 60%).

Concerning the standard deviation of the tree density ( $>20\text{m}$ ) for female at small scales, high heterogeneity values were predicted unfavorable (linear response), whereas at largest scales, intermediate values were the most unfavorable (quadratic response). However, for male models, higher values of standard deviation were unfavorable at all scales. The response to the standard deviation of the proportion of gaps ( $25\text{-}200\text{m}^2$ ) showed that higher values were more favorable at all scales and for both sexes.

Concerning the mean proportion of gaps ( $200\text{-}1000\text{m}^2$ ) a quadratic response curve was observed for male mixed models whereas linear responses were observed for single scale models.

Figure 54 : Response curves for female models in the Jura study area.

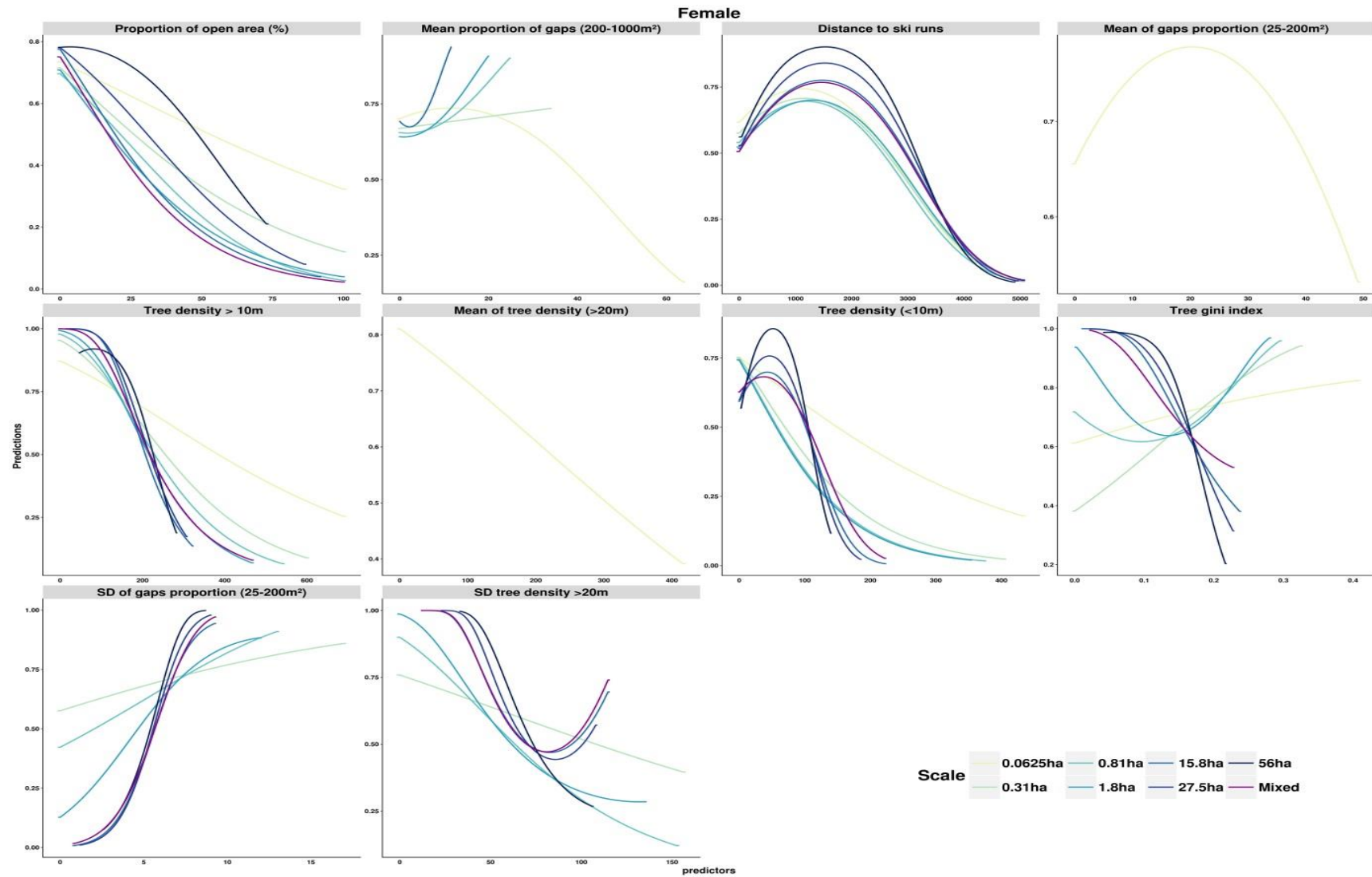
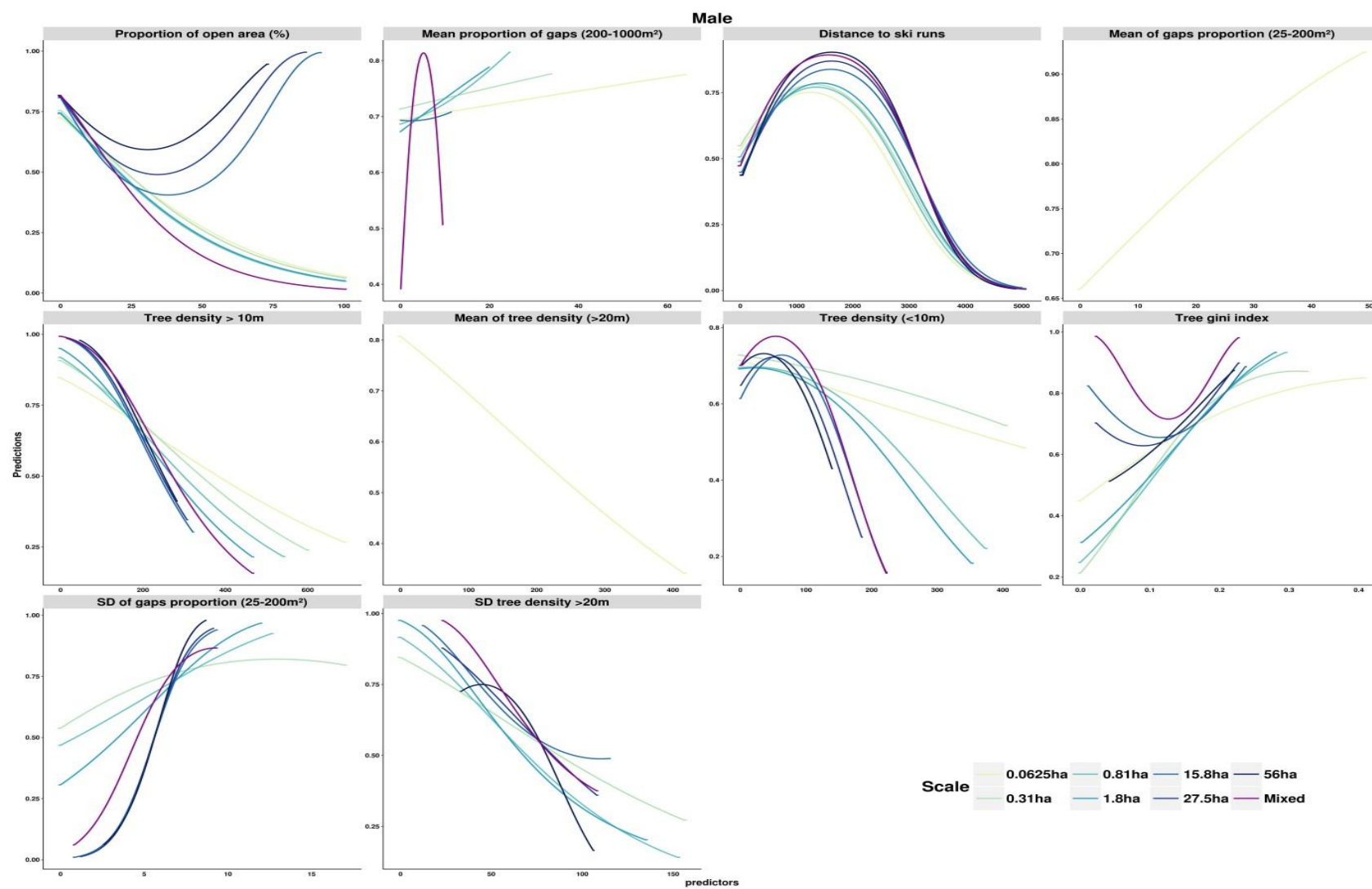


Figure 55 : Response curve for male models in the Jura study area.





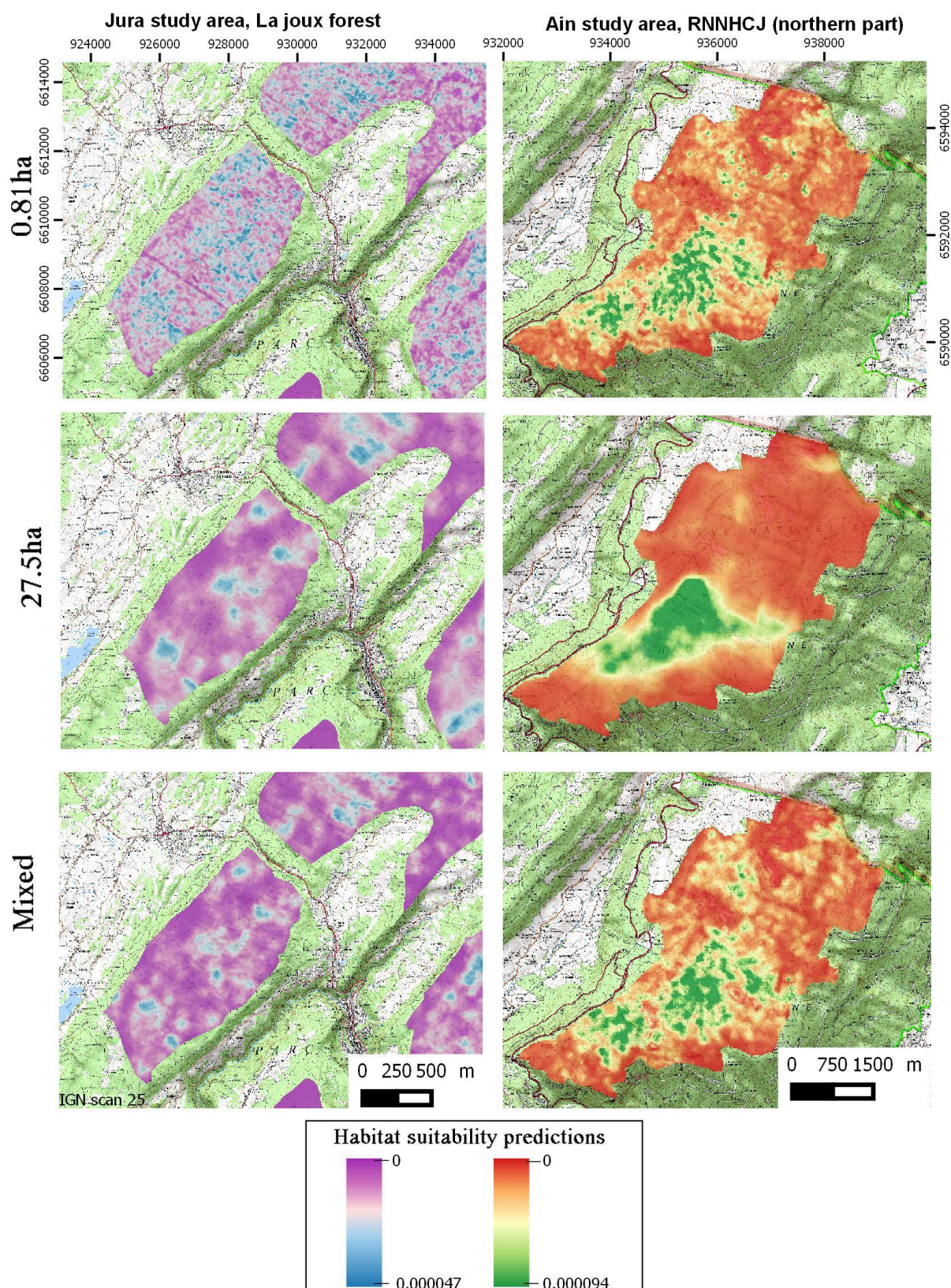


### 3.3.5 Predictions

Predictions were made over the two study areas for each single scale, multi-variables models and for the mixed models. Predictions from models at the smallest scales (micro-habitat and patch orders), showed that the forest is composed of a mixed of favorable and unfavorable habitat patches. In the figure 56, we can observe that females habitat appears more fragmented in the *La Joux* forest than in the northern part of the “*Réserve Naturelle Nationale de la haute Chaîne du Jura*” (RNNHCJ). However, a large northern part of the reserve seems unfavorable for females Capercaillie (the same pattern was observed for male as well). At larger scales, the favorable patches for the species are larger. This can be observed in the figure 56 at the scale of 27.5 ha in both sites. The map created from the mixed model in the Jura study area was very similar to the 27.5 ha scale predicted map. On the contrary, the patchiness aspect of the landscape observed in the Ain study area is conserved with the mixed model predictions.

*Figure 56 : Prediction at two different scales and mixed models for female*

## Predictions at two different scales and mixed models for female

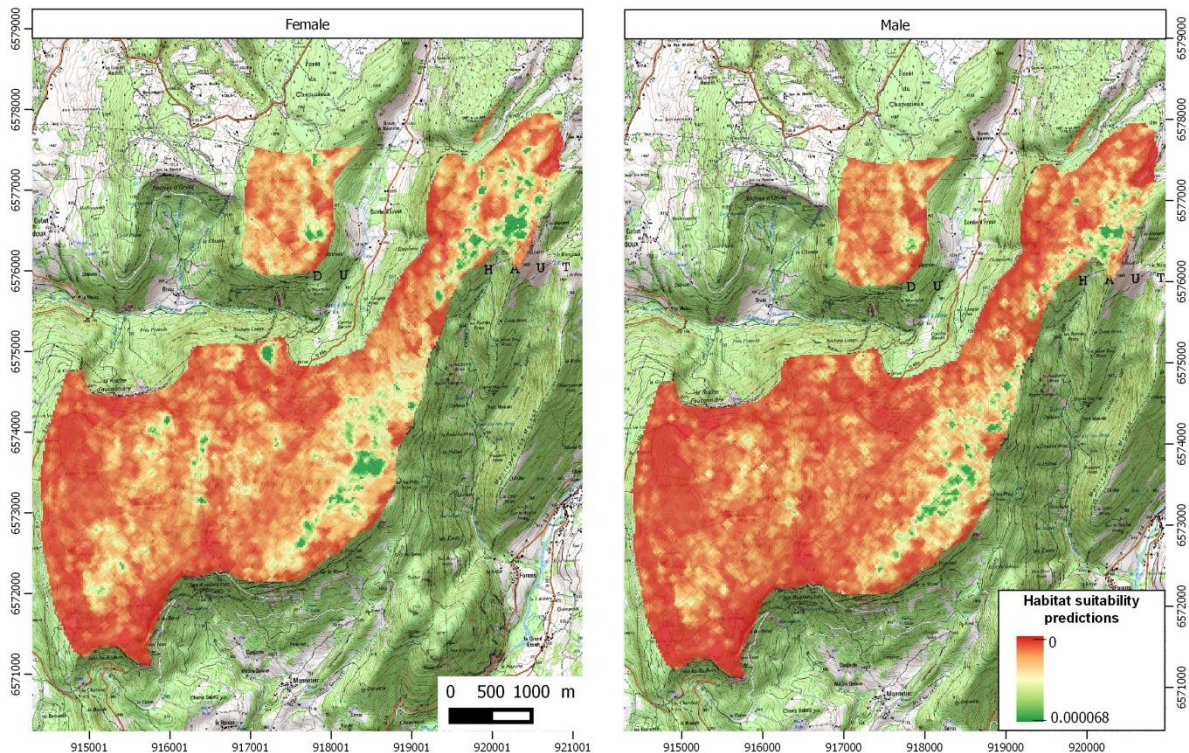






When female and male distributions were compared, slight differences were observed. This is illustrated in the Figure 57 in the *Champfromier* forest (Ain study area). Indeed, if the location of the best habitat was similar between sexes, habitats the most suitable for female were represented by larger patches. Such differences were however not observed in the Jura study area.

*Figure 57 : Comparison of the predicted male and female habitat suitability at the scale of 0.81 ha (Ain study area, Champfromier forest)*



### 3.4 Summer models results

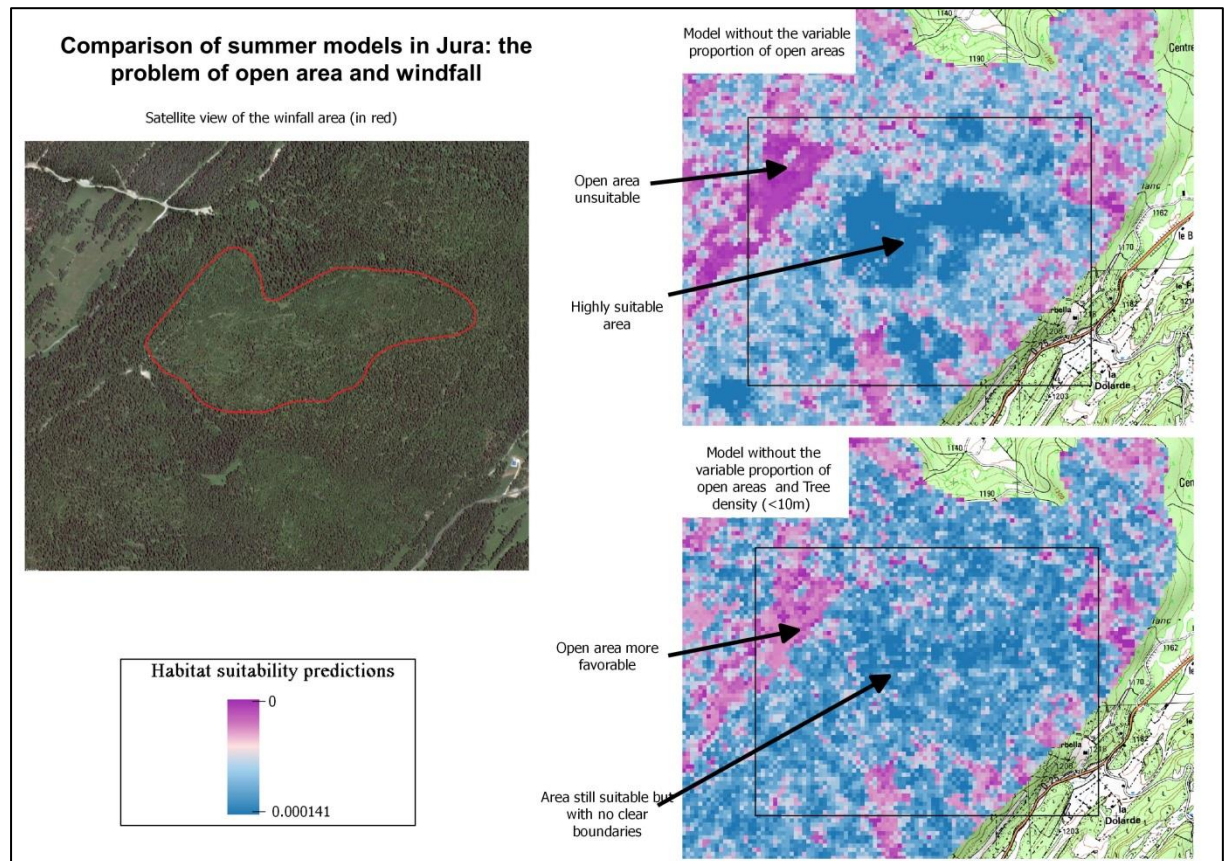
#### 3.4.1 Difficulties with summer models in the Jura study area

The models created over the Jura study area were problematic. Indeed, when the first models were created using the selected variables (same as those used in winter models), the model was favoring old windfall areas, that were characterized by a high density of trees measuring between 5 and 10m (positive linear response) (Figure 58). In addition, the open areas were not predicted as unfavorable, contrary to winter models and summer Ain models. The importance of the variable density of trees (<10m) was confirmed by using the lime R package with the aim to find which variable was determining the final predicted value. In order to improve the relevance of those models, two



variables initially selected were removed: proportion of open areas and the tree density (<10m). Nevertheless, the predicted maps are still not showing those types of habitat as unsuitable. The results in terms of performance, variables contribution and response curves are described hereafter.

*Figure 58 : Comprison of summer models in Jura : The problem of open area and windfall*



### 3.4.2 Mixed model creation

In the Ain study area, all variables were selected at the home range level except for the basal area which was selected at 1.8 ha (patch level). In the Jura study area, only the standard deviation of the tree density (>20 m), the gaps proportion (200-1000 m<sup>2</sup>), the tree gini index and the tree density (>10 m) were selected at the home range level, whereas the other variables were all selected at the patch level (Table 25).

*Table 25 : AUC of single-scale single-variable models*

Area	Level	Scale	Open area	SD Tree density (> 20m)	gaps proportion (200-1000m²)	Tree gini index	Tree density (>10 m)	Tree_density (>10 m)	SD gaps proportion (25-200 m²)
JURA	Micro-Habitat	0.0625ha	0.54	NA	0.47	0.57	0.48	0.59	NA
		0.31ha	0.54	0.53	0.51	0.59	0.52	0.58	0.56
	Patch	0.81ha	0.55	0.55	0.5	0.6	0.52	0.59	0.57
		1.8ha	0.57	0.55	0.49	0.61	0.52	0.6	0.56
	Home range	15_8ha	0.59	0.58	0.49	0.63	0.53	0.6	0.51
		27_5ha	0.59	0.61	0.5	0.63	0.53	0.59	0.52
		56ha	0.59	0.62	0.54	0.63	0.51	0.59	0.53

Area	Level	Scale	Open area	SD Tree density (> 20m)	gaps proportion (200-1000m²)	Tree gini index	Basal area	Tree_density (>10 m)	SD gaps proportion (25-200 m²)
AIN	Micro-Habitat	0.0625ha	0.58	NA	0.56	0.65	0.67	0.59	NA
		0.31ha	0.59	0.6	0.59	0.69	0.69	0.62	0.67
	Patch	0.81ha	0.58	0.66	0.63	0.7	0.71	0.64	0.7
		1.8ha	0.6	0.69	0.63	0.71	0.72	0.65	0.73
	Home range	15_8ha	0.63	0.74	0.66	0.73	0.71	0.66	0.75
		27_5ha	0.63	0.74	0.66	0.73	0.71	0.65	0.76
		56ha	0.65	0.7	0.65	0.72	0.7	0.64	0.73

### 3.4.3 Models performances

Annex I. Models performances were differing between scales for the Ain study area but can be considered as good with AUC values from 0.8 to 0.84 Araújo et al. (2005). For the Jura area, the AUC values can be considered as fair (0.74 to 0.76). The best performance was observed for mixed model in this study area (Table 26).

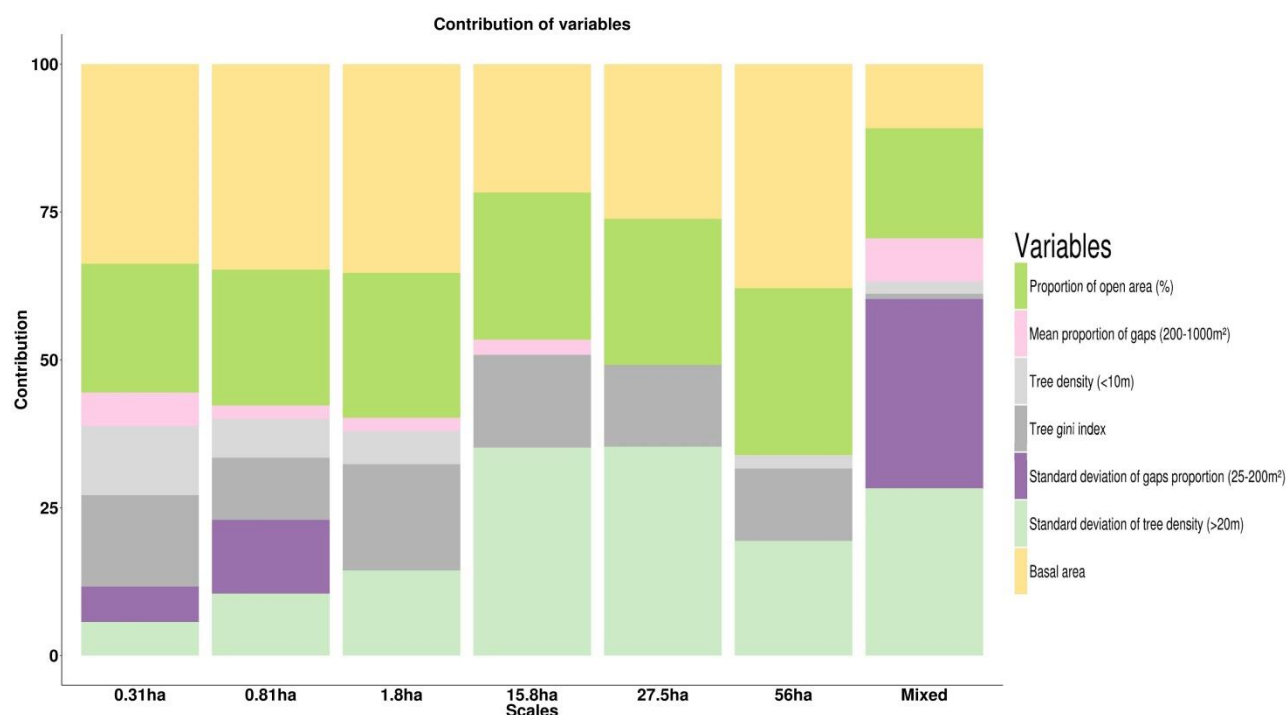
*Table 26 : AUC values for summer models*

Scale	JURA	AIN
0,0625ha	0,75	0,8
0,31ha	0,75	0,81
0,81ha	0,75	0,81
1,8ha	0,74	0,81
15,8ha	0,75	0,83
27,5ha	0,76	0,83
56ha	0,76	0,82
Mixte	0,76	0,84

### 3.4.4 Variables contribution

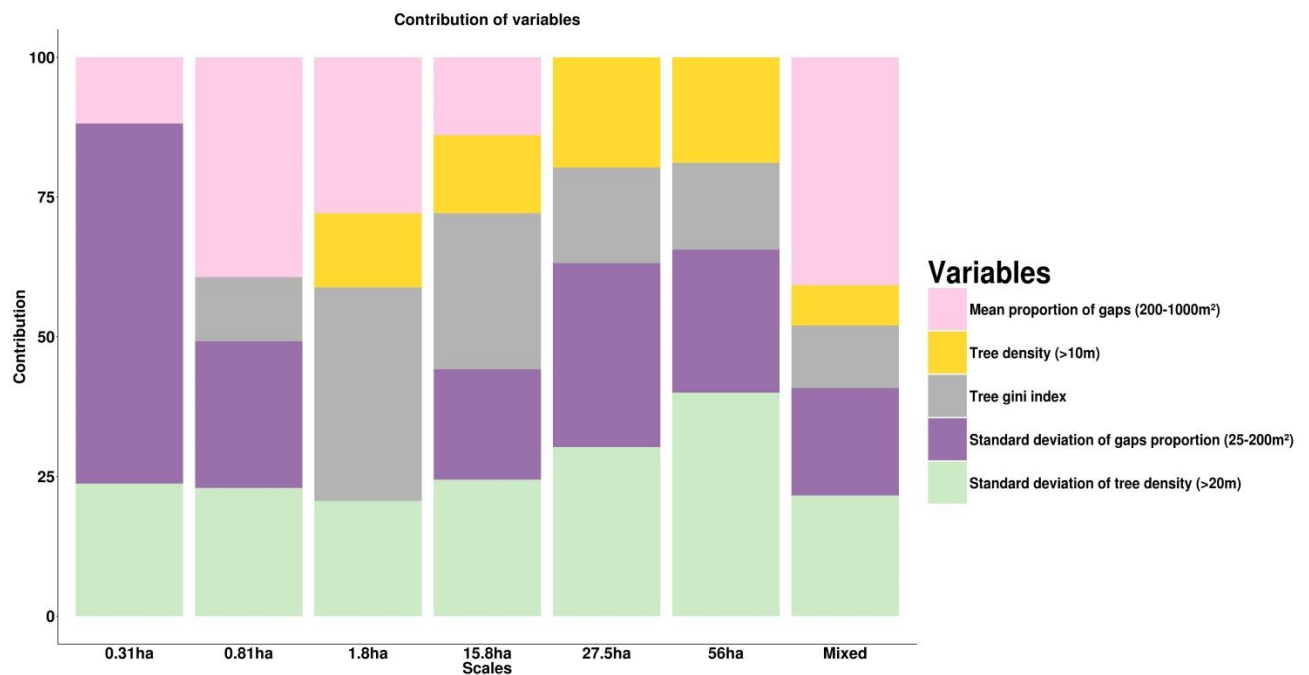
In the Ain study area, the basal area and proportion of open area contributed highly to all models (Figure 59). The standard deviation of the tree density (>20 m) contributed the most to the largest scale models. The standard deviation of gap proportion highly contributed to the mixed model and contributed moderately to small scales models. The tree gini index highly contributed to the 0.31 ha and 1.8 ha scales models.

Figure 59 : Variable contribution for the Ain study area models



In the Jura study area, the variables that contributed to models were different among scales (Figure 60). The standard deviation of tree density (>20 m) contributed to all models. The proportion of gaps (200-1000 m²) contributed to models at 0.81 ha, 1.8 ha and to the mixed model. The standard deviation of the proportion of gaps (25-200 m²) and the tree gini index contributed to all models except at the 1.8 ha scale.



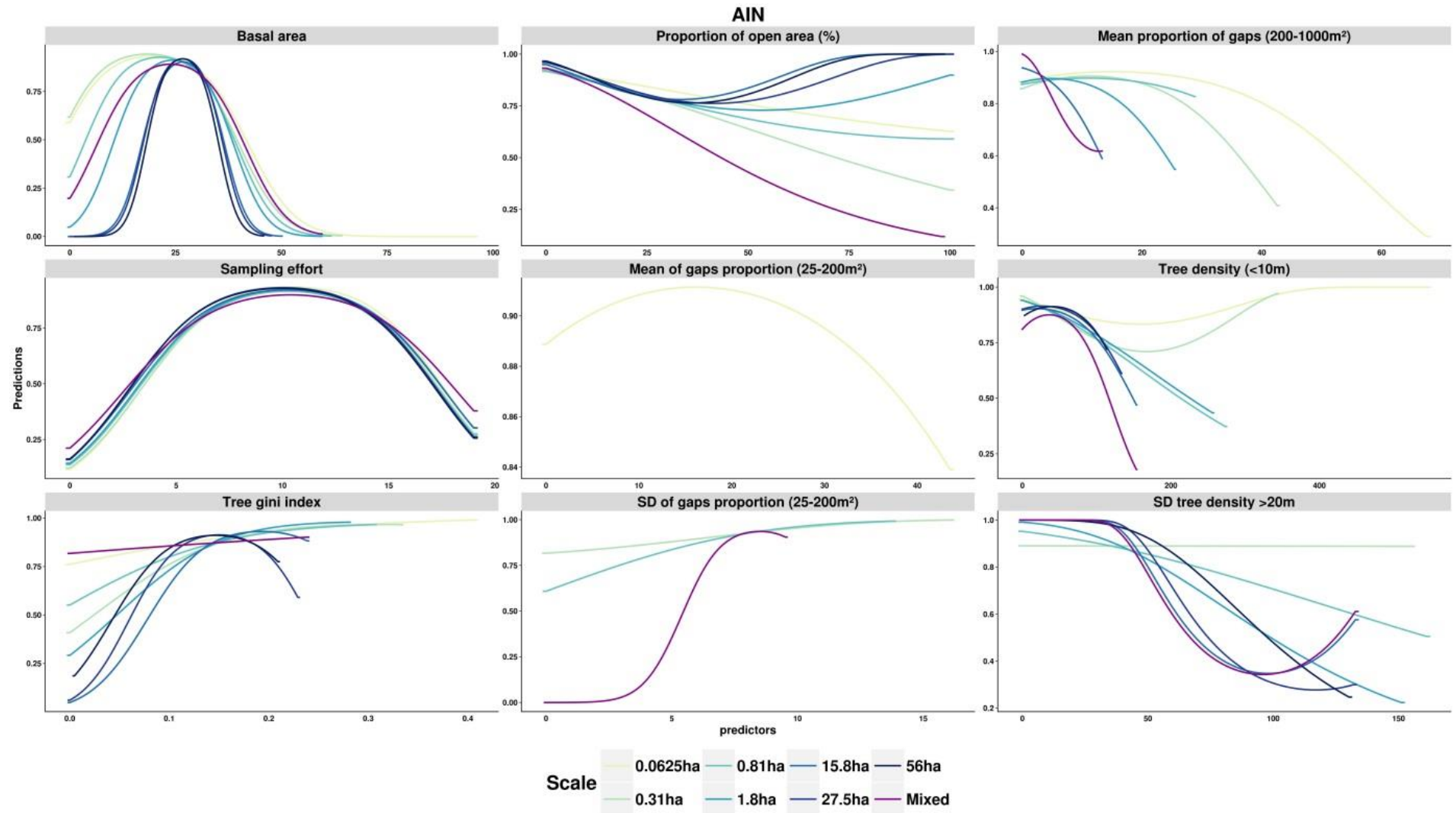
*Figure 60 : Variable contribution in the Jura study area for summer models*

### 3.4.5 Response curves

The response curves of the variables were more similar between scales and mixed models except for the tree gini index and standard deviation of the gap proportion (25-200 m<sup>2</sup>) (Figure 61). The basal area response curves showed that values between 20 to 25m<sup>2</sup>/ha are the most suitable for the species, with lower values for the smallest scales. The highest proportion of open area was predicted as unfavorable at the smallest scales and for the mixed model. However, intermediate values (25%) were predicted as more favorable at the three largest scales. For 5 of the 6 scales and for the mixed model, a high standard deviation value for the density of tree (>20m) was predicted unfavorable. At the smallest scales, a high tree gini index was observed to be more suitable. The mixed model response curve for the standard deviation of the gaps proportion (25-200 m<sup>2</sup>), which has a high contribution to mixed models, showed that a proportion of approximately 8% is the most suitable for Capercaillie.



Figure 61 : Response curves for the Ain study area models







The response curves for the summer Jura models differed between scales for numerous variables (Figure 62). For the standard deviation of the tree density ( $>20\text{m}$ ), at the 0.31 ha and 27.5ha scales, high standard deviation values were unfavorable. However, for the 0.81 ha to 15.8 ha scales, intermediate values were predicted as unsuitable. On the contrary, for the largest scale and the mixed model, those same values were predicted as favorable.

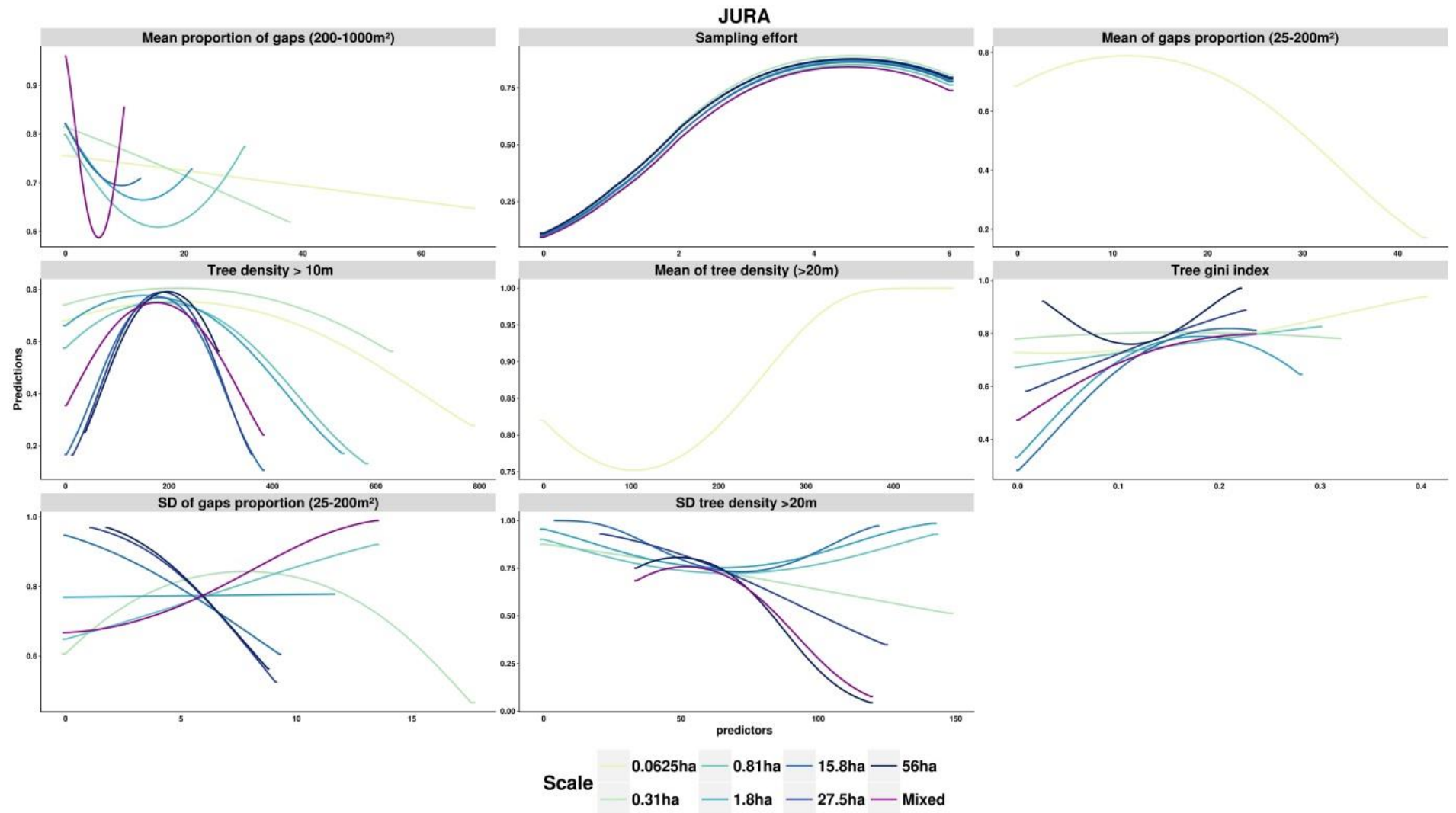
The proportion of gaps (200-1000m<sup>2</sup>) responses to the two smallest scales were linear and showed that a high proportion of gaps was unsuitable. For the other scales and the mixed model, a quadratic response predicted unsuitable habitat for proportion between 10 to 20%. Lower and higher proportions were more suitable compared to those values.

For the standard deviation of the proportion of gaps (25-200m<sup>2</sup>) highly different response curves were observed. Indeed, at small scale an optimum was observed for a value of 7. At the 0.81 ha and for the mixed model, a positive linear response was observed. At the 1.8 ha scale, a monotonous response was observed whereas for all the other cases a native linear response was observed.

The tree gini index response at small scale showed that higher values were more suitable for the species. However, at largest scales an optimum for a value of 0.2 was observed.



Figure 62 : Response curve for the Jura study area models



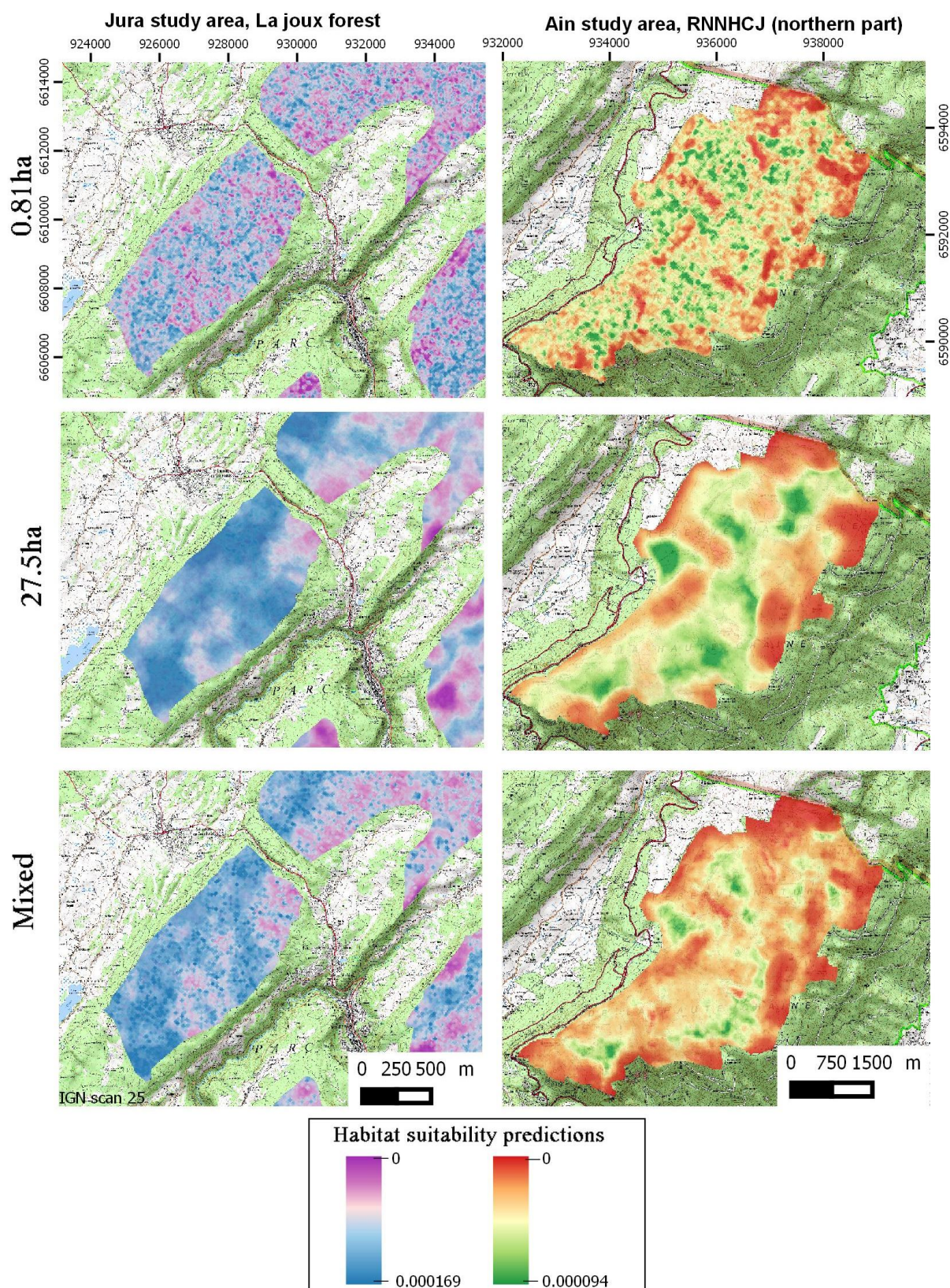


### 3.4.6 Predictions

The predictions for both sexes in summer were very different from the winter models (Figure 63). In general for both areas, summer models seemed to be less spatially restrictive than winter models. In particular, the forest of *La Joux* is almost entirely favorable whereas only several patches were found to be favorable in winter. In the *Réserve Naturelle Nationale de la Haute Chaîne du Jura*, the predictions for the southern part are relatively similar. However the northern part, which was predicted as unfavorable in winter, was predicted as more favorable in comparison to summer models.



*Figure 63 : Predictions at two different scales and for mixed models in summer*





## 4 DISCUSSION

---

In this chapter, Capercaillie distribution models using object-oriented LiDAR variables were produced at multiple scales for both the winter and summer seasons. In the first part we compared PC area-based and object-oriented metrics, and in a second part, we produced final models for both winter and summer for the two study areas (Ain and Jura).

The role of the environmental variables for both models reliability and results transmission was studied in the first part. The use of LiDAR extracted object-oriented variables regarding PC area-based metrics had a limited impact on model performance, accuracy and transferability. Nevertheless, the use of those LiDAR object-oriented variables may still improve communication between researchers and managers, and thus were used to produce the final models in both study areas. In a second step in the Ain study area, the basal area, which is a variable widely used in forestry, was added to the models to reinforce the impact of the results. This variable highly contributed to both winter and summer models. This observation showed that variables collected by managers during forest inventories can be used to predict the distribution of endangered forest species (Teuscher et al., 2013). The consistency of the basal area response curves through the models (between scales and seasons) showed that this variable could be a good predictor for Capercaillie distribution.

The contributions of the other variables that were common between the Ain and Jura, were consistent between scales and sexes but variations were observed between seasons and study areas. Nevertheless, four variables were found to have similar contribution in most models: the distance to ski runs (for winter models), the proportion of open area and the standard deviation of the tree density (> 20 m) and the tree gini index. We remark here that elements describing the presence of unsuitable habitat seemed essential for Capercaillie distribution modelling. As observed in precedent studies, variables describing the horizontal and vertical heterogeneity were important (Graf et al.,



2009b; Zellweger et al., 2013b) however the shape of the response was not always similar between study areas.

Performances varied between scales and sexes but were overall classified from poor to good performance (Araújo et al., 2005). In winter, performances were lower for female models, which may be due to the fact that there are fewer observations compared to males. In addition, one major limitation of our object-oriented metrics is that it cannot yet well describe the understory characteristics. The lack of this information may have a higher impact on female models, as individuals are supposed to use denser understory environment than males. This observation makes the link with the results obtained in Chapter 3, where no amelioration on the sex ratio estimation was found when applying a correction to spatial sampling bias. A difference in detection probability between males and females is likely to have a large influence on models predictions, producing more similar results than expected (Guillera-Arroita, 2016; Lahoz-Monfort et al., 2014).

Summer models showed similar performances in the Jura but higher performances in the Ain study area compared to winter model. Nevertheless, the differences between summer and winter models predictions were important in particular in the Jura study area. It has been shown that the reliability of evaluation metrics can be questioned (Fourcade et al., 2018). Here, summer models in the Jura produced predictions that were not always ecologically relevant (windfall and open areas set as favorable), but this aspect was not reflected by the AUC values. The low number of observations and the use of a less precise sampling bias correction method certainly played a large role in this result. The variables contributing to both seasons in the Ain study area were highly similar, showing again the importance of the variables describing horizontal and vertical heterogeneity for the species. In both areas, the major difference between the models is the use of the distance to ski runs in winter. As this variable contributes for a large part in all winter models and because no variable accounting for human disturbances (hiking and biking activities in summer) were used in summer models, it may be in part responsible of the variation observed in the predictions. As a consequence, summer models produced results based only on the forest structure. Areas that were found to be unfavorable in winter due to disturbances may yet contain habitat that can be structurally favorable for





Capercaillie. Thus, such type of habitat can be detected in summer models. Nevertheless, the differences observed between models confirm the differences existing in the habitat used by Capercaillie between seasons. In particular, the use by females with their chicks of more open areas with herbaceous layers may promote a higher suitability for those areas in summer models than in winter models.

Variations in terms of performance were observed among scales, indicating that the choice of a relevant scale is important to produce reliable models. Those variations were higher for winter models than for summer models. Nevertheless, contrary to the Hazel Grouse models, Capercaillie mixed models had similar or higher results than single scale models, and this in both winter and summer, joining the observations of (Graf et al., 2005). Because the best scales for each variable were different between study areas, the resulting predictions are different. In winter in the Jura study area, the predictions were similar to the largest scales predictions, whereas in the Ain study area predictions were comparable to small scale predictions. As for Hazel Grouse models, because of the lack of similarities between study areas, the best scales for modeling Capercaillie distribution cannot be found. Differences in the landscape between the two study areas and the role of other factors, such as population density and interactions with other species might explain this result (Mateo-Sánchez et al., 2016; Oatway and Morris, 2007; Prokopenko et al., 2017). Nevertheless, even if the mixed models have a better performance, the use of the other scales may still be useful for managers to plan both global and local actions.

## 5 CONCLUSION

---

The role of object-oriented extracted metrics as well as the estimation of variables used in forestry, such as the basal area, did not improve the performance or the transferability of the models but fit better managers conceptual view of the landscape than the commonly used PC area-based metrics. The results obtained with these metrics will allow a better transmission and adoption by managers.



From the different models created (study areas, scales, seasons and sexes), high variations were found indicating that datasets and environmental variables played an important role in models accuracy and performances. We showed the importance of a high horizontal and vertical heterogeneity associated with a basal area ranging between 20 and 30 m<sup>2</sup>/ha. The role of disturbances on individual distribution was also observed with a negative effect on the bird presence in areas close to ski runs (< 1km).



# CHAPTER 6: BEECH REGENERATION DISTRIBUTION MODELS





## 1 INTRODUCTION

---

### 1.1 Habitat loss

Beech regeneration greatly impacts Capercaillie habitat in the Jura Massif, leading to a decrease of the proportion of available suitable habitat. This phenomenon is linked to a change in forestry practices: in the past, beech was used as fuel wood and intensively harvested; but nowadays, the economic interest in this species has declined. In addition, cows pasture activities that long limited understory regeneration are now decreasing in the Jura. As a consequence, intense beech regeneration occurs after logging operations at altitudes lower than 1400 m (Vittoz, 1998) (Figure 64). Yet, bilberry seed germination is affected negatively by shade (Nin et al., 2017). Thus, the understory cover, once suitable for bilberry growing (essential for Capercaillie, Figure 65) is becoming more and more closed and unfavorable.

*Figure 64 : Example of unsuitable habitat due to beech intense regeneration (Groupe Tétrás jura)*



*Figure 65: Example of ideal summer habitat with bilberry ground layer (Groupe Tétrás jura)*



This dynamics can be held up by specific forestry management actions in habitat that are deteriorated (or will soon be) in order to restore a suitable environment (usually the proportion of bilberry cover availability but also the herbaceous layer in general), for the birds (Broome et al., 2014). It consists in removing all beech regeneration to re-open the understory. Such management actions are a part of the “*Plan départemental d’action en faveur des Tétracidés 2014-2017*”,



financed by the departmental council of Ain and as a part of NATURA 2000 network. The combination of both forestry and Capercaillie management in parcels has been tested in the Giron and Champfromier communal forests with the aims to ameliorate the habitat, raise awareness to Capercaillie conservation among managers and evaluate different protocols in the field (*Plan départemental*). This action followed a first project in the natural national reserve of the Chaîne du Haut-Jura (2008-2013) that cleared beech regeneration on 83 plots of 500m<sup>2</sup> (with 82 control plots) (NATURA 2000).

### 1.2 Beech

The European beech is one of the major deciduous tree species found in Europe. This species is dominant in many forests across Europe and is limited in its range by pronounced cold or dry climates (Bolte et al., 2007). Trees measure between 30 and 35 m height, and have a high growth rate. Individuals can live more than 250 years but are usually harvested between 80 and 120 years. Fructification occurs only every 5 to 8 years and the seeds can be conserved for 5 years before germination. The species is not strongly self-pollinating thus the presence of several individuals blooming at the same time is necessary for the reproduction. Beech is a species tolerant to shade, thus regeneration can occur in forest with continuous canopy cover. The species grows preferentially in moist soil where roots can easily penetrate the substrate on a limestone bedrock.

### 1.3 LiDAR and understory modeling

While airborne LiDAR is a powerful tool to capture canopy characteristics, the detection of the understory faces some difficulties (Latifi et al., 2015; Su and Bork, 2007). Indeed, tree foliage intercepts a large part of the laser emissions, thus only very few reach the component present below. The number of points reaching the ground depends highly on the season (leaf-off versus leaf-on), the density of canopy, and the density of laser emissions. Thus, while LiDAR derived measures of the canopy top layers have strong correlation with field observations, this relation is lower for intermediate height layers (Latifi et al., 2015). Nevertheless, multiples studies have tried to address this issue (Hamraz et al., 2017; Latifi et al., 2015; Leutner et al., 2012; Martinuzzi et al., 2009) using



additional multiple variables sources such as hyperspectral or climate data. In particular, LiDAR was used to predict the spatial distribution of 14 understory plants that are food resources for grizzly bear (*Ursus arctos*). The addition of LiDAR variables to climate and forest cover improved the models for eight of the 14 plants species. In this case study, terrain conditions were more important predictors than canopy characteristics, indicating that importance of models variable such as elevation, solar exposition or slope should be taken into account when mapping understory species (Nijland et al., 2014).

### 1.4 Objectives

The aim in this chapter was to create spatial models of the distribution of intense beech regeneration over the two study areas. The resulting predicted maps may be used to plan future actions to limit Capercaillie habitat loss in the Jura Massif. To achieve this objective, LiDAR extracted variables as predictors and field vegetation plots with information on the coverage of the beech regeneration as response variable will be used to create SDMs.

## 2 MATERIALS AND METHODS

---

### 2.1 Environmental variables

LiDAR extracted variables, already used for Capercaillie and Hazel Grouse modeling were also used in this Chapter. Topographic LiDAR extracted variables (elevation, exposition, slope and curvature) were used as well.

### 2.2 Ain study area

As presented in Chapter 2, beech regeneration data over the Ain study area are abundance data recorded as number of individuals by plot. Those data were transformed into binary response presence-absence. When the number of individuals was below 5, the value 0 (absence) was



attributed and all other values were considered as presence. Both types of data were used for modeling.

Those data were collected with a highly precise GNSS receiver, thus with the aim to gain in accuracy, point-cloud based metrics were extracted for each plot with three different radii (10 m, 11.5 m and 14.1 m). The first two diameters were chosen in function of the dataset plot size (for more information see chapter 2 section 3). The radius 14.1 m created a surface equivalent to the surface of a 25\*25 pixel and was thus created to obtain better predictions over the study area. Object-Oriented and topographic variables were not calculated again over the vegetation plots, thus the variables values were extracted from the raster datasets at each location.

### 2.3 Jura study area

Concerning the Jura study area, the dataset containing categorical variables was transformed as well into a binary variable presence-absence. The categories D, C and B were assumed to be presence (regeneration cover > 20%) and the category A to be absence. The dataset containing information on the percentage of cover was used as abundance data.

The LiDAR variables were extracted afterwards at each plot location.

### 2.4 Visualization

As a first step, the identification of differences among the LiDAR variables between presence and absence of beech regeneration plots was explored. Boxplots for each LiDAR variable were thus done with the aim to visually identify potential variables for modeling.

### 2.5 Modeling

Random forest algorithm (Breiman, 2001) and logistic regression were chosen to model the distribution of beech regeneration areas for presence-absence data.

The random forest analysis was conducted using the R package caret (Kuhn, 2008). In addition, the categories “presence-absence” frequencies were imbalanced in the Ain dataset with the minority





category (presence) being the element we would like to predict (Ain: 0.82-0.17, Jura: 0.50-0.49).

Classification algorithms are known to minimize the error rate and underestimate the importance of the minority class in the process (Liaw and Wiener, 2002). Thus, three approaches were used with the aim to overcome this problem in the Ain study area:

- Weight: assigning a higher cost of misclassification of the minority class
- Down-scaling: down sample the majority class to obtain equivalent classes
- Up-scaling: over sample the minority class by replicating the data to obtain equivalent classes.

Both random forest and logistic regression models were evaluated using the AUC.

To analyze abundance type data, a generalized linear model (GLM) with a poisson regression was applied and evaluated using RMSE and linear regression.

Models were created for each area and for the three plot sizes in the Ain area using four different models.

## 3 RESULTS

---

### 3.1 Find potential variables

#### 3.1.1 Ain study area

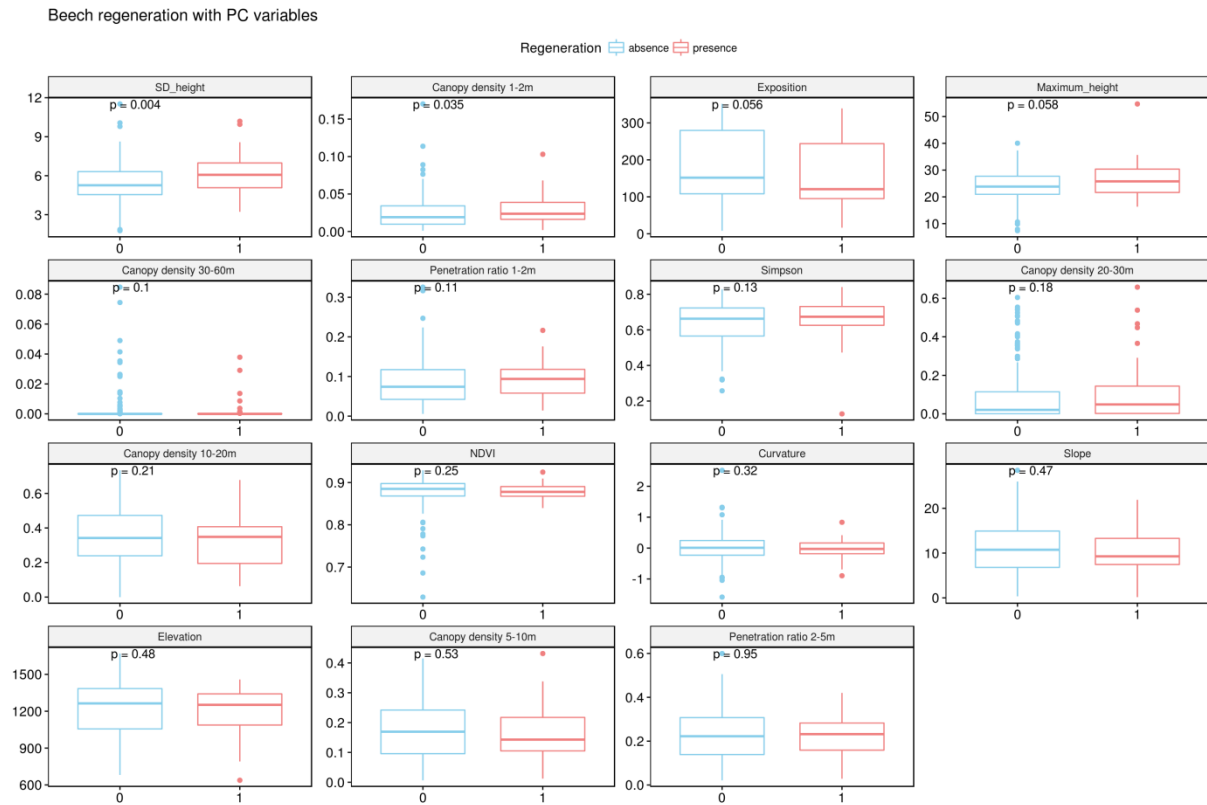
The differences between presence and absence of beech regeneration, regarding the extracted LiDAR variables and the topographic variables were low.

With the locally extracted PC area-based variables, only the variable standard deviation of height and the canopy density 1-2 m were significantly different between plots with beech regeneration and absence plots. Those variables were followed by the exposition and the maximum height from which



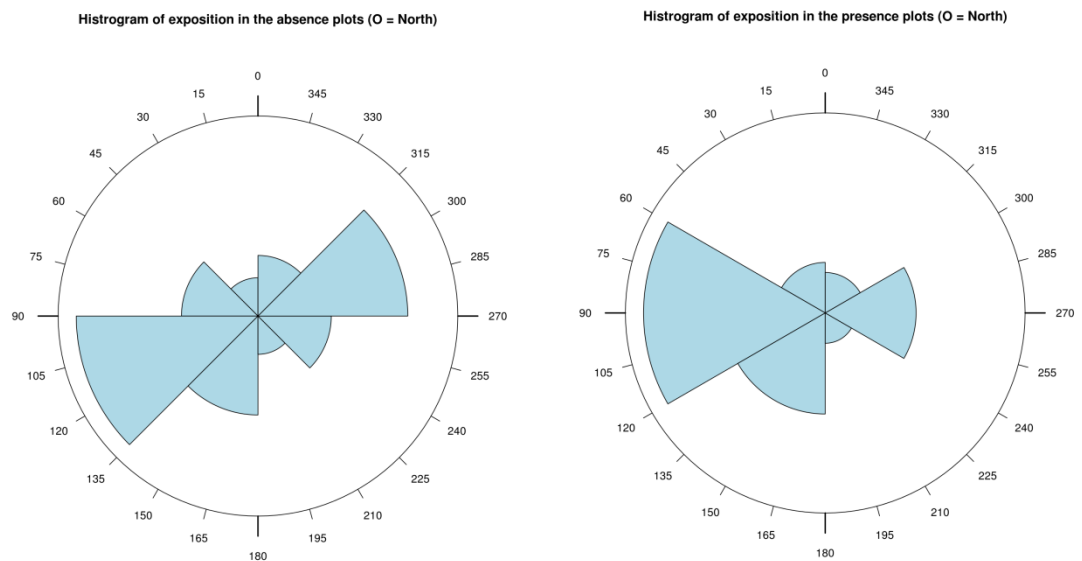
p-value were inferior to 0.06. The other variables were not significantly different between the presence and absence of beech regeneration (Figure 66).

*Figure 66 : Boxplots and p-values (Wilcoxon test) for each PC area based LiDAR variable and topographic variable in the Ain study area.*



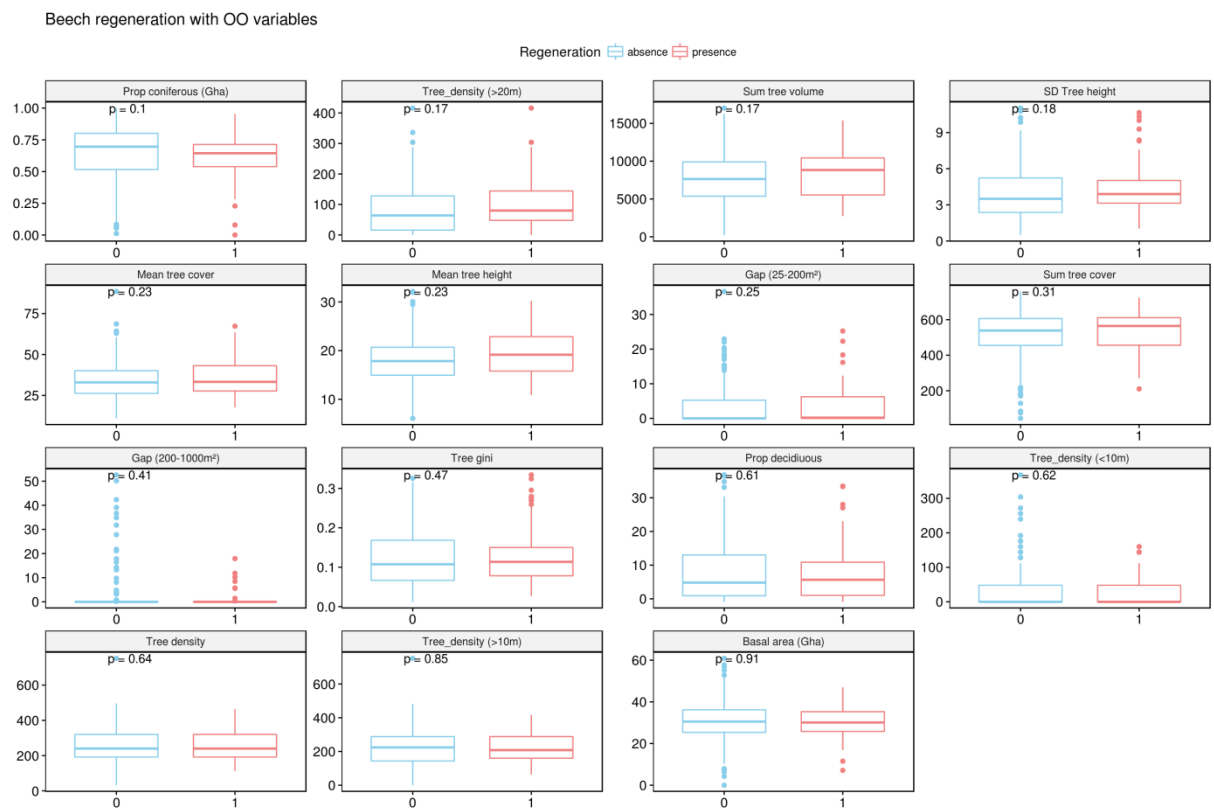
Concerning the exposition variable, the use of a boxplot may not represent best angular exposition data where 0° and 360° are in reality representing the same direction. To ensure that the differences between exposition of presence and absence plots represent a real phenomenon, circular histogram of the plot orientation were made (Figure 67). A high proportion of presence plots were oriented to the west whereas in the absence plots south-west and north-east directions were observed to have the higher proportion.

*Figure 67 : Circular histograms for exposition*



For the OO variables, none were significantly different between plots with or without beech regeneration (Figure 68).

*Figure 68 : Boxplots and  $p$ -values (Wilcoxon test) for each OO LiDAR variable in the Ain study area.*







Thus from those results, seven variables among those extracted over the plot locations (PC area-based metrics) were selected for modeling. An additional variable describing the topographic conditions was also selected (Table 27).

*Table 27 : variables selected for modeling*

Ain
Maximum Height
SD Height
Canopy density 1-2 m
Canopy density 20-30 m
Canopy density 30-60 m
Penetration ratio 1-2 m
Simpson index of heights
Exposition

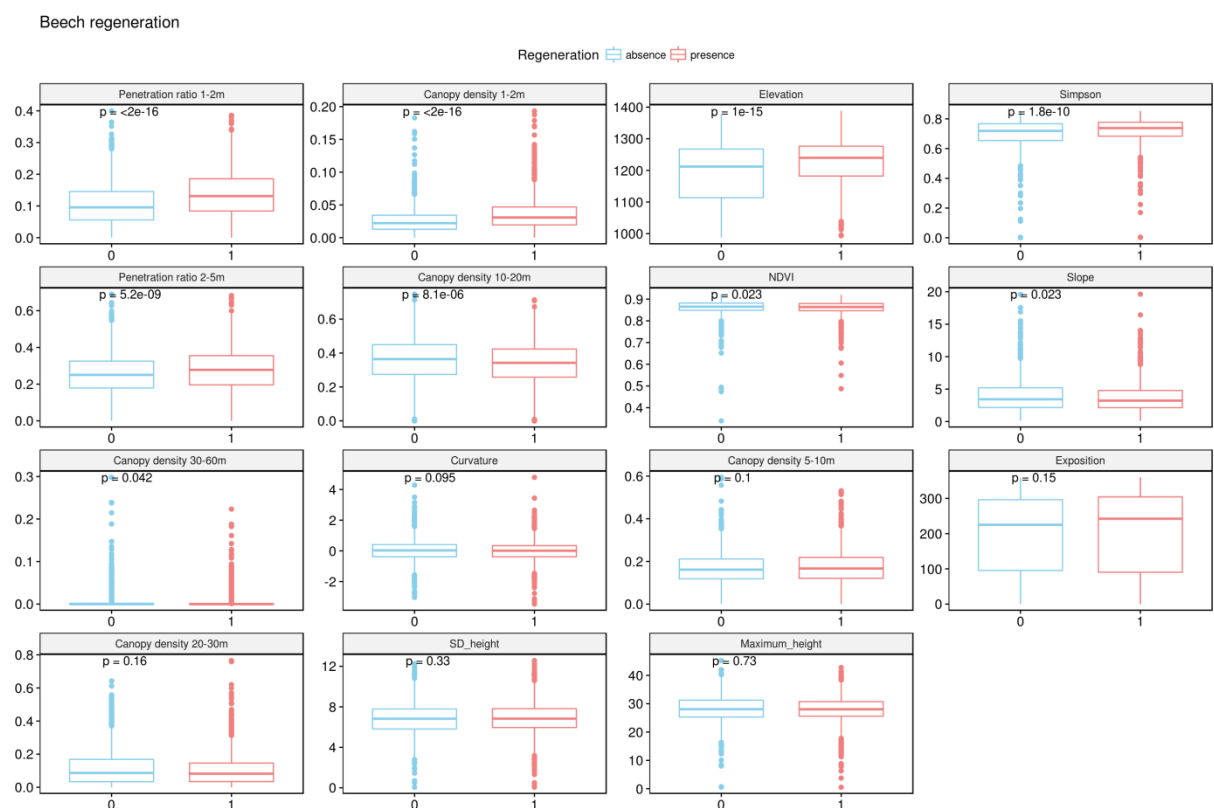


### 3.1.2 Jura study area

In the Jura study area however, multiple variables were found to be significantly different between presence and absence plots. Among those, the penetration ratio 1-2 m and 2-5 m the canopy density 1-2m, the elevation and the Simpson index showed the highest differences. Those variables were followed by the canopy density 10-20 m, the NDVI, the canopy density 30-60 m, and the maximum

(Figure 69).

Figure 69 : Boxplots and  $p$ -values (Wilcoxon test) for each PC area based LiDAR variable and topographic variable in the Jura study area.





Regarding the OO variables, heights were significantly different between presence and absence plots (Figure 70). Among those the Gini index for tree height, the tree density (<10m), the standard deviation of tree height, the sum of tree volume and the mean tree height were selected (Table 28).

Figure 70 : Boxplots and  $p$ -values (Wilcoxon test) for each OO LiDAR variable in the Jura study area.

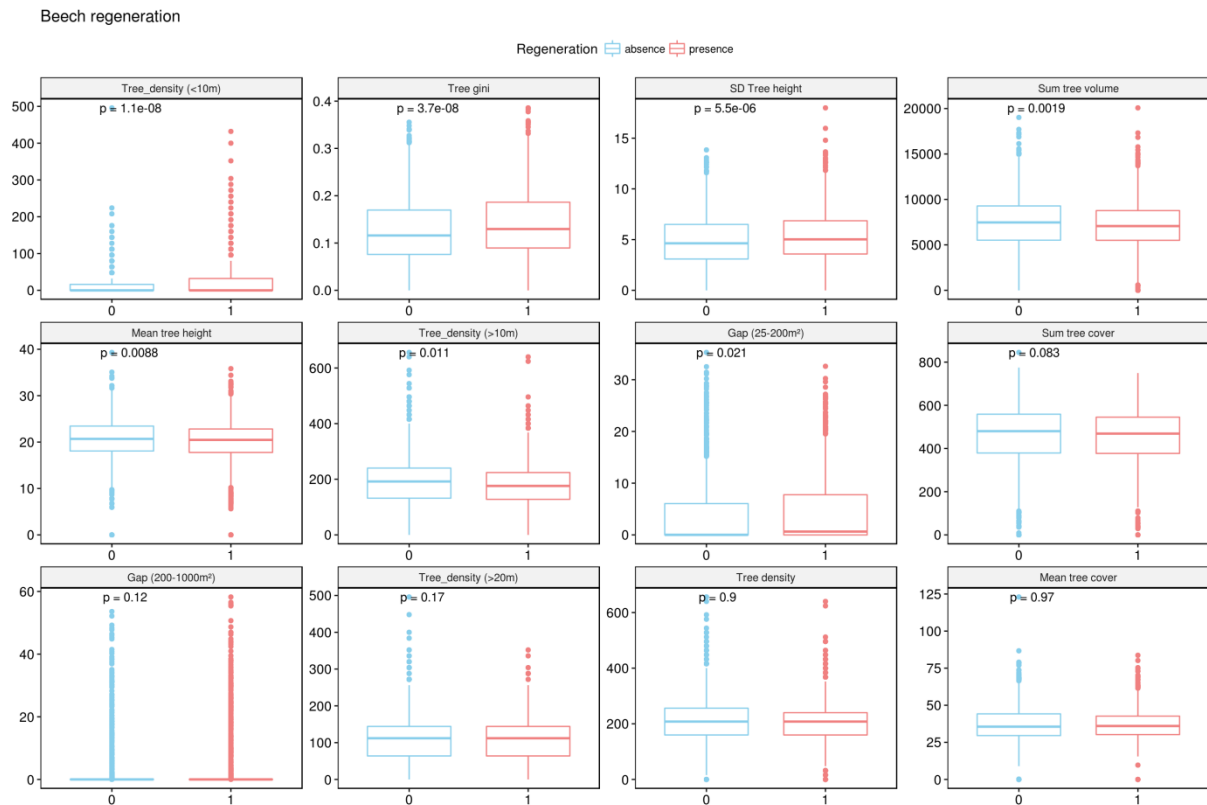


Table 28 : variables selected for modeling

Jura
Slope
Canopy density 1-2 m
Canopy 10-20 m
Penetration ratio 1-2 m
Penetration ratio 2-5 m
NDVI
Tree density (<10 m)
Gini index for tree height
SD of tree height
Simpson index
Elevation
Sum of tree volume
Mean tree height

## 3.2 Models

### 3.2.1 Ain study area

First for the plot radius 10 m, we observed that the classification error for presence is very high for the basic and weighted cases, indicating that the model cannot predict with accuracy the presence of beech regeneration (Table 29). The AUC is in consequence very low with values below 0.6. When applying the down scaling sampling method, the classification error is lower for the presence (but higher for the absence) leading to a higher Out-Of-Bag (OOB) value and a poor performance. Finally the up-scaling sampling method showed the better classification accuracy for both presence and absence and the higher AUC value. Nevertheless, the different models can be classified as failure, because the all AUC values were below 0.6 (Guisan et al., 2017). Both the logistic regression and the GLM with poisson regression on abundance data also failed to produce a performant mode with very low AUC values or higher RMSE values (4.63) and R-squared (0.01).

The same general pattern was observed for the two other plot radius cases, with overall low performance and classification accuracy with the exception of the logistic regression (radius 11.5) and the down scaling random forest model (radius 14.1 m). Both models showed higher AUC values, which are nevertheless classified as poor performances ( $\leq 0.7$ ).

Table 29 : models results

Ain				
Plot radius	Model type	OOB	Classification error	AUC
10 m	RF basic	19.63%	0.97	0.53
	RF weighted	19.18%	0.97	0.51
	RF down	37.18%	0.41	0.54
	RF up	5.83%	0.02	0.55
	Logistic reg.	.	.	0.48
11.5 m	RF basic	18.26%	0.94	0.47
	RF weighted	18.72%	0.92	0.53
	RF down	46.15%	0.49	0.59
	RF up	4.44%	0	0.56
	Logistic reg.	.	.	0.62
14.1	RF basic	20.09%	0.97	0.50
	RF weighted	19.63%	1	0.54
	RF down	44.87%	0.43	0.68
	RF up	5.83%	0.01	0.54
	Logistic reg.	.	.	0.44

Plot radius	Model type	RMSE	R-squared
10 m	GLM poisson	4.63	0.01
11.5 m	GLM poisson	4.68	0.01
14.1	GLM poisson	4.72	0.01

### 3.2.2 Jura study area

In the Jura study area, the basic random forest model showed a moderate classification error for beech regeneration presence category (Table 30). The logistic regression showed the lowest performance and AUC values are classified as fair. GLM with Poisson regression showed low performance with an RMSE of 20.5 and an r-squared of 0.1.

Table 30 : Random Forest models results

Jura				
Pixel width	Model type	OOB	Classification error	AUC
25 m	basic	34.67%	0.39	0.69
	Logistic reg.	.	.	0.65

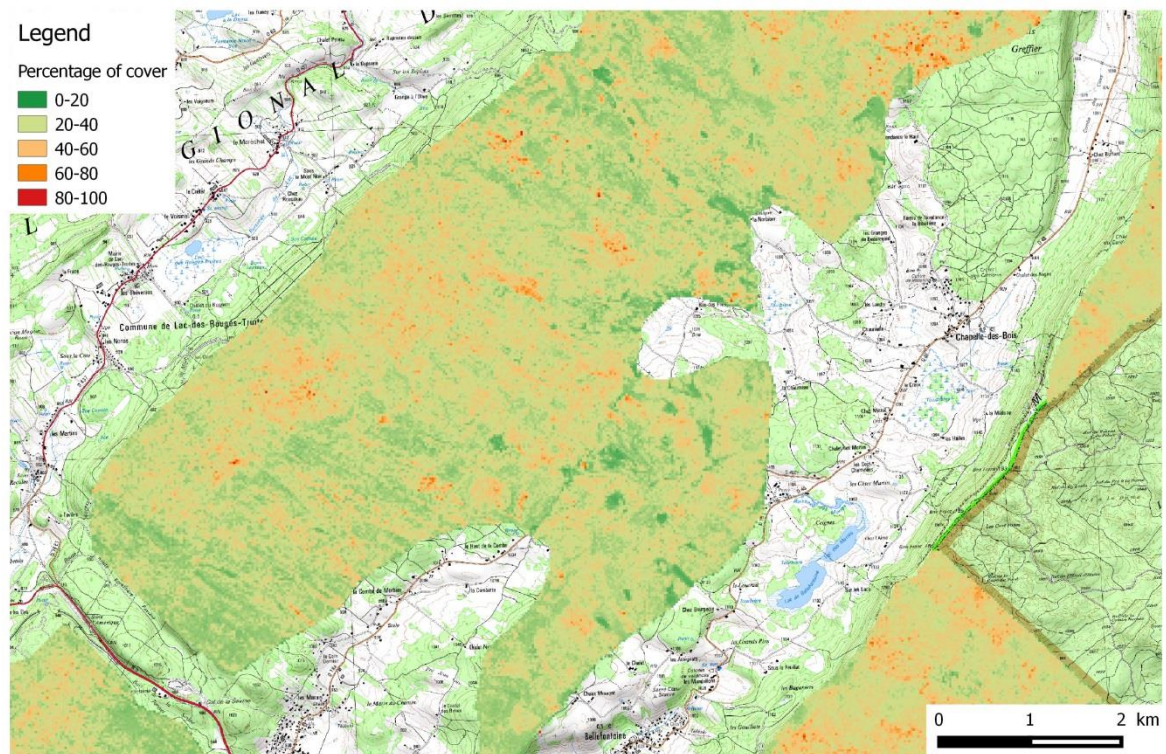
Pixel width	Model type	RMSE	R-squared
10 m	GLM poisson	20.5	0.1



### 3.2.3 Predictions

Predictions were done for the best models in each area. In the Jura study area, the models predicted a large part of the forest to be between a cover of 0 and 40% of beech regeneration (Figure 71). Isolated patches with predictions of higher cover are also observed.

*Figure 71 : Predictions from the GLM Poisson distribution model (Mont Noir Forest, Jura study area)*

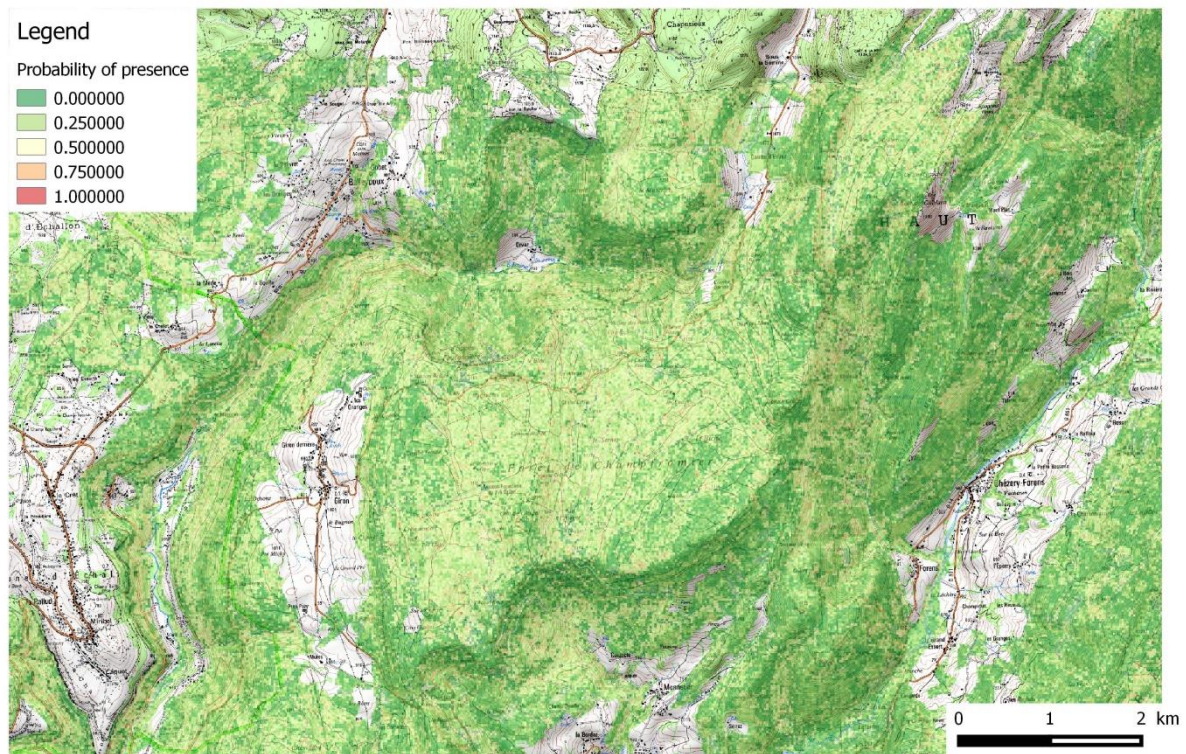


The predicted map from the Random forest model with down scaling, showed that the forest is composed of a matrix constituted of pixel with a probability of the presence of beech regeneration higher than 0.5 (Figure 72). Large patches of low probability of presence are observed but very high probabilities of presence are rare. If this model had the highest AUC value, the classification error is high thus probably leading to a low reliability of the predictions. In addition, this model was not validated by manager regarding their knowledge of the forests in the Ain study area.





*Figure 72 : Predictions from the GLM Poisson distribution model (Champfromier Forest, Ain study area)*



## 4 DISCUSSION

---

First of all, when searching for potential environmental variables, the results were different for each study area. Only few variables showed differences between presence and absence plots in the Ain study area whereas a large part of the tested variables differed in the Jura study area. This might be due to the large difference in sample size (Ain: 274; Jura: 2100) or to the differences between the two datasets. Indeed, beech regeneration is usually present in patches smaller than the plot size, thus in one case the measure of the regeneration cover is approximate, but might be more representative of the whole plot characteristics than the count of the number of individuals on three small sub-plots. Furthermore, in the Jura study area, vegetation plots were sampled in areas where the Capercaillie presence is known, thus this dataset might be overall more homogenous regarding the forest structure than the Ain study area dataset. In the Ain area, selected variables were all from

PC based metrics, which is not surprising as OO variables were not designed to measure directly understory characteristics thus having a less pronounced response to the presence or absence of regeneration.

In the Ain study area, higher canopy density 1-2 m and higher forest stands with a higher heterogeneity were associated to beech regeneration presence plots. In the Jura study area, variables calculated for understory layers (1-2 m and 2-5 m) indicated as well that such denser layer can be associated with beech regeneration presence. These observations confirm that LiDAR can to some extent detected directly understory elements. In addition to those observations, the role of the canopy density was characterized by a lower density of the 10-20 m layer in the Jura and of the 30-60 m layer in the Ain study area indicating that indirect characteristics of the landscape may also help to differentiate regeneration presence or absence.

Concerning the topographic variables, the elevation (Jura), slope (Jura) and the exposition (Ain) were found to be different between presence and absence plots. Lower elevation were associated with regeneration presence plots, which is not surprising as it is known that beech does not occur at the highest elevation (>1400 m) (Vittoz, 1998) but this threshold might soon be higher due to climate change. Beech regeneration presence plots were more exposed to west than other plots in the Ain study area.

Despite those promising primary results, the modeling of the distribution of the beech regeneration was challenging. In the Ain study area, most models failed as only two out of 15 had an AUC higher than 0.6. In the Jura study area however, the observed models performances were higher with one model classified as fair ( $0.8 \geq \text{AUC} \geq 0.7$ ). In this area, the GLM with Poisson distribution applied on abundance data gave low accuracy.

Those results showed that the prediction of understory using LiDAR datasets is difficult, even if some other studies obtained good predictions accuracy. One characteristic of our LiDAR datasets may have played a large role in the difficulty to describe the understory. Indeed, both campaigns were done at a season when deciduous trees still have their leaves on, thus reducing the penetration of the signal





through the canopy. A LiDAR campaign during winter may thus help to better characterize directly the understory. However a thick snow cover may limit the collection of information on the lower vegetation. Nevertheless, the improvement in LiDAR processing may soon allow a better analysis of understory characteristics and even extract shrub objects (Hamraz et al., 2017). However, in this study a mean density of 170 points/m<sup>2</sup> in the LiDAR raw point cloud was used to accurately extracted such objects, which is much denser than our LiDAR datasets (21.3 and 18 points/m<sup>2</sup>).

A second aspect is the effect of imbalanced data. Indeed, the rarity of the plots with an intense beech regeneration process regarding the total number of plots is making it difficult to extract relevant data probability due to strong variations between plots. The effect of the different methods to correct for imbalanced data did not improve models performances in the Ain study area. Nevertheless, a strong positive effect on the classification accuracy with down and up-scaling was observed. These results indicate that those methods can in some cases improved random forest model accuracy. Yet, in the case of down-scaling sampling, the diminution of the number of absence plots may reduce the scope of the environmental values associated with the absence of regeneration. Regarding the up-scaling sampling method, the replication of the small number of presence plots may increase the weight of some environmental variables, which may not truly represent the variables impacting the regeneration.

Finally, the selected variables may not be enough for predicting the regeneration of a specific species. Indeed, multiples characteristics of the environment were not taken into account in our models. The regeneration intensity is not only influenced by the overall forest structure (tree density, canopy cover or exposition), but also by the type of soils, the presence of a reproducing tree. The presence of a few individuals that produce a large amount of seeds can influence greatly the intensity of the regeneration in the specific locality. Similarly, the dispersion of the seeds by vertebrates plays a role on the density and on the extent of the regeneration patch (Zwolak et al., 2016) but cannot be measured at large scales. In addition, it has been demonstrated that there is a highly intra-population variability in reproduction, which in our case potentially blurs the impact of environmental variables (Muratorio et al., 2012).



From those results, accurate and precise maps showing an ongoing intense generation process could not be made for the Ain study area but were created for the Jura study area. Indeed, the predicted maps were aiming to be used to plan efficient management actions (beech removing), thus highly accurate prediction of the presence of regeneration was needed.

## 5 CONCLUSION

---

The objective to produce accurate predictions for future management actions in favor of Capercaillie was achieved only for one of the study area. Despite the high potential of many LiDAR extracted variables to describe environmental structures that may both directly represent the understory structure or indirectly influence the beech regeneration intensity, the models created in the Ain study area failed or performed poorly. In the Jura study area, better results were obtained, probably due to a larger datasets and to a better estimation of the regeneration intensity through a response variable indicating the cover rather than individual count. The causes of the poor performance of many models may be due to the overall rarity of plots with intense regeneration and to the fact that some important variables may have been missed. However, future improvement in LiDAR datasets processing may soon allow a better estimation of the understory characteristics.

Currently field survey is the only method used to detect beech regeneration patches, which is time consuming. The predicted map in the Jury study area will be used to plan future actions of beech removal by helping the identification of areas that need to be surveyed in priority. Those actions will also be supported by the Capercaillie distribution maps created in the previous chapter.

# CHAPTER 7: SYNTHESIS AND PERSPECTIVES



Creative Commons Zero-CC0

This PhD project aimed to produce spatial distribution maps for Hazel Grouse, Capercaillie and beech regeneration to help managers in their decisions for prospective conservation actions or surveys. The specificities of the datasets (presence only, LiDAR) used and the targets species ecological requirements, allowed the exploration of the effects of spatial sampling bias, features interactions, environmental variables choice on models' performance, accuracy and on the result transfer to managers. This Chapter aims to highlight the principal contributions of this work and to propose thoughtful reflections on how our results could be improved in general (e.g. modeling methods) and locally (e.g. protocols design, new data collections, genetic).

## 1 IMPROVING SDMS PERFORMANCES AND UNDERSTAND THEIR LIMITS

---

### 1.1 Spatial sampling bias

#### 1.1.1 Main contributions

The effect of spatial sampling bias, in the specific case of a species constituted of sub-groups with distinctive behaviors has been evaluated using both a virtual framework (virtual species and virtual sampling) and a sexually dimorphic species, the Capercaillie. The estimation of the sex ratio was indeed found to be biased toward one sub-group when no corrections were applied. The two tested corrections methods (targeted background points and distance to trajectories variables) both corrected the parameters and the sex ratio estimation in the virtual study case, but not with the Capercaillie datasets probably due to large differences in the probability of detection between male and females. In addition, the role of sampling design on SDMs accuracy was highlighted, as better results were observed with a combination of systematic and subjective design than with subjective design alone. The confirmation of the sampling bias correction methods efficiency allowed us thereafter to produce reliable models for both target species, including one showing a high sexual dimorphism, using heterogeneous datasets.



### 1.1.2 Perspectives

The obtained results showed the great potential of using a virtual framework to improve our knowledge on SDMs outcomes. Such framework can be used to test new modeling methods and to estimate models' accuracies and errors before application on a real dataset (Miller, 2014; Zurell et al., 2010). A virtual species can also be used to answer a particular ecological question while the initial dataset did not allow it (e.g what is the effect of competition on SDMs? (Godsoe et al., 2017)). The impact of sampling design on SDM prediction accuracy has been to our knowledge mainly investigated using real datasets thus evaluated after the data collection (Edwards Jr et al., 2006; Hirzel et al., 2006; Phillips et al., 2009). The creation of a virtual sampling approach along with real landscape data is promising as different scenarios can be tested even before the implementation of the survey on the study area as proposed by (Hirzel and Guisan, 2002). Thus the combination of different designs may be evaluated to optimize surveys in terms of costs in money and time (Fithian et al., 2015b; Zurell et al., 2010). In addition a careful evaluation of the outcomes regarding the implementation of a sampling design will help to obtain more reliable data for SDMs. The generalization of this approach may allow modelers to ensure that the chosen sampling method has the possibility to predict the species distribution with a high accuracy.

The main limit of the spatial bias corrections applied on models is the general lack of knowledge about the bias occurring in a dataset (Isaac and Pocock, 2015). Therefore the use of variables describing the potential causes of the bias as predictor is an asset as the impact of each variable can be evaluated through its contribution (Varela et al., 2014). In a second step predictions can thus be corrected accordingly to those results. Moreover, the growing availability of GPS tracks may allow in the future the acquisition of more data on the observers' trajectories, thus making the use the targeted background point possible in more projects.

The impact on models' accuracy when a part of the population is largely undetected remains unknown. Imperfect detection assuming that all individuals can be detected with the same probability has been demonstrated to have consequences on models' accuracy (Guillera-Arroita et al., 2015). These consequences were shown to be more important when the imperfect detection

probability was depending on the environment (e.g. lower detection in dense vegetation cover) (Lahoz-Monfort et al., 2014). Yet, this aspect has been poorly taken into consideration in SDMs studies (14% out of 108 studies from 2008 to 2012 mentioned this issue (Yackulic et al., 2013)), and should in the future be further investigated.

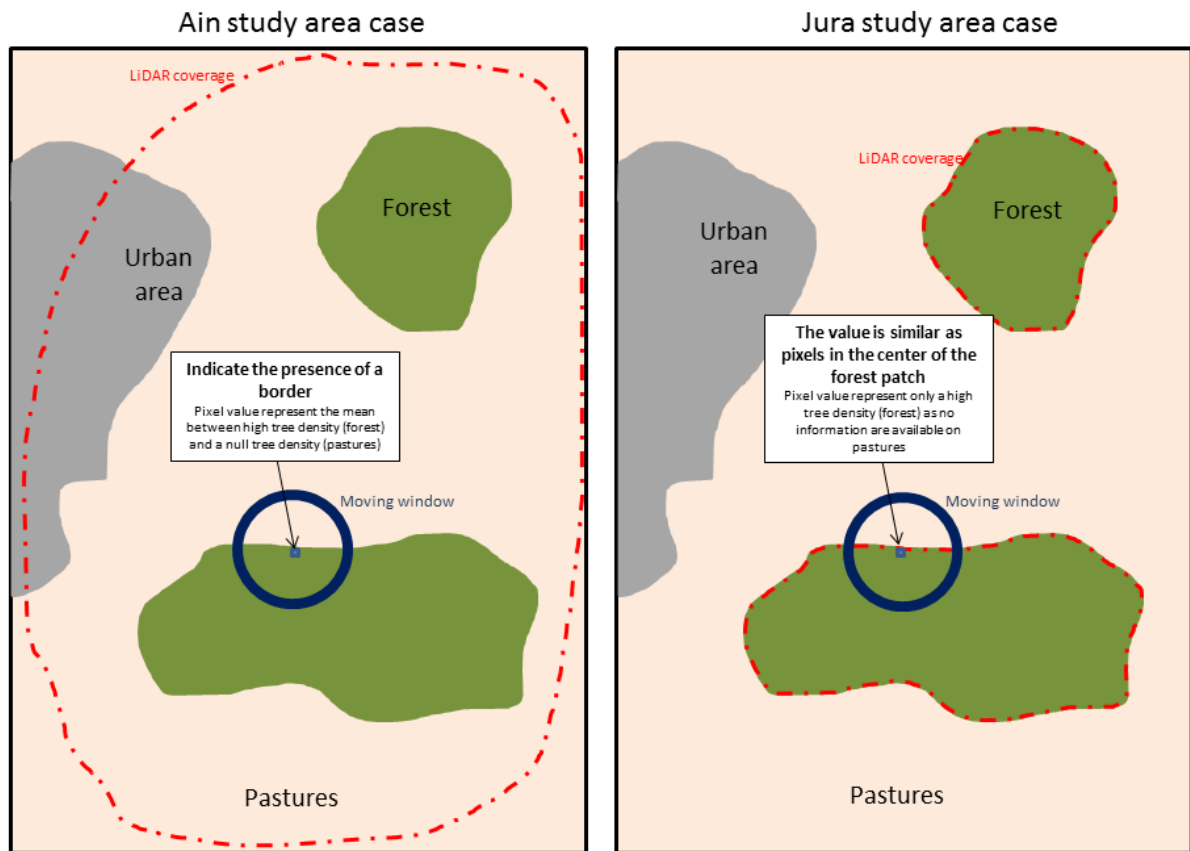
## 1.2 Scale evaluation

### 1.2.1 Main contributions

Single-scale and multi-scales models were compared for two targeted species with the aim to identify at which level environmental variables are perceived by individuals, thus improving the accuracy of the results. Differences in performance, accuracy, response were observed between scales for both Hazel Grouse and Capercaillie models. Unexpectedly, no overall best single scale for modeling was found, showing the high dependency of such result on the study area and/or on the initial choice of the scales. In addition the use of mixed models (multi-scales, multi-variables models) showed higher performances for both winter and summer Capercaillie models but not for Hazel Grouse models. In this work the differences between the LiDAR coverage of each study area may have played a large role in those differences. Indeed, the Jura study area was composed of forest landscape patches only (thus fragmented) whereas the Ain study area represented the full landscape (thus representing one block). By using a moving window to calculate metrics at different scales, the landscape variations outside forest were considered in the Ain area but not in the Jura study area (Figure 73). Our results showed that pseudo-optimized scale selection can be an asset to produce performant models but that the model accuracy is nevertheless dependent of multiple factors such as environmental variables, study area (location and extent) or targeted species ecological requirements.



*Figure 73 : Differences between the Ain and Jura study area when calculating the values at different scales*



### 1.2.2 Perspectives

Similarly to our results, Suárez-Seoane et al. (2013) observed that differences in variables contributions were higher between study areas than between the different scales in the same study area. The choice of relevant scales is challenging for multiple reasons. From our results, it seems yet difficult to compared results obtained from different studies and find the scales that will matter the most for a target species. Thus, the promotion of true optimization that does not involve the a priori selection of a set of scales to identify the best scales should be encouraged, as a choice is likely to influence the results (McGarigal et al., 2016).

The role of the study area extent on SDMs accuracy has been studied mainly at large scales, comparing models using continental or regional observation datasets (Suárez-Seoane et al., 2013; Vale et al., 2014). The use of data providing information on the full or local range of a species was found to give different results in terms of performances, responses curves and spatial predictions

(Vale et al., 2014). It was also shown that the use of coarse scales when working at a regional extent decrease model performance (Suárez-Seoane et al., 2013). Here, the originality of the Jura study area dataset, describing only part of the environment arises new interrogations regarding the choice of scales. Is it preferable to use data that will show only the variations existing in an overall favorable landscape (here forest patches) thus predicting well the distribution in this habitat type? How to measure the cost of missing external elements influencing the species distribution in such case? This questioning should be address in the first place regarding the aim of the study, but the implications of using restrained and fragmented study areas should be further evaluated.

## 1.3 Predictions

### 1.3.1 Main contributions

Spatial predictions were created for Hazel Grouse and Capercaillie in both study areas and for beech regeneration in the Jura study area. The comparison of Capercaillie predictions over an area common to both LiDAR coverages showed a good transferability of models from the Ain study area to the Jura study area with correlation higher than 0.66 with both PC area-based and OO metrics.

The results at fine scales showed that suitable forest habitat is fragmented for both Hazel Grouse and Capercaillie. At large scales suitable habitat patches may be found at low elevation for Hazel Grouse were the presence or absence of the species is not yet documented. Suitable Capercaillie habitat is mainly restricted to higher elevation in winter models, but summer models showed a wider distribution of suitable habitats. The differences between male and female spatial predictions were low. Nevertheless, in the Ain study area female suitable patches were larger indicating less restrictive needs regarding the habitat characteristics.

High differences in predictions were observed between summer and winter Capercaillie models. The interpretation of the responses curves and performance indexes were in some cases inconsistent as illustrated by the observation of incoherent responses shapes associated with high performances. The causes of these differences observed between models among scales, study areas and sexes are probably due to multiple causes. This overall result demonstrates that improvements in terms of





ecological relevance and prediction accuracy are still to be made for summer models. It also confirms that the indexes (AUC and Boyce) used for the selection of the best model need to be interpreted with caution (Fourcade et al., 2018). This was furthermore highlighted by the difference observed between the result obtained with AUC and Boyce index when evaluating the performances of PC area based and OO metrics.

Beech regeneration distribution was predicted with a high accuracy in one study area showing the presence of large patches within the landscape. Such regeneration patches are overall present in every forest where the Capercaillie is present showing that the effect on the habitat availability might be important. Nevertheless, some forests units like Le Risoux, Le Risol or Le Massacre are observed to be more subject to the invasion of beech regeneration than others like the forest of La Joux or Mont Noir.

### 1.3.2 Perspectives

The decision to select the best model regarding the performances among only one (Capercaillie and Hazel Grouse) to three modeling technics (beech regeneration) may be a major source of variability in the predictions (Garcia et al., 2012). The weaknesses of our modeling framework may be overcome by using ensemble modeling (Guisan et al., 2017). Ensemble modeling consists in producing SDMs using several modeling algorithms with the aim to improve the prediction of the species distribution. Such framework can provide an estimation of the uncertainty causes by the algorithms sensitivity differences of to sample size, environmental variable proprieties, sampling bias, thus leading to a better identification of area suitable for the target species (Meller et al., 2014).

## 1.4 Environmental variables

### 1.4.1 Main contributions

Four main variables contributing to the description of Hazel Grouse habitat were identified: the standard deviation of the canopy density, the proportion of open areas, the penetration ratio between 2-5m and the tree density >10m. Those variables describing the characteristics of both the understory and the canopy were evidenced on both sites at all scales, indicating the importance of

those parameters for Hazel Grouse and joining the observation of previous studies (Rechsteiner et al., 2017b; Sachot et al., 2003a; Zellweger et al., 2013b).

Concerning Capercaillie, the effect of human disturbances illustrated by the variables distance to ski runs was found to have a high impact on the species distribution confirming the results of Coppes et al. (2017) and Thiel et al. (2007, 2008). This result was found in both study area and at all scales. The response was also consistent between models showing an optimal distance to ski runs of 1 km for the bird but this observation is higher than the relative avoidance distance (320m) observed by (Coppes et al., 2017). This observation may indicate that the disturbances are impacting the individuals at a wider scale than initially thought. Also, the presence of an optimum indicates that lower altitudes forests are not favorable for the species (a long distance from a ski resort suggests lower elevation areas). Differences in variable contributions and response shapes were observed. Nevertheless, the importance of a high horizontal (Standard deviation of tree density ( $>20\text{m}$ ); standard deviation of the proportion of gaps ( $25\text{-}200\text{m}^2$ )) and vertical heterogeneity (Tree gini index) associated with a basal area ranging between 20 and  $30\text{ m}^2/\text{ha}$  was shown. The variables used to describe a priori unsuitable habitats, such as the proportion of open areas, were also important in models to predict the distribution with accuracy.

#### 1.4.2 Perspectives

LiDAR extracted metrics whether PC area based or OO are highly valuable when working with forest birds species (Tattoni et al., 2012b). Nevertheless, the habitat structure may not be sufficient to describe all the essential aspects required by a species. Both species are depending on specific tree species as food resources. Thus, an essential aspect of the habitat description may have been missed by using mainly structure variables. The addition of variables describing better the species present by using satellite images or field inventories may further improve models (Bae et al., 2014b).

The characterization of understory with LiDAR data is still to be improved. If by using PC area based, estimation on the density of those layers can be done (see Hazel Grouse and Beech regeneration models), the extraction of understory objects is yet challenging. Nevertheless, recent contributions



on vertical stratification of the LiDAR point cloud showed that such extraction can be done using a dense point cloud (170 points/m<sup>2</sup>) (Hamraz et al., 2017). Another study recently proposed a new lead on LiDAR point-cloud analysis by exploring the use of full variability of the data without making a priori assumptions on structural attributes. Few predictors were created using a principal component analysis and used with success to model red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) distributions (Ciuti et al., 2017). However, if this type of approaches can be used to improve models (for birds but also beech regeneration) the advantages of using OO metrics to improve results transmission to managers will be reduced. The whole potential of LiDAR object-oriented metrics is yet to be explored. In this work the extraction of two types of objects was realized, but other types of objects may be extracted in forest (e.g. shrubs, trails (Ferraz et al., 2016)) or other environment (e.g. buildings (Al-Nahas and Shafri, 2018)) from fine scales (e.g. trees) to coarse scales (e.g. vegetation ensemble (Ruiz et al., 2018)).

The role of disturbances highlighted in winter models may also be further explored. In particular, off-tracks skiing or snowshoeing have recently become more popular, but the quantification of this impact cannot simply be estimated by using information on existing trails and ski runs. The emergence of sport applications with GPS support (like ©Strava) or simply the location of online activity (twitter, flickr, facebook) may soon allow researchers to obtain better information on such activities (frequentation and tracks) (Korpilo et al., 2017; Walden-Schreiner et al., 2018). The frequentation of the forest was also evaluated by the Groupe Tétràs Jura, with the installation of camera traps in popular off-tracks locations in winter and might also be used to quantify this impact in future studies. Summer leisure activities may also be an important factor impacting Capercaillie distributions (Coppes et al., 2017). In particular a finer distinction between hiking trails, forest roads and unfrequented trails (which can even become favorable for broods) may improve summer model accuracy. Indeed, our description of gaps based on the total surface provoked the classification of trails and road as open areas. The addition of a description of the shape of large gaps (long and narrow versus circular type shape, soil curvature) may help to better classified large open areas (Ferraz et al., 2016; White et al., 2010).

## 2 IMPROVING TRANSMISSION TO MANAGERS

---

In this work, the improvement in the results transmission to managers was proposed at two different steps of the SDMs creation. First the choice of the variables type before modeling was done to optimize and facilitate the interpretation of the results without decreasing the model performance and accuracy. The use of object-oriented variables gave promising results for future research on endangered species distribution despite the lack of information on understory and allowed us to produce more transferable SMDs results to managers. The improvement in results transmission was not restricted to the extraction of object from the LiDAR point-cloud, as estimations of variables used in forestry such as the basal area by using vegetation inventories were found to be relevant as well in Capercaillie models. Those results show that the production of accurate models combined with the use of variable reflecting manager's points of view was achieved.

Secondly, three-dimensional response curves visualization was used to improve the comprehension of the variable interactions implications. Indeed, if a variable is important for the species it is thus essential to understand the implication for potential management actions. In the case of Hazel Grouse the usefulness of the interacting variables was limited because the interaction itself did not essentially contribute to the models. Nevertheless, such graphical visualization can be used in other cases even between non interacting variables (e.g. to observe the best combination between understory density and tree density or between the density tree (>10m) and the basal area)) (Zurell et al., 2012).

Those two results show that the improvement of the communication between researchers and managers can be considered from the beginning of the process. In addition, within this work, two reports in French and final predicted maps along with LiDAR extracted metrics were produced as raster files which can be used as a basis for future decisions. Nevertheless, in order to plan future management actions those results which can be qualified as "raw" may need further interpretation. In particular, it could useful to subset the results by forests units or even by communal/private forest



plots as the decisions will mainly be taken within administrative units. The creation of interactive tools (interactive maps or graphics) that can be easily used by everyone may also help the transfer of the results to the target audience. The development of the shiny R package (Chang et al., 2018) which produces interactive web applications is making the creation of such tools easy for people who are already familiar with R.

One major aspect which was not taken into account in Capercaillie models are the economics drivers impacting the bird habitat suitability (Braunisch et al., 2012). Indeed, if beech regeneration does have a strong impact on the Capercaillie habitat, it is not the only driver that leads to habitat degradation. Homogenous or young forest stand and a global dense tree density also constituted unsuitable habitat for the species. Propositions for managing such areas will require to take into account the economic and societal aspects in order to be accepted by municipalities and private owners (Di Minin et al., 2017; Naidoo et al., 2006). It was indeed a point participating to managers incomprehension raised by Addison et al. (2013). Question such as: where can we plan management actions at least cost for owners? How to achieve conservation actions given the limited resources at our disposal? Should connectivity between suitable patches be promoted or improvement of large habitat patches will give better results for Capercaillie population persistence? Those questions cannot be fully answered with the models produced in the project. Thus, those aspects need further discussion with managers in order to propose different possible scenarios regarding the results obtained with Capercaillie models.

One way to further ameliorate future recommendations for management actions in favor of Capercaillie may be to combine the results obtained with the Capercaillie and Beech vegetation models as proposed by Braunisch et al. (2016). New Capercaillie models using as one of the predicting variables the results from the beech regeneration model can be used to find the optimal characteristics for the bird. With this model, an optimal distribution can be predicted by setting the beech regeneration at its optimal value. From this result, the potential expansion of favorable habitat could be investigated by comparing it to the original Capercaillie model.

### 3 LOCAL PERSPECTIVES

---

The results obtained for Hazel grouse will be used as a basis for the implementation of a new survey design detached from Capercaillie surveys. The collection of new data on the species may in the future be used to improve SDMs by giving information on areas that were not surveyed before. In particular, differences in vegetation types and species assemblages that may be different between high and low elevation could help to better estimate the species distribution. Such types of differences were not taken into account in our models for two reasons: 1) Observation data came from higher elevation 2) LiDAR extracted variable could only identify the proportion of coniferous trees and not the tree species.

For Capercaillie, the challenge remains to identify with a high accuracy the areas with high beech regeneration. The collection of more data on beech regeneration, in particular in the Ain study area, could be intended to ameliorate SDMs results. The disparity between the different beech regeneration protocols probably participates in the difficulty to create accurate models and the use of observations on regeneration cover seems to be more efficient than the number of stems. The collection of more precise data on regeneration patches might be necessary (presence of reproducing trees, type of soil, patch size) to better understand what is driving this phenomenon. Another perspective would be to map not only the unsuitable understory habitat but also the most suitable one with a model for the distribution of bilberry.

In addition to vegetation models, as Capercaillie may be more sensitive to the absence of bilberry in summer, the improvement of models for this season may be investigated. The low number of observations in this season, due to a sparse survey effort (driven by the difficulty to detect the signs of presence without a snow cover) is probably at the origin of some ecologically irrelevant results we observed in the Jura study area. The collection of data on the species occurrence in summer may benefit future intent of SDMs for this species. The differences in detection between male and female leading to a highly biased observation dataset in favor of males in winter was also a problem



encountered in this project. The continuation of the genetic survey using systematic sampling will soon give more information on the sex ratio and population dynamics. The role of the lek in population structure and individual occurrences may also benefit those data. The acquisition of genetics data by the Groupe Tétrás Jura opens new perspectives to evaluate the effects of landscape characteristics on individual's dispersal and functional connectivity (Milanesi et al., 2016). As our results showed a fragmentation in small patches of good forest habitats the genetic analysis of distances can help to understand the effects of habitat fragmentation on the populations and to identify elements preventing individual movements such as low quality habitat, ski resort, urbans and open areas.

## 4 CONCLUSION

---

This work proposed reliable SDMs predictions for two endangered forest bird and for an unsuitable habitat type: the beech regeneration. The evaluation of the effect the spatial sampling bias and the different corrections tested allowed us to use a heterogeneous datasets for modeling the distribution of the two target species. While the results confirm the usefulness of LiDAR data to map habitat structure for forest bird, the exploration of the use of LiDAR Object-Oriented extracted metrics for endangered species distribution modeling allowed the creation of more transferable models results to managers. The use of a framework to evaluate the role of scale in models and ultimately select that best one showed that this selection is highly dependent of multiple factors such as the study area location and extent, the environmental variable or as the target species requirements. The results obtained are one more step of a long path toward Hazel Grouse and Capercaillie conservation in the Jura massif and numerous ways to broken the predicted maps down into element for futures actions can now be intent with managers.

# BIBLIOGRAPHY

- Åberg, J., Jansson, G., Swenson, J., and Angelstam, P. (1995). The effect of matrix on the occurrence of hazel grouse (*Bonasa bonasia*) in isolated habitat fragments. *Oecologia* 103, 265–269.
- Aberg, J., Swenson, J.E., and Angelstam, P. (2003). The habitat requirements of hazel grouse (*Bonasa bonasia*) in managed boreal forest and applicability of forest stand descriptions as a tool to identify suitable patches. *For. Ecol. Manag.* 175, 437–444.
- Addison, P.F., Rumpff, L., Bau, S.S., Carey, J.M., Chee, Y.E., Jarrad, F.C., McBride, M.F., and Burgman, M.A. (2013). Practical solutions for making models indispensable in conservation decision-making. *Divers. Distrib.* 19, 490–502.
- Aguirre, A.A., Keefe, T.J., Reif, J.S., Kashinsky, L., Yochem, P.K., Saliki, J.T., Stott, J.L., Goldstein, T., Dubey, J.P., Braun, R., et al. (2007). Infectious disease monitoring of the endangered Hawaiian Monk Seal. *J. Wildl. Dis.* 43, 229–241.
- Al-Nahas, F., and Shafri, H. (2018). Pixel-based and object-oriented classifications of airborne LiDAR and high resolution satellite data for building extraction. (IOP Publishing), p. 12032.
- Andreev, A. (1988). Ecological energetics of Palaearctic Tetraonidae in relation to chemical composition and digestibility of their winter diets. *Can. J. Zool.* 66, 1382–1388.
- Andren, H. (1994). Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos* 355–366.
- Angelieri, C.C.S., Adams-Hosking, C., de Barros, K.M.P.M., de Souza, M.P., and McAlpine, C.A. (2016). Using species distribution models to predict potential landscape restoration effects on puma conservation. *PloS One* 11, e0145232.
- Antrop, M. (2004). Landscape change and the urbanization process in Europe. *Landsc. Urban Plan.* 67, 9–26.
- Araújo, M.B., and Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33, 1677–1688.
- Araújo, M.B., Pearson, R.G., Thuiller, W., and Erhard, M. (2005). Validation of species–climate impact models under climate change. *Glob. Change Biol.* 11, 1504–1513.
- Arcos, J.M., Bécares, J., Villero, D., Brotons, L., Rodríguez, B., and Ruiz, A. (2012). Assessing the location and stability of foraging hotspots for pelagic seabirds: an approach to identify marine Important Bird Areas (IBAs) in Spain. *Biol. Conserv.* 156, 30–42.
- Arkle, R.S., Pilliod, D.S., Hanser, S.E., Brooks, M.L., Chambers, J.C., Grace, J.B., Knutson, K.C., Pyke, D.A., Welty, J.L., and Wirth, T.A. (2014). Quantifying restoration effectiveness using multi-scale habitat models: implications for sage-grouse in the Great Basin. *Ecosphere* 5, 1–32.
- Arlettaz, R., Schaub, M., Fournier, J., Reichlin, T.S., Sierro, A., Watson, J.E., and Braunisch, V. (2010). From publications to public actions: when conservation biologists bridge the gap between research and implementation. *BioScience* 60, 835–842.





- Atkinson, C.T., and LaPointe, D.A. (2009). Introduced avian diseases, climate change, and the future of Hawaiian honeycreepers. *J. Avian Med. Surg.* 23, 53–63.
- Austin, M. (1980). Searching for a model for use in vegetation analysis. *Vegetatio* 42, 11–21.
- Bae, S., Reineking, B., Ewald, M., and Mueller, J. (2014a). Comparison of airborne lidar, aerial photography, and field surveys to model the habitat suitability of a cryptic forest species—the hazel grouse. *Int. J. Remote Sens.* 35, 6469–6489.
- Bae, S., Reineking, B., Ewald, M., and Mueller, J. (2014b). Comparison of airborne lidar, aerial photography, and field surveys to model the habitat suitability of a cryptic forest species—the hazel grouse. *Int. J. Remote Sens.* 35, 6469–6489.
- Baltsavias, E.P. (1999). Airborne laser scanning: existing systems and firms and other resources. *ISPRS J. Photogramm. Remote Sens.* 54, 164–198.
- Bergen, K.M., Goetz, S.J., Dubayah, R.O., Henebry, G.M., Hunsaker, C.T., Imhoff, M.L., Nelson, R.F., Parker, G.G., and Radeloff, V.C. (2009). Remote sensing of vegetation 3-D structure for biodiversity and habitat: Review and implications for lidar and radar spaceborne missions. *J. Geophys. Res.-Biogeosciences* 114, G00E06–G00E06.
- Bolte, A., Czajkowski, T., and Kompa, T. (2007). The north-eastern distribution range of European beech—a review. *Forestry* 80, 413–429.
- Bradley, B.A., and Fleishman, E. (2008). Can remote sensing of land cover improve species distribution modelling? *J. Biogeogr.* 35, 1158–1159.
- Braunisch, V., Home, R., Pellet, J., and Arlettaz, R. (2012). Conservation science relevant to action: A research agenda identified and prioritized by practitioners. *Biol. Conserv.* 153, 201–210.
- Braunisch, V., Patthey, P., and Arlettaz, R. (2016). Where to Combat Shrub Encroachment in Alpine Timberline Ecosystems: Combining Remotely-Sensed Vegetation Information with Species Habitat Modelling. *PLoS ONE* 11, e0164318.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.
- Broome, A., Connolly, T., and Quine, C.P. (2014). An evaluation of thinning to improve habitat for capercaillie (*Tetrao urogallus*). *For. Ecol. Manag.* 314, 94–103.
- Brotons, L., Thuiller, W., Araújo, M.B., and Hirzel, A.H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27, 437–448.
- Buffet, N., and Dumont-Dayot, E. (2013). Bird collisions with overhead ski-cables: a reducible source of mortality. *Impacts Ski. Relat. Winter Recreat. Act. Mt. Environ.* 123.
- Cardador, L., Diaz-luque, J.A., Hiraldo, F., Gilardy, J.D., and Tella, J.L. (2017). The effects of spatial survey bias and habitat suitability on predicting the distribution of threatened species living in remote areas. *Bird Conserv. Int.* 1–12.
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2018). shiny: Web Application Framework for R.

- Ciuti, S., Tripke, H., Antkowiak, P., Gonzalez, R.S., Dormann, C.F., and Heurich, M. (2017). An efficient method to exploit LiDAR data in animal ecology. *Methods Ecol. Evol.* n/a-n/a.
- Clark, M.L., Clark, D.B., and Roberts, D.A. (2004). Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sens. Environ.* *91*, 68–89.
- Clavero, M., and García-Berthou, E. (2005). Invasive species are a leading cause of animal extinctions. *Trends Ecol. Evol.* *20*, 110.
- Conde, D.A., Colchero, F., Zarza, H., Christensen Jr., N.L., Sexton, J.O., Manterola, C., Chávez, C., Rivera, A., Azuara, D., and Ceballos, G. (2010). Sex matters: Modeling male and female habitat differences for jaguar conservation. *Biol. Conserv.* *143*, 1980–1988.
- Coppes, J., Ehrlacher Judith, Thiel Dominik, Suchant Rudi, and Braunisch Veronika (2017). Outdoor recreation causes effective habitat reduction in capercaillie *Tetrao urogallus*: a major threat for geographically restricted populations. *J. Avian Biol.* *48*, 1583–1594.
- Crowley, T.J. (2000). Causes of climate change over the past 1000 years. *Science* *289*, 270–277.
- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., and Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology* *88*, 2783–2792.
- Dassot, M., Constant, T., and Fournier, M. (2011). The use of terrestrial LiDAR technology in forest science: application fields, benefits and challenges. *Ann. For. Sci.* *68*, 959–974.
- Di Minin, E., Soutullo, A., Bartesaghi, L., Rios, M., Szephegyi, M.N., and Moilanen, A. (2017). Integrating biodiversity, ecosystem services and socio-economic data to identify priority areas and landowners for conservation actions at the national scale. *Biol. Conserv.* *206*, 56–64.
- Duriez, O., Sachet, J.-M., Ménoni, E., Pidancier, N., Miquel, C., and Taberlet, P. (2007). Phylogeography of the capercaillie in Eurasia: what is the conservation status in the Pyrenees and Cantabrian Mounts? *Conserv. Genet.* *8*, 513–526.
- Easterling, D.R., Meehl, G.A., Parmesan, C., Changnon, S.A., Karl, T.R., and Mearns, L.O. (2000). Climate extremes: observations, modeling, and impacts. *Science* *289*, 2068–2074.
- Edwards Jr, T.C., Cutler, D.R., Zimmermann, N.E., Geiser, L., and Moisen, G.G. (2006). Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecol. Model.* *199*, 132–141.
- Elith, J., and Leathwick, J.R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* *40*, 677.
- Elith, J., Kearney, M., and Phillips, S. (2010). The art of modelling range-shifting species. *Methods Ecol. Evol.* *1*, 330–342.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., and Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* *17*, 43–57.
- Fabianek, F., Simard, M.A., Racine, E.B., and Desrochers, A. (2015). Selection of roosting habitat by male *Myotis* bats in a boreal forest. *Can. J. Zool.* *93*, 539–546.



- Ferraz, A., Mallet, C., and Chehata, N. (2016). Large-scale road detection in forested mountainous areas using airborne topographic lidar data. *ISPRS J. Photogramm. Remote Sens.* *112*, 23–36.
- Ficetola, G.F., Thuiller, W., and Miaud, C. (2007). Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. *Divers. Distrib.* *13*, 476–485.
- Fithian, W., Elith, J., Hastie, T., and Keith, D.A. (2015a). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol. Evol.* *6*, 424–438.
- Fithian, W., Elith, J., Hastie, T., and Keith, D.A. (2015b). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol. Evol.* *6*, 424–438.
- Flaspohler, D.J., Giardina, C.P., Asner, G.P., Hart, P., Price, J., Lyons, C.K., and Castaneda, X. (2010). Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biol. Conserv.* *143*, 280–288.
- Fourcade, Y., Besnard, A.G., and Secondi, J. (2018). Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob. Ecol. Biogeogr.* *27*, 245–256.
- Franklin, J. (2009a). Moving beyond static species distribution models in support of conservation biogeography. *Divers. Distrib.* *16*, 321–330.
- Franklin, J. (2009b). Mapping species distributions: spatial inference and prediction (Cambridge University Press).
- Froidevaux, J.S.P., Zellweger, F., Bollmann, K., Jones, G., and Obrist, M.K. (2016). From field surveys to LiDAR: Shining a light on how bats respond to forest structure. *Remote Sens. Environ.* *175*, 242–250.
- Garcia, R.A., Burgess, N.D., Cabeza, M., Rahbek, C., and Araújo, M.B. (2012). Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Glob. Change Biol.* *18*, 1253–1269.
- Gerber, E., Krebs, C., Murrell, C., Moretti, M., Rocklin, R., and Schaffner, U. (2008). Exotic invasive knotweeds (*Fallopia* spp.) negatively affect native plant and invertebrate assemblages in European riparian habitats. *Biol. Conserv.* *141*, 646–654.
- Gjerde, I., and Wegge, P. (1989a). Spacing pattern, habitat use and survival of capercaillie in a fragmented winter habitat. *Ornis Scand.* 219–225.
- Gjerde, I., and Wegge, P. (1989b). Spacing pattern, habitat use and survival of capercaillie in a fragmented winter habitat. *Ornis Scand.* 219–225.
- Godsoe, W., Franklin, J., and Blanchet, F.G. (2017). Effects of biotic interactions on modeled species' distribution can be masked by environmental gradients. *Ecol. Evol.* *7*, 654–664.
- Goetz, S., Steinberg, D., Dubayah, R., and Blair, B. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* *108*, 254–263.

- Gosselin, F., and Callois, J.-M. (2018). Relationships between human activity and biodiversity in Europe at the national scale: Spatial density of human activity as a core driver of biodiversity erosion. *Ecol. Indic.* *90*, 356–365.
- Graf, R.F., Bollmann, K., Suter, W., and Bugmann, H. (2005). The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. *Landsc. Ecol.* *20*, 703–717.
- Graf, R.F., Mathys, L., and Bollmann, K. (2009a). Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps. *For. Ecol. Manag.* *257*, 160–167.
- Graf, R.F., Mathys, L., and Bollmann, K. (2009b). Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps. *For. Ecol. Manag.* *257*, 160–167.
- Guillera-Arroita, G. (2016). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* *40*, 281–295.
- Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., and Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Glob. Ecol. Biogeogr.* *24*, 276–292.
- Guisan, A., and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecol. Model.* *135*, 147–186.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., and Zimmermann, N.E. (2006). Using niche-based models to improve the sampling of rare species. *Conserv. Biol.* *20*, 501–511.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., et al. (2013). Predicting species distributions for conservation decisions. *Ecol. Lett.* *16*, 1424–1435.
- Guisan, A., Thuiller, W., and Zimmermann, N.E. (2017). *Habitat suitability and distribution models: with applications in R* (Cambridge University Press).
- Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., and Collins, C.D. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* *1*, e1500052.
- Hamraz, H., Contreras, M.A., and Zhang, J. (2017). Vertical stratification of forest canopy for segmentation of understory trees within small-footprint airborne LiDAR point clouds. *ISPRS J. Photogramm. Remote Sens.* *130*, 385–392.
- He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., and Pettorelli, N. (2015). Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.* *1*, 4–18.
- Hijmans, R.J., and van Etten, J. (2014). *raster: Geographic data analysis and modeling. R Package Version 2.*
- Hill, R.A., and Broughton, R.K. (2009). Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne LiDAR data: A case study in lowland Britain. *Isprs J. Photogramm. Remote Sens.* *64*, 223–233.



- Hinsley, S.A., Hill, R.A., Bellamy, P.E., Harrison, N.M., Speakman, J.R., Wilson, A.K., and Ferns, P.N. (2008). Effects of structural and functional habitat gaps on breeding woodland birds: working harder for less. *Landsc. Ecol.* 23, 615–626.
- Hirzel, A., and Guisan, A. (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecol. Model.* 157, 331–341.
- Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., and Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* 199, 142–152.
- Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., and Caldeira, K. (2007). Coral reefs under rapid climate change and ocean acidification. *Science* 318, 1737–1742.
- Huhta, E., Helle, P., Nivala, V., and Nikula, A. (2017). The effect of human-modified landscape structure on forest grouse broods in two landscape types. *Ecosphere* 8.
- Imperio, S., Bionda, R., Viterbi, R., and Provenza, A. (2013). Climate change and human disturbance can lead to local extinction of Alpine rock ptarmigan: New insight from the Western Italian Alps. *PLoS One* 8, e81598.
- Isaac, N.J., and Pocock, M.J. (2015). Bias and information in biological records. *Biol. J. Linn. Soc.* 115, 522–531.
- Jahren, T., Storaas, T., Willebrand, T., Moa, P.F., and Hagen, B.-R. (2016). Declining reproductive output in capercaillie and black grouse—16 countries and 80 years. *Anim. Biol.* 66, 363–400.
- Johnson, D.H. (1980). The comparison of usage and availability measurements for evaluating resource preference. *Ecology* 61, 65–71.
- Johnson, C.J., and Gillingham, M.P. (2005). An evaluation of mapped species distribution models used for conservation planning. *Environ. Conserv.* 32, 117–128.
- Kearney, M. (2006). Habitat, environment and niche: what are we modelling? *Oikos* 115, 186–191.
- Klingberg, J., Konarska, J., Lindberg, F., Johansson, L., and Thorsson, S. (2017). Mapping leaf area of urban greenery using aerial LiDAR and ground-based measurements in Gothenburg, Sweden. *Spec. Featur.* 26, 31–40.
- Knight, A.T., Cowling, R.M., Rouget, M., Balmford, A., Lombard, A.T., and Campbell, B.M. (2008). Knowing but not doing: selecting priority conservation areas and the research–implementation gap. *Conserv. Biol.* 22, 610–617.
- Korpilo, S., Virtanen, T., and Lehtväärtä, S. (2017). Smartphone GPS tracking—Inexpensive and efficient data collection on recreational movement. *Landsc. Urban Plan.* 157, 608–617.
- Koshkina, V., Wang, Y., Gordon, A., Dorazio, R.M., White, M., and Stone, L. (2017). Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. *Methods Ecol. Evol.* 8, 420–430.

- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., and Augeri, D.M. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366–1379.
- Kremen, C., Cameron, A., Moilanen, A., Phillips, S., Thomas, C., Beentje, H., Dransfield, J., Fisher, B., Glaw, F., and Good, T. (2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science* 320, 222–226.
- Kuhn, M. (2008). Caret package. *J. Stat. Softw.* 28, 1–26.
- Kurki, S., Nikula, A., Helle, P., and Linden, H. (2000). Landscape fragmentation and forest composition effects on grouse breeding success in boreal forests. *Ecology* 81, 1985–1997.
- Lahoz-Monfort, J.J., Guillera-Arroita, G., and Wintle, B.A. (2014). Imperfect detection impacts the performance of species distribution models. *Glob. Ecol. Biogeogr.* 23, 504–515.
- Latifi, H., Heurich, M., Hartig, F., Müller, J., Krzystek, P., Jehl, H., and Dech, S. (2015). Estimating over-and understorey canopy density of temperate mixed stands by airborne LiDAR data. *For. Int. J. For. Res.* 89, 69–81.
- Leclercq, B. (1985). Influence des routes et voies de pénétration humaine sur les comportements de grands tétras et de gélinottes dans le Haut-Jura français.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., and Harding, D.J. (2002). Lidar Remote Sensing for Ecosystem Studies Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest to forest, landscape, and global ecologists. *BioScience* 52, 19–30.
- Lengyel, S., Kosztyi, B., Schmeller, D.S., Henry, P.-Y., Kotarac, M., Lin, Y.-P., and Henle, K. (2018). Evaluating and benchmarking biodiversity monitoring: Metadata-based indicators for sampling design, sampling effort and data analysis. *Ecol. Indic.* 85, 624–633.
- Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., and Wegmann, M. (2012). Modelling forest  $\alpha$ -diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. *Remote Sens.* 4, 2818–2845.
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.
- Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., et al. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Adapt. For. For. Manag. Chang. Clim.* 259, 698–709.
- Mac Nally, R. (2000). Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodivers. Conserv.* 9, 655–671.
- Mallet, C., and Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *ISPRS J. Photogramm. Remote Sens.* 64, 1–16.



- Marcer, A., Sáez, L., Molowny-Horas, R., Pons, X., and Pino, J. (2013). Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. *Biol. Conserv.* **166**, 221–230.
- Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., and Vierling, K.T. (2009). Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens. Environ.* **113**, 2533–2546.
- Mateo-Sánchez, M.C., Gastón, A., García-Viñas, J.I., Cuevas, J., López-Leiva, C., Fernández-Landa, A., Algeet-Abarquero, N., Marchamalo, M., Fortin, M.-J., and Saura, S. (2016). Seasonal and temporal changes in species use of the landscape: how do they impact the inferences from multi-scale habitat modeling? *Landsc. Ecol.* **31**, 1261–1276.
- Mayor, S.J., Schneider, D.C., Schaefer, J.A., and Mahoney, S.P. (2009). Habitat selection at multiple scales. *Ecoscience* **16**, 238–247.
- McGarigal, K., Wan, H.Y., Zeller, K.A., Timm, B.C., and Cushman, S.A. (2016). Multi-scale habitat selection modeling: a review and outlook. *Landsc. Ecol.* **31**, 1161–1175.
- Medina, F.M., Bonnaud, E., Vidal, E., Tershy, B.R., Zavaleta, E.S., Josh Donlan, C., Keitt, B.S., Corre, M., Horwath, S.V., and Nogales, M. (2011). A global review of the impacts of invasive cats on island endangered vertebrates. *Glob. Change Biol.* **17**, 3503–3510.
- Melin, M., Mehtätalo, L., Miettinen, J., Tossavainen, S., and Packalen, P. (2016). Forest structure as a determinant of grouse brood occurrence – An analysis linking LiDAR data with presence/absence field data. *Spec. Sect. Drought US For. Impacts Potential Manag. Responses* **380**, 202–211.
- Meller, L., Cabeza, M., Pironon, S., Barbet-Massin, M., Maiorano, L., Georges, D., and Thuiller, W. (2014). Ensemble distribution models in conservation prioritization: from consensus predictions to consensus reserve networks. *Divers. Distrib.* **20**, 309–321.
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., and Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**, 1267–1281.
- Meyer, C.B., and Thuiller, W. (2006). Accuracy of resource selection functions across spatial scales. *Divers. Distrib.* **12**, 288–297.
- Milanesi, P., Holderegger, R., Bollmann, K., Gugerli, F., and Zellweger, F. (2016). Three-dimensional habitat structure and landscape genetics: a step forward in estimating functional connectivity. *Ecology*.
- Miller, J.A. (2014). Virtual species distribution models: Using simulated data to evaluate aspects of model performance. *Prog. Phys. Geogr.* **38**, 117–128.
- Milling, C.R., Rachlow, J.L., Olsoy, P.J., Chappell, M.A., Johnson, T.R., Forbey, J.S., Shipley, L.A., and Thornton, D.H. (2018). Habitat structure modifies microclimate: an approach for mapping fine-scale thermal refuge. *Methods Ecol. Evol.*
- Moilanen, A., Leathwick, J.R., and Quinn, J.M. (2011). Spatial prioritization of conservation management. *Conserv. Lett.* **4**, 383–393.

- Montadert, M., and Léonard, P. (2006). Post-juvenile dispersal of Hazel Grouse *Bonasa bonasia* in an expanding population of the southeastern French Alps. *Ibis* **148**, 1–13.
- Moss, R. (1986). Rain, breeding success and distribution of capercaillie *Tetrao urogallus* and black grouse *Tetrao tetrix* in Scotland. *Ibis* **128**, 65–72.
- Moss, R., Oswald, J., and Baines, D. (2001). Climate change and breeding success: decline of the capercaillie in Scotland. *J. Anim. Ecol.* **70**, 47–61.
- Müller, D., Schröder, B., and Müller, J. (2009). Modelling habitat selection of the cryptic Hazel Grouse *Bonasa bonasia* in a montane forest. *J. Ornithol.* **150**, 717–732.
- Müller, J., Stadler, J., and Brandl, R. (2010). Composition versus physiognomy of vegetation as predictors of bird assemblages: The role of lidar. *Remote Sens. Environ.* **114**, 490–495.
- Müller, J., Brandl, R., Brändle, M., Förster, B., de Araujo, B.C., Gossner, M.M., Ladas, A., Wagner, M., Maraun, M., and Schall, P. (2017). LiDAR-derived canopy structure supports the more-individuals hypothesis for arthropod diversity in temperate forests. *Oikos*.
- Munoz, A., Bock, J., Monnet, J., Renaud, J., Jolly, A., and Riond, C. (2015). Évaluation par validation indépendante des prédictions des paramètres forestiers réalisées à partir de données LiDAR aéroporté. *Rev. Fr. Photogrammétrie Télédétection* **211**, 81–92.
- Muratorio, S., Klein, E., and Gaüzere, J. (2012). Variations individuelles du régime de reproduction et voisinage écologique des arbres-mères chez des populations de hêtre commun. p. np.
- Naidoo, R., Balmford, A., Ferraro, P.J., Polasky, S., Ricketts, T.H., and Rouget, M. (2006). Integrating economic costs into conservation planning. *Trends Ecol. Evol.* **21**, 681–687.
- Nijland, W., Nielsen, S.E., Coops, N.C., Wulder, M.A., and Stenhouse, G.B. (2014). Fine-spatial scale predictions of understory species using climate-and LiDAR-derived terrain and canopy metrics. *J. Appl. Remote Sens.* **8**, 83572.
- Nin, S., Petrucci, W.A., Del Bubba, M., Ancillotti, C., and Giordani, E. (2017). Effects of environmental factors on seed germination and seedling establishment in bilberry (*Vaccinium myrtillus* L.). *Sci. Hortic.* **226**, 241–249.
- Oatway, M., and Morris, D. (2007). Do animals select habitat at small or large scales? An experiment with meadow voles (*Microtus pennsylvanicus*). *Can. J. Zool.* **85**, 479–487.
- Oja, R., Zilmer, K., and Valdmann, H. (2015). Spatiotemporal effects of supplementary feeding of wild boar (*Sus scrofa*) on artificial ground nest depredation. *PloS One* **10**, e0135254.
- P Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J Hijmans, R., Huettmann, F., R Leathwick, J., Lehmann, A., Li, J., and G Lohmann, L. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151.
- Palmeri, S., Powell, G.V., Asner, G.P., and Peres, C.A. (2012). LiDAR measurements of canopy structure predict spatial distribution of a tropical mature forest primate. *Remote Sens. Environ.* **127**, 98–105.





- Pavlacky Jr, D.C., Lukacs, P.M., Blakesley, J.A., Skorkowsky, R.C., Klute, D.S., Hahn, B.A., Dreitz, V.J., George, T.L., and Hanni, D.J. (2017). A statistically rigorous sampling design to integrate avian monitoring and management within Bird Conservation Regions. *PloS One* 12, e0185924.
- Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., and Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., and Blair, M.E. (2017). Opening the black box: an open-source release of Maxent. *Ecography* n/a-n/a.
- Prokopenko, C.M., Boyce, M.S., and Avgar, T. (2017). Extent-dependent habitat selection in a migratory large herbivore: road avoidance across scales. *Landsc. Ecol.* 32, 313–325.
- Pulliam, H.R. (2000). On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- Purkis, S.J. (2018). Remote Sensing Tropical Coral Reefs: The View from Above. *Annu. Rev. Mar. Sci.*
- Räty, M. (1979). Effect of highway traffic on tetraonid densities. *Ornis Fenn.* 56, 169–170.
- Rechsteiner, C., Zellweger, F., Gerber, A., Breiner, F.T., and Bollmann, K. (2017a). Remotely sensed forest habitat structures improve regional species conservation. *Remote Sens. Ecol. Conserv.* n/a-n/a.
- Rechsteiner, C., Zellweger, F., Gerber, A., Breiner, F.T., and Bollmann, K. (2017b). Remotely sensed forest habitat structures improve regional species conservation. *Remote Sens. Ecol. Conserv.* n/a-n/a.
- Reddy, S., and Dávalos, L.M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *J. Biogeogr.* 30, 1719–1727.
- Rhim, S.-J. (2006). Home range and habitat selection of hazel grouse *Bonasa bonasia* in a temperate forest of South Korea. *For. Ecol. Manag.* 226, 22–25.
- Roussel, J.-R., and Auty, D. (2016). lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 1.0.0. <https://github.com/Jean-Romain/lidR>.
- Ruiz, L.Á., Recio, J.A., Crespo-Peremarch, P., and Sapena, M. (2018). An object-based approach for mapping forest structural types based on low-density LiDAR and multispectral imagery. *Geocarto Int.* 33, 443–457.
- Sachot, Perrin, N., and Neet, C. (2003a). Winter habitat selection by two sympatric forest grouse in western Switzerland: implications for conservation. *Biol. Conserv.* 112, 373–382.
- Sachot, S., Perrin, N., and Neet, C. (2003b). Winter habitat selection by two sympatric forest grouse in western Switzerland: implications for conservation. *Biol. Conserv.* 112, 373–382.

- Schadt, S., Revilla, E., Wiegand, T., Knauer, F., Kaczensky, P., Breitenmoser, U., Bufka, L., Červený, J., Koubek, P., and Huber, T. (2002). Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. *J. Appl. Ecol.* 39, 189–203.
- Schatt, J. (1993). Étude du régime alimentaire de la Gelinotte des bois *Bonasa bonasia* dans l'Ain. *Alauda*.
- Schäublin, S., and Bollmann, K. (2011). Winter habitat selection and conservation of Hazel Grouse (*Bonasa bonasia*) in mountain forests. *J. Ornithol.* 152, 179–192.
- Scridel, D., Brambilla, M., Martin, K., Lehtikainen, A., Iemma, A., Matteo, A., Jähnig, S., Caprio, E., Bogliani, G., and Pedrini, P. (2018). A review and meta-analysis of the effects of climate change on Holarctic mountain and upland bird populations. *Ibis*.
- Segelbacher, G., Storch, I., and Tomiuk, J. (2003). Genetic evidence of capercaillie *Tetrao urogallus* dispersal sources and sinks in the Alps. *Wildl. Biol.* 9, 267–273.
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10, 1115–1123.
- Stommel, H. (1963). Varieties of oceanographic experience. *Science* 139, 572–576.
- Storch, I. (1993a). Patterns and strategies of winter habitat selection in alpine capercaillie. *Ecography* 16, 351–359.
- Storch, I. (1993b). Habitat selection by capercaillie in summer and autumn: Is bilberry important? *Oecologia* 95, 257–265.
- Storch, I. (1993c). Habitat selection by capercaillie in summer and autumn: Is bilberry important? *Oecologia* 95, 257–265.
- Storch, I. (1994). Habitat and survival of capercaillie *Tetrao urogallus* nests and broods in the Bavarian Alps. *Biol. Conserv.* 70, 237–243.
- Storch, I. (1995). Annual home ranges and spacing patterns of capercaillie in central Europe. *J. Wildl. Manag.* 392–400.
- Storch, I. (2007). Grouse, Status Survey and Conservation Action Plan 2006-2010.
- Su, J.G., and Bork, E.W. (2007). Characterization of diverse plant communities in Aspen Parkland rangeland using LiDAR data. *Appl. Veg. Sci.* 10, 407–416.
- Suárez-Seoane, S., Virgós, E., Terroba, O., Pardavila, X., and Barea-Azcón, J.M. (2013). Scaling of species distribution models across spatial resolutions and extents along a biogeographic gradient. The case of the Iberian mole *Talpa occidentalis*. *Ecography* 37, 279–292.
- Tattoni, C., Rizzolli, F., and Pedrini, P. (2012a). Can LiDAR data improve bird habitat suitability models? *Ecol. Model.* 245, 103–110.
- Tattoni, C., Rizzolli, F., and Pedrini, P. (2012b). Can LiDAR data improve bird habitat suitability models? *Ecol. Model.* 245, 103–110.



- Teuscher, M., Brandl, R., Foerster, B., Hothorn, T., Roesner, S., and Mueller, J. (2013). Forest inventories are a valuable data source for habitat modelling of forest species: an alternative to remote-sensing data. *Forestry* 86, 241–253.
- Thiel, D., Ménoni, E., BRENOT, J., and Jenni, L. (2007). Effects of recreation and hunting on flushing distance of capercaillie. *J. Wildl. Manag.* 71, 1784–1792.
- Thiel, D., Jenni-Eiermann, S., Braunisch, V., Palme, R., and Jenni, L. (2008). Ski tourism affects habitat use and evokes a physiological stress response in capercaillie *Tetrao urogallus*: a new methodological approach. *J. Appl. Ecol.* 45, 845–853.
- Vale, C.G., Tarroso, P., and Brito, J.C. (2014). Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara–Sahel transition zone. *Divers. Distrib.* 20, 20–33.
- Varela, S., Anderson, R.P., García-Valdés, R., and Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography* 37, 1084–1091.
- Villero, D., Pla, M., Camps, D., Ruiz-Olmo, J., and Brotons, L. (2017). Integrating species distribution modelling into decision-making to inform conservation actions. *Biodivers. Conserv.* 1–21.
- Vincent, P., and Haworth, J. (1983). Poisson regression models of species abundance. *J. Biogeogr.* 153–160.
- Vitousek, P.M., Mooney, H.A., Lubchenco, J., and Melillo, J.M. (1997). Human domination of Earth's ecosystems. *Science* 277, 494–499.
- Vittoz, P. (1998). Flore et végétation du parc jurassien: typologie, écologie et dynamique des milieux.
- Vogeler, J.C., Hudak, A.T., Vierling, L.A., Evans, J., Green, P., and Vierling, K.I.T. (2014). Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. *Remote Sens. Environ.* 147, 13–22.
- Walden-Schreiner, C., Leung, Y.-F., and Tateosian, L. (2018). Digital footprints: Incorporating crowdsourced geographic information for protected area management. *Appl. Geogr.* 90, 44–54.
- Wegge, P., and Kastedalen, L. (2007). Pattern and causes of natural mortality of capercaillie, *Tetrao urogallus*, chicks in a fragmented boreal forest. (JSTOR), pp. 141–151.
- Wegge, P., and Kastedalen, L. (2008). Habitat and diet of young grouse broods: resource partitioning between Capercaillie (*Tetrao urogallus*) and Black Grouse (*Tetrao tetrix*) in boreal forests. *J. Ornithol.* 149, 237–244.
- Wegge, P., and Rolstad, J. (1986). Size and spacing of capercaillie leks in relation to social behavior and habitat. *Behav. Ecol. Sociobiol.* 19, 401–408.
- Weisberg, P.J., Dilts, T.E., Becker, M.E., Young, J.S., Wong-Kone, D.C., Newton, W.E., and Ammon, E.M. (2014). Guild-specific responses of avian species richness to LiDAR-derived habitat heterogeneity. *Acta Oecologica-Int. J. Ecol.* 59, 72–83.

White, R.A., Dietterick, B.C., Mastin, T., and Strohman, R. (2010). Forest roads mapped using LiDAR in steep forested terrain. *Remote Sens.* 2, 1120–1141.

Wiens, J.A. (1989). Spatial scaling in ecology. *Funct. Ecol.* 3, 385–397.

Wiles, G.J., Bart, J., Beck, R.E., and Aguon, C.F. (2003). Impacts of the brown tree snake: patterns of decline and species persistence in Guam's avifauna. *Conserv. Biol.* 17, 1350–1360.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., and Veran, S. (2013). Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4, 236–243.

Zellweger, F., Braunisch, V., Baltensweiler, A., and Bollmann, K. (2013a). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *For. Ecol. Manag.* 307, 303–312.

Zellweger, F., Braunisch, V., Baltensweiler, A., and Bollmann, K. (2013b). Remotely sensed forest structural complexity predicts multi species occurrence at the landscape scale. *For. Ecol. Manag.* 307, 303–312.

Zellweger, F., Morsdorf, F., Purves, R.S., Braunisch, V., and Bollmann, K. (2014). Improved methods for measuring forest landscape structure: LiDAR complements field-based habitat assessment. *Biodivers. Conserv.* 23, 289–307.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., and Schröder, B. (2010). The virtual ecologist approach: simulating data and observers. *Oikos* 119, 622–635.

Zurell, D., Elith, J., and Schröder, B. (2012). Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Divers. Distrib.* 18, 628–634.

Zwolak, R., Bogdziewicz, M., Wróbel, A., and Crone, E.E. (2016). Advantages of masting in European beech: timing of granivore satiation and benefits of seed caching support the predator dispersal hypothesis. *Oecologia* 180, 749–758.

# ANNEXES

## 1 CONTENT

---

<b><u>ANNEX I. RESUME EN FRANÇAIS DES CHAPITRES.....</u></b>	<b><u>266</u></b>
<b><u>ANNEX III. LIDAR METRICS CODE .....</u></b>	<b><u>276</u></b>
<b><u>ANNEX IV. SUPPLEMENTARY MATERIAL ARTICLE 1 : DISTRIBUTION MODELS OF DIMORPHIC SPECIES: CORRECTING SPATIAL SAMPLING BIAS IMPROVES ESTIMATES OF HABITAT ASSOCIATION AND SEX RATIO .....</u></b>	<b><u>289</u></b>
<b><u>ANNEX V. SUPPLEMENTARY MATERIAL ARTICLE 2: EFFECT OF INTERACTION TERMS ON SPECIES DISTRIBUTION MODELS PREDICTIONS INTERPRETATION AT MULTIPLE SCALES. ....</u></b>	<b><u>293</u></b>
<b><u>ANNEX VI. SUPPLEMENTARY MATERIAL ARTICLE 3: MAPPING AND MANAGING BIRD HABITAT WITH LIDAR-BASED MULTI-SCALE OBJECT-ORIENTED INDICATORS.....</u></b>	<b><u>302</u></b>
<b><u>ANNEX VII. FRENCH REPORT ON HAZEL GROUSE DISTRIBUTIONS MODELS.....</u></b>	<b><u>306</u></b>
<b><u>ANNEX IX. FRENCH REPORT ON CAPERCAILLIE DISTRIBUTION MODELS.....</u></b>	<b><u>351</u></b>
<b><u>ANNEX X. FILES PRODUCED .....</u></b>	<b><u>411</u></b>
<b><u>ANNEX XI. PRESENTATIONS .....</u></b>	<b><u>411</u></b>
<b><u>ANNEX XII. OUTREACH .....</u></b>	<b><u>411</u></b>

## ANNEX I. RÉSUMÉ EN FRANÇAIS DES CHAPITRES

### RÉSUMÉ DU CHAPITRE 1 : INTRODUCTION

Les impacts des activités humaines ont des conséquences dramatiques sur les écosystèmes, altérant leur intégrité et fonctionnalité durablement (Vitousek et al. 1997). Dans le contexte d'une érosion globale de la biodiversité des actions pour la gestion durable des écosystèmes sont nécessaires. La mise en place des mesures de gestion est un défi car elles doivent protéger efficacement les milieux et les espèces des nombreux impacts (fragmentation de l'habitat, espèces invasives, changement climatique), tout en maintenant l'utilisation de ces milieux par les populations (agriculture, sylviculture, activités de loisirs). La mise en place de ces actions est particulièrement importante dans le cas de la protection d'espèces en danger d'extinction qui sont caractérisées par de petites populations. Ainsi la rapidité et l'efficacité des mesures est essentielle pour assurer avec succès la conservation de ces espèces, mais les types (long ou court terme) et les localisations (surface, emplacement) des actions doivent être évalués en fonction des coûts en terme d'argent et de temps (Moilanen, Leathwick, and Quinn 2011). Ainsi, la difficulté de prioriser les zones pour effectuer des actions en faveur de la conservation d'espèces est un problème inhérent à chaque projet et qui est au cœur de ce travail de thèse.

Le futur de deux espèces emblématiques des forêts du massif Jurassien (France), le Grand Tétras (*Tetrao urogallus*) et la Gelinotte des bois (*Bonasa bonasia*) dépend, pour la première, des futures actions mises en place pour la conservation de son habitat et, pour la seconde, d'une meilleure connaissance de la distribution des populations. Les deux espèces sont menacées localement par de nombreux facteurs, notamment par la perte et la fragmentation de leur habitat, ainsi que par les perturbations causées par les activités humaines (ski et randonnée notamment). Ces espèces sont inféodées à des habitats forestiers constitués de strates verticales et horizontales hétérogènes correspondant à des besoins qui varient au fil des saisons. La présence de myrtilles dans le sous-bois est aussi un élément essentiel de l'habitat du Grand Tétras, lui permettant en particulier de faire des réserves avant l'hiver. Cependant, la récente progression de tâches de régénération du hêtre,

obstruant la lumière pour la végétation au sol diminue la disponibilité d'habitat où la myrtille est présente. Les pratiques forestières ont une grande influence sur la composition en espèces et sur la structure des forêts, ainsi l'adaptation des pratiques forestières aux besoins vitaux de chaque espèce peut permettre le maintien des populations dans le massif jurassien.

Les modèles de distribution d'espèces (SDMs) sont largement utilisés en appui aux mesures de conservation (Franklin 2009a). Ces modèles décrivent les relations existant entre les observations d'une espèce et des variables environnementales, permettant d'obtenir les réponses de l'espèce à son environnement et de prédire spatialement la qualité de l'habitat. L'utilisation toujours plus large des SDMs est couplée avec l'émergence de la télédétection qui permet d'obtenir des caractéristiques du paysage à large échelle et à très fine résolution. Ces dernières années, de nombreuses études ont montré la pertinence de l'utilisation de la technique LiDAR (Light Detection And Ranging) pour extraire des variables d'environnement dans les milieux forestiers (Bae et al. 2014; Rechsteiner et al. 2017). Le survol LiDAR des forêts où sont présentes les deux espèces (Jura, Doubs, Ain), rend possible l'extraction de variables d'environnement permettant de connaître précisément la structure tridimensionnelle du couvert végétal.

Cependant, si les SDMs ont déjà été utilisés avec succès pour la mise en place de mesures de conservation (création d'une zone protégée), la grande majorité des résultats apportés par les publications scientifiques ne sont pas utilisés par les gestionnaires dans la mise en place des mesures de conservation (Guisan et al. 2013). Ce fossé entre les chercheurs et les gestionnaires est bien connu et provient d'une incompréhension mutuelle, d'une part des besoins des gestionnaires et, d'autre part, de la difficulté à interpréter et utiliser les résultats obtenus par les chercheurs par ces derniers. Ces difficultés sont liées aux différences entre la vision du terrain et des problématiques locales par les gestionnaires et les types de résultat pouvant être obtenu avec des SDMs. La complexité des modèles et les incertitudes liées aux résultats sont aussi un frein à leur utilisation. Ainsi pour assurer l'utilisation des SDMs dans le cas du Grand Tétras et de la Gélinoite des bois, trois points principaux identifiés comme essentiels constituent la structure de cette thèse :



- **Les résultats doivent être fiables même en utilisant des données non-standardisées**
  - Importance du choix des variables environnementales (Johnson and Gillingham 2005; Fourcade, Besnard, and Secondi 2018).
  - Importance des caractéristiques des jeux de données d'observation, potentiellement biaisés dans l'espace et le temps (Cardador et al. 2017; Phillips et al. 2009).
  - Importance du choix des échelles spatiales et temporelles (Arkle et al. 2014; Graf et al. 2005).
- **Les résultats des modèles doivent correspondre à la vision conceptuelle des gestionnaires et doivent être faciles à interpréter.**
  - Nécessité d'une collaboration approfondie avec les gestionnaires pour mieux comprendre les enjeux et les besoins.
  - Utilisation de variables environnementales pertinentes du point de vue du gestionnaire.
- **La communication des résultats doit être accessible pour chacun**
  - Production de rapports décrivant les résultats en dehors des publications scientifiques.

## RESUME DU CHAPITRE 2 : MATERIEL

---

### Jeux de données d'observations

Une base de données fournie par le Groupe Tétras Jura, forte de 6713 observations pour le Grand Tétras et de 6854 observations pour la Gélinotte des bois, a été utilisée. Les données ont été triées en trois catégories à l'aide des dates, des noms des observateurs et de la distance aux tracés : prospections (hiver ou été), battues (été) et protocole inconnu. Seules les données les plus récentes

(depuis 2007), localisées avec un GPS ont été utilisées pour la construction des modèles. Les données de battues n'ont pas été utilisées car les oiseaux étant « chassés » par les observateurs, leur localisation ne correspond pas forcément à l'endroit où ils étaient au départ (comportement de fuite).

### Jeux de données LiDAR

Les données LiDAR proviennent de deux campagnes différentes. La première s'est déroulée en automne 2014 et couvre 626 km<sup>2</sup> du département de l'Ain. La seconde s'est déroulée en été 2016, et couvre les zones d'intérêt forestières pour la Gélinoite des bois et du Grand Tétràs du Jura et du Doubs (431km<sup>2</sup>). Il faut souligner ici que, à cause de plusieurs différences entre ces campagnes (technique, saison) des modèles seront proposés pour chaque zone séparément.

Les variables LiDAR ont été calculées grâce aux *packages* LidR (Roussel and Auty 2016) et LidarRtree (en développement par Jean-Matthieu Monnet) sur le logiciel R. Deux types de variables ont été extraits, les métriques nuages de points basés sur la surface et les métriques orientées-objet. Les métriques dites « nuages de points basés sur la surface » sont des valeurs comme la moyenne de la hauteur, la densité des points, l'écart-type de la hauteur, calculées sur une surface définie de 25m\*25 m (taille du pixel choisie). En revanche, les métriques orientées-objet sont constituées d'éléments précis existant dans le paysage tel que les arbres ou les trouées. Ces objets sont extraits du nuage de points grâce à des caractéristiques définies à l'avance à une résolution de 0,5 m. Les arbres ont été extraits avec la méthode de détection des sommets. Les trouées ont été définies comme étant les zones d'une hauteur de canopée inférieure à 1 m, entourées d'arbres mesurant moins de la moitié de sa largeur. Par la suite, la moyenne résumant les caractéristiques des objets a été calculée pour une surface de 25m\*25 m (taille du pixel). Enfin, à partir des objets « arbre » de la zone de l'Ain et grâce aux données des placettes permanentes de l'ONF (localisation des arbres sur chaque placette), la proportion de résineux par rapport aux feuillus a pu être estimée (Eysn et al. 2015).

## Jeux de données sur la régénération du hêtre

Trois jeux de données ont été utilisés : deux collectés par le Groupe Tétras Jura (Jura et Doubs) et un troisième provenant des placettes d'inventaires permanentes de l'ONF (Ain). Les premiers jeux de données représentent 1380 placettes où le couvert de la régénération du hêtre a été mesuré et 1276 placettes où une note (de A à D) a été attribuée suivant l'intensité de la régénération du hêtre. Les données de l'ONF sont constituées d'un dénombrement des tiges sur trois sous-placettes de 1,5 m de rayon.

## RESUME DU CHAPITRE 3: METHODES DE MODELISATION

---

Les SDMs reposent sur différents postulats qui, s'ils ne sont pas respectés peuvent mener à des modèles peu performants et peu fiables :

- L'espèce étudiée doit être à l'équilibre, ainsi les relations avec l'environnement ne changent pas au cours du temps ou de l'espace.
- La méthode statistique choisie doit être appropriée aux types de données.
- Les données d'observations ne doivent pas être sujettes à un biais d'échantillonnage (spatial ou temporel) et sont indépendantes.
- Les variables environnementales ayant un impact sur la distribution de l'espèce sont toutes disponibles à toutes les échelles pertinentes pour la modélisation et ne contiennent pas d'erreurs.

Cependant, les données d'observations sont généralement collectées sans protocoles définis, entraînant un biais d'échantillonnage spatial. Ce biais, s'il n'est pas corrigé, entraîne la création de modèles qui représentent les caractéristiques de l'effort d'échantillonnage, plutôt que la distribution de l'espèce.

Dans ce chapitre l'effet du biais d'échantillonnage a été exploré avec une espèce virtuelle constituée de deux sous-groupes avec des habitats différents et avec le cas d'une espèce dimorphique : le Grand

Tétras. Trois protocoles d'échantillonnages et deux méthodes de correction ont été appliqués à l'espèce virtuelle, tandis que des données provenant d'échantillonnages systématiques et subjectifs (biaisés) ont été utilisés pour le Grand Tétras. Les modèles non corrigés pour le biais d'échantillonnage ont montré des résultats biaisés pour l'estimation des paramètres et du *sexe ratio*, mais les deux méthodes de corrections appliquées ont permis de meilleures estimations de ces deux variables. L'effet du protocole d'échantillonnage montre que de meilleures estimations sont faites pour les cas aléatoires et systématiques mais que la combinaison de différents protocoles (systématiques et subjectifs) peut aussi améliorer les résultats en y apportant une correction. Concernant les données d'observations du Grand Tétras, les méthodes de correction n'ont pas permis d'obtenir un *sexe ratio* plus équilibré. En revanche le *sexe ratio* était plus proche de 50% avec le protocole systématique qu'avec le protocole subjectif. Ces résultats montrent l'importance de corriger le biais d'échantillonnage spatial dans le cas d'espèces constituées de groupes avec des besoins d'habitats différents (sexe, âge).

Si le choix des variables environnementales est important, les échelles spatiales qui correspondent à la manière dont les individus perçoivent leur environnement est aussi essentiel. Ainsi, sept échelles différentes ont été choisies pour créer des modèles : sept modèles à échelle unique et variables multiples et un modèle mixte à échelles multiples et variables multiples. La performance des différents modèles a ensuite été comparée pour la Gelinotte des bois et le Grand Tétras dans les chapitres suivants, avec l'hypothèse que les modèles mixtes donneront de meilleures performances.

## RESUME DU CHAPITRE 4: LA GELINOTTE DES BOIS

Les modèles de distribution pour la Gelinotte des bois créés à l'aide de données LiDAR comme variables environnementales ont permis de produire des cartes de prédictions à plusieurs échelles. Quatre variables contribuant aux modèles dans chacune des zones d'études et à chaque échelle ont été observées, montrant que les caractéristiques structurales du milieu sont importantes pour l'espèce, en particulier pour la strate entre 5 m et 7 m. Les performances des modèles aux différentes échelles ont montré que les modèles mixtes ne sont pas les plus performants,

contrairement à l'hypothèse de départ. Cependant, les cartes produites prenant en compte les variables à différentes échelles permettent d'associer des effets locaux (végétation) à des effets à plus grande échelle (perturbations, fragmentation), qu'il est intéressant de prendre en compte. La limitation principale à l'interprétation des courbes de réponses provient du fait que seuls deux types de réponses peuvent être modélisés, linéaire et quadratique. Ceci a pour effet de rendre certaines interprétations non compatibles avec nos connaissances de l'écologie de l'espèce. Par exemple, des zones de densité d'arbres de 0 données favorables. La création de graphiques en trois dimensions pour visualiser au mieux l'effet de variables en interaction sur les prédictions a montré des résultats très différents suivant les zones d'études et les échelles. Ce type de graphique peut néanmoins permettre une meilleure compréhension des caractéristiques de l'habitat menant à des valeurs de prédiction favorables ou non.

La mise en place d'un nouveau protocole de prospection pour la Gélinoite des bois permettra d'accumuler de nouvelles données, à la fois sur sa présence ou son absence, et, dans le futur, d'améliorer notre connaissance de la distribution de l'espèce dans le massif Jurassien. Ces données permettront aussi d'améliorer d'éventuelles futures modélisations de l'habitat de la Gélinoite des bois.

## RESUME DU CHAPITRE 5: LE GRAND TETRAS

---

L'utilisation de variables LiDAR orientées-objet peut faciliter grandement l'interprétation des résultats par les gestionnaires. Elles sont cependant peu utilisées pour la création de modèles de distribution d'espèces en danger d'extinction. Dans ce chapitre elles ont été utilisées dans les modèles de distribution du Grand Tétrás. Les modèles orientés-objet créés à plusieurs échelles ont montré des performances légèrement inférieures aux modèles créés à partir des variables nuages de points (basées sur la surface), avec des différences plus prononcées à grande échelles. Les modèles orientés-objet étaient soit autant, soit plus transférables à une autre zone d'étude que les modèles nuages de points (basés sur la surface), avec des corrélations entre les prédictions supérieures à 0,66 (évaluées sur une zone commune aux deux sites d'études). Ainsi l'utilisation de variables orientées-

objet, dans notre cas les arbres et les trouées, montre des résultats prometteurs pour des modèles de distribution d'espèces dans une optique de conservation et a un grand potentiel pour améliorer la transmission des résultats aux gestionnaires.

Les modèles de distribution pour le Grand Tétrás créés à l'aide de données LiDAR orientées-objet comme variables environnementales ont permis de produire des cartes à plusieurs échelles avec des performances allant de faibles à bonnes. Les performances des modèles aux différentes échelles ont montré que les modèles mixtes sont, dans le cas du Grand Tétrás, plus performants. Cependant, les cartes des modèles mixtes présentent entre elles de larges différences, car l'aspect de la carte dépend en grande partie de l'échelle sélectionnée pour chaque variable. Or ces échelles optimales étaient différentes entre les zones d'études, montrant que le choix d'une échelle pertinente reste difficile. Peu de différences entre les sexes ont été observées pour la contribution des variables et les courbes de réponse. Le rôle du dérangement (pistes de ski) dans la distribution de l'espèce semble important car cette variable contribue à tous les modèles d'hiver à toutes les échelles et sur les deux zones d'études. En revanche, des différences entre les zones d'études ont été observées pour la contribution des variables ne permettant pas, comme pour la Gélinothe des bois, de proposer les variables LiDAR les plus pertinentes pour cette espèce. Nos résultats montrent cependant l'importance d'une grande hétérogénéité verticale et horizontale associée à une surface terrière située entre 20 et 30 m<sup>2</sup>/ha dans la composition d'un habitat favorable pour le Grand Tétrás.

## RESUME DU CHAPITRE 6: LA REGENERATION DU HETRE

---

Des modèles ont été créés afin de prédire la distribution des zones avec une régénération intense du hêtre, qui est un habitat défavorable pour le Grand Tétrás, mais aussi pour la Gélinothe des bois. Les variables environnementales utilisées dans les modèles ont été sélectionnées parmi des variables LiDAR ayant montré un fort potentiel. Pour cela les placettes avec une présence ou une absence de régénération du hêtre ont été comparées pour chaque variable LiDAR. Ainsi, des variables décrivant directement les caractéristiques du sous étage (variables nuage de points) entre 1 et 2 m ainsi que la

hauteur et l'hétérogénéité de la canopée ont été sélectionnées. Deux variables topographiques décrivant l'exposition, la pente et l'altitude ont aussi été utilisées.

Trois algorithmes différents ont été utilisés pour créer ces modèles : random forest, régression logistique, et un modèle linéaire généralisé (distribution de poisson). Pour la zone d'étude de l'Ain, les mauvaises performances des modèles n'ont pas permis de créer une carte de prédiction suffisamment fiable. En revanche, dans la zone d'étude du Jura, de meilleures performances ont été observées.

Les résultats montrent qu'il est encore difficile de prédire avec une grande précision la distribution de la végétation du sous-étage en utilisant des variables LiDAR. Dans notre cas, l'utilisation de variables décrivant les caractéristiques du sol ou les espèces présentes pourrait améliorer les résultats obtenus. Néanmoins, les progrès rapides dans l'analyse et l'extraction d'objets pourront bientôt permettre une meilleure caractérisation du sous-étage avec le LiDAR (Hamraz, Contreras, and Zhang 2017).

## RESUME DU CHAPITRE 7: SYNTHESE ET PERSPECTIVES

Ce travail de thèse avait pour objectif de produire des cartes de prédiction de la distribution de la Gélinothe des bois, du Grand Tétrás et de la régénération du hêtre dans le massif Jurassien pour aider les gestionnaires dans leurs décisions de conservation, restauration d'habitat et mise en place de nouveaux plans échantillonnages. Les spécificités des espèces étudiées et des jeux de données utilisés ont permis l'exploration des effets du biais d'échantillonnage, des interactions entre variables, du choix des variables et des échelles sur la performance et la précision des modèles ainsi que sur le transfert des résultats aux gestionnaires. Les perspectives pour améliorer les résultats obtenus sont nombreuses.

En ce qui concerne l'amélioration des méthodes de modélisation, on peut remarquer le potentiel d'utilisation des espèces virtuelles pour étudier des questions spécifiques, tester de nouveaux algorithmes ou de nouvelles méthodes d'échantillonnages adaptées une zone d'étude. Le choix des échelles d'analyse est difficile et est dépendant de la zone d'étude (localisation, étendue), ainsi

l'utilisation d'un cadre d'optimisation du choix des échelles peut être poursuivi dans la recherche de l'obtention de prédictions plus fiables.

Pour l'amélioration du transfert et de l'utilisation des résultats provenant de SDMs, l'utilisation de variables orientées-objet a montré un grand potentiel pour participer à une meilleure transmission des résultats aux gestionnaires. Un effort dans la communication des résultats pourrait encore être fait, notamment grâce à l'émergence d'applications en ligne (R shiny par exemple) permettant à chacun d'avoir accès aux données, résultats, cartes et graphiques pouvant aider à la décision.

L'acquisition depuis 2016 par le Groupe Tétras Jura de données génétiques pour le Grand Tétras ouvre de nouvelles perspectives qui pourront venir compléter les résultats obtenus, notamment pour évaluer l'effet de la fragmentation de l'habitat sur la structure des populations et la dispersion des individus, mais aussi pour estimer de manière plus précise le *sexe ratio*.



## ANNEX II. LIDAR METRICS CODE

### Extraction of LiDAR metrics : R code

- Introduction
- Charge needed packages
- Set working directory
- Point clouds normalization using delaunay interpolation
- LiDAR metrics
  - Set parameters
  - Calculate metrics
  - Two-dimensions (2D) Canopy Height Model metrics
  - Gap segmentation metrics
  - Tree segmentation metrics
  - One-dimension (1D) height metrics
- Merge all dataframes together
- Calculate cumulated metrics
- Save all rasters
- Multiple tiles processing code

### Introduction

The code used to extract metrics from LiDAR point clouds use two main packages: `lidR` and `LidarRtree`. This code calculates multiple metrics from LiDAR normalized point clouds.

All heights and surface thresholds were defined for our study case but can be change in function of the need.

Here we will present the different steps using one tile as an example. However, for multi-tiles processing, the code is presented at the end. The steps are the same, only parallel processing and the use of the buffer is added. The buffer is loaded to prevent edge effects in tree segmentation and CHM processing.

### Charge needed packages

```
library(lidR)
library(lidarRtree)
library(raster)
library(imager)
library(reldist)
library(vegan)
library(foreach)
#library(doMC)
library(doParallel)

library(rasterVis) #for raster data visualisation
```

### Set working directory

Raw LiDAR data are contained in the `./DATA/` folder. Parallelized processing of normalized las tiles contained in a `./DATA/NORM/` subdirectory. Outputs rasters of metrics in a `./OUTPUT/` subdirectory.

```
base_path<-"./media/anouk/DATAPART1/These/Data/LIDAR/TEST_code_JM/"

setwd(paste0(base_path, "WORKFLOW/"))

#Laz files folder
LIDARtarget_path <- paste0(base_path, "WORKFLOW/DATA/")
LIDARNORMtarget_path <- paste0(base_path, "WORKFLOW/DATA/NORM/")
LIDARoutput_path<-paste0(base_path, "WORKFLOW/OUTPUT/")
```

### Point clouds normalization using delaunay interpolation

Raw LiDAR data are normalized using the function `lasnormalize` from the `lidR` package. Here is an exemple with two tiles.

```
#Create base name for files and new folder ./data
laz_base<-paste0(LIDARNORMtarget_path, "12601_TetrasJura_SemisLidar_")

#Extract laz_file from DATA folder
laz_file<-list.files(LIDARTarget_path, pattern=".laz") # to loop over all the files
laz_file<-paste0(LIDARTarget_path, laz_file)

#Register number of needed core for analysis
registerDoParallel(cores=2)

# Create nomalized las files in ./data folder
foreach (a=laz_file, .packages=c( "lidR")) %dopar% {

  tile_id <- rev(strsplit(gsub("\\.laz", "", a), "_")[[1]][1])
  lidar<-lidR::readLAS(a)

  b<-lasnormalize(lidar, method="deelaunay")
  writeLAS(b, file=paste0(laz_base, tile_id, "NORM", ".las"))
}
```

## LiDAR metrics

LiDAR files should be normalized, with non-overlapping rectangular tiles, aligned on a multiple of resolution, points should not lie on the northern or eastern border when such borders are common with adjacent tiles

## Set parameters

Thoses parameters can be change regarding the aim of the study.

```
# Set number of cores to use for parallel processing
#doMC::registerDoMC(15)
registerDoParallel(cores=15)

# Set resolution for final metrics map in meters
resolution <- 25

# Set resolution for Canopy Height Model (CHM)
resCHM <- 0.5

# Set buffer size (in meters) for CHM
# Remark : 20 m is better for gaps metrics / 10 m is enough for tree metrics
bsize <- 20

# Set height threshold for removing highest points (unwanted)
h.points <- 60

# points classes to retain for analysis (Here vegetation (5), ground (2) and non-classified (0))
class.points <- c(0,2,5)

# Specify ground class
class.ground <- 2

# Define height breaks for penetration ratio and density metrics
breaksH <- c(-Inf,0,0.5,1,2,5,10,20,30,60,Inf)

# Define height bin names for metrics
n.breaksH <- gsub("-", "", paste("H.nb", breaksH[c(-length(breaksH))], "_", breaksH[c(-1)], sep=""))

# Set surface breaks for gap metrics (in squaremeters)
breaksGapSurface <- c(20, 200, 1000, Inf)

# set surface bin names fro gaps metrics
n.breaksG <- gsub("-", "", paste("G.s", breaksGapSurface[c(-length(breaksGapSurface))], "_", breaksGapSurface[c(-1)], sep=""))

# Set percentiles of height distribution for metrics
percent <- c(0.10,0.25,0.5,0.75,0.9)

# create catalog of LAS files
cata <- lidR::catalog(paste0(LIDARNORMtarget_path))

#Select only one file metrics calculation example
cata<-cata[2,]
```

## Calculate metrics

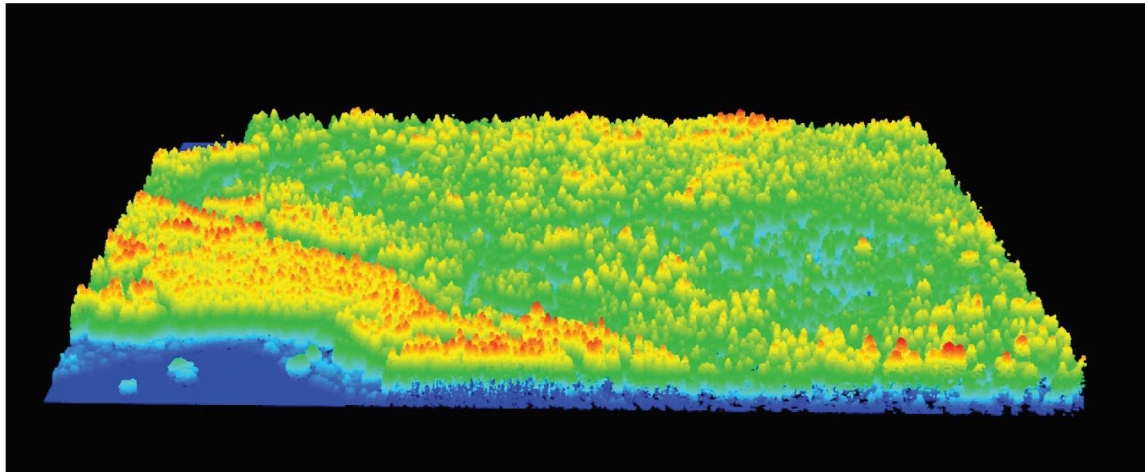
First, the normalize LAS file is loaded and cleaned up for unwanted point (height treshold)

```
# Load one tile extent
a <- try(lidR::catalog_queries(cata, (cata$Max.X+cata$Min.X)/2, (cata$Max.Y+cata$Min.Y)/2, (cata$Max.X-cata$Min.X)/2+bsize, (cata$Max.Y-cata$Min.Y)/2+bsize, mc.cores = 1))

a <- a[[1]]

# Remove unwanted point classes, and points higher than height threshold
a <- a %>% lasfilter(is.element(Classification, class.points) & Z<=h.points)

# set negative heights to 0
a@data$Z[a@data$Z<0] <- 0
```



LiDAR point cloud

## Two-dimensions (2D) Canopy Height Model metrics

---

Point density between selected height thresholds

Penetration ratio between selected height thresholds

Canopy height Models smoothed at 1, 2, 4, 8 and 16 pixels kernel

Canopy height Models standard deviation

Canopy height Models proportion

```
#####

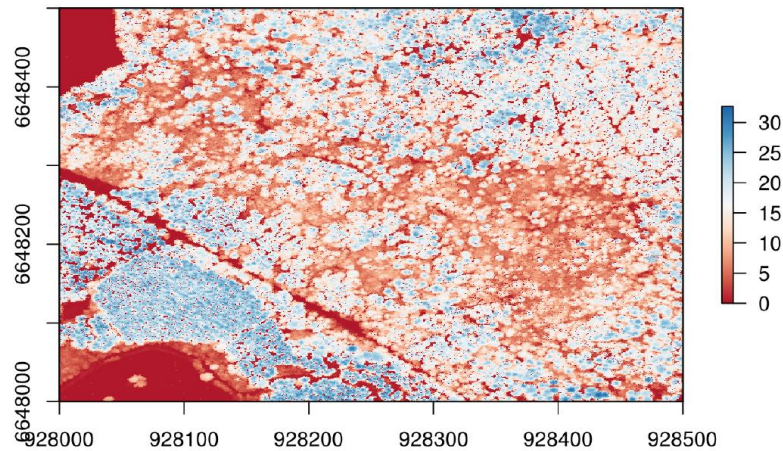
# Compute 2D CHM metrics

# Compute CHM
dsm <- lidaRtRee::points2DSM(a,res=resCHM)

# Replace NA, low and high values by 0
dsm[is.na(dsm) | dsm<0 | dsm>h.points] <- 0

# Convert to imager
dsm.c <- imager::as.cimg(raster::as.matrix(dsm))
```

Calculated DSM (resolution = 0.5m)



```
# Multiscale smoothing

st <- list()

# Non linear filtering
st[[1]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "Closing", nlSize = 5, sigmap = 0)[[1]]

# smoothing

# kernel sigmap in pixels
st[[2]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "none", nlSize = 5, sigmap = 1)[[2]]
st[[3]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "none", nlSize = 5, sigmap = 2)[[2]]
st[[4]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "none", nlSize = 5, sigmap = 4)[[2]]
st[[5]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "none", nlSize = 5, sigmap = 8)[[2]]
st[[6]] <- lidaRtRee::demfiltering(dsm.c, nlFilter = "none", nlSize = 5, sigmap = 16)[[2]]

# Convert results to raster
st <- lapply(st,function(x){x <- raster::raster(as.matrix(x));extent(x) <- extent(dsm);return(x)})
st <- stack(st)
```

```
st2<-st
names(st2) <- c("CHM.0", "CHM.1", "CHM.2", "CHM.4", "CHM.8", "CHM.16")

# Custom theme (from rasterVis package)
my.theme <- RdBuTheme()

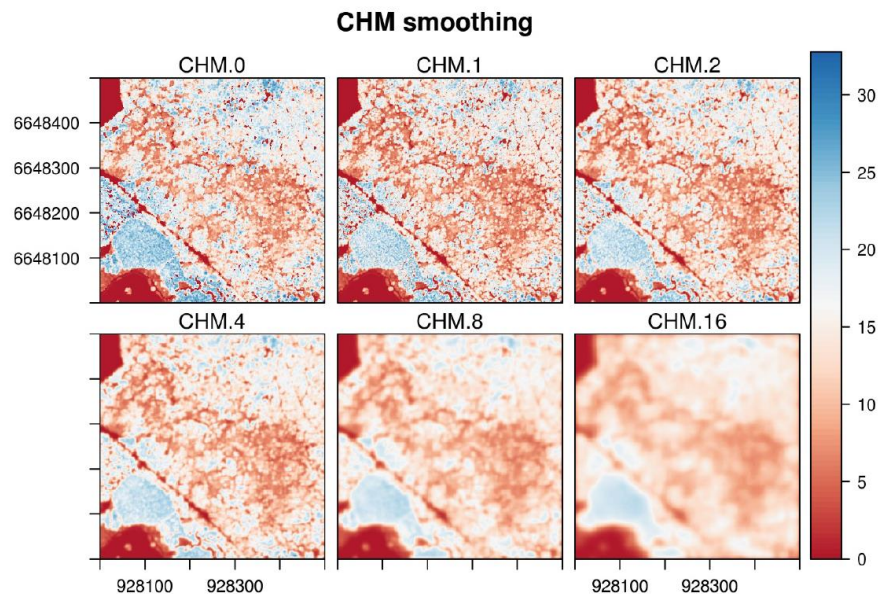
# Find the min and max values
my.min <- min(st2$CHM.0@data@min)
my.max <- max(st2$CHM.0@data@max)

# Customize the colorkey
my.at <- seq(my.min, my.max, length.out=length(my.theme$regions$col)-1)

my.key <- list(at=my.at, col=my.theme$regions$col)

levelplot(st2, par.settings=my.theme, at=my.at, colorkey=my.key, main="CHM smoothing")
```





```
# Convert to data.frame
st <- as.data.frame(raster::rasterToPoints(st))

# Remove pixels outside of tile (for multiple tiles processing)
#st <- st[st$x>=cata[i,"Min.X"] & st$x<=cata[i,"Max.X"] & st$y>=cata[i,"Min.Y"] & st$y<=cata[i,"Max.Y"],]

# Compute coordinates of cell center at metrics resolution
st$X <- round((st$x-resolution/2)/resolution)*resolution+resolution/2
st$Y <- round((st$y-resolution/2)/resolution)*resolution+resolution/2

# Add split factor
dummy <- factor(paste(st$X, st$Y, sep="_"))

# Compute metrics by factor
dummy <- lapply(split(st, dummy), FUN=function(x){c(sd(x$layer.1), sd(x$layer.2), sd(x$layer.3), sd(x$layer.4),
sd(x$layer.5), sd(x$layer.6), mean(x$layer.1), sum(x$layer.1<0.5), sum(x$layer.1<1), sum(x$layer.1>5), sum(x
$layer.1>10),
sum(x$layer.1>20), sum(x$layer.1<5)-sum(x$layer.1<1), s
um(x$layer.1<5)-sum(x$layer.1<2)))})

# Create data.frame
temp2d <- data.frame(matrix(unlist(dummy), nrow=length(dummy), byrow=T))
names(temp2d) <- c("CHM0.sd", "CHM1.sd", "CHM2.sd", "CHM4.sd", "CHM8.sd", "CHM16.sd", "CHM.mean", "CHM.PercI
nf0.5", "CHM.PercInf1", "CHM.PercSup5", "CHM.PercSup10", "CHM.PercSup20", "CHM.Perc1_5", "CHM.Perc2_5")

# Divide some metrics by number of cells to compute proportion
temp2d[,8:ncol(temp2d)] <- temp2d[,8:ncol(temp2d)]/(resolution/resCHM)^2

# Multiply some metrics by 100 to convert to percentage
temp2d[,8:ncol(temp2d)] <- temp2d[,8:ncol(temp2d)]*100

# Add factor for merging with other metrics
temp2d$merge <- names(dummy)

head(temp2d)
```

```
##      CHM0.sd   CHM1.sd   CHM2.sd   CHM4.sd   CHM8.sd   CHM16.sd   CHM.mean
## 1  0.4750724  0.3577934  0.279228  0.238486  0.2411888  0.2227496  0.147328
## 2  2.0417390  1.8224832  1.681732  1.473005  1.1611503  0.7643707  2.910080
## 3  5.7902894  5.3086402  5.083927  4.675219  4.0982912  3.2926146  6.842668
## 4  5.5968207  6.0170112  5.108027  4.118623  3.6142666  3.1851267  20.488272
## 5  7.1466114  6.0238234  5.641743  5.004316  4.1891314  3.1391484  4.976456
## 6  7.7957879  6.9318817  5.947954  4.592096  3.4096652  2.5718756  15.662468
##      CHM.PercInf0.5  CHM.PercInf1  CHM.PercSup5  CHM.PercSup10  CHM.PercSup20
## 1           91.40         97.00           0.32           0.00           0.00
## 2           21.92         31.96          13.60           0.00           0.00
## 3           0.00          0.00          34.00          20.08           6.36
## 4           0.12          0.12          96.60          93.52          63.52
## 5          46.44          50.04          31.08          18.44           7.76
## 6          12.88          13.40          84.04          79.88          28.24
##      CHM.Perc1_5  CHM.Perc2_5      merge
## 1           2.68           0.76  928012.5_6648012.5
## 2          54.40          50.44  928012.5_6648037.5
## 3          65.96          65.12  928012.5_6648062.5
## 4           3.28           3.28  928012.5_6648087.5
## 5          18.64          15.84  928012.5_6648112.5
## 6           2.56           1.84  928012.5_6648137.5
```

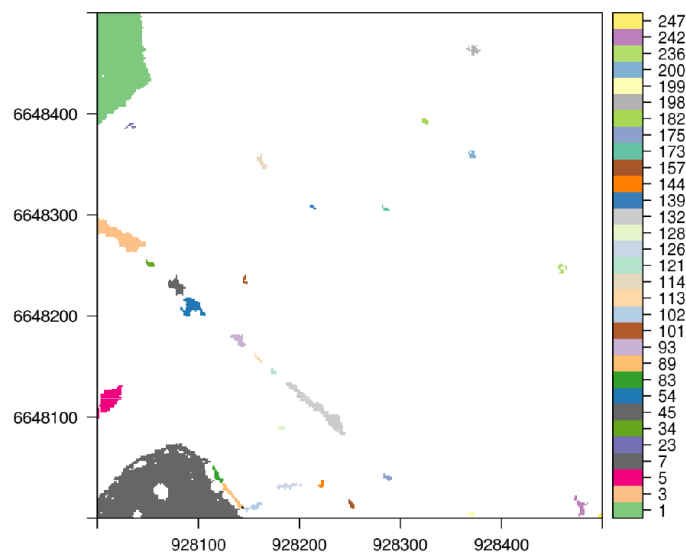
## Gap segmentation metrics

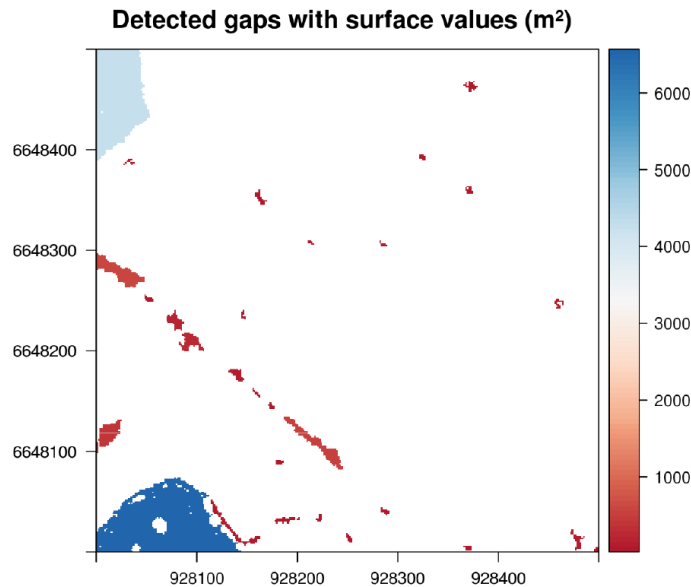
Percentage of gaps of different surface categories

```
#####
# compute gap metrics

gaps <- gap.detection(dsm, ratio=2, gap.max.height=1, min.gap.surface=min(breaksGapSurface), closing.height.
bin=1, nlFilter="Median", nls=3)
```

Detected gaps with labels ID





```
# Crop results to tile size (for multiple tiles processing)
#gaps.labels <- crop(gaps[[1]], extent(c(cata["Min.X"], cata["Max.X"], cata["Min.Y"], cata["Max.Y"])))
#gaps.surface <- crop(gaps[[2]], extent(c(cata["Min.X"], cata["Max.X"], cata["Min.Y"], cata["Max.Y"])))

gaps.surface<-gaps[[2]]
# Convert to data.frame
st <- as.data.frame(raster::rasterToPoints(gaps.surface))

# Compute coordinates of cell center at metrics resolution
st$X <- round((st$x-resolution/2)/resolution)*resolution+resolution/2
st$Y <- round((st$y-resolution/2)/resolution)*resolution+resolution/2

# Split factor
dummy <- factor(paste(st$X, st$Y, sep="_"))

# Compute metrics by factor
dummy <- lapply(split(st, dummy), FUN=function(x){hist(x$layer, breaks=breaksGapSurface, plot=F)$counts*((res
CHM/resolution)^2)*100})

# Create data.frame
tempGap <- data.frame(matrix(unlist(dummy),nrow=length(dummy),byrow=T))
tempGap$test <- apply(tempGap,1,sum)
names(tempGap) <- c(n.breaksG, paste("G.s", min(breaksGapSurface), "_Inf", sep=""))
tempGap$merge <- names(dummy)

head(tempGap)
```

```
##      G.s20_200 G.s200_1000 G.s1000_Inf G.s20_Inf      merge
## 1           0          0.00      97.92      97.92 928012.5_6648012.5
## 2           0          0.00      33.92      33.92 928012.5_6648037.5
## 3           0          0.40       0.00       0.40 928012.5_6648087.5
## 4           0         53.40       0.00      53.40 928012.5_6648112.5
## 5           0         10.16       0.00      10.16 928012.5_6648137.5
## 6           0          3.24       0.00       3.24 928012.5_6648262.5
```

## Tree segmentation metrics

### Tree densities at different height thresholds

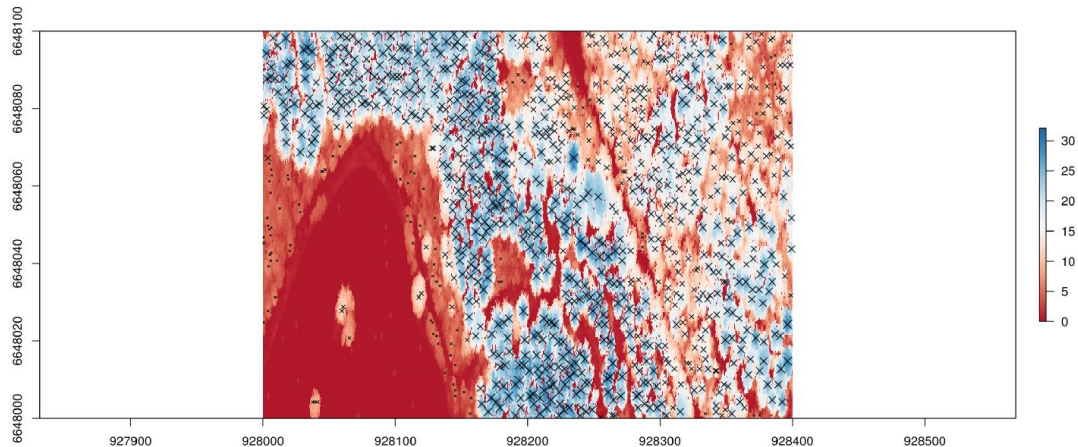


```
#####

# Compute tree level metrics

# Tree top detection with a minimum height of 5 meters
test <- lidaRtRee::treetopdetection(dsm, hmin=5)

# Extraction to data.frame
segms <- lidaRtRee::treeExtraction(test[[1]],test[[2]],test[[3]])
```



```
# Remove trees outside of tile (for multiple tiles processing)
#segms <- segms[segms$x>=cata[i, "Min.X"] & segms$x<=cata[i, "Max.X"] & segms$y>=cata[i, "Min.Y"] & segms$y<
=cata[i, "Max.Y"],]

# Compute coordinates of cell center at metrics resolution
segms$x <- round((segms$x-resolution/2)/resolution)*resolution+resolution/2
segms$y <- round((segms$y-resolution/2)/resolution)*resolution+resolution/2

# Split factor
dummy <- factor(paste(segms$x, segms$y, sep="_"))

# Compute metrics by factor
dummy <- lapply(split(segms, dummy), FUN=function(x){c(mean(x$h), sd(x$h), gini(x$h), length(x$h), sum(x$h<10)
, sum(x$h>10), sum(x$h>20), sum(x$h>30), mean(x$s), sum(x$s), sum(x$v))})

# Create data.frame
tempTree <- data.frame(matrix(unlist(dummy),nrow=length(dummy),byrow=T))
names(tempTree) <- c("Tree.meanH", "Tree.sdH", "Tree.giniH", "Tree.density", "Treeinf10.density", "Tree10.den
sity", "Tree20.density", "Tree30.density", "Tree.meanS", "Tree.sumS", "Tree.sumV")

# Convert some metrics to number of trees per ha
tempTree[,4:8] <- tempTree[,4:8]/(resolution)^2*10000

# Add factor for merging with other metrics
tempTree$merge <- names(dummy)

# Replace fixel filled with NA by 0 : area with no trees are set to 0 presence of trees, to be distinct from
area not covered by LiDAR
tempTree[is.na(tempTree)] <- 0
```

## One-dimension (1D) height metrics

### Number of ground returns

First pulse mean intensity

Number of vegetation and unclassified point between height thresholds

Maximum height

Mean height

Mean height at defined quantile

```
#####  
  
# Compute 1D height metrics  
  
# Remove points outside initial tile extent (for multiple tiles processing)  
#a <- a %>% lidR::lasfilter(X <= cata[i,"Max.X"] & X >= cata[i,"Min.X"] & Y >= cata[i,"Min.Y"] & Y <= cata[i,  
  "Max.Y"])  
#  
  
# first set of metrics  
templd1 <- a %>% lidR::grid_metrics(as.list(c(as.vector(quantile(Z,probs=percent)),  
  sum(Classification==2),  
  mean(Intensity [ReturnNumber == 1])  
)),res = resolution)  
names(templd1) <- c("X","Y", paste("H.p",percent*100,sep=""), "nbSol", "I.mean.1stpulse")  
  
# Remove metrics outside of tile (for multiple tiles processing)  
#templd1 <- templd1[templd1$X>=cata[i,"Min.X"] & templd1$X<=cata[i,"Max.X"] & templd1$Y>=cata[i,"Min.Y"] &  
templd1$Y<=cata[i,"Max.Y"],]  
  
# Add factor for merging  
templd1$merge <- paste(templd1$X,templd1$Y,sep="_")  
  
# Vegetation-only metrics  
a <- a %>% lidR::lasfilter(Classification!=class.ground)  
  
templd2 <- a %>% lidR::grid_metrics(as.list(c(mean(Z),  
  max(Z),  
  sd(Z),  
  length(Z),  
  hist(Z, breaks=breaksH, right=F, plot=F)$counts)  
),res = resolution)  
names(templd2) <- c("X","Y","H.mean","H.max","H.sd","nbVeg",n.breaksH)  
  
# remove metrics outside of tile (for multiple tiles processing)  
#templd2 <- templd2[templd2$X>=cata[i,"Min.X"] & templd2$X<=cata[i,"Max.X"] & templd2$Y>=cata[i,"Min.Y"]  
& templd2$Y<=cata[i,"Max.Y"],]  
  
# Add factor for merging  
templd2$merge <- paste(templd2$X,templd2$Y,sep="_")  
templd2$X <- templd2$Y <- NULL  
  
head(templd2)
```

```
##      H.mean H.max      H.sd nbVeg H.nbInf_0 H.nb0_0.5 H.nb0.5_1 H.nb1_2
## 1: 0.2547558 5.50 0.6767391 2866      0      2458      239      88
## 2: 2.1126492 9.58 3.1131578 3401      0      2260      12      29
## 3: 2.5908661 7.97 1.7116757 8706      0      1470      766     1032
## 4: 1.8229920 9.34 2.7717938 3242      0      2102      118      49
## 5: 0.1472309 2.16 0.2566539 2694      0      2423      223      47
## 6: 5.5419225 26.26 5.5033585 10684      0      404      458     1308
##      H.nb2_5 H.nb5_10 H.nb10_20 H.nb20_30 H.nb30_60 H.nb60_Inf
## 1:      72      9      0      0      0      0
## 2:     218     882      0      0      0      0
## 3:    4794     644      0      0      0      0
## 4:     318     655      0      0      0      0
## 5:       1       0      0      0      0      0
## 6:    5653     993    1472     396      0      0
##      merge
## 1: 928012.5_6648012.5
## 2: 928037.5_6648012.5
## 3: 928012.5_6648037.5
## 4: 928062.5_6648012.5
## 5: 928037.5_6648037.5
## 6: 928012.5_6648062.5
```

## Merge all dataframes together

```
## merge two sets of 1d metrics, keeping all cells from 1st set
temp <- base::merge(temp1d1,temp1d2,all.x=T, by="merge")
temp <- data.frame(temp)

## merge 1d and 2d metrics, keeping all 1d cells (without buffer)
temp <- base::merge(temp,temp2d, all.x=T, by="merge")

## merge also gaps metrics, keeping all 1d cells (without buffer)
temp <- base::merge(temp,tempGap, all.x=T, by="merge")

## merge also tree metrics, keeping all 1d cells (without buffer)
temp <- base::merge(temp, tempTree, all.x=T, by="merge")

# remove merge field
temp <- temp[,-1]

# Replace pixel filled with NA by 0
temp[is.na(temp)] <- 0

metrics<-temp

head(metrics)
```

```
##      X      Y H.p10 H.p25 H.p50  H.p75  H.p90 nbSol I.mean.1stpulse
## 1 928012.5 6648012 0.000 0.00 0.00 0.020 0.170 4183 148.45291
## 2 928012.5 6648038 0.010 0.50 2.48 3.760 4.680 934 69.16062
## 3 928012.5 6648062 1.030 2.24 3.56 5.150 14.760 190 48.77620
## 4 928012.5 6648088 1.740 9.87 17.48 21.425 24.122 297 41.36993
## 5 928012.5 6648112 0.040 0.18 2.27 6.725 16.358 508 74.39711
## 6 928012.5 6648138 0.101 8.25 14.68 18.500 21.890 475 58.52660
##      H.mean H.max      H.sd nbVeg H.nbInf_0 H.nb0_0.5 H.nb0.5_1 H.nb1_2
## 1 0.2547558 5.50 0.6767391 2866 0 2458 239 88
## 2 2.5908661 7.97 1.7116757 8706 0 1470 766 1032
## 3 5.5419225 26.26 5.5033585 10684 0 404 458 1308
## 4 15.7713734 29.59 7.5116389 9662 0 386 126 238
## 5 5.5928690 28.68 6.8827497 4587 0 1495 140 303
## 6 14.1362396 30.52 7.0015787 6377 0 600 23 44
##      H.nb2_5 H.nb5_10 H.nb10_20 H.nb20_30 H.nb30_60 H.nb60_Inf CHM0.sd
## 1 72 9 0 0 0 0 0.4750724
## 2 4794 644 0 0 0 0 2.0417390
## 3 5653 993 1472 396 0 0 5.7902894
## 4 565 921 4022 3404 0 0 5.5968207
## 5 1020 657 668 304 0 0 7.1466114
## 6 203 699 3685 1117 6 0 7.7957879
##      CHM1.sd CHM2.sd CHM4.sd CHM8.sd CHM16.sd CHM.mean CHM.PercInf0.5
## 1 0.3577934 0.279228 0.238486 0.2411888 0.2227496 0.147328 91.40
## 2 1.8224832 1.681732 1.473005 1.1611503 0.7643707 2.910080 21.92
## 3 5.3086402 5.083927 4.675219 4.0982912 3.2926146 6.842668 0.00
## 4 6.0170112 5.108027 4.118623 3.6142666 3.1851267 20.488272 0.12
## 5 6.0238234 5.641743 5.004316 4.1891314 3.1391484 4.976456 46.44
## 6 6.9318817 5.947954 4.592096 3.4096652 2.5718756 15.662468 12.88
##      CHM.PercInf1 CHM.PercSup5 CHM.PercSup10 CHM.PercSup20 CHM.Perc1_5
## 1 97.00 0.32 0.00 0.00 2.68
## 2 31.96 13.60 0.00 0.00 54.40
## 3 0.00 34.00 20.08 6.36 65.96
## 4 0.12 96.60 93.52 63.52 3.28
## 5 50.04 31.08 18.44 7.76 18.64
## 6 13.40 84.04 79.88 28.24 2.56
##      CHM.Perc2_5 G.s20_200 G.s200_1000 G.s1000_Inf G.s20_Inf Tree.meanH
## 1 0.76 0 0.00 97.92 97.92 5.090000
## 2 50.44 0 0.00 33.92 33.92 6.308000
## 3 65.12 0 0.00 0.00 0.00 8.733636
## 4 3.28 0 0.40 0.00 0.40 26.233333
## 5 15.84 0 53.40 0.00 53.40 16.504444
## 6 1.84 0 10.16 0.00 10.16 22.571667
##      Tree.sdH Tree.giniH Tree.density Treeinf10.density Tree10.density
## 1 0.0000000 0.0000000 16 16 0
## 2 0.9603437 0.08266047 240 240 0
## 3 7.1159585 0.30810867 176 144 32
## 4 2.8229618 0.05421432 336 0 336
## 5 8.4086817 0.27020032 144 48 96
## 6 4.5061569 0.10700485 288 0 288
##      Tree20.density Tree30.density Tree.meanS Tree.sumS Tree.sumV
## 1 0 0 6.75000 6.75 31.0875
## 2 0 0 18.53333 278.00 1288.6400
## 3 32 0 27.61364 303.75 2515.8075
## 4 304 0 31.85714 669.00 13960.9525
## 5 48 0 17.47222 157.25 2003.9400
## 6 224 16 29.56944 532.25 9664.9525
```

## Calulate cumulated metrics

```
metrics$H.simpson<-vegan::diversity(metrics[,n.breaksH[c(-1,-length(n.breaksH))]], index="simpson")

for ( i in n.breaksH[c(-1,-length(n.breaksH))])
{
  metrics[,paste(i,"relative_density",sep="")] <- metrics[,i] / (metrics$nbVeg + metrics$nbSol)
}

cumS <- t(apply(metrics[,c("nbSol",n.breaksH[-length(n.breaksH)])],1,cumsum))
dummy <- log(cumS[,-1])
dummy <- t(apply(dummy, 1, diff))
dummy <- 1 - exp(-dummy)
dummy <- as.data.frame(dummy)
names(dummy) <- paste(names(dummy),"ratio",sep="")
#
for (i in names(dummy))
{
  metrics[,i] <- dummy[,i]
}
```

## Save all rasters

```
for (i in names(metrics))
{
  print(i)

  #Set CRS for metrics rasters
  dummy <- rasterFromXYZ(metrics[,c("X","Y",i)],res=25, crs="+proj=lcc +lat_1=49 +lat_2=44 +lat_0=46.5 +lon_0=3 +x_0=700000 +y_0=6600000 +ellps=GRS80 +towgs84=0,0,0,0,0,0 +units=m +vunits=m +no_defs")

  writeRaster(dummy,file=paste0(LIDAROutput_path, "raster_",i,".tif",sep=""),overwrite=T)
}
```

# ANNEX III. SUPPLEMENTARY MATERIAL ARTICLE 1 : DISTRIBUTION MODELS OF DIMORPHIC SPECIES: CORRECTING SPATIAL SAMPLING BIAS IMPROVES ESTIMATES OF HABITAT ASSOCIATION AND SEX RATIO

## Supplementary material

### Appendix A1

Table A1: Parameter values for the virtual species sub-groups 1 (female) and 2 (male)

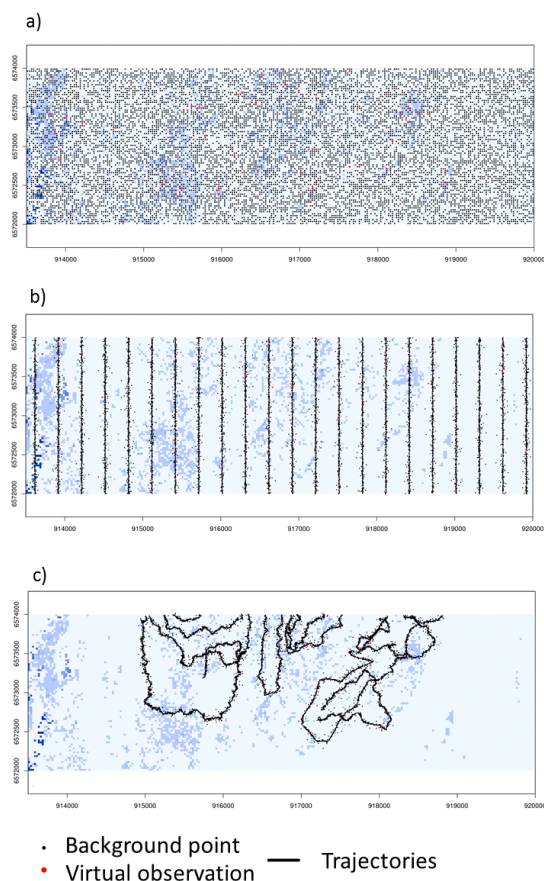
	Slope	Canopy density 10-20m	Simpson index	Intercept
Sub-group 1	-0,5	-0.3	-1	-3.8
Sub-group 2	0.01	-0.02	0.08	-1,0

## Appendix A2

Table A1: Sampling design tested on the virtual species

Type	Name	Description
Random	Random	Random points generation
Systematic	Transects	Sampling along transects
Non-systematic	Subjective	Real survey tracks

Fig. A2. Overview of the three-sampling designs for the virtual species case study. a) represents the random sampling design and random background point, b) represents the transects sampling design and targeted background point sampling, c) represent the subjective sampling design and targeted background points.



## Appendix A3

Fig. A1: Parameter estimations (100 models) for virtual sub-group 2 (male)

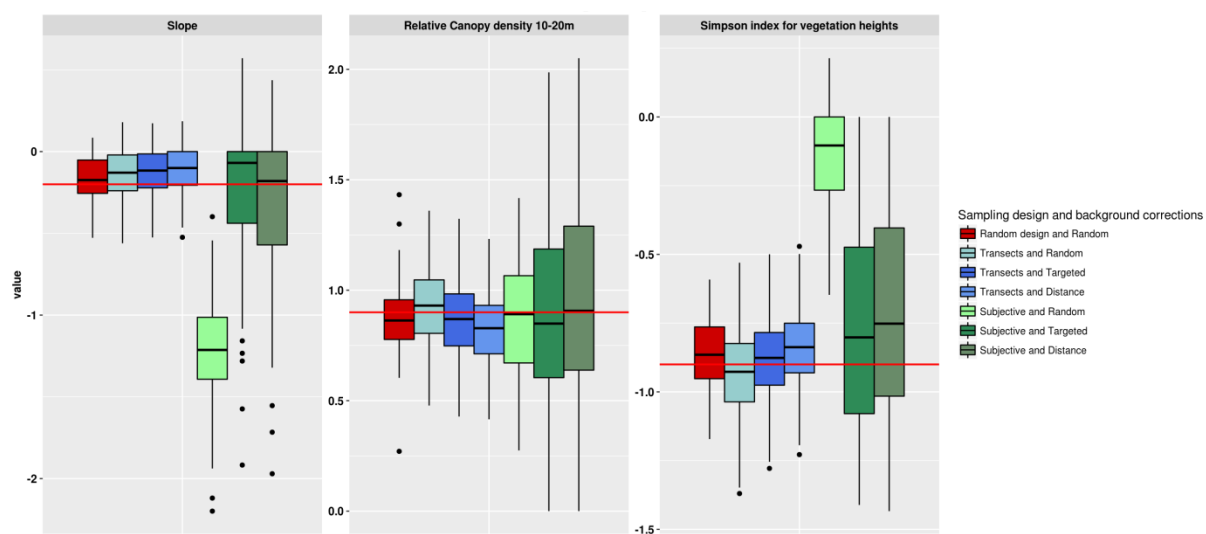


Table A2: Wilcoxon test p-values on parameter estimation mean values for the virtual sub-group 2 (male)

Sub-group 2				
Transects and Random	Random/Targeted	0.49	0.65	0.55
Transects and Targeted	Targeted/Distance	0.65	0.43	0.9
Transects and Distance	Random/Distance	0.29	0.23	0.6
Subjective and Random	Random/Targeted	2.1e-05	1.9e-05	0.017
Subjective and Targeted	Targeted/Distance	0.91	0.34	0.98
Subjective and Distance	Random/Distance	3.1e-05	5e-07	0.019



Table A3: P-values from the Wilcoxon test on the ratio estimation distribution.

Case	Estimated ratio
Transects and Random	1.8e-10
Transects and Targeted	1
Transects and Distance	1
Subjective and Random	1.7e-31
Subjective and Targeted	0.47
Subjective and Distance	0.0012

## ANNEX IV. SUPPLEMENTARY MATERIAL ARTICLE 2: EFFECT OF INTERACTION TERMS ON SPECIES DISTRIBUTION MODELS PREDICTIONS INTERPRETATION AT MULTIPLE SCALES.

### Wildlife Biology

#### Appendix 1

Table A1: Contribution of the interacting terms in models

Scale	Jura		Ain	
	Interaction 1	Interaction 2	Interaction 1	Interaction 2
0.31 ha	Yes	No	Yes	No
0.81 ha	No	No	Yes	No
1.8 ha	No	No	Yes	No
7 ha	No	Yes	Yes	Yes
15.8 ha	Yes	Yes	Yes	No
27.5 ha	Yes	No	Yes	No

Figure A1.1. 3D surface response curves for the two study area at the 0.31ha scale.

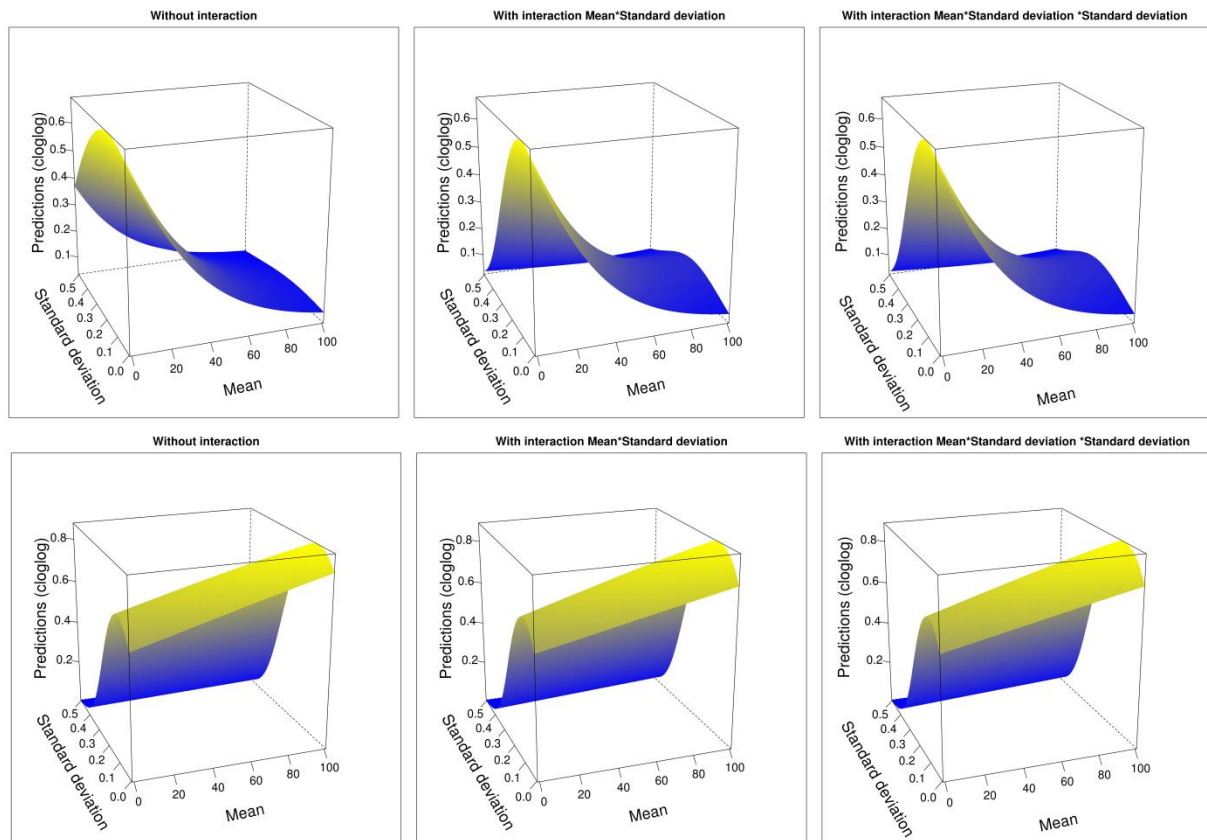


Figure A1.2 3D surface response curves for the two study area at the 0.81ha scale.

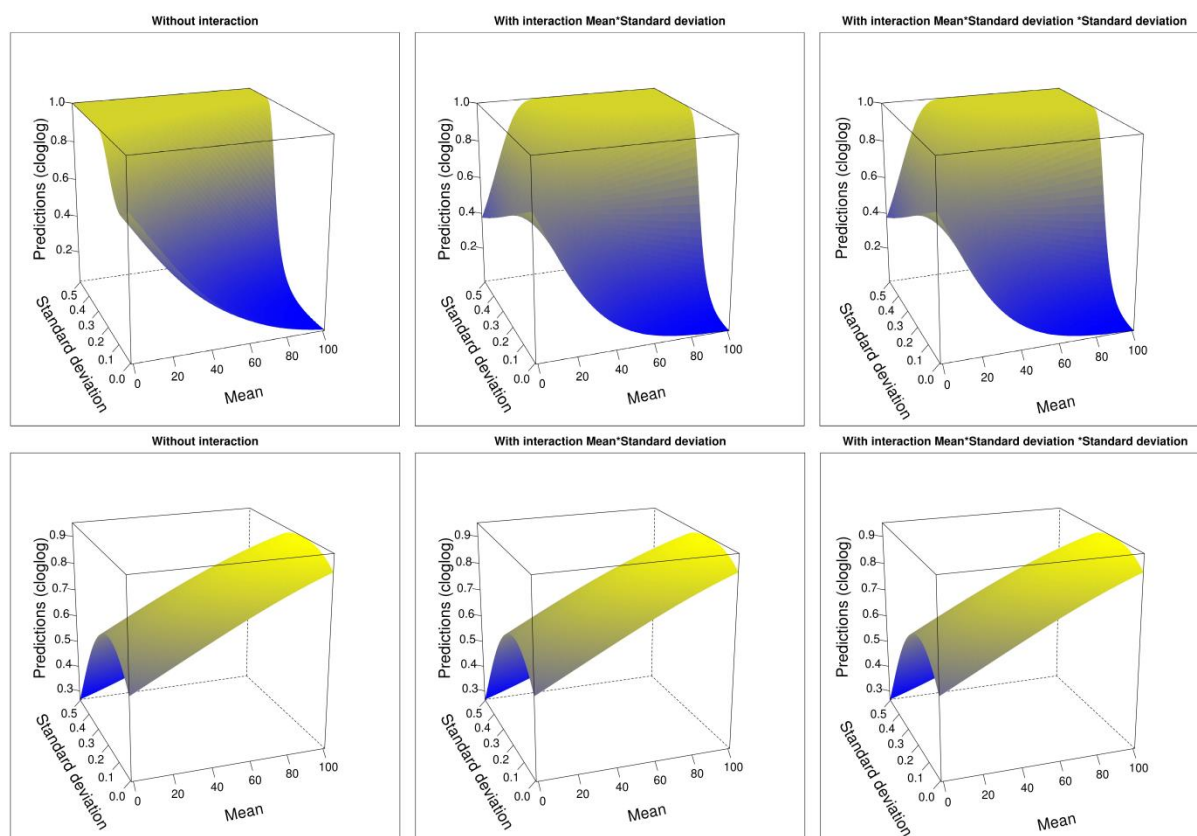


Figure A1.3. 3D surface response curves for the two study area at the 1.8ha scale.

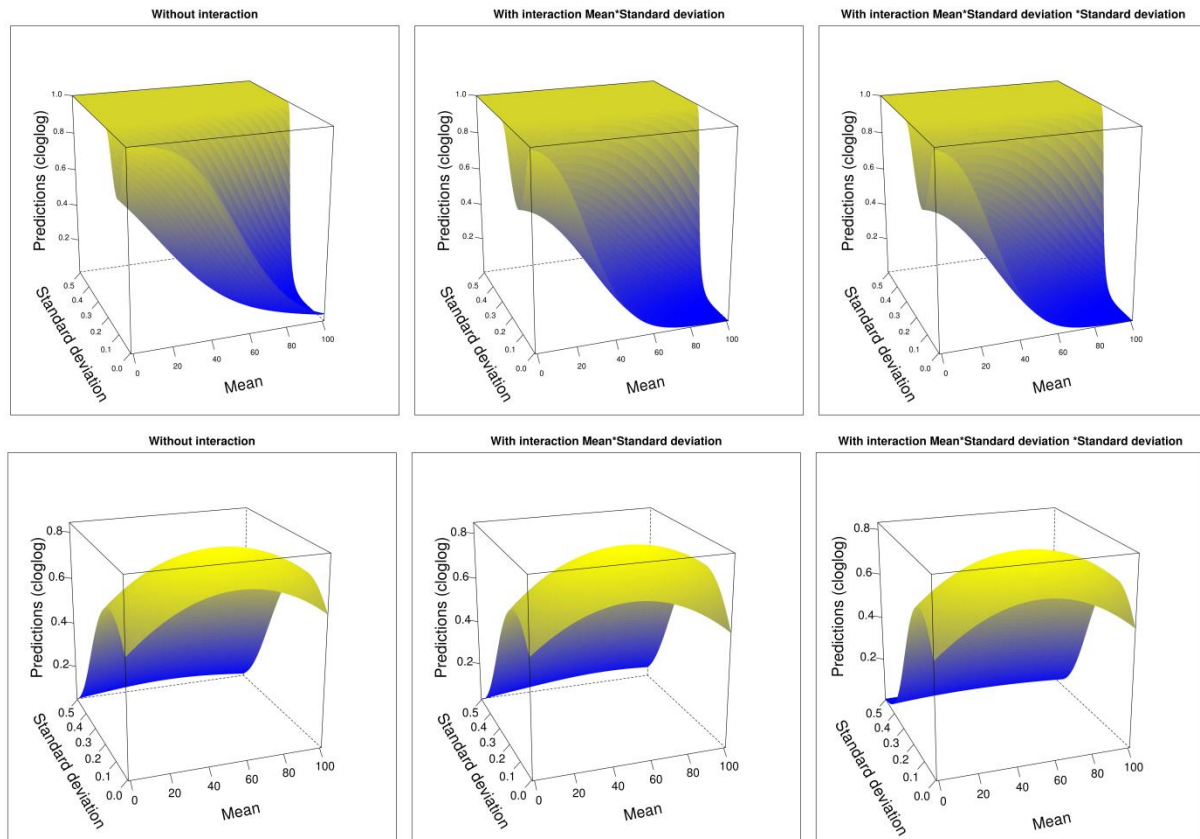
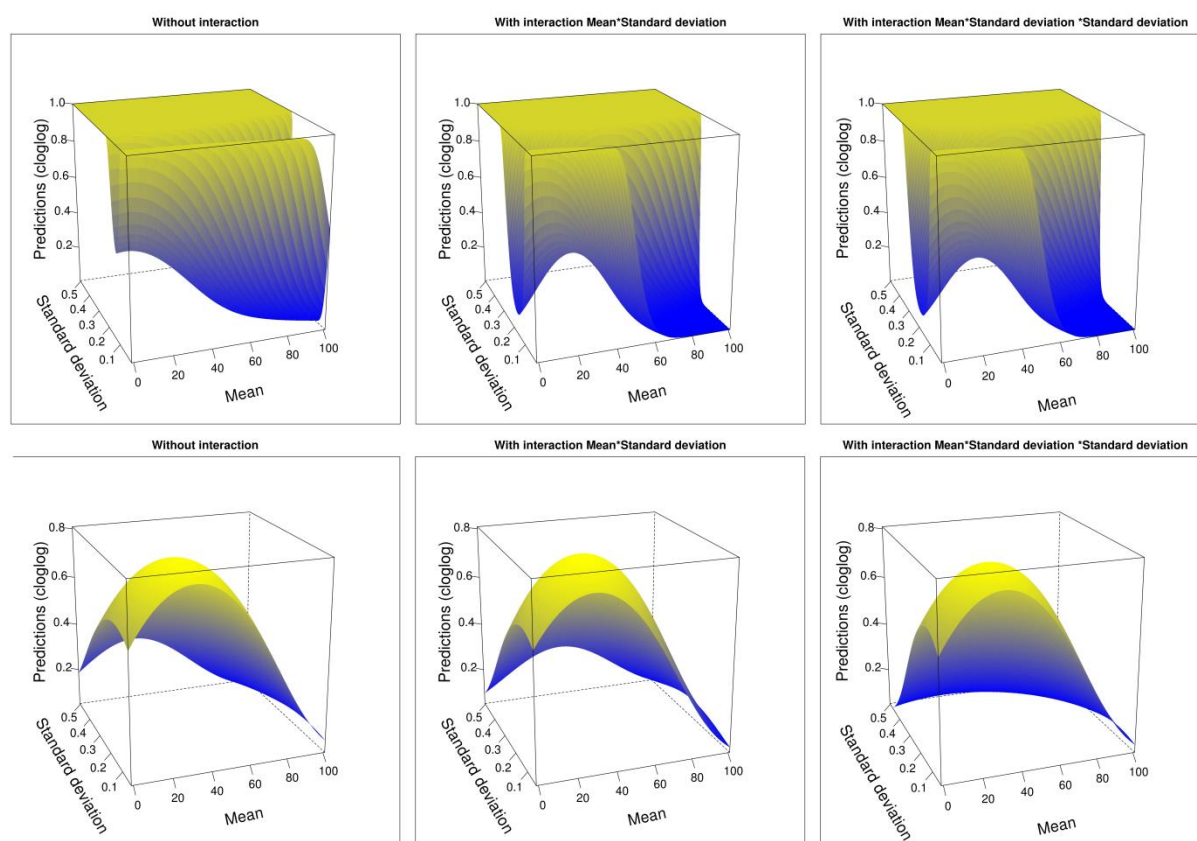


Figure A1.4 3D surface response curves for the two study area at the 15.8ha scale.



## Appendix 2

Table A2.1

Area	scale	Grassland	Prop. resG	Prop. resGsd	Middle_gap	Tree_gini	Tree_density10	Tree_densityinf10	Ratio0205sd	Ratio0205	Ratio0501sd	Canopy2030sd	Buildings	Road
Ain	0.0625ha	0.58	0.56	NA	0.49	0.53	0.56	0.56	NA	0.64	NA	NA	0.51	0.57
	0.31ha	0.6	0.59	0.58	0.46	0.53	0.58	0.57	0.52	0.68	0.48	0.52	0.51	0.57
	0.81ha	0.6	0.59	0.57	0.48	0.52	0.57	0.57	0.5	0.69	0.54	0.53	0.51	0.57
	1.8ha	0.59	0.6	0.56	0.51	0.52	0.57	0.56	0.49	0.7	0.55	0.5	0.51	0.57
	7ha	0.57	0.59	0.46	0.54	0.54	0.54	0.54	0.5	0.68	0.54	0.48	0.54	0.57
	15.8ha	0.56	0.59	0.47	0.56	0.54	0.52	0.54	0.5	0.67	0.55	0.47	0.55	0.57
	27.5ha	0.55	0.58	0.53	0.57	0.53	0.5	0.54	0.49	0.66	0.57	0.45	0.57	0.57

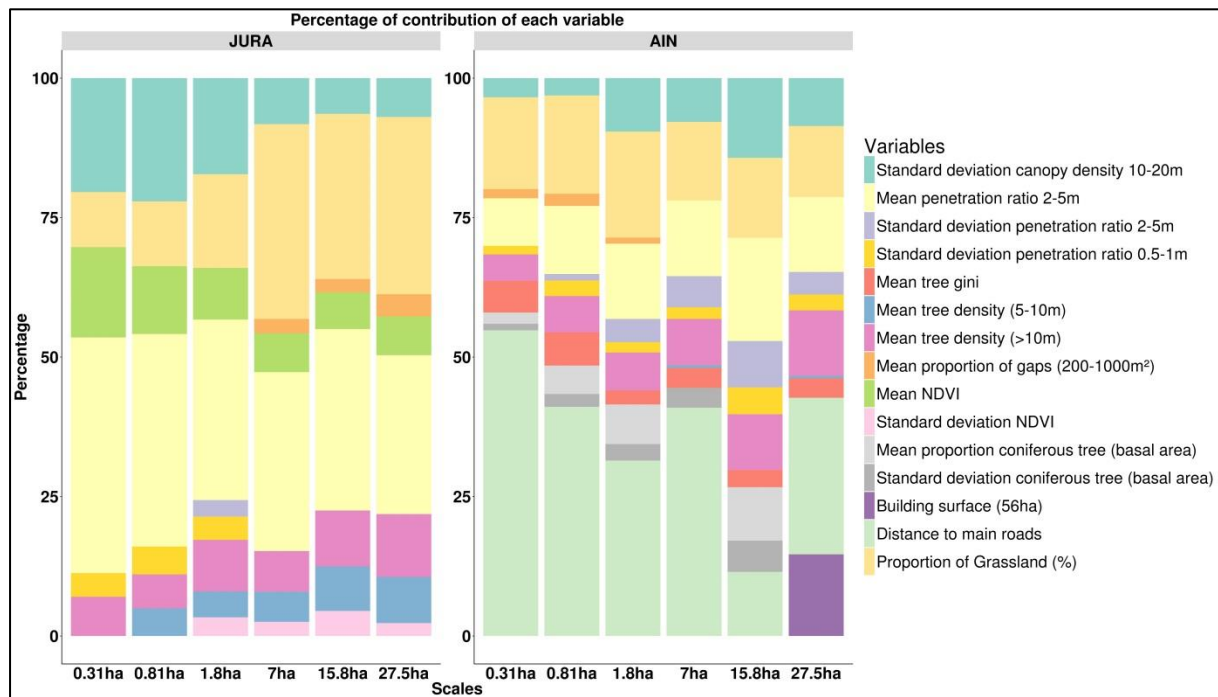
Table A2.2

Area	scale	Grassland	NDVI	NDVIstd	Middle_gap	Tree_gini	Tree_density10	Tree_densityinf10	Ratio0205sd	Ratio0205	Ratio0501sd	Canopy2030sd
Jura	0,0625ha	0.5	0.55	NA	0.5	0.53	0.49	0.52	NA	0.57	NA	NA
	0.31ha	0.49	0.57	0.51	0.48	0.54	0.48	0.54	0.52	0.59	0.5	0.56
	0.81ha	0.49	0.57	0.52	0.49	0.54	0.48	0.55	0.53	0.6	0.5	0.55
	1.8ha	0.5	0.57	0.52	0.5	0.53	0.52	0.54	0.53	0.6	0.49	0.55
	7ha	0.53	0.55	0.49	0.5	0.52	0.52	0.53	0.55	0.57	0.51	0.52
	15.8ha	0.53	0.54	0.49	0.5	0.52	0.52	0.52	0.55	0.57	0.52	0.51
	27.5ha	0.53	0.54	0.5	0.5	0.52	0.53	0.49	0.55	0.56	0.52	0.51



## Appendix 3

Figure A3. Contributions of all variables for each study area and each scale.



## Appendix 4

R code will be available at: <https://gitlab.irstea.fr/anouk.glad/effect-of-interaction>

## ANNEX V. SUPPLEMENTARY MATERIAL ARTICLE 3: MAPPING AND MANAGING BIRD HABITAT WITH LiDAR-BASED MULTI-SCALE OBJECT- ORIENTED INDICATORS

### Remote sensing in ecology and conservation

#### Appendix 1

Figure A1.1 Correlation (Pearson) between AUC and Boyce index for each case.

Case	Correlation
Female PC	0.72
Female OO	0.41
Male PC	0.24
Male OO	0.92

Figure A1.2. Boyce index plot, with for x axis the habitat suitability and for the y axis the Predicted/Expected ratio. (female models)

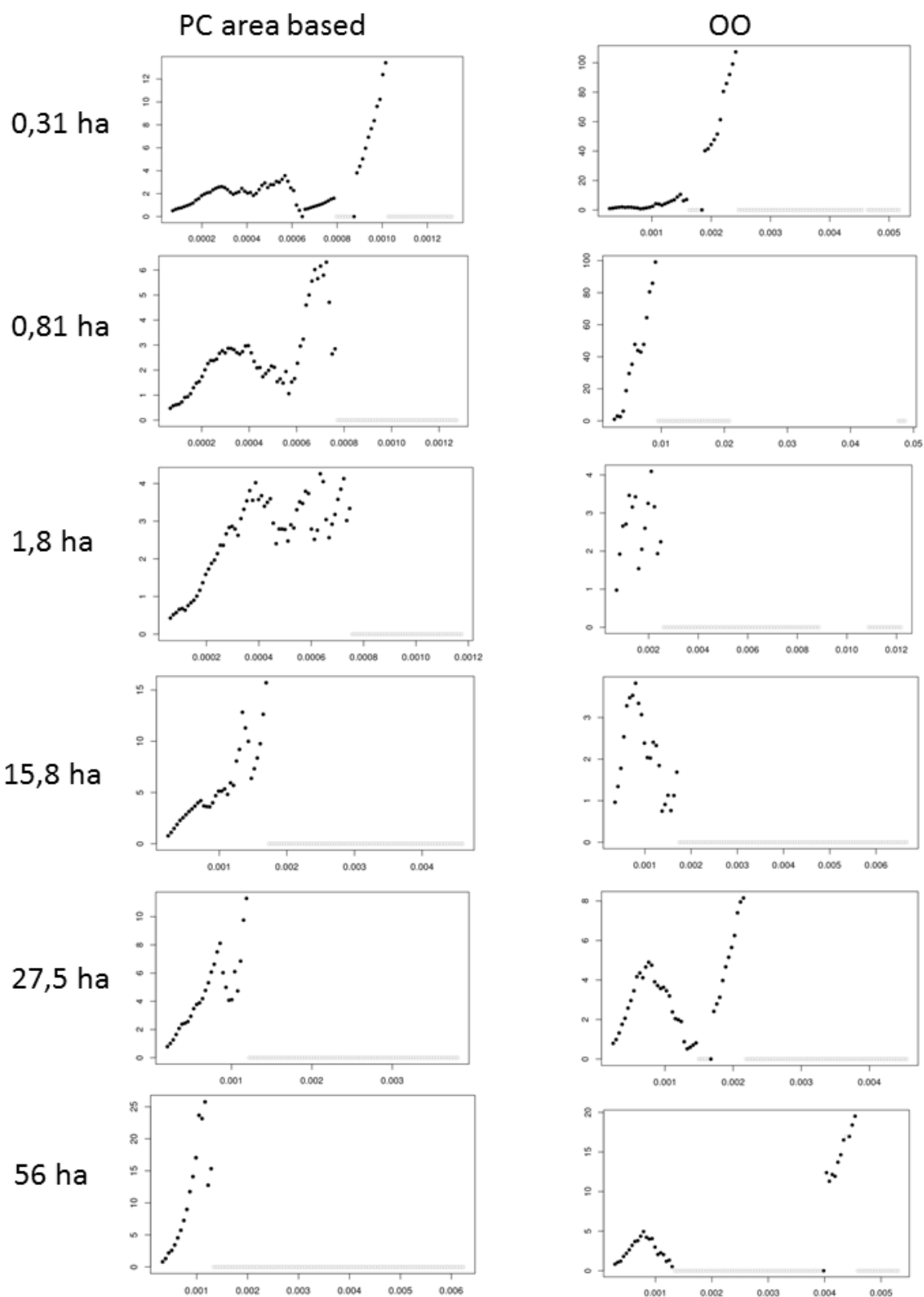


Figure A1.3. Boyce index plot, with for x axis the habitat suitability and for the y axis the Predicted/Expected ratio. (male models)

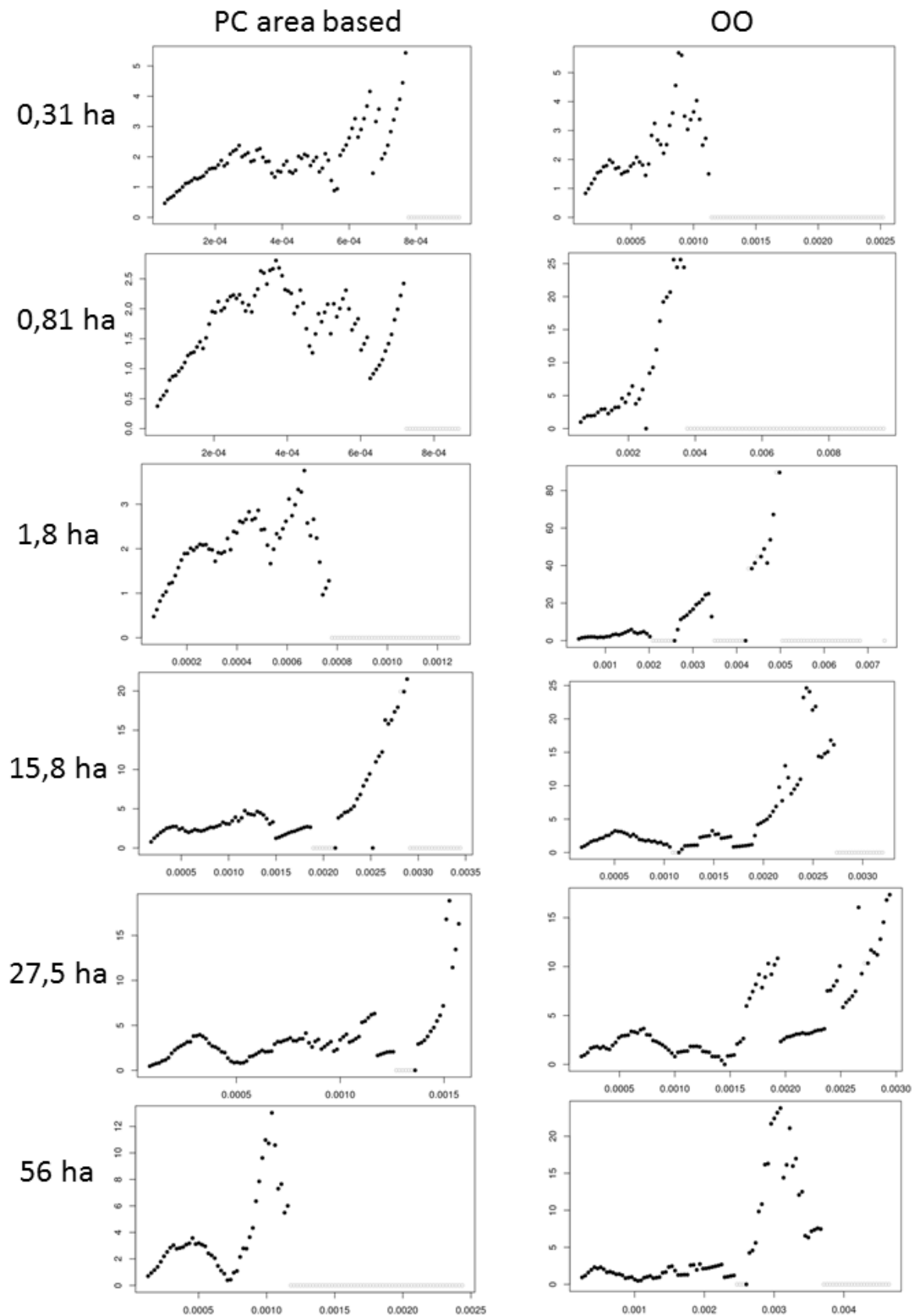


Figure A2.1. Response curves to the major contributing variable distance do ski runs (female)

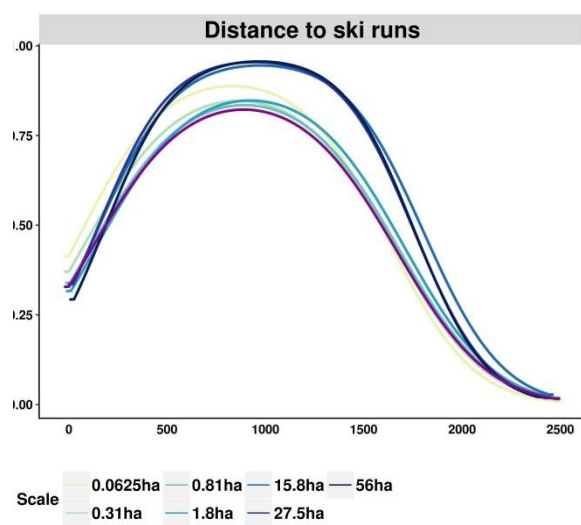
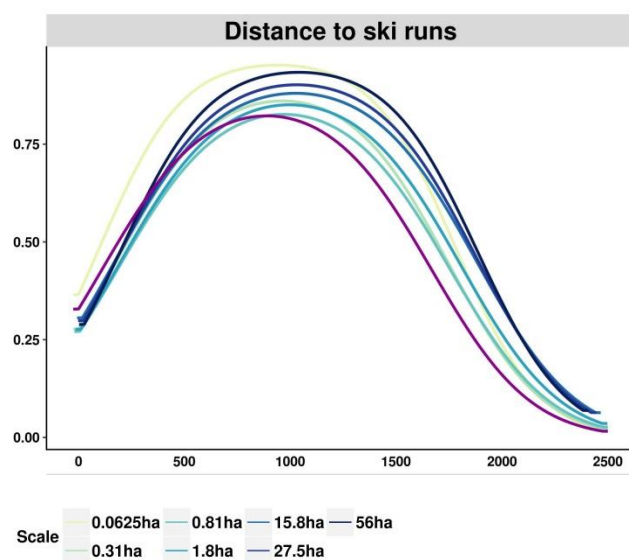
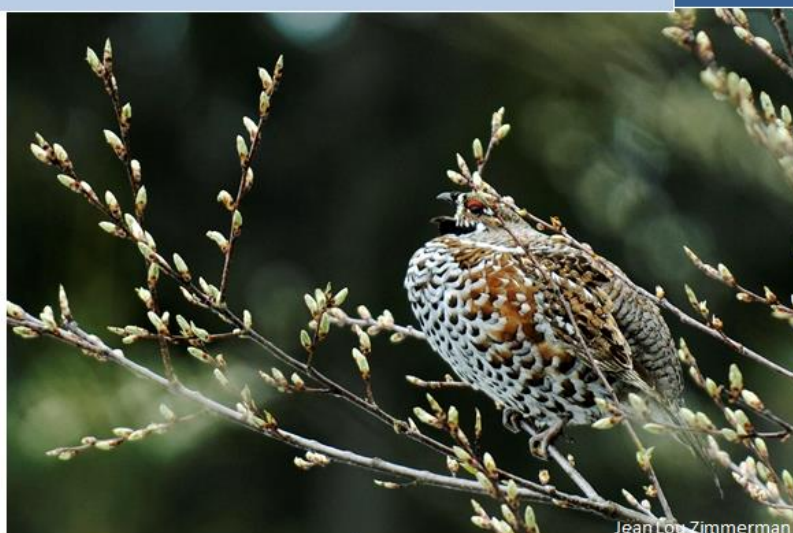


Figure A2.1. Response curves to the major contributing variable distance do ski runs (male)



## ANNEX VI. FRENCH REPORT ON HAZEL GROUSE DISTRIBUTIONS MODELS

### Modèles de distributions de la Gélinotte des bois dans le massif Jurassien à partir de données LiDAR



Jean-Lou Zimmerman



région **BOURGOGNE  
FRANCHE-COMTÉ**



Glad Anouk  
Groupe Tétràs Jura, Irstea centre de  
Grenoble unité LESSEM  
30/03/2018

Note : Ce rapport constitue un résumé des travaux réalisés pendant la thèse (collaboration avec l'IRSTEA Grenoble LESSEM, encadrée par Björn Reineking et Jean-Matthieu Monnet et financée par le conseil départemental de l'Ain). Il présente les différentes méthodes mises en œuvre et les résultats obtenus dans la réalisation des modèles spatiaux pour la Gélinothe des Bois. Pour plus de détails concernant les divers aspects présentés ici, se référer aux chapitres et articles correspondant dans la thèse.



## Remerciements

Je souhaite ici remercier tous ceux qui ont participé au projet, en particulier mes encadrants Björn Reineking et Jean-Matthieu Monnet pour leur aide tout au long de la thèse. Je remercie aussi Alexandra Depraz et Marc Montadert qui ont apporté leurs connaissances sur la Gélinothe afin de m'aider dans la réalisation des modèles.

## Documents associés :

### Cartes de prédictions au format raster pour deux zones d'études à différentes échelles.

Les cartes de prédiction sont rangées dans le dossier « ProjetQGIS\_cartes/DATA/Gelinotte » par échelle d'analyse dans des sous-dossiers séparés. Les fichiers sont aussi nommés avec un selon la zone d'étude et de l'échelle d'analyse.

Nom données de prédictions brutes : Cartesbrutes/hiver /Zoneétude\_GEL\_Echelle\_exponential

Nom données de prédictions catégorisées : Classes /hiver/Zoneétude\_GEL\_Echelle

Un projet QGIS nommé CARTES\_MODELE.qgs permet de voir directement les cartes classifiées avec un fond de carte.

### Cartes des variables LiDAR utilisées pour les modèles au format raster :

Ces données sont rangées dans le dossier « ProjetQGIS\_cartes/DATA/Gelinotte /Variables » par échelle d'analyse et par type (moyenne ou déviation standard) dans des sous-dossiers séparés.

Type de variable	Base nom fichier	Nom d'utilisation
Nuage de points	sdH.nb20_30relative_density	Standard deviation de la densité de la canopée 20-30m
	sdH.nb0.5_1ratio	Standard deviation du ratio de pénétration 0,5-1m
	MeanH.nb2_5ratio	Ratio de pénétration 2-5m
	sdH.nb2_5ratio	Standard deviation du ratio de pénétration 2-5m
Orienté-objet	Meantreeinf10.density	Densité des arbres (<10m)
	Meantree10.density	Densité des arbres (>10m)
	Meantree.giniH	Indice de gini
	MeanG.s200_1000	Pourcentage de trouées de taille moyenne
	MeanG.s1000_Inf	Pourcentage de zones ouvertes
Données ONF placettes et LiDAR (Ain seulement)	sdprop_conifG2	standard deviation de la proportion de conifères (Gha)
	meanprop_conifG2	Proportion de conifères (Gha)
Données satellite sentinel (Theia Cnes) (Jura seulement)	sdndvi	Standard déviation NDVI
	meanndvi	NDVI
Aurtes (Ain seulement)	bati_surfaceratio	Ratio de la surface des batiments
	MeanRoad	Distance aux routes principales

<b><u>1</u></b>	<b><u>INTRODUCTION .....</u></b>	<b><u>312</u></b>
<b><u>2</u></b>	<b><u>METHODES.....</u></b>	<b><u>315</u></b>
2.1	JEUX DE DONNEES DES OBSERVATIONS DE GELINOTTE DES BOIS .....	315
2.2	JEUX DE DONNEES LIDAR .....	315
2.3	CALCUL DES VARIABLES LIDAR.....	316
2.4	SELECTION DES VARIABLES .....	317
2.5	ECHELLES SPATIALES .....	320
2.6	MODELISATION .....	321
2.7	METHODE DE CONSTRUCTION DES MODELES MULTI-ECHELLES .....	324
<b><u>3</u></b>	<b><u>RESULTATS.....</u></b>	<b><u>326</u></b>
3.1	PERFORMANCE DES MODELES .....	326
3.1.1	MODELES A ECHELLE UNIQUE ET VARIABLES MULTIPLES .....	326
3.1.2	MODELE A ECHELLES MULTIPLES ET VARIABLES MULTIPLES (MIXTE).....	326
3.2	CONTRIBUTION DES VARIABLES .....	328
3.2.1	CONTRIBUTION DE TOUTES LES VARIABLES.....	328
3.2.2	CONTRIBUTION DES VARIABLES COMMUNES .....	329
3.2.3	CONTRIBUTION DES VARIABLES A PETITE ECHELLE .....	330
3.3	COURBES DE REPONSES .....	331
3.4	INTERPRETATION DES CARTES DE PREDICTION .....	335
3.5	CREATION DE CATEGORIES POUR LES CARTES DE PREDICTIONS .....	341
3.6	VALIDATION EN FORET DE CHAMPFROMIER .....	342
<b><u>4</u></b>	<b><u>DISCUSSION .....</u></b>	<b><u>345</u></b>
<b><u>5</u></b>	<b><u>BIBLIOGRAPHIE .....</u></b>	<b><u>348</u></b>

## 1 INTRODUCTION

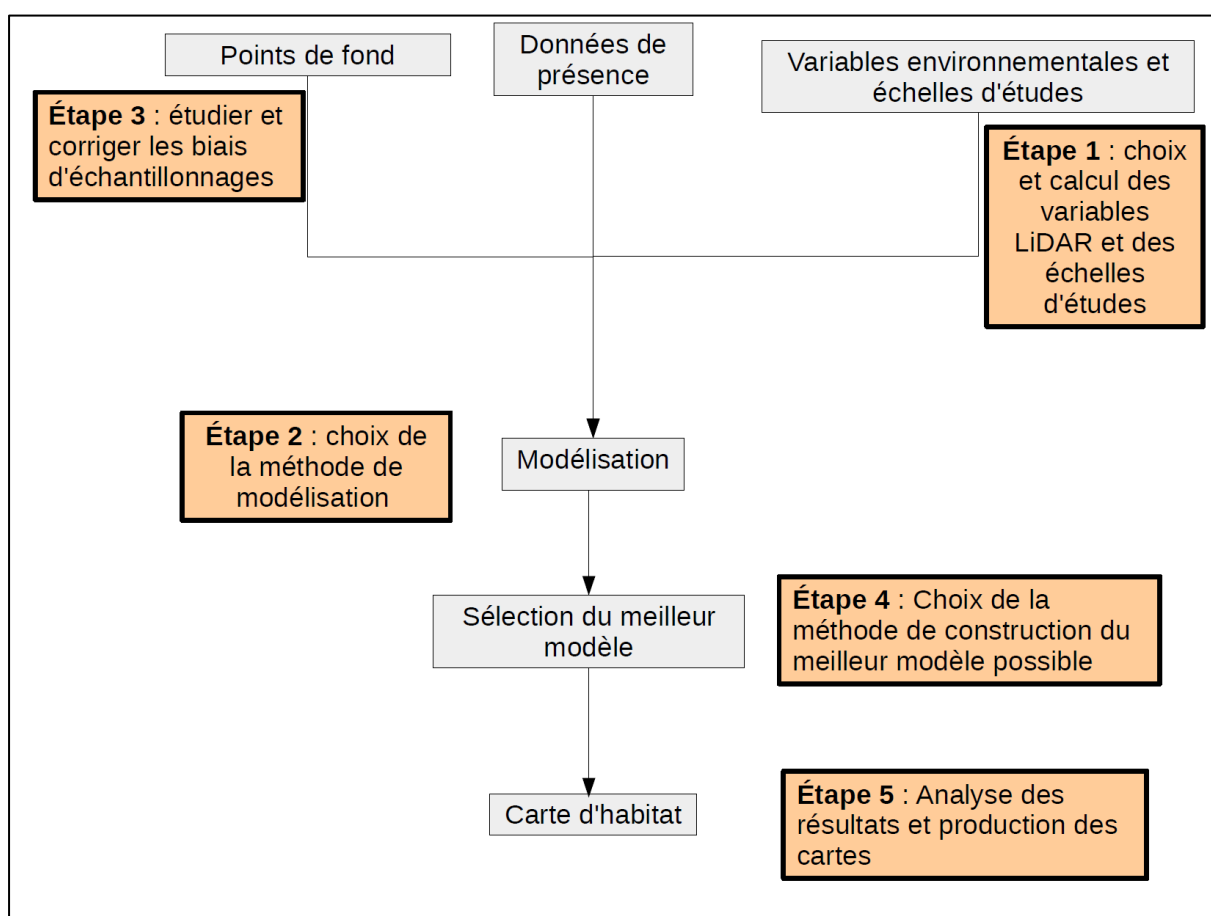
---

Dans le massif Jurassien, les populations de Gélinothe des bois sont en déclin. Cependant la distribution complète de l'espèce sur ce vaste territoire est aujourd'hui méconnue en particulier dans les zones de basse altitude (Bellon 2011). Le Groupe Tétraz Jura dans sa mission d'étude et de protection des galliformes de montagne présent dans le massif souhaite acquérir pour les années futures de nouvelles données sur les populations de Gélinothe des bois Jurassiennes. La zone étant très vaste, la nécessité d'optimiser le plan de prospection selon la richesse potentielle des massifs (probabilité de présence forte, incertaine, faible) et de maîtriser les moyens d'inventaire a été soulignée. Dans ce cadre, l'idée de développer des modèles de distribution d'espèces a été proposée. Les modèles de distribution d'espèce, aussi appelés modèles de niche, sont aujourd'hui communément utilisés en écologie et biologie de la conservation (Franklin 2009b). Ces modèles permettent de relier les présences d'individus aux caractéristiques du paysage où ils ont été observés. Ainsi, ces modèles permettent de connaître la réponse de l'espèce pour chaque variable utilisée et de faire des prédictions pour des zones où les données sur l'espèce sont manquantes. On obtient alors une carte de prédiction qui montre les zones où l'habitat est favorable ou non à la persistance de l'espèce (Elith et al. 2011). De nombreuses étapes et données sont nécessaires pour construire un modèle de distribution d'espèces (Figure 1).

Les caractéristiques du paysage peuvent être représentées par de nombreuses variables dans l'espace et le temps. Récemment, le développement des systèmes d'acquisition d'image, notamment satellite, a permis de couvrir des zones de plus en plus larges. Ces dernières années, de nombreuses études ont montré la pertinence de l'utilisation de la technique **LiDAR** pour extraire des variables d'environnement dans les milieux forestiers (Bae et al. 2014; Rechsteiner et al. 2017). Grâce au survol LiDAR des forêts du massif (Jura, Doubs, Ain), des variables

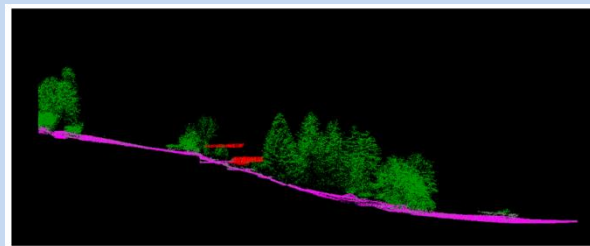
d'environnement permettant de connaître précisément la structure tridimensionnelle du couvert végétal ont été calculées pour les modèles de distribution d'espèces. Le calcul et le choix de ces variables n'est que la première étape de la méthode permettant de construire ces modèles. En tout, cinq étapes majeures et successives ont été nécessaires à l'élaboration des modèles (figure 1). Ces différentes étapes sont décrites en détails dans les sections: méthodes et résultats.

*Figure 1 : Les étapes de la création d'un modèle de distribution d'espèces.*



## **LiDAR**

Le LIDAR (ou télédétection par laser) dont le nom provient de l'anglais **Light Detection And Ranging**, est une technique qui permet, grâce à l'émission d'une onde laser et à la réception de signaux de cette même onde réfléchis par les objets qu'elle rencontre, de calculer la distance entre l'appareil et l'objet. Cette distance est calculée à partir du temps écoulé entre l'émission et la réception du signal. L'appareil peut être utilisé au sol, mais il démontre surtout son avantage lorsqu'il est embarqué à bord d'un avion ou d'un hélicoptère. Dans ce cas, le laser est dirigé vers le sol, ce qui permet de calculer la distance entre l'avion et les objets rencontrés par l'onde laser. Ainsi, en connaissant l'altitude du vol et la position de l'avion à la verticale de ce point grâce au système GPS, il est possible de connaître la hauteur des objets (Maisons, arbres, buissons ...) à la surface du sol. Ces acquisitions par avion permettent à la fois de couvrir une très grande surface et d'avoir un nuage de points en 3 dimensions (latitude, longitude et hauteur) très précis (résolution jusqu'à 15cm grâce à l'envoi de 10 impulsions laser par m<sup>2</sup>).



Mais le plus grand avantage d'utiliser une onde laser, c'est qu'elle peut pénétrer le couvert végétal. Ainsi, une impulsion peut en réalité renvoyer plusieurs signaux de retour : une première partie du faisceau laser est réfléchi par le feuillage mais le reste traverse et est ensuite réfléchi par l'objet situé au-dessous (une autre partie du feuillage, une branche, le tronc, un buisson, et même le sol). Ainsi, grâce à une seule émission, on peut recevoir jusqu'à 4 ou 5 signaux de retour. Cela permet d'avoir non seulement des informations sur la hauteur de la canopée, mais aussi de produire dans une certaine mesure une cartographie des sous-bois et des reliefs du sol.

## 2 MÉTHODES

---

### 2.1 Jeux de données des observations de Gélinothe des Bois

Les données des prospections d'hiver récoltées entre 2007 et 2016 par le Groupe Tétraz Jura ont été utilisées. Les observations sont en majorité des indices de présence localisés en utilisant un GPS. Les tracés des prospections de chaque observateur a aussi été enregistré et permet de savoir très précisément quelles zones ont été prospectées (effort d'échantillonnage). Un total de 1100 localisations a été utilisé pour faire les modèles de distribution d'espèces (961 dans le Jura et le Doubs et 139 dans le département de l'Ain).

### 2.2 Jeux de données LiDAR

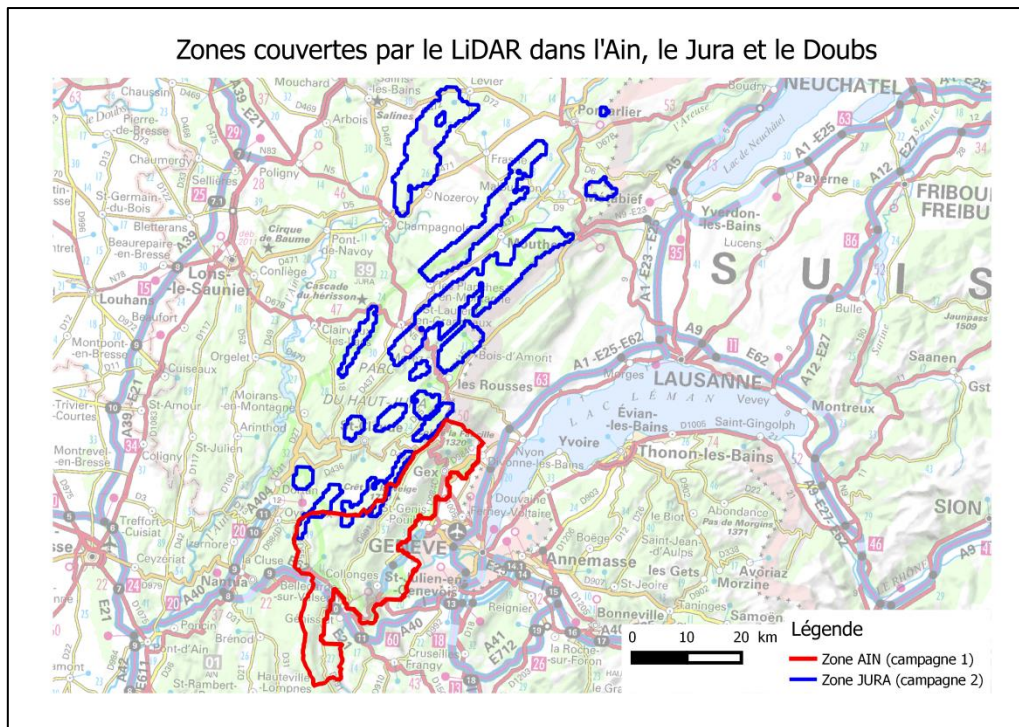
Les données LiDAR proviennent de deux campagnes différentes. La première s'est déroulée en automne 2014 et couvre 626km<sup>2</sup> du département de l'Ain. La seconde s'est déroulée en été 2016, et couvre les zones d'intérêts pour la Gélinothe du Jura et du Doubs (431km<sup>2</sup>). Il faut souligner ici que, à cause de plusieurs différences entre ces campagnes, des modèles seront proposé pour chaque zone séparément.

Le deuxième point est le fait que la zone du département de l'Ain recouvre une surface continue tandis que la deuxième campagne cible les zones forestières. Ainsi, une attention particulière doit être accordée, lors de l'interprétation des résultats du modèle aux variations présente du paysage de chaque zone.

Pour la suite du rapport et pour des raisons pratiques, les modèles construits avec le LiDAR de la première campagne seront appelés « modèles de l'Ain », et ceux issus des données de la deuxième campagne « modèles du Jura ».



Figure 2 : Les couvertures LiDAR des deux zones d'études du massif Jurassien

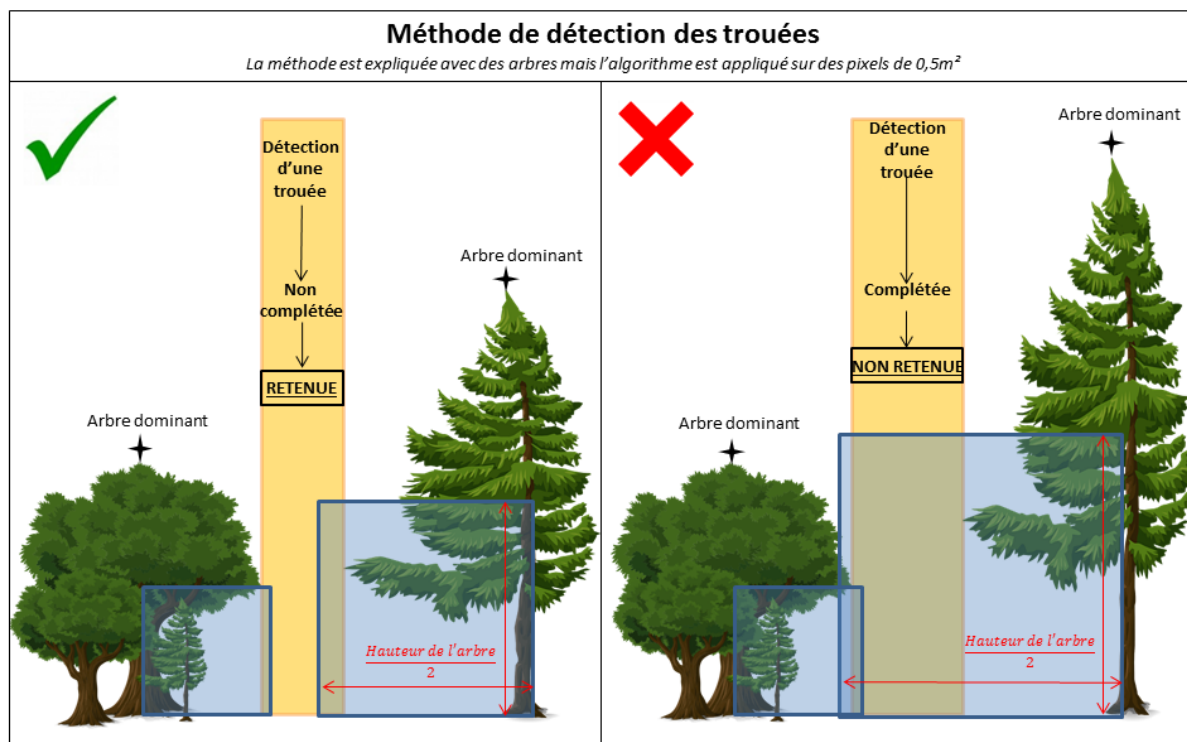


### 2.3 Calcul des variables LiDAR

Les variables LiDAR ont été calculées grâce aux packages LidR (Roussel and Auty 2016) et LidarRtree (en développement par Jean-Matthieu Monnet) sur le logiciel R. Deux types de variables ont été extraits, les métriques nuages de points et les métriques orientées-objet. Les métriques dites nuage de point sont des valeurs comme la moyenne de la hauteur, la densité des points, l'écart-type de la hauteur calculées sur une surface définie de 25m\*25m (taille du pixel choisie). En revanche, les métriques orientées-objet sont constituées d'éléments précis existant dans le paysage tel que les arbres ou les trouées. Ces objets sont extraits du nuage de point grâce à des caractéristiques définies à l'avance à une résolution de 0.5m. Les arbres ont été extraits avec la méthode de détection des sommets. Les trouées ont été définies comme étant les zones d'une hauteur de canopée inférieure à 1m, entourée d'arbres mesurant moins de la moitié de sa largeur (Figure 3). Par la suite, la moyenne résumant les caractéristiques des objets a été calculée pour une surface de 25m\*25m (taille du pixel). Enfin à partir des objets arbre de la zone de l'Ain et grâce aux données des placettes permanentes de l'ONF (localisation

des arbres sur chaque placette), la proportion de résineux par rapport aux feuillus a pu être estimée (Eysn et al. 2015).

Figure 3 : Illustration de la méthode utilisée pour la détection des trouées.



## 2.4 Sélection des variables

Les variables ont été sélectionnées après une première étape de recherche bibliographique et de consultation experts. Elles ont été choisies afin de représenter les différents éléments du paysage potentiellement importants pour la Gélinothe des bois (Tableau 1).

La densité du couvert est un élément essentiel pour cette espèce. En effet, une couverture de la canopée modérée a souvent été déterminée comme étant favorable pour l'oiseau (Hofstetter et al. 2015; Zellweger et al. 2014). Dans ce but, nous avons sélectionné plusieurs variables pouvant représenter ces variations dans le paysage : la densité des arbres de hauteur de 5 à 10m, la densité des arbres de hauteur supérieure à 10m ainsi que la variable pourcentage de trouées de taille moyenne (200 à 1000m<sup>2</sup>).

La Gélinothe des bois favorise un paysage hétérogène horizontalement, en particulier concernant le sous étage, alternant des zones de densité végétale plus ou moins fortes (Mathys et al. 2006). Pour étudier cet aspect, trois variables ont été retenues : la déviation standard de la densité de la canopée entre 10 et 20m et la déviation standard du ratio de pénétration entre 0.5-1m et 2-5m.

De plus, la Gélinothe des bois est communément associée à des forêts bien structurées verticalement (Schäublin and Bollmann 2011). L'ajout d'une variable indice de Gini pour la hauteur des arbres a donc été décidé.

La Gélinothe des bois n'est pas présente dans les milieux ouverts. Pour tenir compte de cet aspect dans les modèles, la variable proportion de zones ouvertes a été sélectionnée. Ces zones ouvertes peuvent représenter aussi bien des zones d'alpages que des champs ou des zones urbaines.

Enfin, dans le but d'incorporer dans les modèles l'effet de proportion des espèces végétales forestières sur la Gélinothe des bois, deux autres variables ont été calculées. En effet, la Gélinothe est principalement présente en France dans des forêts mixtes à dominante conifères (Sachot, Perrin, and Neet 2003). Pour la zone de l'Ain, la proportion de résineux ainsi que la déviation standard de cette proportion ont été utilisées. En revanche, pour la zone d'étude du Jura, cette proportion n'a pas pu être calculée. Ainsi, l'indice NDVI (Normalized Difference Vegetation Index) a été utilisé dans le but de mesurer les variations de productions dans le couvert végétal (données du 17/07/2016, Copernicus Sentinel-2 2016 par le CNES pour le centre THEIA).

Le dérangement par les activités humaines a aussi été souligné comme pouvant influencer la présence de l'espèce, même si cela affecte la Gélinothe des bois dans une moindre mesure comparé au Grand Tétras (Räty 1979; Leclercq 1985). Ainsi pour la zone d'étude de l'Ain qui

représente le paysage dans sa totalité, deux variables ont été ajoutées, afin d’avoir des prédictions plus fiables.

- La distance aux routes principales (Institut national de l'information géographique et forestière IGN Route 500®), où uniquement les axes locaux à nationaux reliant les principaux centre urbains et infrastructures ont été retenus.
- Le ratio de la surface des bâtiments à différentes échelles spatiales afin d’incorporer la présence des centres urbains (IGN BDTOPO®).

*Tableau 1 : Variables sélectionnées*

Type de variable	Nom	Description	Indice	Importance pour la Gélinotte	References
Nuage de points	Standard deviation de la densité de la canopée 20-30m	Densité des points 20 et 30m (déviati on standard)	Indice de l'hétérogénéité horizontale de la couverture de la canopée	L'Habitat sera favorable si la surface couverte par la végétation est modérée (40% -60%) (>5m), de grand disparités indiquent un paysage hétérogène	Mathys 2000, Hofstetter 2015
	Standard deviation du ratio de pénétration 0,5-1m	Ratio de pénétration 0,5 et 1m (déviati on standard)	Indice de l'hétérogénéité horizontale de la densité du sous bois	Présence de Myrtilles ou de strate herbacée (élevage des jeunes), une très forte densité de point pourrait indiquer un couvert non favorable	Mathys 2006, Aberg 2003, Zellwger 2014, Katjoch 2012 , Sachot 2003
	Ratio de pénétration 2-5m	Ratio de pénétration 2 et 5m	Indice de densité de la strate arbustive basse	Abris et zones d'alimentation pour la Gélinotte (si présence de saules, sorbier)	Mathys 2006, Schäublin 2011, Montadert (pers. Com)
	Standard deviation du ratio de pénétration 2-5m	Ratio de pénétration 2 et 5m (déviati on standard)	Indice de l'hétérogénéité horizontale de densité de la strate arbustive basse	Abris et zones d'alimentation pour la Gélinotte (si présence de saules, sorbier)	Mathys 2006, Schäublin 2011, Montadert (pers. Com)
Orienté-objet	Densité des arbres (<10m)	Densité des arbres mesurant moins de 10m (nombre/ha)	Indice du couvert de la strate arbustive	L'Habitat sera favorable si la surface couverte par la végétation est modérée (40% -60%) (>5m)	Mathys 2000, Hofstetter 2015
	Densité des arbres (>10m)	Densité des arbres mesurant plus de 10m (nombre/ha)	Indice du couvert de la canopée	L'Habitat sera favorable si la surface couverte par la végétation est modérée (40% -60%) (>5m)	Mathys 2000, Hofstetter 2015
	Indice de gini	Indice de Gini pour la hauteur des arbres	Indice d'hétérogénéité des hauteurs	Les habitats hétérogènes sont favorisés	Mathys 2006, Zellweger 2014, Schäublin 2011
	Pourcentage de trouées de taille moyenne	Proportion de trouées larges de 200m <sup>2</sup> à 1000m <sup>2</sup> (Hauteur <1m et la demi hauteur des arbres autour plus petite que la demi-largeur de la trouée)	Pourcentage de moyenne trouées	Présence de trouées où la myrtille peut être présente (attention, les zones découvertes telle que les routes/prairies sont aussi mises en valeurs)	Mathys 2006, Aberg 2003, Zellwger 2014, Katjoch 2012, Montadert, Sachot 2003
	Pourcentage de zones ouvertes	Proportion de trouées supérieures à 1000m <sup>2</sup> (Hauteur <1m et la demi hauteur des arbres autour plus petite que la demi-largeur de la trouée)	Pourcentage de zones ouvertes	Défavorable pour l'espèce	A.Depraz (pers. Com)
Données ONF placettes et LiDAR (Ain seulement)	standard deviation de la proportion de conifères (Gha)	Proportion de conifères en surface terrière (parmi les arbres de diamètre > 17,5, données de l'ONF) (déviati on standard)	Indice de l'hétérogénéité horizontale de la composition en essences de l'habitat	Préfère les habitat mixtes, avec présence de buissons, perches de sorbier et d'une futaie résineuse	Mathys 2006, Aberg 2003, Montadert (pers. Com)
	Proportion de conifères (Gha)	Proportion de conifères en surface terrière (parmi les arbres de diamètre > 17,5, données de l'ONF)	Indice de la composition en essences de l'habitat	Préfère les habitat mixtes, avec présence de buissons, perches de sorbier et d'une futaie résineuse	Mathys 2006, Aberg 2003, Montadert (pers. Com)
Données satellite sentinel (Theia Cnes) (Jura seulement)	Standard déviati on NDVI	Indice de la végétation par différence normalisée (déviati on standard)	Indice de l'hétérogénéité horizontale de la composition en essences de l'habitat	Préfère les habitat mixtes, avec présence de buissons, perches de sorbier et d'une futaie résineuse	Mathys 2006, Aberg 2003, Montadert (pers. Com)
	NDVI	Indice de la végétation par différence normalisée	Indice de la composition en essences de l'habitat	Préfère les habitat mixtes, avec présence de buissons, perches de sorbier et d'un futaie résineuse	Mathys 2006, Aberg 2003, Montadert (pers. Com)
Autres (Ain seulement)	Ratio de la surface des bâtiments	Surface totale des bâtiments sur la surface de chaque échelle	Activités humaines	Défavorable pour l'espèce dans une certaine mesure	Räty 1979, Leclerq 1985
	Distance aux routes principales	Distance aux principales routes	Dérangement	Défavorable pour l'espèce dans une certaine mesure	Räty 1979, Leclerq 198

## 2.5 Echelles spatiales

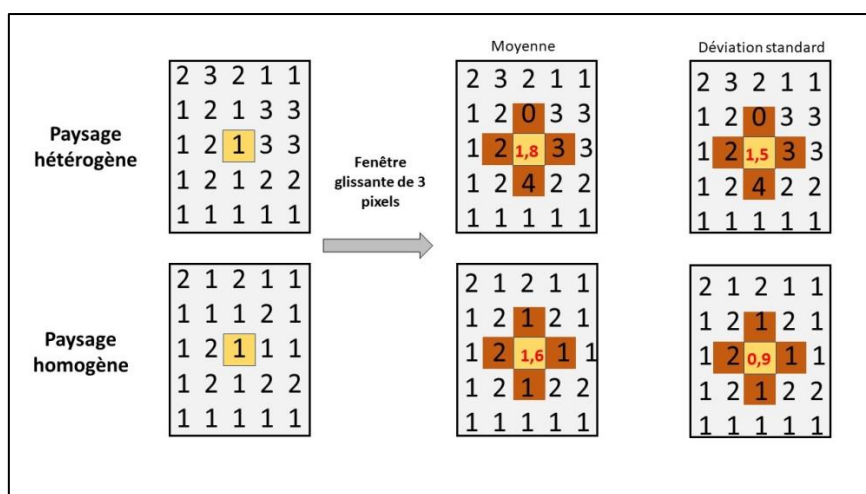
Sept échelles spatiales ont été sélectionnées pour les modèles de la Gélinotte des bois (Tableau 2). Les deux plus larges ont été sélectionnées pour représenter la surface potentielle d'un domaine vital. Trois échelles ont été sélectionnées pour représenter le niveau de la tâche d'habitat. Enfin, deux autres échelles représentent une évaluation de l'environnement au niveau du micro-habitat.

Chaque variable a ainsi été calculée à nouveau à l'aide d'une fenêtre glissante (moyenne et déviation standard). Pour les zones de bordures, les valeurs réelles ont été répliquées là où il n'y en avait pas afin de limiter l'effet de bord. Cependant, cela part du principe que le milieu de chaque côté de la bordure est similaire. Une zone tampon de 25m a été appliquée par la suite afin de minimiser cet effet. La plus petite échelle (0.0625ha) représente les données brutes par pixel de 25\*25m. Ainsi, pour cette échelle, les valeurs de déviation standard ne sont pas connues.

*Tableau 2: Les échelles sélectionnées pour les modèles de la Gélinotte des bois.*

Gélinotte des bois			
Niveau	definition	Echelles	Objectifs
1	Niveau domaine vital	27,5ha	Prendre en compte les différents habitats sur une échelle large
		15,8ha	
2	Niveau de la tâche d'habitat	7ha	Identifier localement des zones favorables, dans un objectif de gestion
		1,8ha	
		0,81ha	
3	Niveau micro-habitat	0,31ha	
		0,0625ha	

Figure 4 : Schéma du calcul des variables par fenêtres glissantes.



## 2.6 Modélisation

La méthode de modélisation utilisée est Maxent, réalisée à l'aide du package R Maxnet (Phillips et al. 2017). Cette méthode permet d'utiliser des données d'espèces appelées « présence uniquement », car elles ne contiennent pas de localisation où l'absence de l'espèce a été contrôlée. 10 000 points de fond ont été créés dans le but de donner au modèle des informations sur le paysage présent dans les zones prospectées. Ces points de fond ont été localisés le long des tracés de prospection, dans le but d'éviter un **effet du biais d'échantillonnage** sur les résultats du modèle.

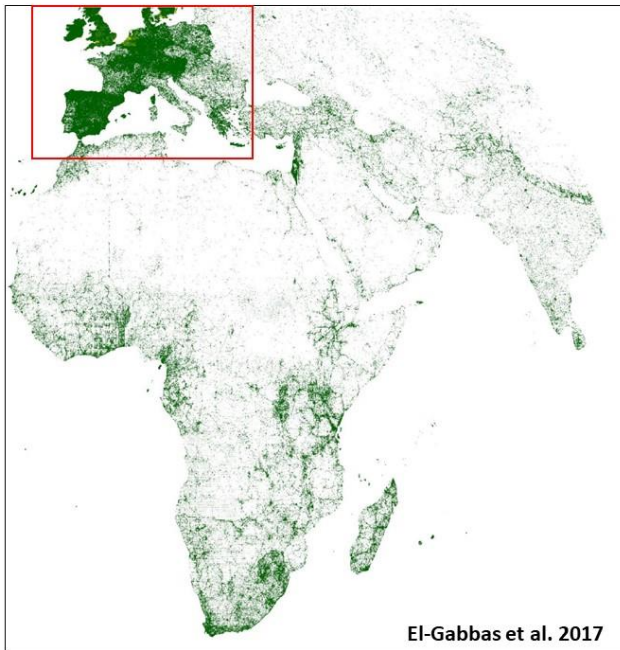
La corrélation entre les variables sélectionnées précédemment est évaluée 2 à 2, et lorsque la valeur est supérieure à 0.7, l'une des deux variables n'est pas conservée dans le modèle (à l'exception de la variable proportion de zones ouvertes). Ces variables sont ensuite normalisées par la formule suivante :

$$\text{Valeur pixel normalisée} = \frac{\text{valeur pixel} - \text{moyenne paysage}}{\text{écart type paysage}}$$



## Biais d'échantillonnage

Le biais d'échantillonnage est connu pour influencer les résultats des modèles de distributions d'espèces (Kramer-Schadt et al. 2013; El-Gabbas et Dormann, 2017). En effet, si l'échantillonnage n'est pas systématique ou aléatoire une partie du paysage peut être prospectée de manière plus intense qu'une autre. C'est notamment le cas des données participatives où l'effort d'échantillonnage ne peut être contrôlé. Ainsi, si l'effet de ce biais n'est pas corrigé, le modèle estimera de bonne qualité les zones qui ont été favorisées par l'échantillonnage plutôt que les zones favorables à l'espèce. La difficulté est de réussir à estimer ce biais alors même que généralement les informations disponibles sont très réduites.



Carte produite avec toutes les données terrestre enregistrées dans la base de données GBIF (Global Biodiversity Information Facility)

On observe ici un biais d'échantillonnage important où l'**Europe** possède une très grande densité d'observations. Hors, il existe aussi une très grande biodiversité en Afrique ou au Moyen-Orient. Ils apparaissent ici avec beaucoup moins d'observations car ils ont été moins souvent échantillonnés donnant l'impression fausse d'une moins forte présence d'animaux et végétaux terrestres.

Pour chaque modèle, la valeur **Area Under Curve (AUC)** a été calculée. Cette valeur permet de comparer les performances des modèles entre elles. En effet, cette valeur est calculée en effectuant un test de prédiction du modèle sur une partie du jeu de donnée (présences et points de fond), et permet donc d'évaluer la capacité du modèle à prédire la qualité de l'habitat. De plus, pour tous les modèles, la contribution de chaque variable a été calculée, en additionnant les valeurs betas du modèle. Seules les variables avec une contribution supérieure à 0.05 sont présentées dans les résultats. Les courbes de réponses sont calculées en prédisant la qualité de l'habitat pour une large gamme de valeurs d'une variable, pendant que toutes les autres sont fixées à leur valeur moyenne.

Plusieurs composantes ont été sélectionnées pour ces modèles. La première est la composante linéaire, qui permet de contraindre la réponse de l'espèce à une variable à une réponse linéaire. Cependant pour plus de réalisme, car les réponses des individus à leur environnement sont complexes, la composante quadratique a été aussi sélectionnée. De plus, une attention particulière a été apportée à l'introduction d'interactions entre deux variables grâce à la composante « produit ». En effet, en caractérisant le paysage par des variables distinctes, le lien qui existe entre elles disparaît. Par exemple, la diversité du paysage peut être caractérisée par la variation du couvert forestier entre 20m et 30m. Cependant, toutes les zones où aucune canopée n'a été détectée, ne peuvent pas être forcément classées dans une même catégorie : jeune forêt, plantation, pâturages, végétation basse, zones urbaines peuvent constituer ces zones, rendant l'interprétation des résultats difficile. L'introduction d'une interaction entre la déviation standard du couvert de la canopée entre 20 et 30m et la variable pourcentage de zones ouvertes permet ainsi de faire la distinction lors de l'interprétation des résultats, entre les zones avec un couvert forestier, les zones sans couvert forestier que sont les pâturages ou clairières et les zones sans couvert forestier qui ne sont ni l'un ni l'autre.



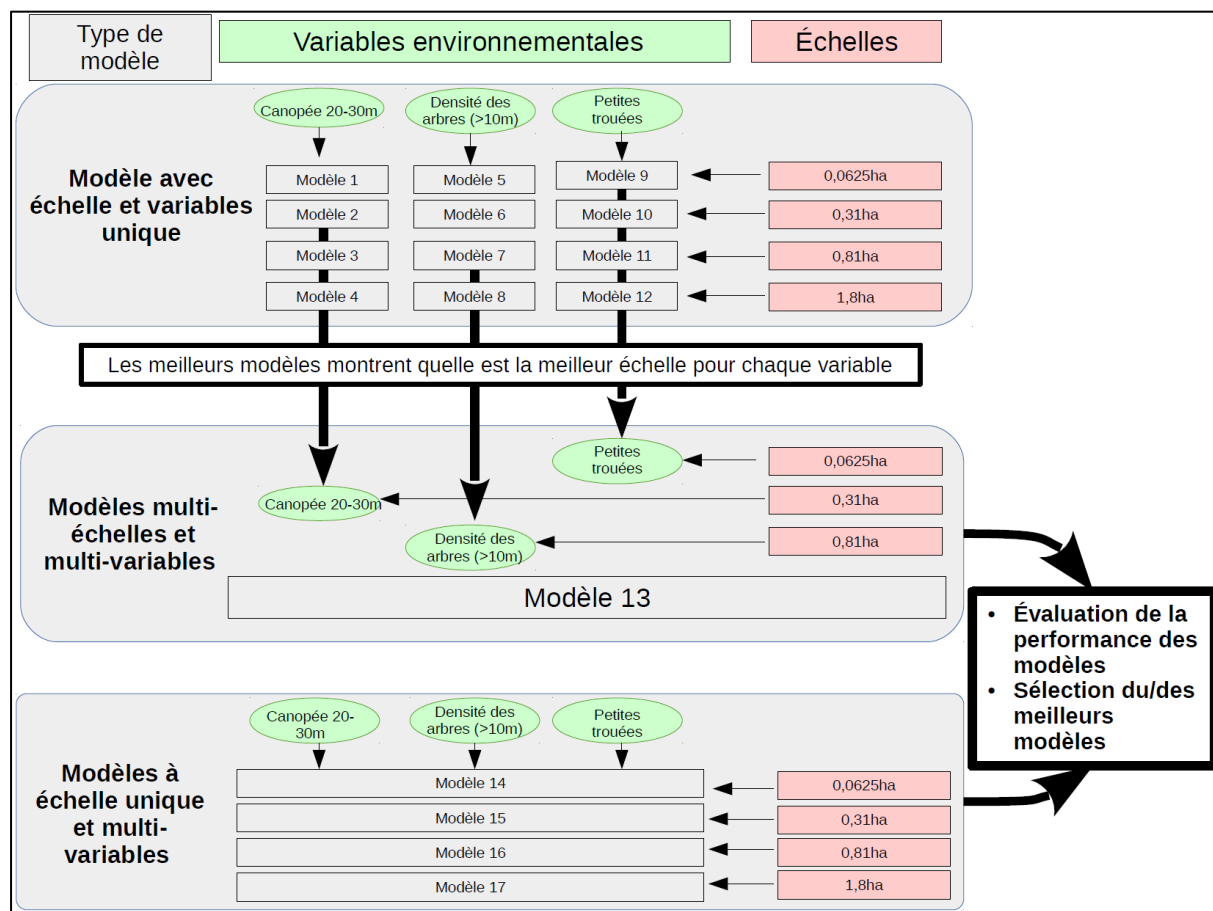
### **L'Area Under Curve (AUC)**

Cette valeur située entre 0 et 1 est calculée en effectuant un test de prédiction du modèle sur une partie du jeu de donnée (présences et points de fond). Elle nous permet de savoir si le modèle prédit la qualité de l'habitat mieux que s'il le faisait de manière aléatoire (valeur de 0.5). Ainsi plus la valeur est proche de 1 plus la performance du modèle est grande.

## **2.7 Méthode de construction des modèles multi-échelles**

Dans l'optique de produire le meilleur modèle possible, la comparaison de modèles à différentes échelles a été faite dans chacune des zones d'études. De plus, deux méthodes ont été utilisées et comparées, celles-ci ont été décrites récemment par McGarigal et al. (2016). La première étape fût de créer des modèles à une seule échelle spatiale mais avec toutes les variables choisies récemment. Ces modèles seront appelés par la suite modèles à échelle-unique et variables multiples. La performance des modèles à ces différentes échelles a ensuite été comparée. Dans un second temps, l'élaboration d'un modèle mixte (multi-échelles et multi-variables) a été faite. Pour cela, des modèles pour chaque échelle et chaque variable ont été créés individuellement. Pour chaque variable, l'échelle à laquelle le modèle montrait une meilleure performance a été conservée pour le modèle mixte final.

Figure 5 : Schéma de la méthode de sélection du meilleur modèle à partir de différentes échelles d'étude (les numéros des modèles sont indicatifs et ne correspondent pas aux modèles délivrés).



## 3 RÉSULTATS

---

### 3.1 Performance des modèles

#### 3.1.1 Modèles à échelle unique et variables multiples

Les modèles de la zone d'étude de l'Ain montrent une meilleure performance que ceux réalisés dans la zone d'étude du Jura (Tableau 3). Dans la zone de l'Ain, le meilleur modèle a été observé pour une échelle de 15.8ha avec une valeur AUC de 0.75. En revanche, pour la zone du Jura, les échelles de 0.81 et 1.8 ha ont donnés de meilleurs résultats.

*Tableau 3 : Résultat AUC pour les modèles à échelle unique et variables multiples.*

Echelle	Zone d'étude	AUC
0.0625ha	JURA	0.59
0.31ha		0.61
0.81ha		0.62
1.8ha		0.62
7ha		0.61
15.8ha		0.61
27.5ha		0.61
0.0625ha	AIN	0.67
0.31ha		0.68
0.81ha		0.7
1.8ha		0.7
7ha		0.74
15.8ha		0.75
27.5ha		0.72

#### 3.1.2 Modèle à échelles multiples et variables multiples (mixte)

Tout d'abord, des modèles à échelle et variable unique ont été réalisés (Tableau 4). Pour certaines variables, une échelle unique donne de meilleures performances, tandis que pour d'autres au contraire plusieurs échelles montrent la même performance. Dans ce cas, la plus petite échelle parmi celles qui ont montré les meilleures performances a été retenue. De plus on remarque que les échelles sélectionnées sont différentes suivant la zone d'étude.

Tableau 4 : Valeurs d'AUC pour les modèles à échelle et variable uniques

Zone d'étude	Echelle	Grassland	Prop_resG	Prop_resGsd	Middle_gap	Tree_gini	Tree_density10	Tree_densityinf10	Ratio0205sd	Ratio0205	Ratio0501sd	Canopy2030sd	Buildings	Road
Ain	0,0625ha	0.58	0.56	NA	0.49	0.53	0.56	0.56	NA	0.64	NA	NA	0.51	0.57
	0,31ha	0.6	0.59	0.58	0.46	0.53	0.58	0.57	0.52	0.68	0.48	0.52	0.51	0.57
	0,81ha	0.6	0.59	0.57	0.48	0.52	0.57	0.57	0.5	0.69	0.54	0.53	0.51	0.57
	1,8ha	0.59	0.6	0.56	0.51	0.52	0.57	0.56	0.49	0.7	0.55	0.5	0.51	0.57
	7ha	0.57	0.59	0.46	0.54	0.54	0.54	0.54	0.5	0.68	0.54	0.48	0.54	0.57
Jura	15,8ha	0.56	0.59	0.47	0.56	0.54	0.52	0.54	0.5	0.67	0.55	0.47	0.55	0.57
	27,5ha	0.55	0.58	0.53	0.57	0.53	0.5	0.54	0.49	0.66	0.57	0.45	0.57	0.57
	0,0625ha	0.5	0.55	NA	0.5	0.53	0.49	0.52	NA	0.57	NA	NA	0.51	0.57
	0,31ha	0.49	0.57	0.51	0.48	0.54	0.48	0.54	0.52	0.59	0.5	0.56	0.51	0.57
	0,81ha	0.49	0.57	0.52	0.49	0.54	0.48	0.55	0.53	0.6	0.49	0.55	0.51	0.57
Jura	1,8ha	0.5	0.57	0.52	0.5	0.53	0.52	0.54	0.53	0.6	0.49	0.55	0.51	0.57
	7ha	0.53	0.55	0.49	0.5	0.52	0.52	0.53	0.55	0.57	0.51	0.52	0.51	0.57
	15,8ha	0.53	0.54	0.49	0.5	0.52	0.52	0.52	0.55	0.57	0.52	0.51	0.51	0.57
	27,5ha	0.53	0.54	0.5	0.5	0.52	0.53	0.49	0.55	0.56	0.52	0.51	0.51	0.57

Avec les variables sélectionnées pour chaque zone d'étude, deux modèles mixtes ont été réalisés (Tableau 5). Le modèle de la zone de l'Ain a une valeur AUC de 0.7 et celui de la zone du Jura de 0.62.

*Tableau 5: Meilleures échelles sélectionnées pour les modèles mixtes.*

Variable	Ain	JURA
Grassland	0,31ha	7ha
NDVI	NA	0,31ha
NDVIsd	NA	0,31ha
Middle_gap	27,5ha	0,0625ha
Tree_gini	7ha	0,31ha
Tree_density10	0,31ha	27,5ha
Tree_densityinf10	0,31ha	0,81ha
Ratio0205sd	0,31ha	7ha
Ratio0205	1,8ha	0,81ha
Ratio0501sd	27,5ha	15,8ha
Canopy2030sd	0,81ha	0,31ha
Prop_resG	0,31ha	NA
Prop_resGsd	0,31ha	NA
Buildings	27,5ha	NA
Road	0,0625ha	NA

### 3.2 Contribution des variables

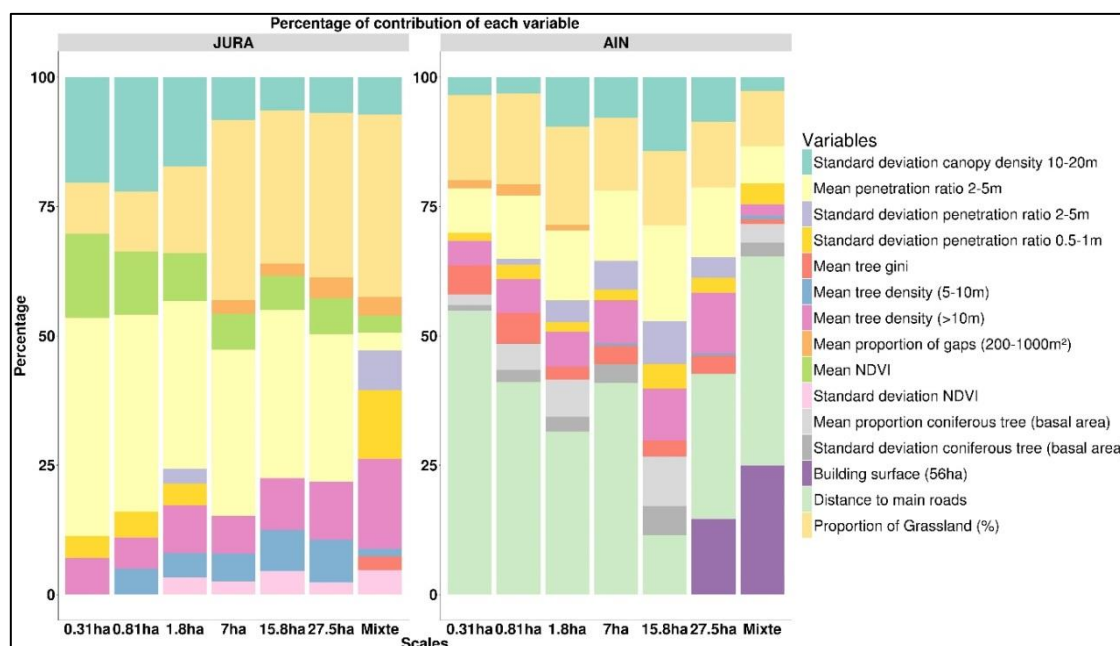
La contribution des différentes variables aux modèles permet de savoir quels sont les éléments du paysage qui influence le plus la présence de la Gélinoite des bois. Ici, les résultats seront présentés en trois temps. Tout d'abord, les contributions de toutes les variables de chaque modèle seront présentées. Dans un second temps, on étudiera plus en détails les contributions des variables communes aux modèles. Enfin, la contribution des variables pour la plus petite échelle sera décrite (les modèles à l'échelle 0.0625ha sont composés de variables différentes).

#### 3.2.1 Contribution de toutes les variables

On remarque que pour le modèle du Jura, les variables qui participent le plus au modèle sont le ratio pénétration moyen entre 2 et 5m, la moyenne du NDVI, le pourcentage moyen de zones ouvertes et la déviation standard de la densité de la canopée entre 20 et 30 m. Pour le modèle mixte en revanche les variables, pourcentage moyen de zones ouvertes, densité moyenne des

arbres (>10m), la déviation standard entre 0.5 et 1 m. Pour les modèles de l'Ain on remarque que la variable distance aux routes principales contribue beaucoup aux modèles. À la plus large échelle et dans le modèle mixte, le ratio de la surface des bâtiments participe aussi dans une grande proportion aux modèles.

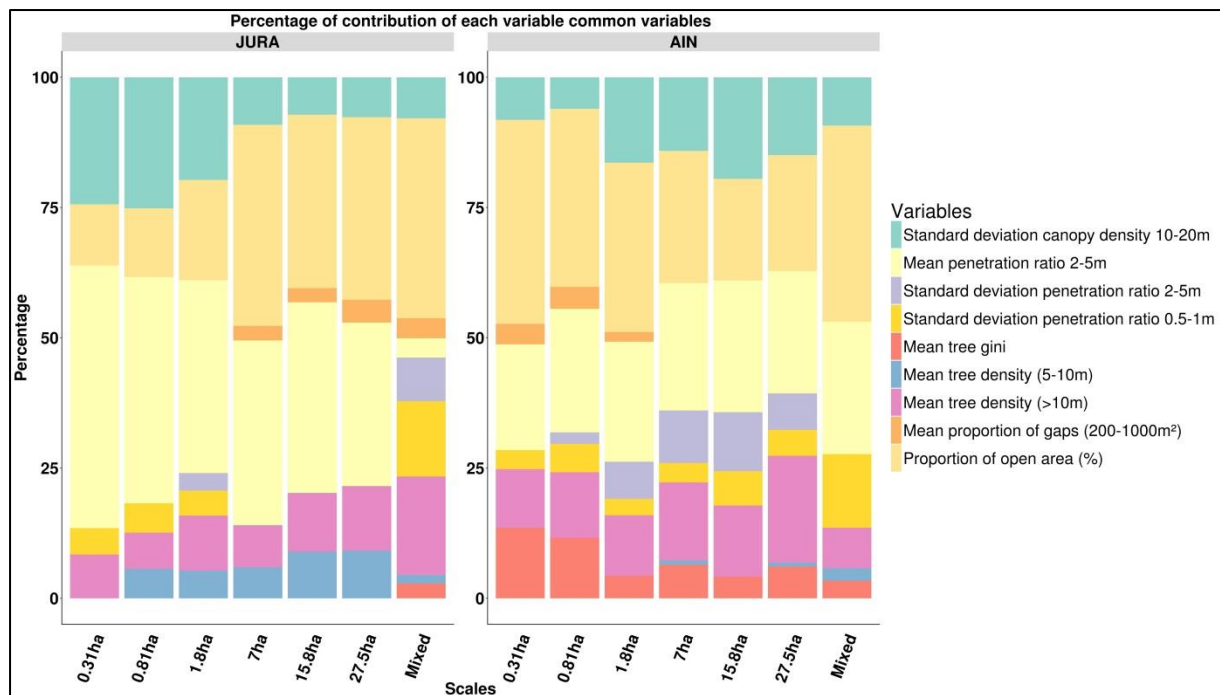
*Figure 6 : Contribution aux modèles de toutes les variables (attention pour l'Ain à l'échelle 7ha la proportion moyenne de conifère n'est pas incluse, de même pour l'échelle 27.5ha ou la moyenne et la déviation standard ont été enlevé (corrélation >0.7))*



### 3.2.2 Contribution des variables communes

Quand on compare la contribution des variables communes aux deux modèles, on remarque que quatre variables contribuent de façon importante aux modèles : le ratio pénétration moyen entre 2 et 5m, la densité moyenne des arbres (>10m), le pourcentage moyen de zones ouvertes et la déviation standard de la densité de la canopée entre 20 et 30 m. En revanche on observe que la densité des arbres entre 5 et 10 m, contribue aux modèles du Jura mais pas à seulement très peu ceux de l'Ain. Inversement, la variable Indice de Gini pour la hauteur des arbres contribue aux modèles de l'Ain mais pas à ceux du Jura (sauf pour le modèle mixte).

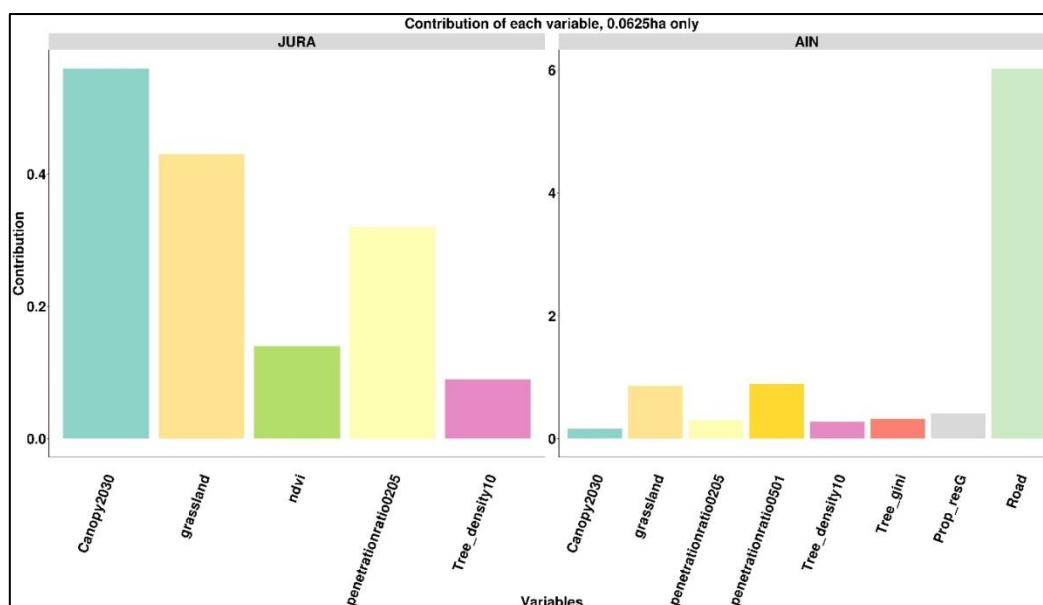
Figure 7 : Contribution aux modèles des variables communes aux modèles



### 3.2.3 Contribution des variables à petite échelle

Pour les modèles à petite échelle, on remarque que les variables qui contribuent le plus dans la zone du Jura sont la densité de la canopée entre 10-20m, le pourcentage de zones ouvertes et le ratio de pénétration entre 2 et 5m. Pour la zone de l'Ain, la variables distance aux routes participe le plus, suivit dans une moindre proportion du pourcentage de zones ouvertes et le ratio de pénétration entre 0.5 et 1m.

Figure 8 : Contribution des variables pour l'échelle 0.0625ha



### 3.3 Courbes de réponses

Les courbes de réponses aux variables qui contribuent le plus aux modèles décrits dans le paragraphe précédent ont été calculées (figure 9). Elles permettent de comprendre comment l'espèce répond aux différentes valeurs de chaque variables, et ainsi de savoir quelles sont les valeurs favorables ou non. Ces courbes de réponses sont présentées pour les sept échelles afin d'en observer l'effet sur la forme de la réponse. Les réponses possibles sont de deux types : linéaire ou quadratique. Associées aux courbes de réponses, les courbes de densité des valeurs présentes dans le paysage permettent de remettre dans le contexte les courbes de réponses.

- Moyenne du ratio de pénétration

Dans les deux zones d'études, le ratio de pénétration 2-5m présente la même forme de réponse avec des valeurs optimales autour de 0.3. Une valeur de 0 représente une couche cible sans végétation. Une grande valeur de ce ratio représente un cas où la couche cible est plus dense que la somme des couches en dessous. Plus la valeur est grande plus la différence est forte. Les courbes de densités dans la zone d'étude de l'Ain montre deux pics de valeurs, l'un à 0 l'autre à 0.18. En revanche, dans la zone d'étude du Jura, un seul pic est observé autour de 0.25. Il est



intéressant de remarquer que les réponses concordent dans les deux zones d'études pour cette variable qui justement décrit l'espace dans la zone de 2-5 m qui est importante pour la Gélinothe.

- Déviation standard de la densité de la canopée 20-30 m

En revanche, la réponse observée est différente dans chaque zone pour la déviation standard de la densité de la canopée 20-30 m. Dans la zone de l'Ain, les valeurs basses (zones peu hétérogènes) sont favorisées à petites échelles, mais les valeurs hautes sont plus favorables à large échelle. Dans la zone du Jura, les valeurs les plus favorables se trouvent pour une déviation standard de 0 (zone homogène pour la canopée) à toutes les échelles. On remarque en observant les courbes de densité que les paysages des deux zones d'études sont différents. Dans la zone de l'Ain, la plupart des valeurs sont entre 0 et 0.05, montrant la présence de zones très homogènes, mais des zones hétérogènes sont aussi présentes. Dans le Jura un deuxième pic est observé entre 0.05 et 0.1 montrant aussi la présence zones hétérogènes.

- Moyenne de la densité des arbres supérieurs à 10 m

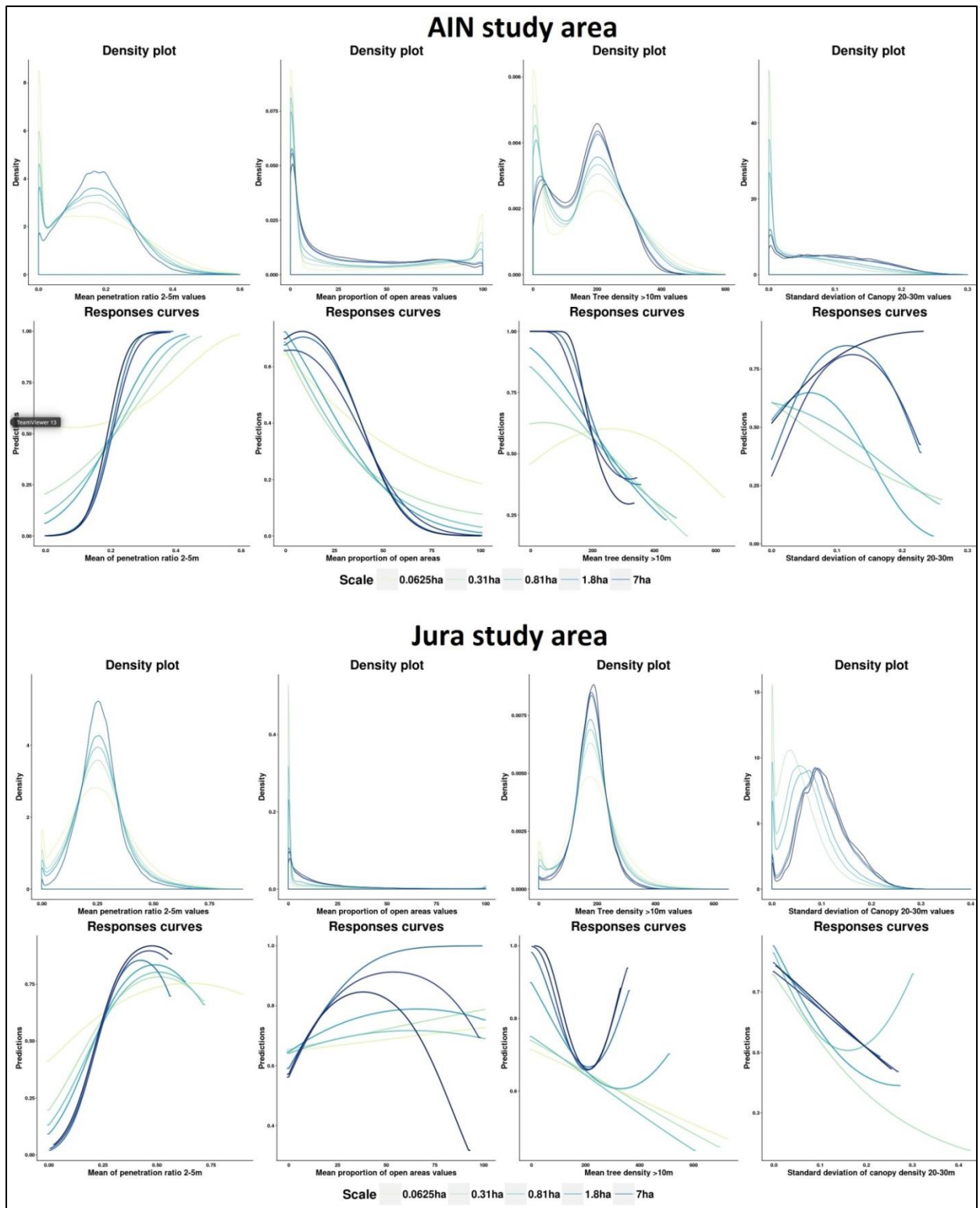
Dans la zone d'étude de l'Ain la réponse de l'espèce à la densité des arbres (>10m) est différente suivant les échelles. A l'échelle de 0.31ha, la densité de 200 arbres par ha est favorable pour l'espèce. En revanche, à plus large échelle une densité de 0 à 100 arbres par ha est préférable. Dans le Jura, à petite échelle des densités d'arbres basses sont préférable (réponse linéaire), mais à large échelle une densité de 200 arbres par ha est montrée comme non favorable. On observe ici un effet des réponses quadratiques, qui pour les valeurs extrême hautes ou basses donneront souvent des prédictions extrêmes (faibles ou hautes) de par la forme même de la courbe. Quand ces observations sont remises dans un contexte global, on observe par exemple dans le Jura que le pic de densité se situe autour de 200 arbres par ha, laissant peu de valeurs extrêmes supérieures ou inférieures. A l'inverse dans l'Ain, un pic de densité 0 indique la

présence importante de zones sans forêt dans la zone d'étude. Le fait qu'à large échelle, ces zones ouvertes peuvent être prédites comme favorables est contre intuitif au vu des connaissances des préférences de l'espèce. Mais il ne faut pas oublier que les prédictions sont faites à partir des valeurs de plusieurs variables, dont la proportion de zones ouvertes.

- Proportion moyenne des zones ouvertes

Dans la zone d'étude de l'Ain, on observe clairement et à toute les échelles la même réponse, qui montre que les proportions de 0% d'espaces ouverts sont plus favorables. Dans la zone d'étude du Jura, une réponse linéaire à petite échelle montre des zones avec une proportion de zones ouvertes de 100% comme plus favorable. Cependant, les courbes de densités nous montrent que de telles zones ouvertes sont très peu présentes dans cette zone d'étude et de plus les prospections sont faites en zone forestière. Ainsi, une réponse positive à une faible proportion de trouée par rapport à des zones où la proportion est de 0% peut engendrer ce type de réponses linéaires en absence de données sur les zones ouvertes. En revanche à large échelle, on observe des réponses quadratiques avec une valeur moyenne favorable de 30% d'espaces ouverts (sauf pour l'échelle de 7ha).

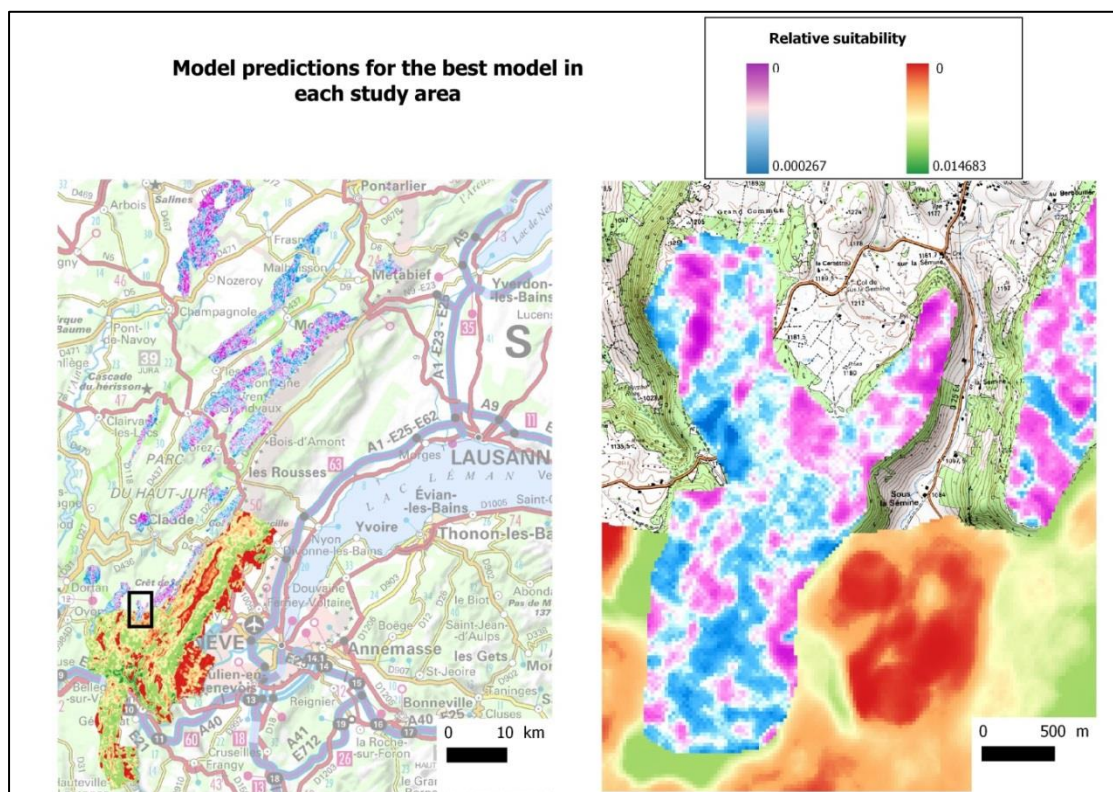
Figure 9 : Courbes de réponses pour les deux zones d'études



## Interprétation des cartes de prédiction

Des cartes de prédictions ont été réalisées pour chaque modèle. Les cartes furent créées en utilisant le mode exponentiel de Maxent (figure 10). Les valeurs de chaque pixel représentent la valeur relative de la qualité de l'habitat par rapport à l'habitat présent dans le paysage. Ainsi, la comparaison des valeurs des pixels entre elles permet de hiérarchiser les habitats présents du meilleur au moins bon. Cependant, la meilleure valeur ne représente pas forcément la valeur du meilleur habitat de manière générale. Si dans le paysage, le meilleur habitat présent est un habitat de qualité moyenne du point de vue de l'espèce, il peut quand même avoir la valeur maximale parmi les prédictions du modèle. La valeur proposée en tant que telle est de plus dépendante du nombre d'observation et du nombre de point de fond donnés au modèle. Ainsi, une même valeur dans chacune des zones d'étude ne correspond pas à la même qualité de l'habitat. Il faut donc utiliser les résultats des deux zones d'études séparément.

*Figure 10: Cartes de prédiction des meilleurs modèles pour chacune des zones d'études. Pour le Jura les couleurs rose indiquent les zones les moins favorables, tandis que pour la zone d'Ain ces zones sont montrées en rouge.*



#### i. Comment les variables participent-elles aux valeurs prédites localement?

Si l'étude de la contribution des variables et des courbes de réponses montre à l'échelle du modèle les effets des variables sur la distribution de la Gélinothe des bois, l'attribution d'une valeur de prédiction basse ou haute peut être dû localement à l'effet d'une ou plusieurs variables. L'utilisation de « Local Interpretable Model-Agnostic Explanation » (LIME) permet pour des valeurs spécifiques de prédictions, de savoir quelles variables et dans quelle mesure elles ont contribué à cette valeur. C'est particulièrement intéressant pour comprendre l'attribution de valeurs prédites extrêmes. Ainsi, le package R LIME a été utilisé pour observer cet effet local des variables pour 20 valeurs prédites parmi les plus hautes (quantile 0.999) et plus basse (quantile 0.001) du paysage. Dans la Figure 11 sont présentés les résultats obtenus pour 20 valeurs parmi les plus hautes à l'échelle de 0.0625ha dans la zone du Jura. Ainsi, on remarque que pour certaines valeurs (exemples des cas 2, 5, 9, 13), toutes les variables ont un effet positif, expliquant ainsi la valeur haute de la prédiction. Dans d'autres cas (1, 3, 6, 7 par exemple), une ou deux variables ont un effet négatif, sans pour autant que la valeur prédite ne soit basse. Ainsi, les variables telles que le ratio de pénétration 2-5m, NDVI et la densité de la canopée entre 20 et 30m ont une influence positive très forte si leur valeur locale est favorable.

Au contraire, pour des valeurs prédites parmi les plus basses du paysage (figure 12), on observe dans tous les cas une contribution négative de ces mêmes variables expliquant ainsi la valeur basse de la prédiction. Dans certains cas on observe une contribution positive de la valeur NDVI (cas 10, 14 et 15 par exemple), mais celle-ci ne compense pas une valeur non favorable des autres variables.

Dans un deuxième exemple, le cas d'un modèle de l'Ain pour l'échelle de 27.5ha est étudié (figure 13). Pour 20 valeurs parmi les plus hautes, on remarque dans une grande majorité que toutes les variables participent positivement (Ratio de pénétration 2-5m, index de Gini de hauteur des arbres, standard déviation de la densité de la canopée). Cependant, on remarque

aussi le cas 10 ou seulement la valeur du ratio de pénétration 2-5m a une contribution positive. Cette contribution est suffisante pour attribuer une valeur de prédiction élevée alors que les autres variables participent négativement à la valeur de prédiction. On montre donc ici une limite du modèle qui peut attribuer des valeurs de prédiction proches mais qui dans les fait correspondent à un environnement local différent.

Figure 11 : 20 valeurs prédites parmi les 0.01% meilleures dans la zone du Jura (0.0625ha)

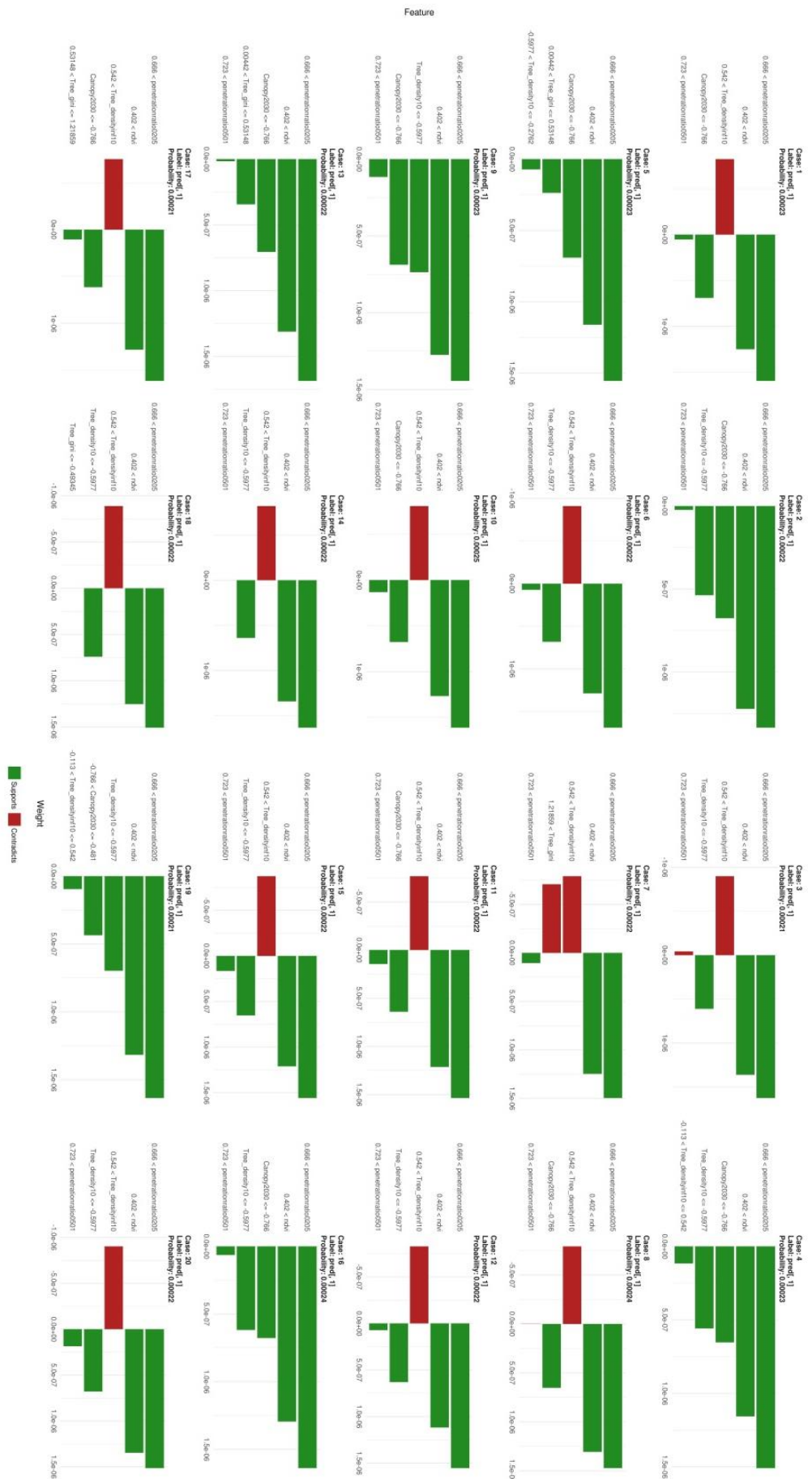




Figure 12: 20 valeurs prédites parmi les 0.01% plus mauvaise dans la zone du Jura (0.0625ha)

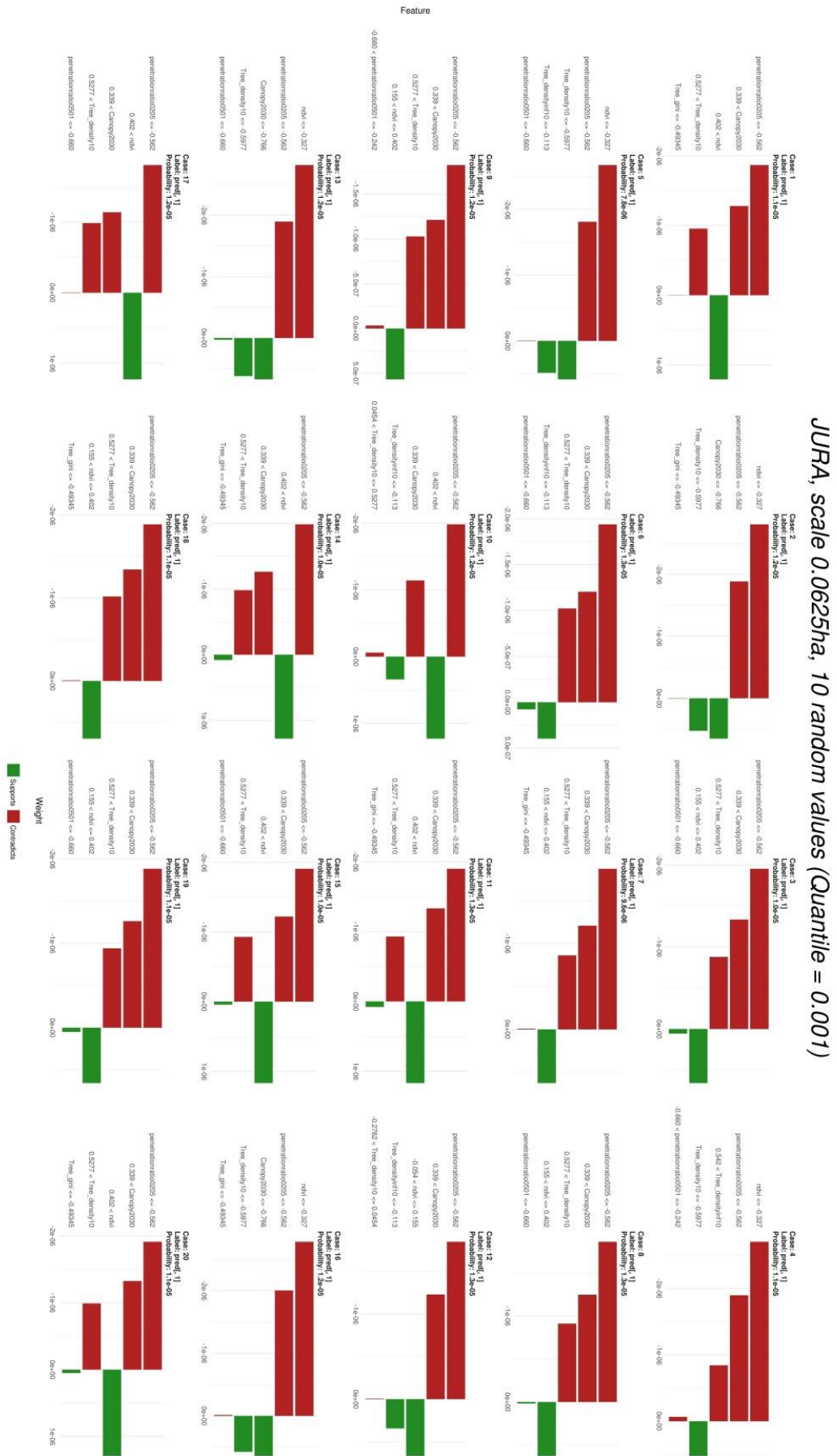
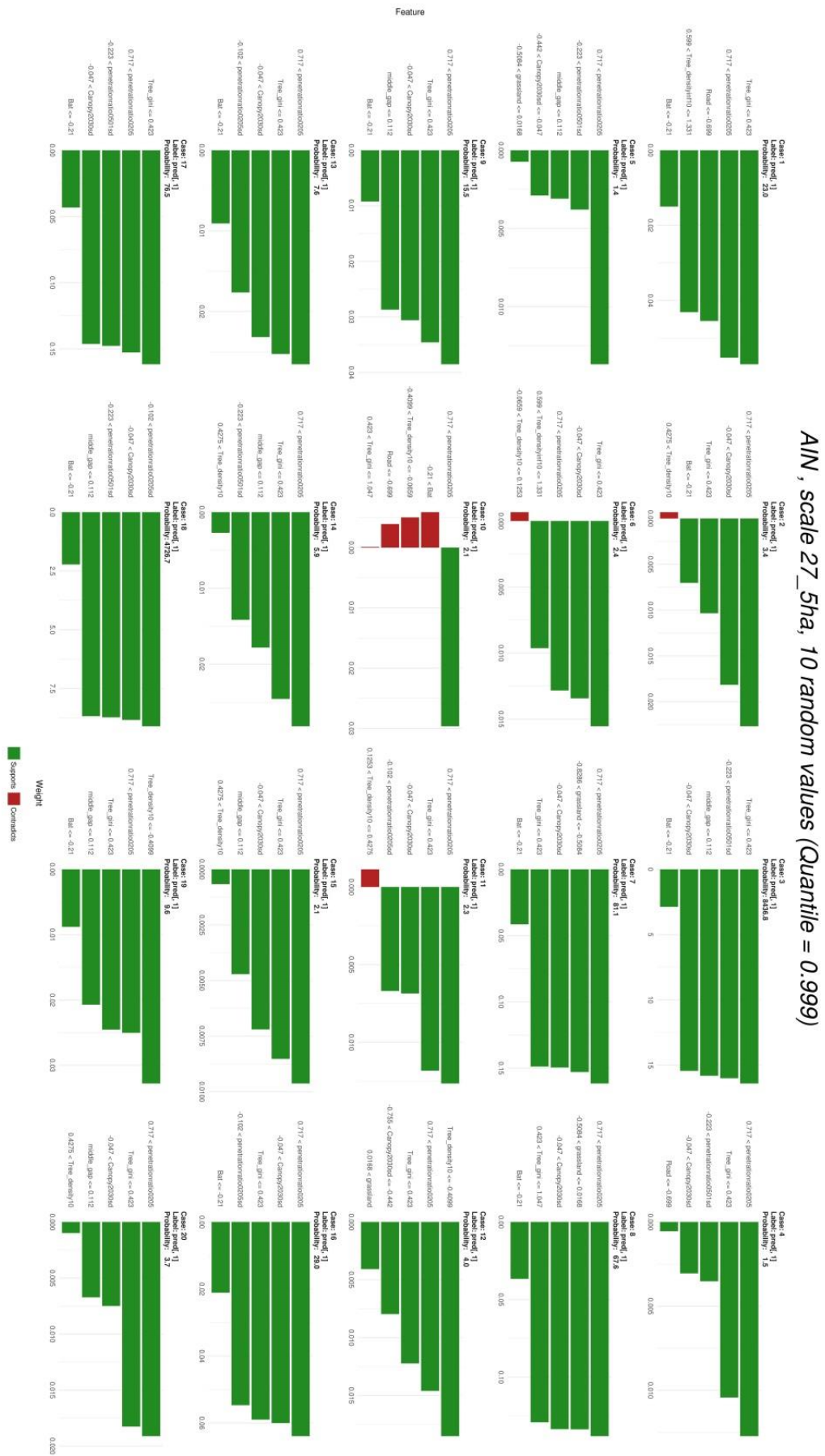




Figure 13 : 20 valeurs prédites parmi les 0.01% meilleures dans la zone de l'Ain (27.5ha)



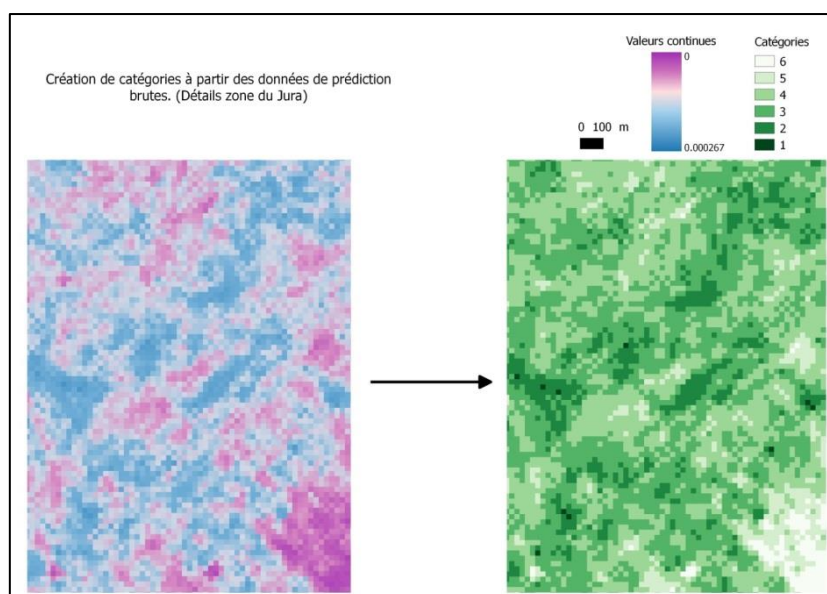
### 3.4 Création de catégories pour les cartes de prédictions

Dans le but de proposer des cartes de prédictions plus facilement interprétables, des cartes simplifiées ont été créées à partir des cartes de prédiction des modèles. Ainsi, les valeurs de prédictions ont été catégorisées en 6 classes définies par les valeurs seuil des quantiles 0.10, 0.25, 0.50, 0.75, 0.90 (Figure 14, Tableau 6). De même que pour les valeurs de prédictions les catégories ne peuvent s'interpréter que dans le contexte défini de la zone d'étude. Ainsi, les valeurs de catégorie 1 dans la zone de l'Ain et la zone de Jura ne peuvent pas être comparées, chacune représentant les meilleures valeurs dans la zone d'étude uniquement.

*Tableau 6 : Catégories créées pour les cartes de prédiction*

Seuils des quantiles	Catégorie	Signification
0-0,10	6	Les 10% des valeurs les plus basses du paysage
0,1-0,25	5	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,1
0,25-0,50	4	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,25
0,50-0,75	3	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,50
0,75-0,90	2	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,75
0,90-1	1	Les 10% des valeurs les plus hautes du paysage

*Figure 14 : Exemple de catégorisation des prédictions pour une échelle de 0.31ha.*



### 3.5 Validation en forêt de Champfromier

Le lien entre les résultats des modèles et la qualité de l'habitat à dire d'expert sur le terrain a été évaluée. Neuf placettes représentant les différentes catégories prédites par les modèles ont été sélectionnées et visitées en Forêt de Champfromier. Une bonne adéquation entre les valeurs des modèles a été observée. Ainsi, les placettes P3 et P17 classées en catégories 1 ou 2 par les modèles (échelles 0.0625ha, 0.31ha, 0.81ha) ont été catégorisées en habitat favorable A ou B par le protocole de classification des habitats de la Gélinothe du Groupe tétras Jura (Figure 15). En revanche, la placette P24 était, elle classifiée en catégorie 6 à la plus petite échelle mais en 1 puis 2 aux plus larges échelles. La bonne qualité de l'habitat a été confirmée sur le terrain, montrant l'importance de la prise en compte des différentes échelles des modèles pour apprécier la qualité générale de l'habitat. En effet, la taille du pixel 25\*25m peut d'une part chevaucher des patches de bonne et mauvaise qualité, dégradant ainsi la note globale, qui sera mieux appréciée à plus large échelle. Il faut aussi tenir compte de l'imprécision du GPS sur le terrain qui ne nous permet pas de rejoindre avec une très grande précision la placette. La placette P14, localisée dans une zone forestière générale d'une qualité moyenne d'après le modèle, a été prédite en catégorie 4 à la plus petite échelle (Figure 15). Sur cette placette, un

envahissement par la régénération du hêtre à été observé, montrant clairement un habitat très défavorable pour l'espèce. Si ce résultat peut paraître étonnant, il prend du sens quand lorsque qu'on regarde les résultats des modèles dans leur globalité. En effet, en faisant des prédictions sur l'ensemble du territoire, les classes les plus mauvaises (6 et 5) sont en très grande majorité des milieux urbains, cultures et larges zones ouvertes. Il paraît alors logique qu'un milieu forestier même dégradé soit en comparaison plus accueillant pour un individu, d'où l'attribution par le modèle d'une valeur moyenne pour l'habitat. Ainsi, pour le modèle de l'Ain, les valeurs des catégories de 4 à 6 représentent des habitats non favorables pour la Gélinoite des bois. Ces catégories, doivent encore une fois être évaluées séparément pour les modèles du Jura, en effet, comme les zones de forêt uniquement ont été survolées, les catégories les plus basses (6 et 5) pour ces modèles devraient représenter des zones forestières non favorables pour l'oiseau.

*Figure 15 : Trois placettes visitées en forêt de Champfromier*

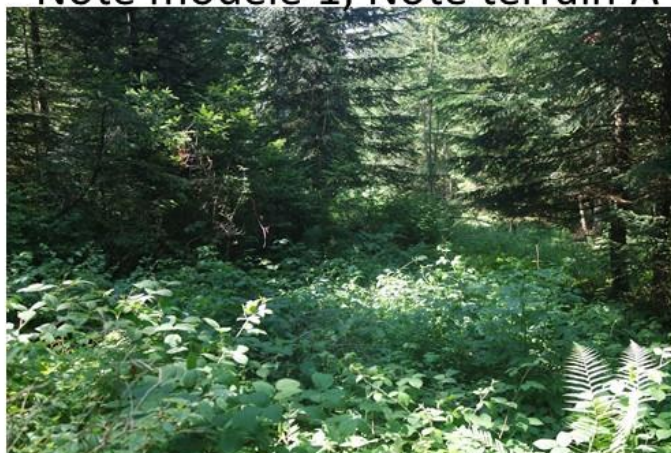
P3

Note modèle 2, Note terrain B



P17

Note modèle 1, Note terrain A



P14

Note modèle 4, Note terrain D





Cette validation en forêt de Champfromier nous a aussi permis de constater une limite du modèle concernant la discrimination entre des zones ouvertes, les routes et les chemins forestiers. En effet, avec la définition mise en place pour la détection des objets trouées, les routes et chemin de par leurs longueurs ont des surfaces très grande, et entrent de ce fait dans la catégorie zones ouvertes. Ces zones ouvertes sont prédites défavorables par les modèles. Hors les chemins forestiers, en particulier ceux présentant une strate herbacées fournie, sont en réalité des habitats favorables pour les nichées. Ainsi, une attention particulière doit être portée sur ces zones lors de l'interprétation du modèle.

## 4 DISCUSSION

---

Les modèles de distribution pour la Gélinothe des bois créés à l'aide de données LiDAR comme variables environnementales ont permis de produire des cartes à plusieurs échelles. Les cartes produites représentent la distribution potentielle de l'espèce, où l'habitat est favorable suivant les critères pris en compte dans le modèle. De nombreux autres processus tels que la capacité de dispersion, la dynamique des populations ou la compétition avec d'autres espèces peuvent restreindre dans la réalité la présence de population même dans des habitats de structure et composition favorables. Ces effets sont encore très peu pris en compte dans la réalisation de modèles de distribution d'espèces (Guisan and Thuiller 2005). Les plus petites échelles donnent des résultats détaillés sur la qualité relative de l'habitat. Il est cependant probable que la précision de ces modèles soit moins bonne que ceux à plus large échelle étant donné une diversité potentiellement plus importante localement. L'étude des variables contribuant à la valeur de prédiction montre d'ailleurs que des prédictions fortes peuvent être dues à l'association positive de plusieurs variables ou seulement d'une seule, décrivant donc des environnements différents. Les performances des modèles aux différentes échelles ont montré que les modèles mixtes ne sont pas plus performants. Cependant, les cartes produites prenant

en compte les variables à différentes échelles permettent d'associer des effets locaux (végétation) à des effets à plus larges échelles (perturbations, fragmentation) qu'il est intéressant de prendre en compte. Une autre limitation de ces modèles est la date de survol LiDAR, qui est déjà suffisamment ancienne pour la zone de l'Ain (2014) pour que des éléments de végétation aient changés (chablis, coupes, croissance et régénération de la végétation). La déclinaison des cartes en 6 catégories permet une interprétation plus facile des prédictions et ainsi de cibler de futures zones de prospection. Les cartes non-catégorisées peuvent être utilisées pour avoir une vision plus fine de la qualité de l'habitat dans une zone précise. Concernant l'interprétation des courbes de réponses, la limitation principale provient du fait que deux types de réponses peuvent être modélisés, linéaire et quadratique. Ceci a pour effet de rendre certaines interprétations non compatibles avec nos connaissances de l'écologie de l'espèce. Par exemple, des zones de densité d'arbres de 0 données favorables. Cependant, les prédictions finales proviennent des variations combinées des variables, ainsi la présence d'une variable pourcentage de milieu ouvert permet de compenser dans une certaine mesure des estimations mauvaises aux valeurs extrêmes des autres variables. Ainsi, les zones de pâtures et champs apparaissent bien comme peu favorables par rapport au milieu forestier pour la Gélinothe des bois dans les cartes de prédictions finales.

La sélection initiale des variables ne couvre pas toutes les variations présentes dans l'habitat. En particulier le LiDAR ne donne que peu d'information sur les espèces végétales présentes dans le milieu en particulier pour le sous étage, ce qui pose un problème pour la caractérisation de l'habitat d'une espèce telle que la Gélinothe des bois. Ceci peut être amélioré avec l'ajout de données satellites qui associées au LiDAR peut permettre une meilleure caractérisation de la structure et la de composition des zones forestières (Bae et al. 2014).

La validation sur le terrain a permis de confirmer la bonne performance des modèles pour prédire la qualité de l'habitat de la Gélinothe des bois dans le Jura. Une attention particulière

doit cependant être apportée lors de l'interprétation des cartes aux différences entre les zones d'études.

La mise en place d'un nouveau protocole de prospection pour la Gélinothe permettra d'accumuler de nouvelles données sur l'espèce à la fois sur leur présence et leur absence, et dans le futur d'améliorer notre connaissance de la distribution de l'espèce dans le massif Jurassien. Ces données permettront aussi d'améliorer d'éventuelles futures modélisations de l'habitat de la Gélinothe des bois en mettant en lumière les limites des prédictions des modèles de distribution d'espèces par rapport à la diversité des habitats présents.



## 5 BIBLIOGRAPHIE

---

Arkle, Robert S, David S Pilliod, Steven E Hanser, Matthew L Brooks, Jeanne C Chambers, James B Grace, Kevin C Knutson, David A Pyke, Justin L Welty, and Troy A Wirth. 2014. "Quantifying Restoration Effectiveness Using Multi-scale Habitat Models: Implications for Sage-grouse in the Great Basin." *Ecosphere* 5 (3): 1–32.

Bae, Soyeon, Bjoern Reineking, Michael Ewald, and Joerg Mueller. 2014. "Comparison of Airborne Lidar, Aerial Photography, and Field Surveys to Model the Habitat Suitability of a Cryptic Forest Species—the Hazel Grouse." *International Journal of Remote Sensing* 35 (17): 6469–89.

Bellon, B. 2011. "Évolution de La Répartition Communale Du Petit Gibier de Montagne En France Décennie 2000-2009." *Supplément Faune Sauvage*, no. 290.

Cardador, Laura, José A. Diaz-luque, Fernando Hiraldo, James D. Gilardy, and José L. Tella. 2017. "The Effects of Spatial Survey Bias and Habitat Suitability on Predicting the Distribution of Threatened Species Living in Remote Areas." *Bird Conservation International*, 1–12.

El-Gabbas, Ahmed, and Carsten F Dormann. n.d. "Improved Species-occurrence Predictions in Data-poor Regions: Using Large-scale Data and Bias Correction with Down-weighted Poisson Regression and Maxent." *Ecography*.

Elith, Jane, Steven J Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, and Colin J Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17 (1): 43–57.

Eysn, Lothar, Markus Hollaus, Eva Lindberg, Frédéric Berger, Jean-Matthieu Monnet, Michele Dalponte, Milan Kobal, Marco Pellegrini, Emanuele Lingua, and Domen Mongus. 2015. "A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space." *Forests* 6 (5): 1721–47.

Fourcade, Yoan, Aurelien G Besnard, and Jean Secondi. 2018. "Paintings Predict the Distribution of Species, or the Challenge of Selecting Environmental Predictors and Evaluation Statistics." *Global Ecology and Biogeography* 27 (2): 245–56.

Franklin, Janet. 2009a. "Moving beyond Static Species Distribution Models in Support of Conservation Biogeography." *Diversity and Distributions* 16 (3): 321–30.

———. 2009b. "Moving beyond Static Species Distribution Models in Support of Conservation Biogeography." *Diversity and Distributions* 16 (3): 321–30.

Graf, Roland F, Kurt Bollmann, Werner Suter, and Harald Bugmann. 2005. "The Importance of Spatial Scale in Habitat Models: Capercaillie in the Swiss Alps." *Landscape Ecology* 20 (6): 703–17.

Guisan, Antoine, and Wilfried Thuiller. 2005. "Predicting Species Distribution: Offering More than Simple Habitat Models." *Ecology Letters* 8 (9): 993–1009.

Guisan, Antoine, Reid Tingley, John B. Baumgartner, Ilona Naujokaitis-Lewis, Patricia R. Sutcliffe, Ayesha I. T. Tulloch, Tracey J. Regan, et al. 2013. "Predicting Species Distributions for Conservation Decisions." *Ecology Letters* 16 (12): 1424–35. <https://doi.org/10.1111/ele.12189>.

Hamraz, Hamid, Marco A Contreras, and Jun Zhang. 2017. "Vertical Stratification of Forest Canopy for Segmentation of Understory Trees within Small-Footprint Airborne LiDAR Point Clouds." *ISPRS Journal of Photogrammetry and Remote Sensing* 130: 385–92.

Hofstetter, Lea, Raphaël Arlettaz, Kurt Bollmann, and Veronika Braunisch. 2015. "Interchangeable Sets of Complementary Habitat Variables Allow for Flexible, Site-Adapted Wildlife Habitat Management in Forest Ecosystems." *Basic and Applied Ecology* 16 (5): 420–33.

Johnson, Chris J, and Michael P Gillingham. 2005. "An Evaluation of Mapped Species Distribution Models Used for Conservation Planning." *Environmental Conservation* 32 (2): 117–28.

Kramer-Schadt, Stephanie, Jürgen Niedballa, John D Pilgrim, Boris Schröder, Jana Lindenborn, Vanessa Reinfelder, Milena Stillfried, Ilja Heckmann, Anne K Scharf, and Dave M Augeri. 2013. "The Importance of Correcting for Sampling Bias in MaxEnt Species Distribution Models." *Diversity and Distributions* 19 (11): 1366–79.

Leclercq, B. 1985. "Influence Des Routes et Voies de Pénétration Humaine Sur Les Comportements de Grands Tétras et de Gélinittes Dans Le Haut-Jura Français." Actes du colloque "Routes et faune sauvage", Strasbourg.

Mathys, Lukas, Niklaus E Zimmermann, Niklaus Zbinden, and Werner Suter. 2006. "Identifying Habitat Suitability for Hazel Grouse *Bonasa Bonasia* at the Landscape Scale." *Wildlife Biology* 12 (4): 357–66.

McGarigal, Kevin, Ho Yi Wan, Kathy A Zeller, Brad C Timm, and Samuel A Cushman. 2016. "Multi-Scale Habitat Selection Modeling: A Review and Outlook." *Landscape Ecology* 31 (6): 1161–75.

Moilanen, Atte, John R Leathwick, and John M Quinn. 2011. "Spatial Prioritization of Conservation Management." *Conservation Letters* 4 (5): 383–93.

Phillips, Steven J., Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, and Mary E. Blair. 2017. "Opening the Black Box: An Open-Source Release of Maxent." *Ecography*, May, n/a-n/a. <https://doi.org/10.1111/ecog.03049>.

Phillips, Steven J, Miroslav Dudík, Jane Elith, Catherine H Graham, Anthony Lehmann, John Leathwick, and Simon Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data." *Ecological Applications* 19 (1): 181–97.

Räty, M. 1979. "Effect of Highway Traffic on Tetraonid Densities." *Ornis Fennica* 56 (4): 169–70.

Rechsteiner, Christian, Florian Zellweger, Anatole Gerber, Frank T. Breiner, and Kurt Bollmann. 2017. "Remotely Sensed Forest Habitat Structures Improve Regional Species Conservation." *Remote Sensing in Ecology and Conservation*, n/a-n/a. <https://doi.org/10.1002/rse2.46>.

Roussel, Jean-Romain, and David Auty. 2016. *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R Package Version 1.0.0*. <https://github.com/Jean-Romain/lidR>.

Sachot, Sebastien, Nicolas Perrin, and Cornelis Neet. 2003. "Winter Habitat Selection by Two Sympatric Forest Grouse in Western Switzerland: Implications for Conservation." *Biological Conservation* 112 (3): 373–82.

Schäublin, Sabrina, and Kurt Bollmann. 2011. "Winter Habitat Selection and Conservation of Hazel Grouse (*Bonasa Bonasia*) in Mountain Forests." *Journal of Ornithology* 152 (1): 179–92.

Zellweger, Florian, Felix Morsdorf, Ross S Purves, Veronika Braunsch, and Kurt Bollmann. 2014. "Improved Methods for Measuring Forest Landscape Structure: LiDAR Complements Field-Based Habitat Assessment." *Biodiversity and Conservation* 23 (2): 289–307.

## ANNEX VIII. FRENCH REPORT ON CAPERCAILLIE DISTRIBUTION MODELS

### Modèles de distributions du Grand Tétras dans le massif Jurassien à partir de données LiDAR



Alexandra Deprez



région **BOURGOGNE  
FRANCHE-COMTE**



Glad Anouk

Groupe Tétràs Jura, Irstea centre de

Grenoble unité LESSEM

01/09/2018



Note : Ce rapport constitue un résumé des travaux réalisés pendant la thèse (collaboration avec l'IRSTEA Grenoble LESSEM, encadrée par Björn Reineking et Jean-Matthieu Monnet et financée par le conseil départemental de l'Ain). Il présente les différentes méthodes mise en œuvre et les résultats obtenus dans la réalisation des modèles spatiaux pour le Grand Tétrás. Pour plus de détails concernant les divers aspects présentés ici, se référer aux chapitres et articles correspondant dans la thèse.

## Remerciements

Je souhaite ici remercier tous ceux qui ont participé au projet, en particulier mes encadrants Björn Reineking et Jean-Matthieu Monnet pour leur aide tout au long de la thèse. Je remercie aussi Alexandra Depraz et Marc Montadert qui ont apporté leurs connaissances sur la Gélinotte afin de m'aider dans la réalisation des modèles.

## Documents associés :

### Cartes de prédictions au format raster pour deux zones d'études et deux saisons à différentes échelles.

Les cartes de prédiction sont rangées dans le dossier « Projet\_QGIS\_cartes/DATA /GrandTetras / » puis saisons (hiver, été) et par sexe dans des sous-dossiers séparés.

Nom données de prédictions brutes :

Cartesbrutes /Zoneétude \_Sexe\_ saison \_echelle \_exponential

Nom données de prédictions catégorisées :

Classes/Zoneétude \_Sexe\_ saison \_echelle

### Résumé des résultats des modèles par parcelles.

Deux fichiers shapefiles contenant les données résumés par parcelles (échelle 0.0625ha). Ces données sont rangées dans le dossier « Projet\_QGIS\_cartes/DATA /Parcelles.

- Parcelles\_AIN\_ete.shp
- Parcelles\_JURA\_ete.shp
- Parcelles\_AIN\_hiver.shp
- Parcelles\_JURA\_hiver.shp

Deux pages en ligne permettent de créer des graphiques pour visualiser la qualité des parcelles pour chaque commune en été.

[https://gladgtj.shinyapps.io/Parcelles\\_Jura/](https://gladgtj.shinyapps.io/Parcelles_Jura/)

[https://gladgtj.shinyapps.io/Parcelles\\_Ain/](https://gladgtj.shinyapps.io/Parcelles_Ain/)

<b>I. INTRODUCTION .....</b>	<b>357</b>
<b>II. METHODES .....</b>	<b>360</b>
1. JEUX DE DONNEES DES OBSERVATIONS DU GRAND TETRAS .....	360
2. JEUX DE DONNEES LIDAR .....	360
3. CHOIX DU TYPE DE VARIABLE A EXTRAIRE .....	361
4. CALCUL DES VARIABLES LIDAR.....	362
5. SELECTION DES VARIABLES .....	363
6. ECHELLES SPATIALES .....	366
7. MODELISATION .....	367
8. METHODE DE CONSTRUCTION DES MODELES MULTI-ECHELLES .....	370
<b>III. RESULTATS DES MODELES HIVER .....</b>	<b>372</b>
1. PERFORMANCE DES MODELES .....	372
II. MODELE A ECHELLES UNIQUE ET VARIABLE UNIQUE .....	372
III. MODELES A ECHELLE UNIQUE ET VARIABLES MULTIPLES ET MODELES MIXTES .....	372
2. CONTRIBUTION DES VARIABLES .....	373
I. CONTRIBUTION DE TOUTES LES VARIABLES .....	373
II. CONTRIBUTION DES VARIABLES A PETITE ECHELLE .....	375
3. COURBES DE REPONSES .....	377
<b>IV. RESULTATS DES MODELES ETE .....</b>	<b>383</b>
4. DIFFICULTES POUR LA ZONE DU JURA .....	383
5. MODELES PERFORMANCE .....	384
III. MODELES A ECHELLES MULTIPLES ET VARIABLES MULTIPLES (MIXTE) .....	384
IV. MODELES A ECHELLE UNIQUE ET VARIABLES MULTIPLES ET MODELES MIXTES .....	385
6. CONTRIBUTION DES VARIABLES .....	385
7. COURBES DE REPONSE .....	387
<b>V. INTERPRETATION DES CARTES DE PREDICTION.....</b>	<b>392</b>
8. COMMENT LES VARIABLES PARTICIPENT-ELLES AUX VALEURS PREDITES LOCALEMENT?.....	394
9. CREATION DE CATEGORIES POUR LES CARTES DE PREDICTIONS .....	398
10. VALIDATION EN FORET DE CHAMPFROMIER .....	399
11. CLASSER ET PRIORISER LES PARCELLES.....	402
V. METHODE DE DISTANCE SIMPLE .....	404
VI. METHODE UTILISANT LA SURFACE TERRIERE (AIN SEULEMENT) .....	404



<b><u>VI.</u></b>	<b><u>DISCUSSION .....</u></b>	<b><u>406</u></b>
-------------------	--------------------------------	-------------------

<b><u>VII.</u></b>	<b><u>BIBLIOGRAPHIE .....</u></b>	<b><u>409</u></b>
--------------------	-----------------------------------	-------------------

## I. INTRODUCTION

---

Dans le massif Jurassien, les populations de Grand Tétras sont en déclin. Cependant la distribution complète de l'espèce sur ce vaste territoire est aujourd'hui méconnue en particulier dans les zones de basse altitude (Bellon 2011). Le Groupe Tétras Jura, dans sa mission d'étude et de protection des galliformes de montagne présents dans le massif, souhaite cartographier les habitats pour le Grand Tétras dans le but de mener des actions pour sa protection. En effet, la mise en place de travaux forestier dans les zones aujourd'hui peu favorables à l'espèce, notamment à cause de la régénération du hêtre, pourront dans le futur permettre à l'espèce de se maintenir et de recoloniser certains espaces du massif jurassien.

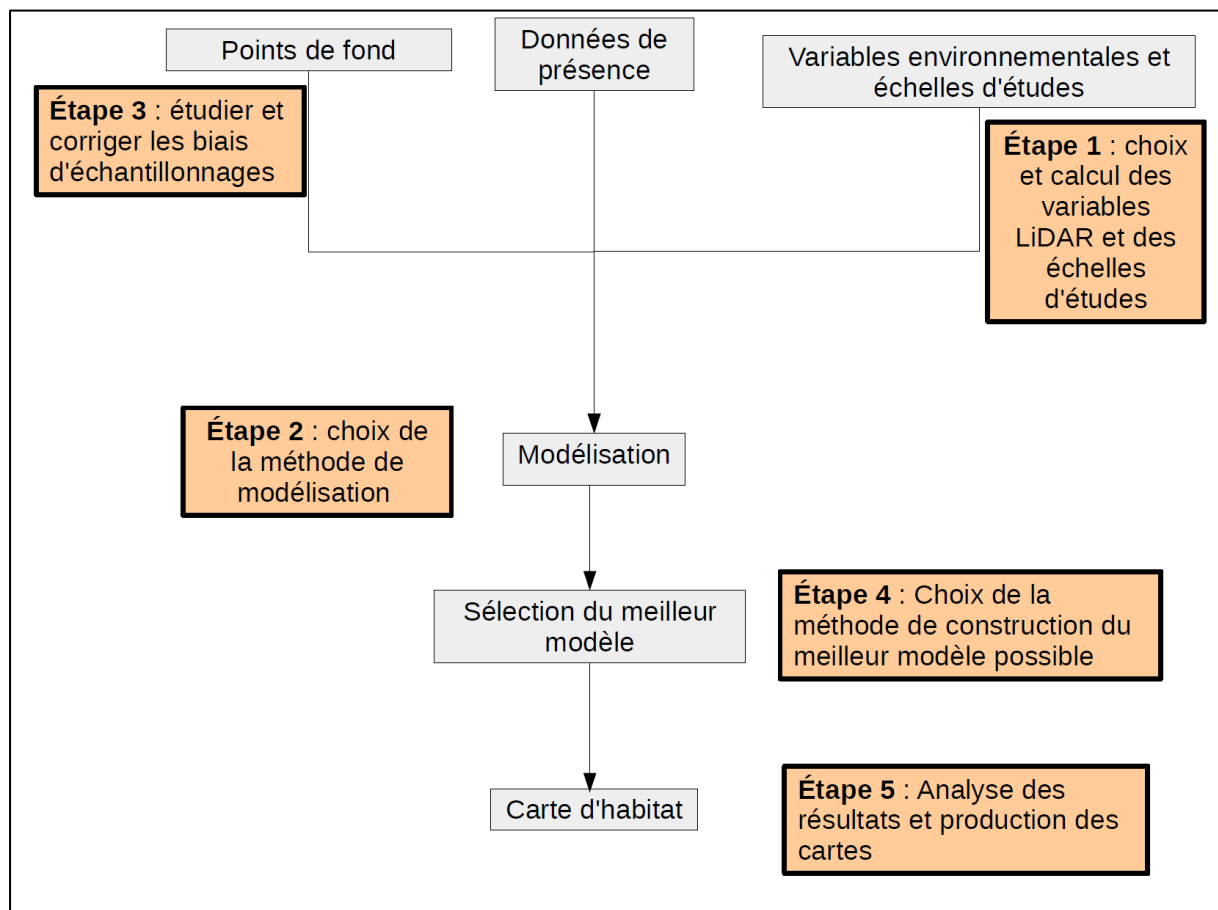
Dans ce cadre, l'idée de développer des modèles de distribution d'espèces a été proposée. Un modèle de distribution d'espèces, aussi appelé modèle de niche est aujourd'hui communément utilisé en écologie et biologie de la conservation (Franklin 2009). Ces modèles permettent de relier les présences d'individus aux caractéristiques du paysage où ils ont été observés. Ainsi, ces modèles permettent de connaître la réponse de l'espèce pour chaque variable utilisée et de faire des prédictions pour des zones où les données sur l'espèce sont manquantes. On obtient alors une carte de prédiction qui montre les zones où l'habitat est favorable ou non à la persistance de l'espèce (Elith et al. 2011). De nombreuses étapes et données sont nécessaires pour construire un modèle de distribution d'espèces (Figure 1).

Les caractéristiques du paysage peuvent être représentées par de nombreuses variables dans l'espace et le temps. Récemment, le développement des systèmes d'acquisition d'image, notamment satellite, a permis de couvrir des zones de plus en plus larges. Ces dernières années, de nombreuses études ont montré la pertinence de l'utilisation de la technique **LiDAR** pour extraire des variables d'environnement dans les milieux forestiers (Bae et al. 2014; Rechsteiner et al. 2017).

Grâce au survol LiDAR des forêts du massif (Jura, Doubs, Ain), des variables d'environnement permettant de connaître précisément la structure tridimensionnelle du couvert végétal ont été calculées pour les modèles de distribution d'espèces. Le calcul et le choix de ces variables n'est que la première étape de la méthode permettant de construire ces modèles.

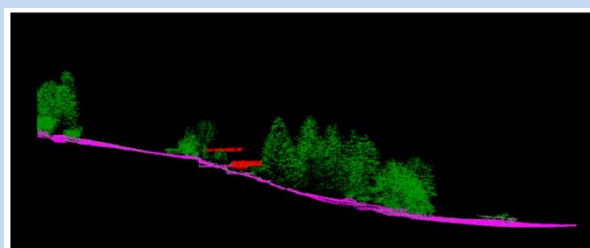
En tout, cinq étapes majeures et successives ont été nécessaires à l'élaboration des modèles (figure 1). Ces différentes étapes sont décrites en détails dans les sections: méthodes et résultats.

*Figure 1 : Les étapes de la création d'un modèle de distribution d'espèces.*



## LIDAR

Le LIDAR (ou télédétection par laser) dont le nom provient de l'anglais **L**ight **D**etection **A**nd **R**anging, est une technique qui permet grâce à l'émission d'une onde laser et à la réception de signaux de cette même onde réfléchis par les objets qu'elle rencontre, de calculer la distance entre l'appareil et l'objet. Cette distance est calculée à partir du temps écoulé entre l'émission et la réception du signal. L'appareil peut être utilisé au sol, mais il démontre surtout son avantage lorsqu'il est embarqué à bord d'un avion ou d'un hélicoptère. Dans ce cas, le laser est dirigé vers le sol ce qui permet de calculer la distance entre l'avion et les objets rencontrés par l'onde laser. Ainsi, en connaissant l'altitude du vol et la position de l'avion à la verticale de ce point grâce au système GPS, il est possible de connaître la hauteur des objets (Maisons, arbres, buissons ...) à la surface du sol. Ces acquisitions par avion, permettent à la fois de couvrir une très grande surface et d'avoir un nuage de points en 3 dimensions (latitude, longitude et hauteur) très précis (résolution jusqu'à 15cm grâce à l'envoi de 10 impulsions laser par m<sup>2</sup>).



Mais le plus grand avantage d'utiliser une onde laser, c'est qu'elle peut pénétrer le couvert végétal. Ainsi, une impulsion peut en réalité renvoyer plusieurs signaux de retour : une première partie du faisceau laser est réfléchi par le feuillage mais le reste traverse et est ensuite réfléchi par l'objet situé au-dessous (une autre partie du feuillage, une branche, le tronc, un buisson, et même le sol). Ainsi, grâce à une seule émission, on peut recevoir jusqu'à 4 ou 5 signaux de retour. Cela permet d'avoir non seulement des informations sur la hauteur de la canopée mais aussi de produire dans une certaine mesure une cartographie des sous-bois et des reliefs du sol.

## II. MÉTHODES

---

### 1. Jeux de données des observations du Grand Tétras

Les données des prospections d'hiver récoltées entre 2007 et 2018 par le Groupe Tétras Jura ont été utilisées. Les observations sont en majorité des indices de présence, localisés en utilisant un GPS. Quand cela était possible, le sexe a été assigné aux observations. Le tracé des prospections de chaque observateur a aussi été enregistré et permet de savoir très précisément quelles zones ont été prospectées (effort d'échantillonnage). Un total de 1484 localisations a été utilisé pour faire les modèles de distribution d'espèces (1105 dans le Jura et le Doubs dont 691 coqs et 414 poules et 379 dans le département de l'Ain dont 207 coqs et 172 poules).

En été, les données utilisées proviennent de quelques prospections d'été et d'observations opportunistes communiquées au Groupe Tétras Jura. Pour garder une bonne précision des données, seules les observations prises avec un GPS ont été utilisées. Ainsi, 134 observations (53 poules et 51 coqs) dans la zone du Jura et 150 observations (74 poules et 33 coqs) dans l'Ain ont été utilisées pour faire les modèles.

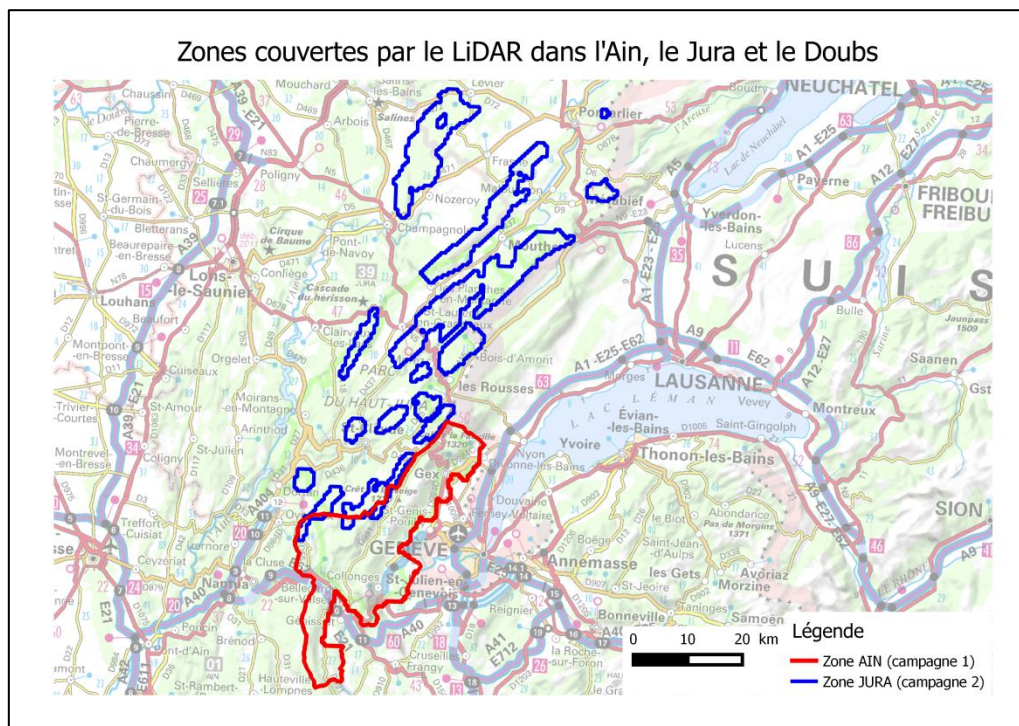
### 2. Jeux de données LiDAR

Les données LiDAR proviennent de deux campagnes différentes. La première s'est déroulée en automne 2014 et couvre 626km<sup>2</sup> du département de l'Ain. La seconde s'est déroulée en été 2016, et couvre les zones d'intérêts pour la Gelinotte du Jura et du Doubs (431km<sup>2</sup>). Il faut souligner ici que, en raison de plusieurs différences entre ces campagnes, des modèles seront proposés pour chaque zone séparément.

Une deuxième différence entre les campagnes est le fait que la zone du département de l'Ain recouvre une surface continue tandis que la deuxième campagne cible les zones forestières. Ainsi, une attention particulière doit être accordée lors de l'interprétation des résultats du modèle aux variations présentes dans le paysage de chaque zone.

Pour la suite du rapport et pour des raisons pratiques, les modèles construits avec le LiDAR de la première campagne seront appelés « modèles de l'Ain », et ceux issus des données de la deuxième campagne « modèles du Jura ».

*Figure 2 : Les couvertures LiDAR des deux zones d'études du massif Jurassien*



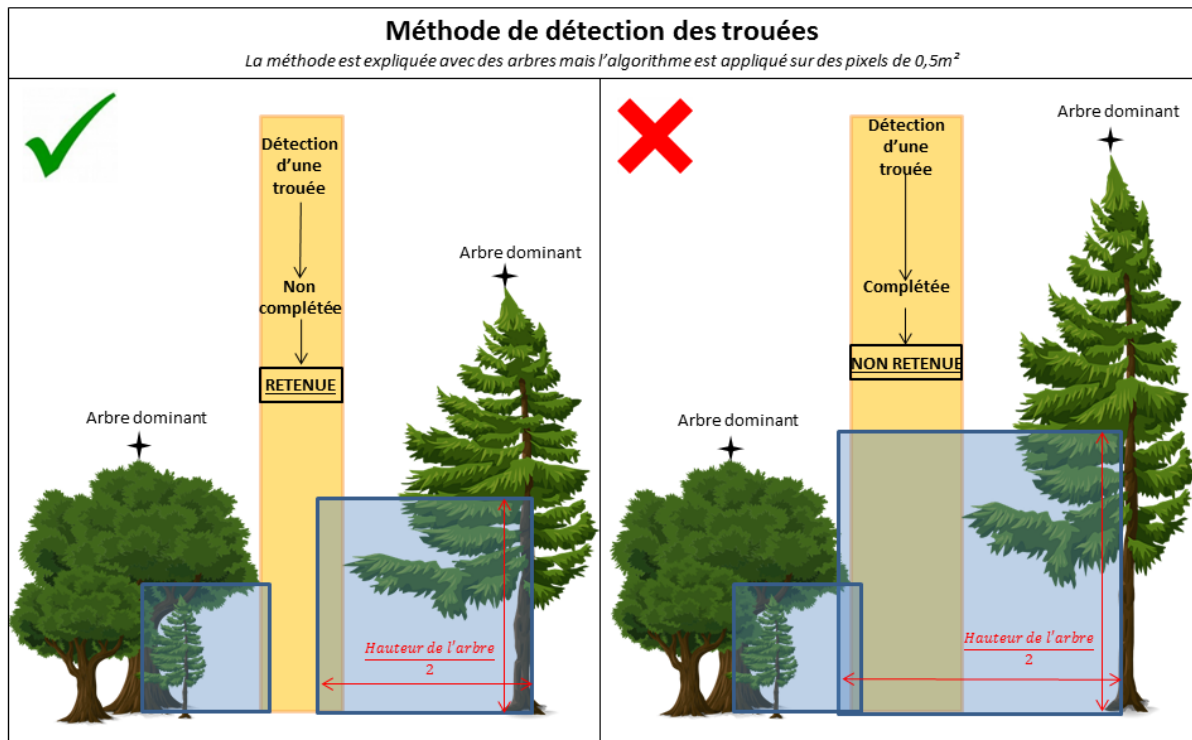
### 3. Choix du type de variable à extraire

Les variables LiDAR, communément utilisées jusqu'à aujourd'hui pour les modèles de distribution d'espèces, utilisent des statistiques résumant les caractéristiques du nuage de point sur une surface déterminée, qui sont parfois difficile à mettre en lien avec ce que l'on observe en forêt. Dans le cas du Grand Tétrás, la compréhension des effets des variables sur l'espèce est nécessaire car cela permettra de mieux comprendre quels sont les conditions nécessaires à la présence de l'espèce et ainsi d'aider à la réalisation d'actions de conservation pour l'espèce. Ainsi, pour les modèles de distribution du Grand Tétrás, des variables orienté-objet, représentant la densité des arbres ou la présence de trouées ont été calculées. De plus, pour certains modèles, l'estimation de la surface terrière a pu être ajoutée comme variable.

#### 4. Calcul des variables LiDAR

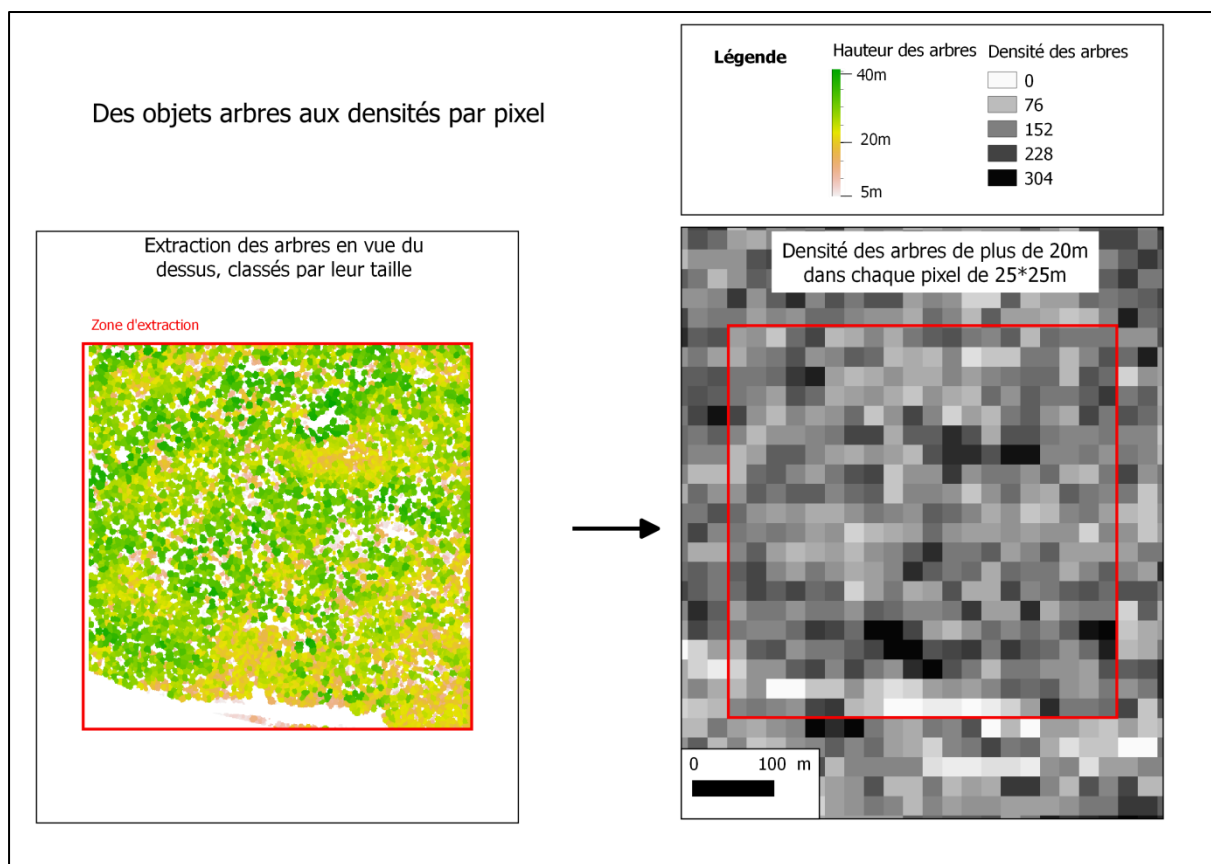
Les variables LiDAR ont été calculées grâce aux packages LidR (Roussel et Auty 2016) et LidarRtree (en développement par Jean-Matthieu Monnet) sur le logiciel R. Les métriques orientées-objet sont constituées d'éléments précis existant dans le paysage tels que les arbres ou les trouées. Ces objets sont extraits du nuage de point grâce à des caractéristiques définies à l'avance à une résolution de 0.5m. Les arbres ont été extraits avec la méthode de détection des sommets. Les trouées ont été définies comme étant les zones d'une hauteur de canopée inférieure à 1m, entourées d'arbres mesurant moins de la moitié de sa largeur (Figure 3).

*Figure 3 : Illustration de la méthode utilisée pour la détection des trouées.*



Par la suite, la moyenne résumant les caractéristiques des objets a été calculée pour une surface de 25m\*25m (taille du pixel). Enfin à partir des objets arbre de la zone de l'Ain et grâce aux données des placettes permanentes de l'ONF (localisation des arbres sur chaque placette), la surface terrière a pu être estimée (Eysn et al. 2015).

*Figure 4 : Illustration de la création d'une variable densité d'arbres (>20m) à partir des arbres détectés*



## 5. Sélection des variables

Les variables ont été sélectionnées après une première étape de recherche bibliographique et de consultation experts. Elles ont été choisies afin de représenter les différents éléments du paysage potentiellement importants pour le Grand Tétras (Tableau 1). Les mêmes variables ont été utilisées pour les modèles hiver et été à l'exception de la distance aux pistes de ski.

La densité du couvert est un élément essentiel pour cette espèce. En effet, une couverture de la canopée modérée (30-81%) a souvent été déterminée comme étant favorable pour l'oiseau (Storch 1993; Bollmann, Weibel, et Graf 2005; Hofstetter et al. 2015). Dans le but de prendre cet aspect en compte plusieurs variables ont été sélectionnées : la densité des arbres de hauteur de 5 à 10m, la densité des arbres de hauteur supérieure à 10m ainsi que la variable pourcentage de trouées de taille moyenne (200 à 1000m<sup>2</sup>).



Le Grand Tétras favorise un paysage hétérogène horizontalement, en particulier parce que des types d'habitat différents sont nécessaires au cours de l'année (Graf, Mathys, et Bollmann 2009). Pour étudier cet aspect, deux variables ont été retenues: la déviation standard de la proportion de petites trouées (25-200m<sup>2</sup>) et la déviation standard de la densité des arbres (>20m).

Le Grand Tétras est aussi communément associé à des forêts bien structurées verticalement (Bollmann, Weibel, et Graf 2005; Graf et al. 2007). L'ajout d'une variable indice de Gini pour la hauteur des arbres a donc été décidé. Le Grand Tétras n'étant pas présent dans les milieux ouverts, la variable proportion de zones ouvertes a été ajoutée aux modèles.

De plus, pour la zone de l'Ain, une variable utilisée souvent pour décrire les peuplements, la surface terrière, a été ajoutée à la place de la variable densité des arbres (>10m) qui est utilisée dans le Jura.

Le dérangement par les activités humaines a été montré dans de nombreuses études. L'hiver est la période la plus critique pour les individus, et les activités de sport d'hiver influencent grandement la présence et impactent leur survie (Thiel et al. 2007, 2008). Ainsi, une variable distance aux pistes de ski a été ajoutée aux modèles d'hiver.

Tableau 1 : Variables sélectionnées

Type de variable	Nom	Description	Indice	Importance pour le Grand Tétrás	References
Orienté-objet	Densité des arbres (<10m)	Densité des arbres mesurant moins de 10m (nombre/ha)	Indice du couvert de la strate arbutive	L'Habitat sera favorable si la surface couverte par la végétation est modérée (30-81%) (>5m)	bollmann 2005, hofstetter 2015
	Densité des arbres (>10m)	Densité des arbres mesurant plus de 10m (nombre/ha)	Indice du couvert de la canopée	L'Habitat sera favorable si la surface couverte par la végétation est modérée (30-81%) (>5m)	bollmann 2005, hofstetter 2015
	Indice de gini	Indice de Gini pour la hauteur des arbres	Indice d'hétérogénéité des hauteurs	Les habitats hétérogènes sont favorisés	Bollmann 2005, Graf 2007,
	Pourcentage de trouées de taille moyenne	Proportion de trouées ayant une surface de 200m <sup>2</sup> à 1000m <sup>2</sup> (Hauteur <1m et la demi hauteur des arbres autour plus petite que la demi-largeur de la trouée)	Pourcentage de moyenne trouées	Présence de trouées où la myrtille peut être présente (attention, les zones découvertes telle que les routes/prairies sont aussi mises en valeurs)	Bollmann 2005, Graf 2007, Storch 1993(a and b), Storch 2002, Storch 1994, Sachot 2003, Wegge 1996
	Pourcentage de zones ouvertes	Proportion de trouées supérieures à 1000m <sup>2</sup> (Hauteur <1m et la demi hauteur des arbres autour plus petite que la demi-largeur de la trouée)	Pourcentage de zones ouvertes	Zones où la surface découverte de l'objet est supérieure à 1000m <sup>2</sup> , comprends aussi les zones de prés-bois et lisières favorables aux nichées et les zones non forestières défavorables (couverture 100%)	bollmann 2005, hofstetter 2015
	Standard déviation proportion de petites trouées	Standard déviation de la proportion de trouée ayant une surface entre 25 et 200m <sup>2</sup> (Hauteur <1m et la demi hauteur des arbres autour plus petite que la demi-largeur de la trouée)	Indice d'hétérogénéité du couvert	Les habitats hétérogènes sont favorisés	Graf 2009
	Standard déviation de la densité des arbres (>20m)	Standard déviation de la densité des arbres mesurant plus de 20m (nombre/ha)	Indice d'hétérogénéité du couvert	Les habitats hétérogènes sont favorisés	Graf 2009
	Surface terrière (Gha)	Surface terrière en m <sup>2</sup> /ha (parmi les arbres de diamètre > 17,5, données de l'ONF)	Indice considérée comme un indicateur pertinent du renouvellement des futaies irrégulières.	Des milieux avec une surface terrière moyenne indiquant un habitat plus favorable	A. Depraz pers. com.
	Distance aux pistes de ski	Distance aux pistes de ski	Dérangement	Défavorable pour l'espèce dans une certaine mesure	Thiel 2008
	Autres				

## 6. Echelles spatiales

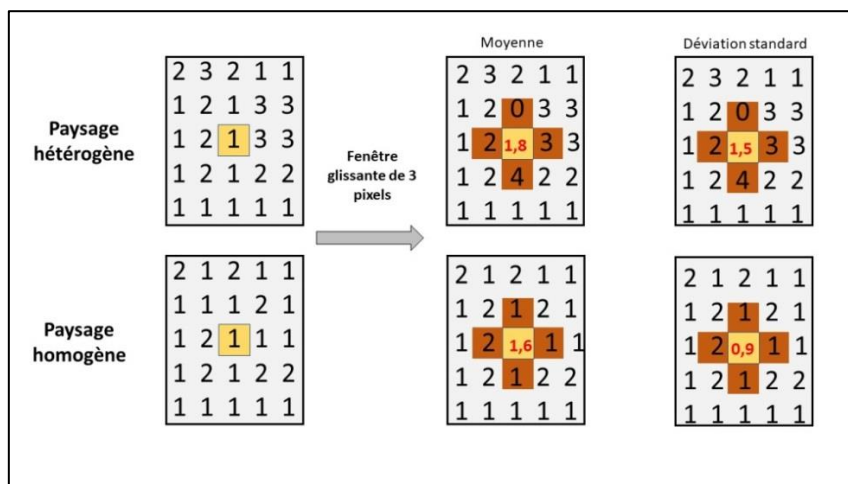
Sept échelles spatiales ont été sélectionnées pour les modèles du Grand Tétras (Tableau 2). Les trois plus larges ont été sélectionnées pour représenter la surface potentielle d'un domaine vital. Deux échelles ont été sélectionnées pour représenter le niveau de la tâche d'habitat. Enfin, deux autres échelles représentent une évaluation de l'environnement au niveau du micro-habitat.

Chaque variable a ainsi été calculée à nouveau à l'aide d'une fenêtre glissante (moyenne et déviation standard (Figure 5). Pour les zones de bordures, les valeurs réelles ont été répliquées là où il n'y en avait pas afin de limiter l'effet de bord. Cependant, cela part du principe que le milieu de chaque côté de la bordure est similaire. Une zone tampon de 25m a été appliquée par la suite afin de minimiser cet effet. La plus petite échelle (0.0625ha) représente les données brutes par pixel de 25\*25 m. Ainsi pour cette échelle, les valeurs de déviation standard ne sont pas connues.

*Tableau 2: Les échelles sélectionnées pour les modèles du Grand Tétras.*

Grand Tétras			
Niveau	Définition	Echelles	Objectifs
1	Niveau domaine vital	56ha	Prendre en compte les différents habitats sur une échelle large
		27,5ha	
		15,8ha	
2	Niveau de la tâche d'habitat	1,8ha	Identifier localement des zones favorables dans un objectif de gestion
		0,81ha	
3	Niveau micro-habitat	0,31ha	
		0,0625ha	

*Figure 5 : Schéma du calcul des variables par fenêtres glissantes.*



## 7. Modélisation

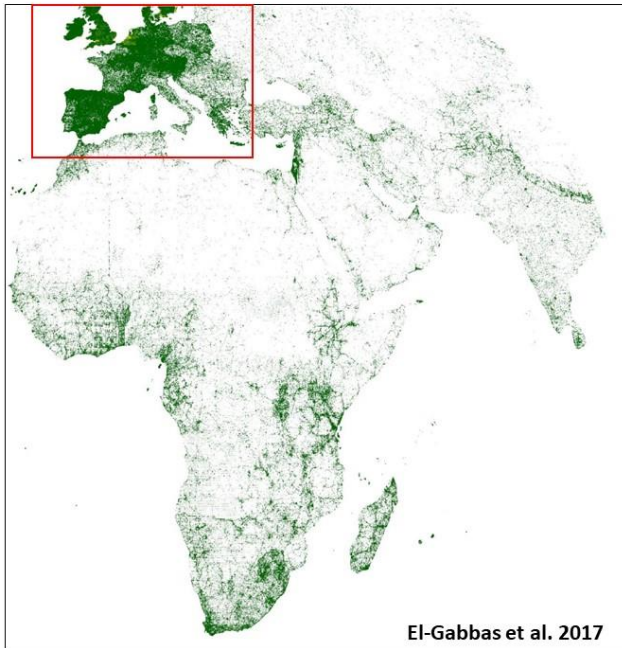
La méthode de modélisation utilisée est Maxent. Elle est réalisée à l'aide du package R Maxnet (Phillips et al. 2017). Cette méthode permet d'utiliser des données d'espèces appelées « présence uniquement », car elles ne contiennent pas de localisation où l'absence de l'espèce a été contrôlée. 10 000 points de fond ont été créés dans le but de donner au modèle des informations sur le paysage présent dans les zones prospectées. Ces points de fond ont été localisés le long des tracés de prospection pour les modèles hiver, dans le but d'éviter un **effet du biais d'échantillonnage** sur les résultats du modèle. En revanche, n'ayant pas ce type de tracé pour les données été, l'effort d'échantillonnage a été estimé en combinant les observations Gelinotte des bois et Grand Tétras. Avec les informations sur la date de l'observation et le nom de l'observateur, un nombre de visite par pixel de 2km de côté a été calculé. Cette information a été ajoutée comme variable dans les modèles été, combinée avec 10000 points de fond répartis aléatoirement.

La corrélation entre les variables sélectionnées précédemment est évaluée 2 à 2, et lorsque la valeur est supérieure à 0.7, l'une des deux variables n'est pas conservée dans le modèle (à l'exception de la variable proportion de zones ouvertes). Ces variables sont ensuite normalisées par la formule suivante :

$$\text{Valeur pixel normalisée} = \frac{\text{valeur pixel} - \text{moyenne paysage}}{\text{écart type paysage}}$$

## Biais d'échantillonnage

Le biais d'échantillonnage est connu pour influencer les résultats des modèles de distributions d'espèces (Kramer-Schadt et al. 2013; El-Gabbas et Dormann, 2017). En effet, si l'échantillonnage n'est pas systématique ou aléatoire une partie du paysage peut être prospectée de manière plus intense qu'une autre. C'est notamment le cas des données participatives où l'effort d'échantillonnage ne peut être contrôlé. Ainsi, si l'effet de ce biais n'est pas corrigé, le modèle estimera de bonne qualité les zones qui ont été favorisées par l'échantillonnage plutôt que les zones favorables à l'espèce. La difficulté est de réussir à estimer ce biais alors même que généralement les informations disponibles sont très réduites.



Carte produite avec toutes les données terrestre enregistrées dans la base de données GBIF (Global Biodiversity Information Facility)

On observe ici un biais d'échantillonnage important où l'**Europe** possède une très grande densité d'observations. Hors, il existe aussi une très grande biodiversité en Afrique ou au Moyen-Orient. Ils apparaissent ici avec beaucoup moins d'observations car ils ont été moins souvent échantillonnés donnant l'impression fausse d'une moins forte présence d'animaux et végétaux terrestres.

Pour chaque modèle, la valeur **Area Under Curve (AUC)** a été calculée. Cette valeur permet de comparer les performances des modèles entre elles. En effet, cette valeur est calculée en effectuant un test de prédiction du modèle sur une partie du jeu de donnée (présences et points de fond), et permet donc d'évaluer la capacité du modèle à prédire la qualité de l'habitat. De plus, pour tous les modèles, la contribution de chaque variable a été calculée, en additionnant les valeurs betas du modèle. Seules les variables avec une contribution supérieure à 0.05 sont présentées dans les résultats. Les courbes de réponses sont calculées en prédisant la qualité de l'habitat pour une large gamme de valeurs d'une variable, pendant que toutes les autres sont fixées à leur valeur moyenne.

Plusieurs composantes ont été sélectionnées pour ces modèles. La première est la composante linéaire, qui permet de contraindre la réponse de l'espèce à une variable à une réponse linéaire. Cependant pour plus de réalisme, car les réponses des individus à leur environnement sont complexes, la composante quadratique a été aussi sélectionnée. De plus, une attention particulière a été apportée à l'introduction d'interactions entre deux variables grâce à la composante « produit ». En effet, en caractérisant le paysage par des variables distinctes, le lien qui existe entre elles disparaît. Par exemple, la diversité du paysage peut être caractérisée par la variation du nombre d'arbres supérieurs à 20m de hauteur. Cependant, toutes les zones où aucun arbre supérieur à 20m n'a été détecté peuvent ne pas être forcément classées dans une même catégorie : jeune forêt, plantation, pâturages, végétation basse et zones urbaines peuvent constituer ces zones, rendant l'interprétation des résultats difficile. L'introduction d'une interaction entre la déviation standard de la densité d'arbre (>20m) et la variable pourcentage de zones ouvertes permet ainsi de faire la distinction lors de l'interprétation des résultats, entre les zones avec un couvert forestier, les zones sans couvert forestier que sont les pâturages ou clairières et les zones sans couvert forestier qui ne sont ni l'un ni l'autre.

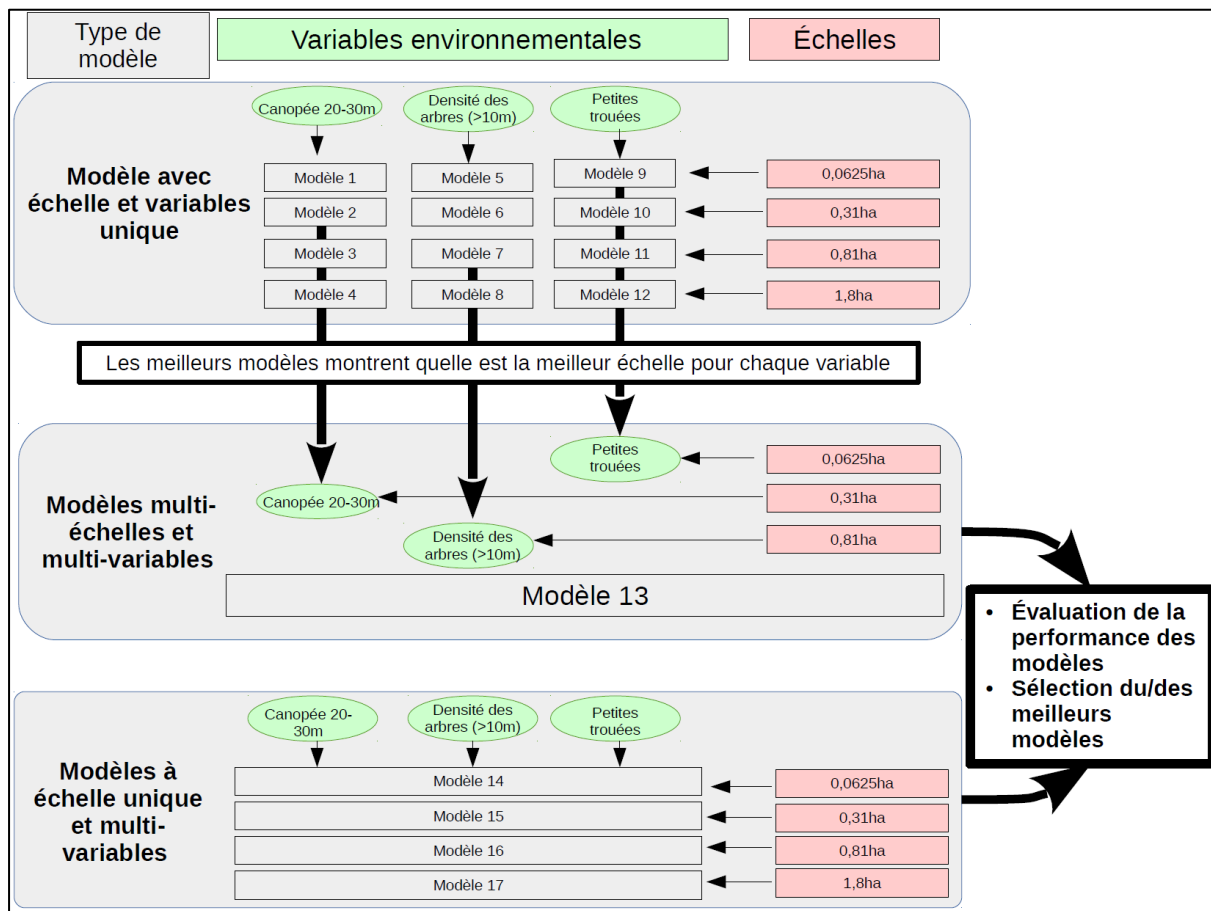
### L'Area Under Curve (AUC)

Cette valeur située entre 0 et 1 est calculée en effectuant un test de prédiction du modèle sur une partie du jeu de donnée (présences et points de fond). Elle nous permet de savoir si le modèle prédit la qualité de l'habitat mieux que s'il le faisait de manière aléatoire (valeur de 0.5). Ainsi plus la valeur est proche de 1 plus la performance du modèle est grande.

## 8. Méthode de construction des modèles multi-échelles

Dans l'optique de produire le meilleur modèle possible, la comparaison de modèles à différentes échelles a été faite dans chacune des zones d'études (Figure 6). De plus, deux méthodes ont été utilisées et comparées, celles-ci ont été décrites récemment par McGarigal et al. (2016). La première étape fût de créer des modèles à une seule échelle spatiale mais avec toutes les variables choisies récemment. Ces modèles seront appelés par la suite modèles à échelle-unique et variables multiples. La performance des modèles à ces différentes échelles a ensuite été comparée. Dans un second temps, l'élaboration d'un modèle mixte (multi-échelles et multi-variables) a été faite. Pour cela, des modèles pour chaque échelle et chaque variable ont été créés individuellement. Pour chaque variable, l'échelle à laquelle le modèle montrait une meilleure performance a été conservée pour le modèle mixte final.

Figure 6 : Schéma de la méthode de sélection du meilleur modèle à partir de différentes échelles d'étude (les numéros des modèles sont indicatifs et ne correspondent pas aux modèles délivrés).





### III. RÉSULTATS DES MODÈLES HIVER

#### 1. Performance des modèles

##### ii. Modèle à échelles unique et variable unique

Tout d'abord, des modèles à échelle et variable unique ont été fait (Tableau 3). Pour certaines variables, une échelle unique donne de meilleures performances, tandis que pour d'autres au contraire plusieurs échelles montrent la même performance. Dans ce cas, la plus petite échelle parmi celles qui ont montré les meilleures performances a été choisie. De plus on remarque que les échelles sélectionnées sont différentes suivant la zone d'étude et le sexe.

*Tableau 3: Valeurs d'AUC pour les modèles à échelle et variable uniques*

Area	Sex	Scale	Open area	Tree_density20sd	Middle_gap	Tree_gini	Tree_density10	Tree_densityinf10	small_gapsd
JURA	Female	0.0625ha	0.54	NA	0.53	0.56	0.57	0.51	NA
		0.31ha	0.56	0.56	0.55	0.59	0.59	0.51	0.59
		0.81ha	0.57	0.61	0.56	0.6	0.6	0.53	0.62
		1.8ha	0.58	0.63	0.57	0.61	0.61	0.55	0.67
		15_8ha	0.56	0.65	0.62	0.62	0.58	0.59	0.69
		27_5ha	0.52	0.65	0.63	0.63	0.58	0.59	0.68
		56ha	0.49	0.64	0.63	0.63	0.58	0.59	0.68
	Male	0.0625ha	0.56	NA	0.53	0.61	0.57	0.56	NA
		0.31ha	0.58	0.59	0.55	0.64	0.6	0.6	0.62
		0.81ha	0.6	0.62	0.56	0.66	0.6	0.62	0.63
		1.8ha	0.6	0.65	0.58	0.67	0.61	0.63	0.66
		15_8ha	0.59	0.66	0.63	0.68	0.56	0.65	0.7
		27_5ha	0.58	0.67	0.65	0.69	0.56	0.65	0.7
		56ha	0.55	0.66	0.66	0.69	0.56	0.65	0.7

Area	Sex	Scale	Open area	Tree_density20sd	Middle_gap	Tree_gini	Gha	Tree_densityinf10	small_gapsd
AIN	Female	0.0625ha	0.51	NA	0.53	0.59	0.5	0.55	NA
		0.31ha	0.52	0.54	0.57	0.64	0.51	0.59	0.62
		0.81ha	0.53	0.56	0.58	0.66	0.54	0.59	0.63
		1.8ha	0.51	0.56	0.6	0.68	0.53	0.6	0.68
		15_8ha	0.5	0.54	0.66	0.65	0.5	0.57	0.65
		27_5ha	0.49	0.55	0.67	0.64	0.5	0.56	0.63
		56ha	0.45	0.54	0.68	0.61	0.54	0.51	0.61
	Male	0.0625ha	0.48	NA	0.53	0.56	0.50	0.57	NA
		0.31ha	0.53	0.57	0.52	0.63	0.53	0.64	0.61
		0.81ha	0.53	0.58	0.48	0.64	0.51	0.65	0.65
		1.8ha	0.51	0.61	0.53	0.65	0.51	0.67	0.66
		15_8ha	0.52	0.64	0.64	0.65	0.51	0.68	0.67
		27_5ha	0.53	0.66	0.66	0.64	0.51	0.67	0.66
		56ha	0.56	0.63	0.63	0.58	0.53	0.62	0.61

##### iii. Modèles à échelle unique et variables multiples et modèles mixtes

Les modèles de la zone d'étude de l'Ain montrent une meilleure performance que ceux réalisés dans la zone d'étude du Jura (Tableau 4). Les modèles pour les poules sont de même manière généralement moins performants que les modèles pour les coqs. Les modèles mixtes ont donné les

meilleures performances sauf pour les modèles pour les poules dans la zone du Jura où une échelle de 27.5ha donne une meilleure performance.

*Tableau 4 : Résultats des valeurs AUC pour les modèles*

Echelle	AIN		JURA	
	Female	Male	Female	Male
0,0625ha	0,69	0,74	0,63	0,68
0,31ha	0,72	0,76	0,66	0,71
0,81ha	0,72	0,78	0,69	0,73
1,8ha	0,73	0,78	0,72	0,74
15,8ha	0,72	0,78	0,73	0,75
27,5ha	0,73	0,78	0,73	0,75
56ha	0,73	0,75	0,72	0,74
Mixte	0,75	0,79	0,74	0,77

## 2. Contribution des variables

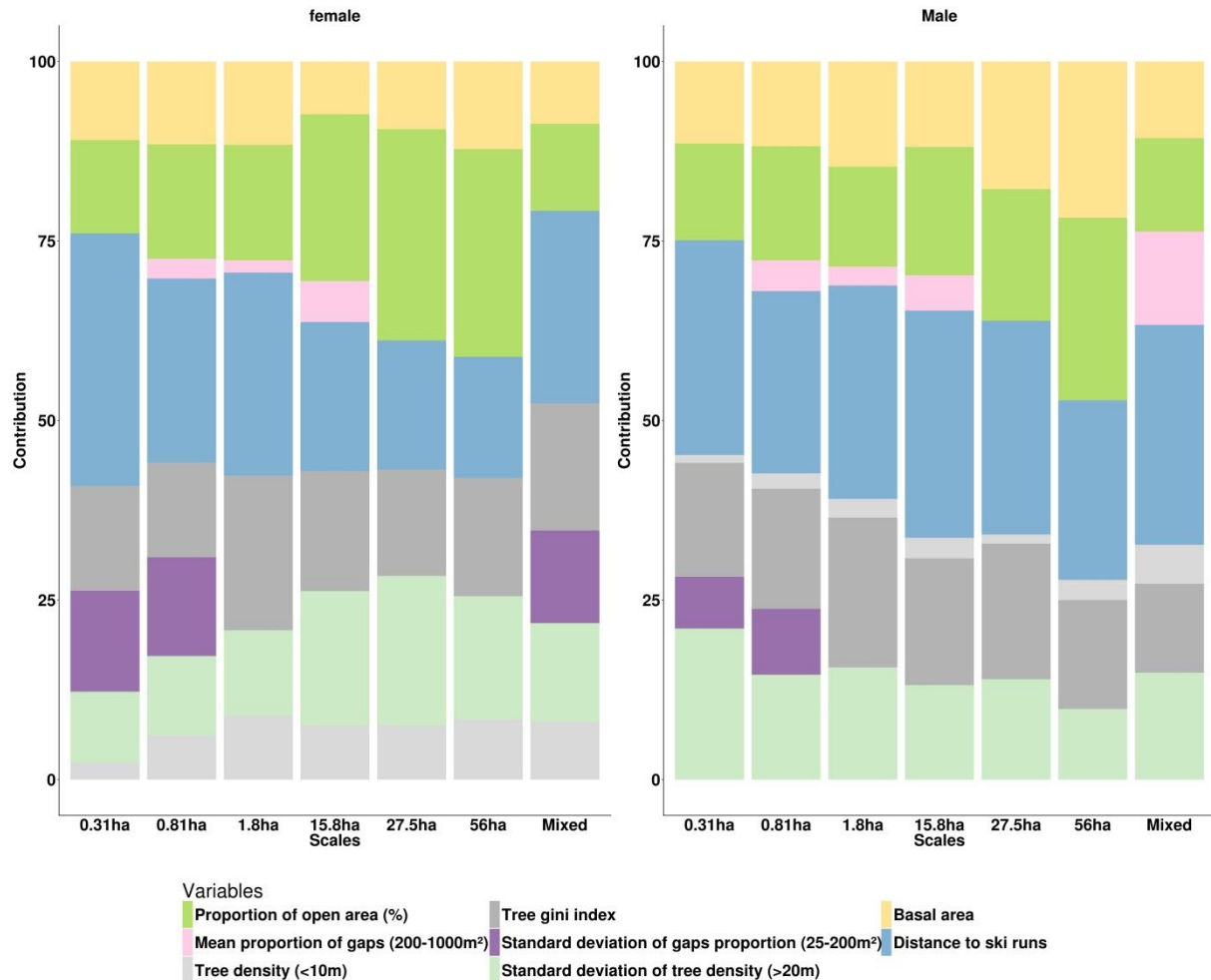
La contribution des différentes variables aux modèles permet de savoir quels sont les éléments du paysage qui influencent le plus la présence du Grand Tétrás. Ici, les résultats seront présentés en deux temps. Tout d’abord, les contributions de toutes les variables de chaque modèle seront présentées. Dans un second temps, la contribution des variables pour la plus petite échelle sera décrite (les modèles à l’échelle 0.0625ha sont composés de variables différentes).

### i. Contribution de toutes les variables

Dans les deux zones d’étude, la distance aux pistes de ski contribue à chaque échelle des modèles (Figure 7). Dans la zone de l’Ain, les variables indice de Gini, surface terrière, proportion de zones ouvertes et déviation standard de la densité d’arbres (>20m) sont importantes dans les modèles malgré des variations suivant les échelles. La déviation standard de la proportion de trouées (25-200m<sup>2</sup>) contribue dans les modèles à petite échelle uniquement, pour les coqs et les poules. La densité des arbres (<10m) contribue aux modèles des poules mais contribue peu aux modèles pour les coqs. La proportion moyenne de trouées (200-1000 m<sup>2</sup>) contribue aux modèles aux échelles 0,81 ha, 1,8 ha et 15,8 ha pour les deux sexes. Dans les modèles mixtes, les différentes variables sont plus homogènes. Des différences entre les modèles mixtes coqs et poules sont observés. En effet, la variable déviation standard de la proportion de trouées (25-200m<sup>2</sup>) contribue uniquement au modèle

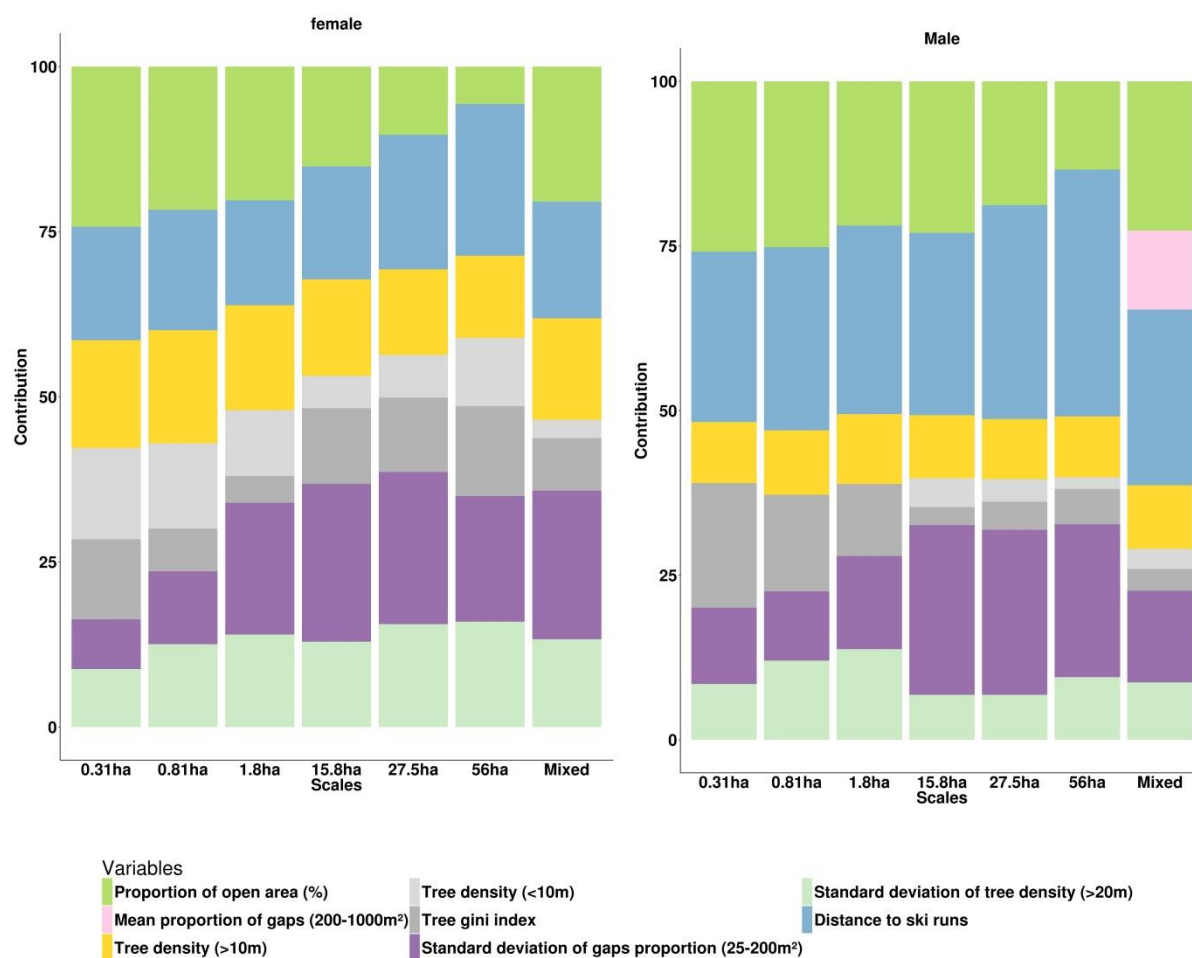
de poules, tandis que la variable proportion moyenne de trouées (200-1000 m<sup>2</sup>) contribue uniquement pour les modèles de coqs.

*Figure 7 : Contribution aux modèles des variables dans la zone de l'Ain*



Dans la zone d'étude du Jura (Figure 8), la proportion de zones ouvertes, la densité d'arbres (>10m) et la déviation standard de la proportion de trouées (25-200m<sup>2</sup>) contribuent le plus aux modèles des deux sexes. La variable indice de Gini contribue le plus pour les modèles de poules, mais contribue aussi aux modèles de coqs à l'échelle 0.31ha. La variable proportion moyenne de trouées (200-1000m<sup>2</sup>) contribue uniquement pour les modèles de coqs.

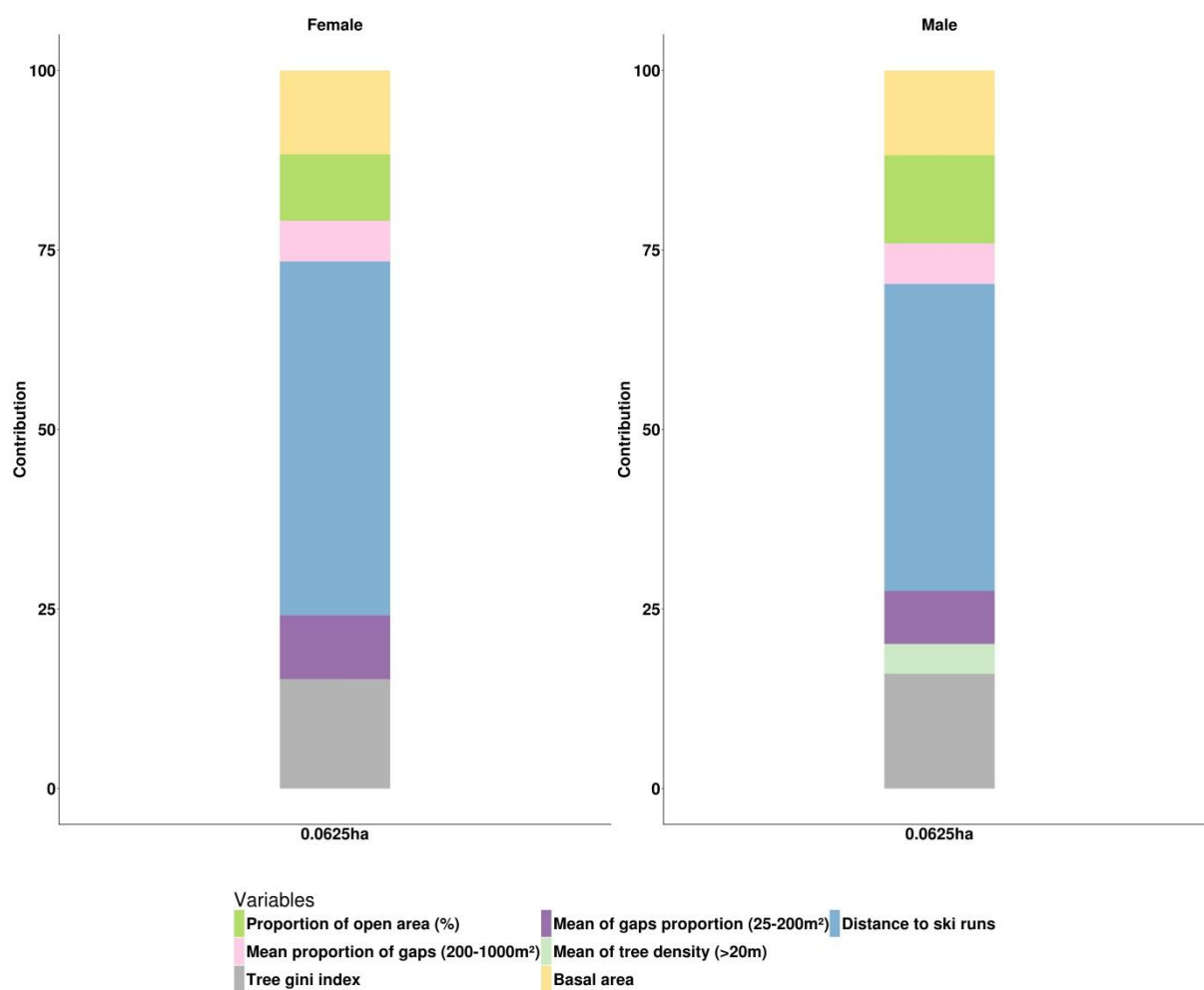
Figure 8 Contributions des variables pour la zone du Jura.



## ii. Contribution des variables à petite échelle

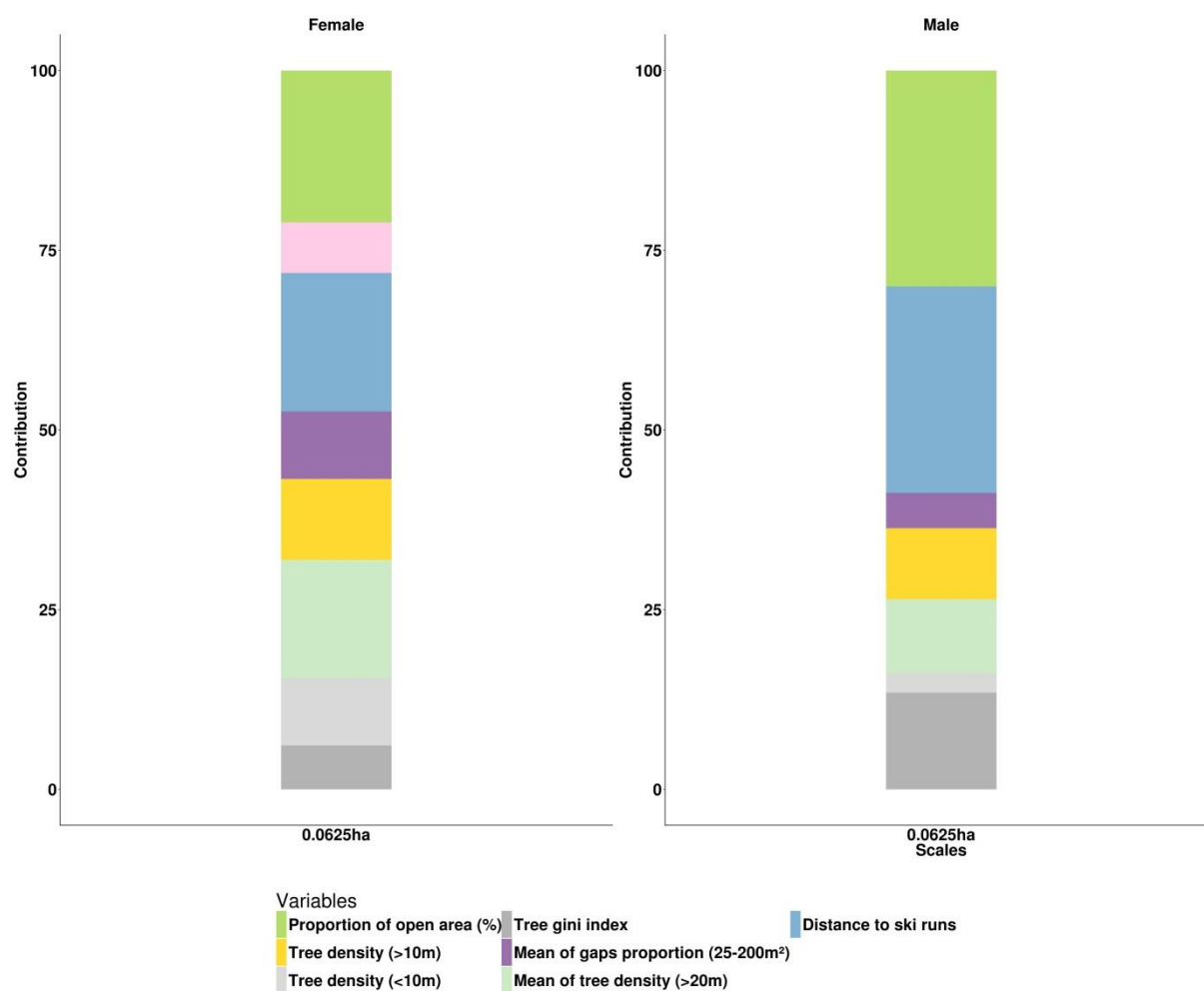
À la plus petite échelle (0.0625ha), dans la zone de l'Ain, la contribution des variables est similaire entre les sexes (Figure 9). La variable distance aux pistes de ski contribue le plus aux modèles, suivie de l'indice de Gini, la surface terrière et la proportion de zone ouvertes.

Figure 9: Contribution des variables à l'échelle 0.0625ha pour la zone de l'Ain



Dans la zone du Jura, la variable qui contribue le plus est aussi la distance aux pistes de ski avec un plus grand impact pour les modèles coqs (Figure 10). Elle est suivie de la proportion de zones ouvertes et de la densité d'arbres (>20m). Pour les modèles de coqs, l'indice de Gini ainsi que la densité d'arbres (>10m) contribue aux modèles tandis que la densité d'arbres (<10m), la proportion de trouées (25-200m²) et la densité d'arbre (>10m) contribuent aux modèles de poules.

Figure 10: Contribution des variables à l'échelle 0.0625ha dans la zone du Jura.



### 3. Courbes de réponses

Les courbes de réponses aux variables qui contribuent le plus aux modèles décrits dans le paragraphe précédent ont été calculées (figure 10 et 12). Elles permettent de comprendre comment l'espèce répond aux différentes valeurs de chaque variable, et ainsi de savoir quelles sont les valeurs favorables ou non. Ces courbes de réponse sont présentées pour les sept échelles et pour les modèles mixtes afin d'en observer l'effet sur la forme de la réponse. Les réponses possibles sont de deux types : linéaires ou quadratiques. Les réponses observées sont très similaires entre coqs et poules à l'exception de deux variables : la déviation standard de la densité des arbres (>20m) et la

déviation standard de la proportion de trouée (25-200m<sup>2</sup>). Seules les réponses des cinq variables qui contribuent le plus aux modèles sont présentées.

- Distances aux pistes de skis

Les réponses sont similaires entre les échelles et montrent une distance optimale aux pistes de 1km. Ainsi, les zones éloignées des pistes de ski sont défavorables. Cela est dû à une forte association des domaines skiable à un environnement favorable pour le Grand Tétrás (zone forestière en montagne de haute altitude). Plus on s'éloigne de ces conditions, moins l'habitat est favorable pour l'oiseau. La proximité avec les pistes de ski à un effet défavorable pour les individus du fait du dérangement occasionné.

- Indice de Gini

Aux trois plus larges échelles, l'indice de Gini avec une valeur de 0.15 ou plus sont favorables. Aux échelles plus petites, un optimum pour des valeurs autour de 0.2 est observé. Ceci indique qu'une certaine hétérogénéité de la hauteur de la végétation est nécessaire, en particulier à large échelle. Cela correspond à nos connaissances de l'oiseau qui favorise les milieux hétérogènes.

- Surface terrière

Les courbes de réponse de la variable surface terrière montre que des valeurs intermédiaires sont les plus favorables pour le Grand Tétrás (entre 30 et 50 m<sup>2</sup>/ha). A l'échelle de 0.0625ha, la réponse est différente et un plateau (valeurs favorables) est atteint à 50m<sup>2</sup>.ha. Dans ce cas, les réponses sont concordantes entre les échelles, et montre que pour l'espèce, une surface terrière trop faible (qui caractérise des milieux plus ouverts) n'est pas favorable. De même, les zones très denses avec une surface terrière importante sont aussi défavorables.

- Proportion de zones ouvertes

Les courbes de réponses à la variable proportion de zones ouvertes, montrent qu'à large échelle, une proportion dans le paysage de 25% de zones ouvertes est le plus favorable. En revanche, cette

réponse forme un plateau pour les valeurs supérieures à 25%, prédisant de ce fait des habitats favorables pour le Grand Tétrás pour les proportions de zones ouvertes très grandes. A plus petite échelle, les courbes de réponses montrent un habitat plus favorable pour de grandes proportions de zones ouvertes dans le paysage. Ici, les courbes de réponse montrent les limitations du modèle. En effet, si le fait qu'une absence totale de zones ouvertes caractérise des milieux non favorables aux oiseaux est plus en accord avec nos connaissances (zones de prés-bois, lisières favorables aux nichées), il est connu que les milieux totalement ouverts ne sont pas favorables pour le Grand Tétrás. La limitation des réponses possibles restreintes aux seules linéaires et quadratiques, ne permet pas au modèle de proposer une réponse plus complexe qui serait plus proche de la réalité. De plus, à large échelle, la probabilité de présence de zones ouvertes dans l'habitat augmente du fait de la fragmentation des zones forestières en îlot distincts. Il faut cependant souligner ici que sur les cartes de prédiction finales, les zones ouvertes apparaissent bien comme défavorables, indiquant que d'autres variables permettent de compenser le manque de complexité des réponses individuelles.



Figure 11 : Courbes de réponse pour les poules dans la zone de l'Ain

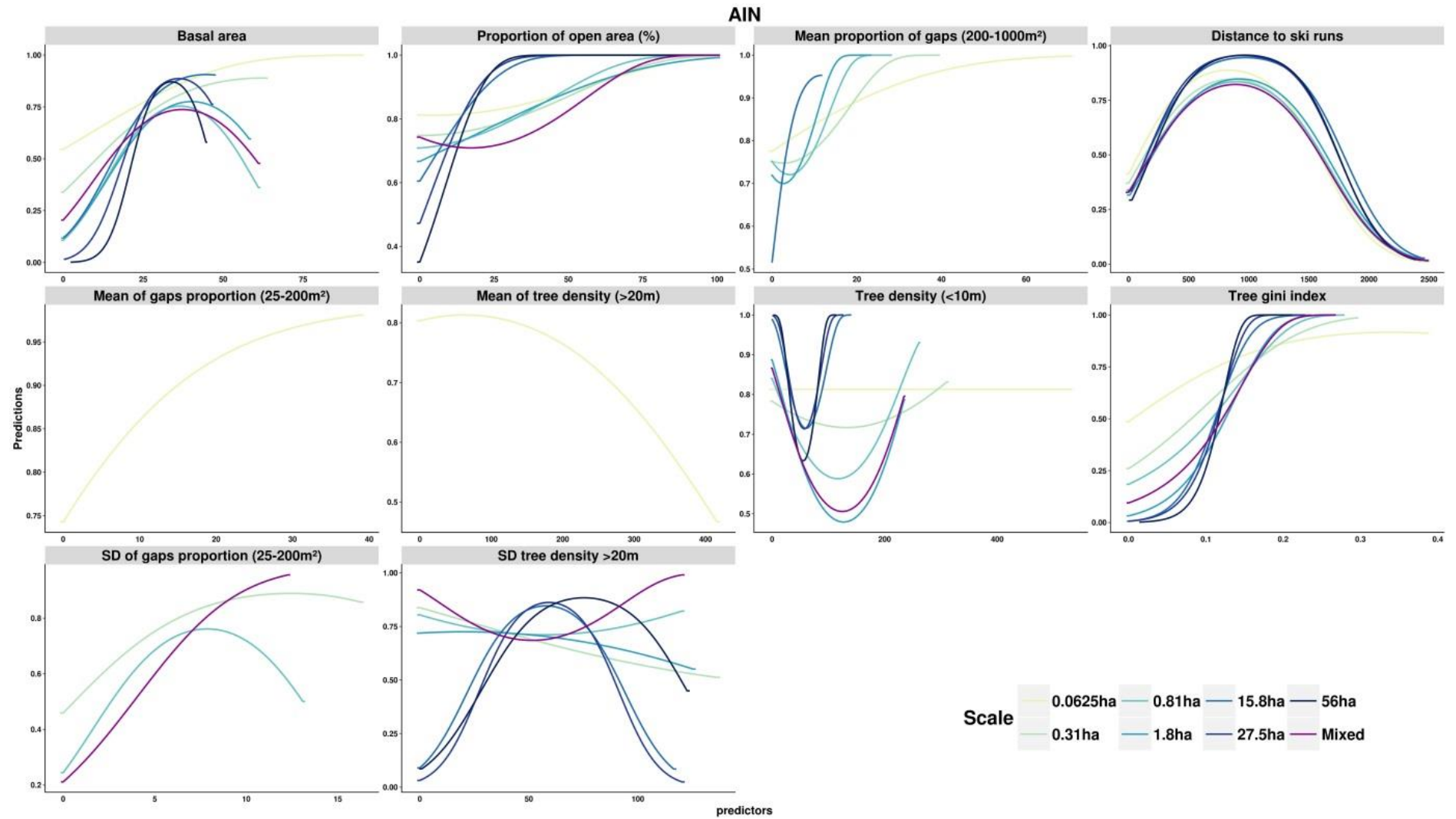
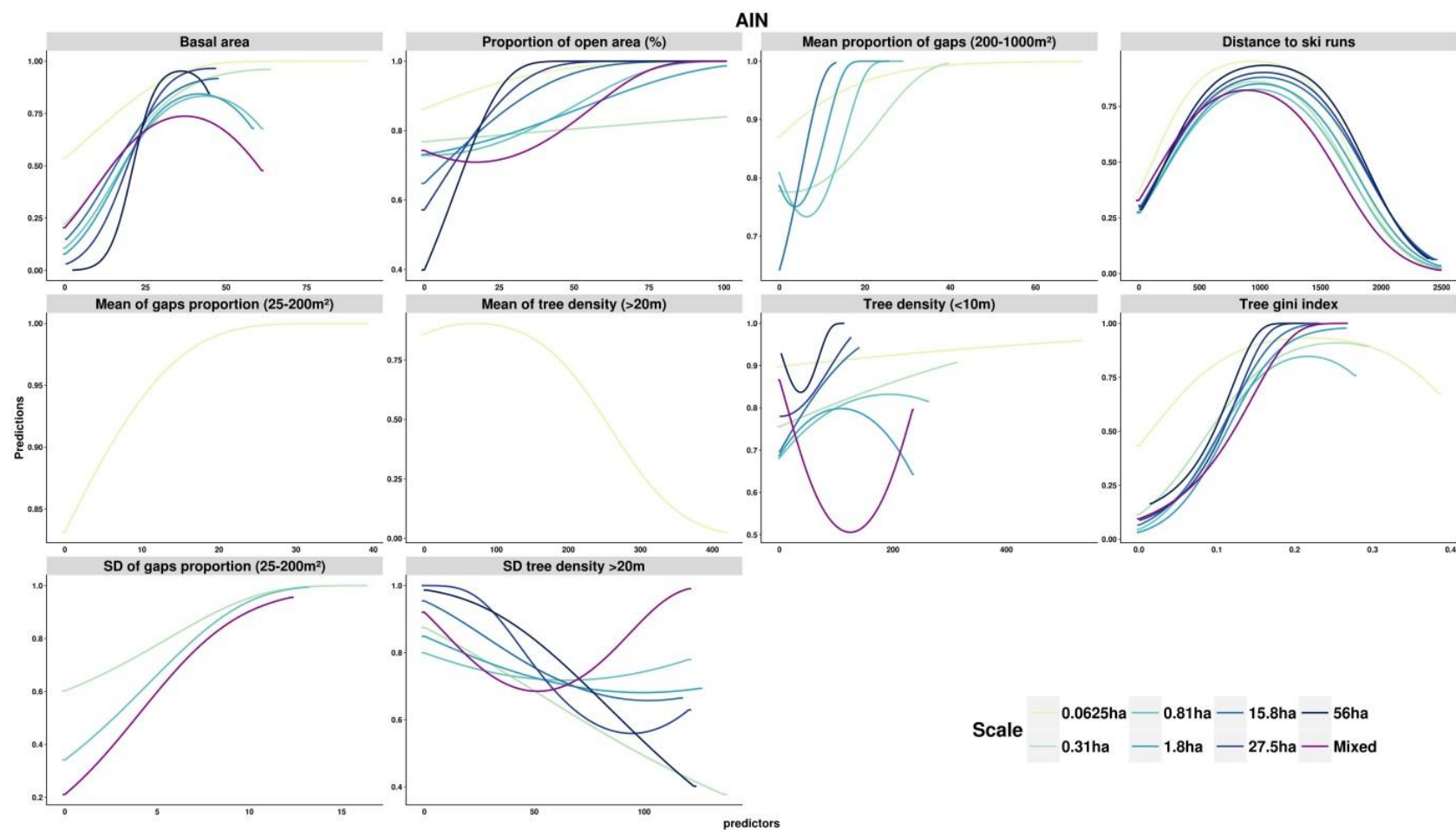


Figure 12: Courbes de réponse pour les coqs dans la zone de l'Ain



Pour les modèles du Jura, les réponses sont différentes entre les sexes pour les variables proportion de zones ouvertes, densité d'arbres (>10m) et l'indice de Gini.

- Distance aux pistes de ski

Comme pour les modèles de l'Ain, une distance optimale de 1km est observée.

- Proportion de zones ouvertes

Pour les poules, de fortes proportions de zone ouvertes sont non-favorables à toutes les échelles. La même observation est faite pour les modèles de coqs pour les petites échelles uniquement. A large échelle, une proportion intermédiaire (25 à 60%) est prédite comme défavorable. Ici contrairement à la zone de l'Ain, on observe une réponse plus cohérente avec nos connaissances de l'espèce pour les poules. En revanche, la réponse observée pour les modèles coqs rejoint les observations de l'Ain. Dans le cas des modèles de la zone Jura, les zones ouvertes sont prédites comme défavorables sur les cartes de prédiction.

- Déviation standard de la densité des arbres (>20m)

Pour les modèles de poules, les plus hautes valeurs de déviation standard sont défavorables à petite échelle (réponse linéaire). Par contre, à large échelle, les valeurs moyennes sont les moins favorables (réponse quadratique). Au contraire, pour les coqs, des valeurs hautes de déviation standard sont défavorables à toutes les échelles. Une haute valeur de déviation standard indique qu'une densité hétérogène d'arbres supérieurs à 20m est favorable au Grand Tétras. La réponse à large échelle des modèles pour les poules ne correspond pas nos connaissances de l'importance d'une mosaïque d'habitat. La collection de plus de données sur l'habitat des poules en hiver pourrait permettre de corriger cette réponse, cependant la limitation du type de réponse possible sera toujours un élément important à prendre en compte.

- Déviation standard de la proportion de trouées (25-200m<sup>2</sup>)

Les courbes de réponses montrent que les plus hautes valeurs sont les plus favorables à toutes les échelle et pour les deux sexes. Pour cette variable, les résultats sont clairs et montrent que, pour la zone du Jura, une mosaïque d'habitat est nécessaire. On peut supposer au vu de nos connaissances qu'une alternance entre les zones avec un couvert de canopée faible (grande proportion de trouées) et des zones de canopée plus dense est favorable pour l'espèce.

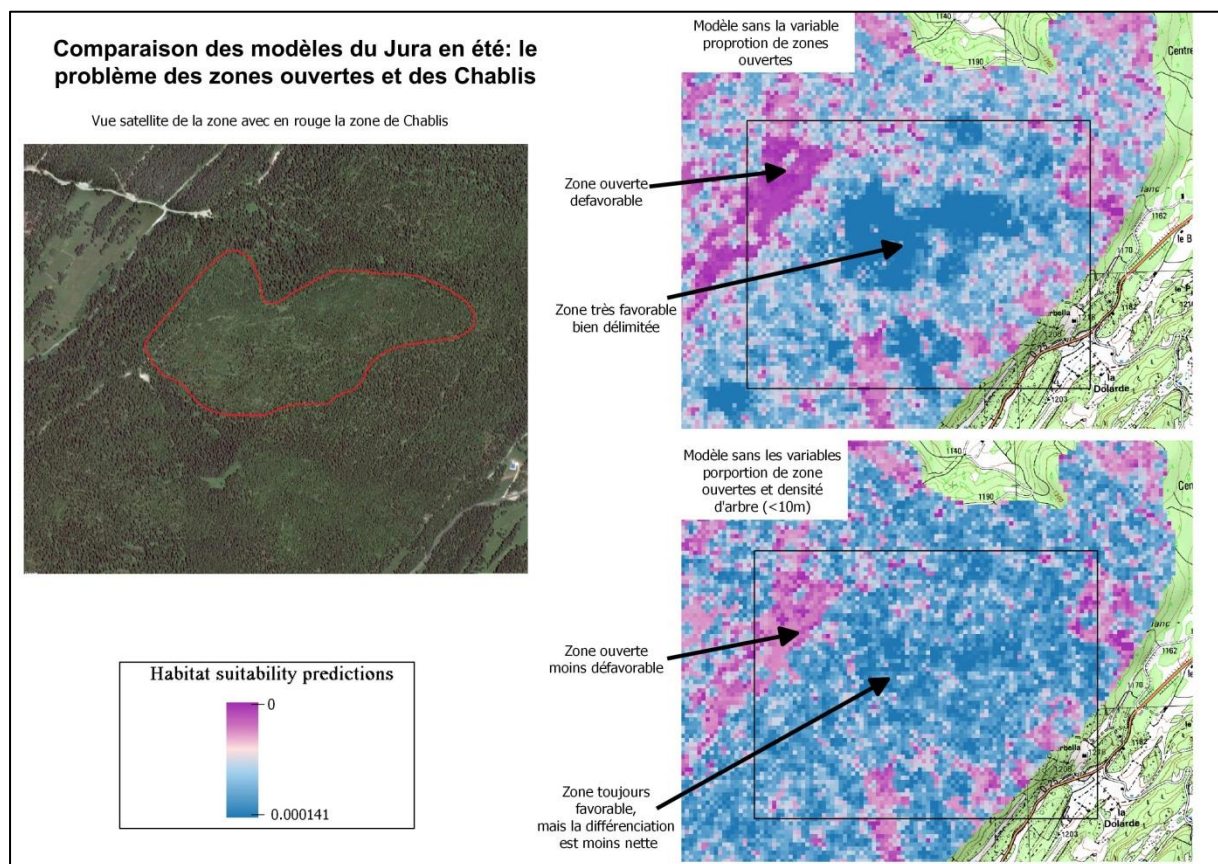
## IV. RÉSULTATS DES MODÈLES ÉTÉ

---

### 4. Difficultés pour la zone du Jura

Les modèles d'été créés pour la zone du Jura posent plusieurs difficultés. En effet, dans la première version du modèle utilisant les variables sélectionnées, ils favorisaient les zones de chablis aujourd'hui en phase de régénération et qui sont caractérisées par une forte densité d'arbre entre 5 et 10m (réponse linéaire positive) (Figure 13). De plus, les zones ouvertes n'étaient pas prédites comme défavorables. Afin de pallier à ces problèmes, les variables proportion de zones ouvertes et densité d'arbres (<10m) ont été retirées des modèles. Malgré cela, les résultats restent peu satisfaisants pour cette zone, prédisant toujours, mais dans une moindre mesure, les zones de chablis comme étant favorables. Les résultats de ces modèles sont tout de même présentés ci-dessous.

Figure 13 : Illustration du problème des zones de chablis



## 5. Modèles performance

### iii. Modèles à échelles multiples et variables multiples (mixte)

Dans la zone de l'Ain, toutes les variables ont été sélectionnées au niveau du domaine vital (larges échelles), à l'exception de la surface terrière (échelle 10.8ha). Dans la zone du Jura, seulement les variables déviation standard de la densité d'arbres (>20m), proportion de trouées (200-1000m<sup>2</sup>), l'indice de Gini et la densité d'arbre (>10m) sont sélectionnées au niveau domaine vital. Toutes les autres variables sont sélectionnées au niveau du patch d'habitat.

Tableaux 5 et 6 : Valeurs des AUC pour les modèles à variable et échelle unique

Area	Scale	Open area	SD Tree_density20	gaps (200-1000m <sup>2</sup> )	Tree_gini	Tree_density10	Tree_densityinf10	SD gaps (>1000m <sup>2</sup> )
JURA	0.0625ha	0.54	NA	0.47	0.57	0.48	0.59	NA
	0.31ha	0.54	0.53	0.51	0.59	0.52	0.58	0.56
	0.81ha	0.55	0.55	0.5	0.6	0.52	0.59	0.57
	1.8ha	0.57	0.55	0.49	0.61	0.52	0.6	0.56
	15.8ha	0.59	0.58	0.49	0.63	0.53	0.6	0.51
	27.5ha	0.59	0.61	0.5	0.63	0.53	0.59	0.52
	56ha	0.59	0.62	0.54	0.63	0.51	0.59	0.53

Area	Scale	Open area	SD Tree_density20	gaps (200-1000m <sup>2</sup> )	Tree_gini	Gha	Tree_densityinf10	SD gaps (>1000m <sup>2</sup> )
AIN	0.0625ha	0.58	NA	0.56	0.65	0.67	0.59	NA
	0.31ha	0.59	0.6	0.59	0.69	0.69	0.62	0.67
	0.81ha	0.58	0.66	0.63	0.7	0.71	0.64	0.7
	1.8ha	0.6	0.69	0.63	0.71	0.72	0.65	0.73
	15.8ha	0.63	0.74	0.66	0.73	0.71	0.66	0.75
	27.5ha	0.63	0.74	0.66	0.73	0.71	0.65	0.76
	56ha	0.65	0.7	0.65	0.72	0.7	0.64	0.73

#### iv. Modèles à échelle unique et variables multiples et modèles mixtes

Pour les modèles de l'Ain, toutes les valeurs d'AUC montrent de bonnes performances (0.8 à 0.84) (Tableau 7). La meilleure performance a été observée pour le modèle mixte pour cette zone. Pour les modèles Jura, les performances sont moins bonnes (0.75 à 0.76) et trois modèles (27.5 ha, 56 ha et mixte) ont les plus hautes valeurs d'AUC.

Table 7: Valeurs d'AUC pour les modèles étés

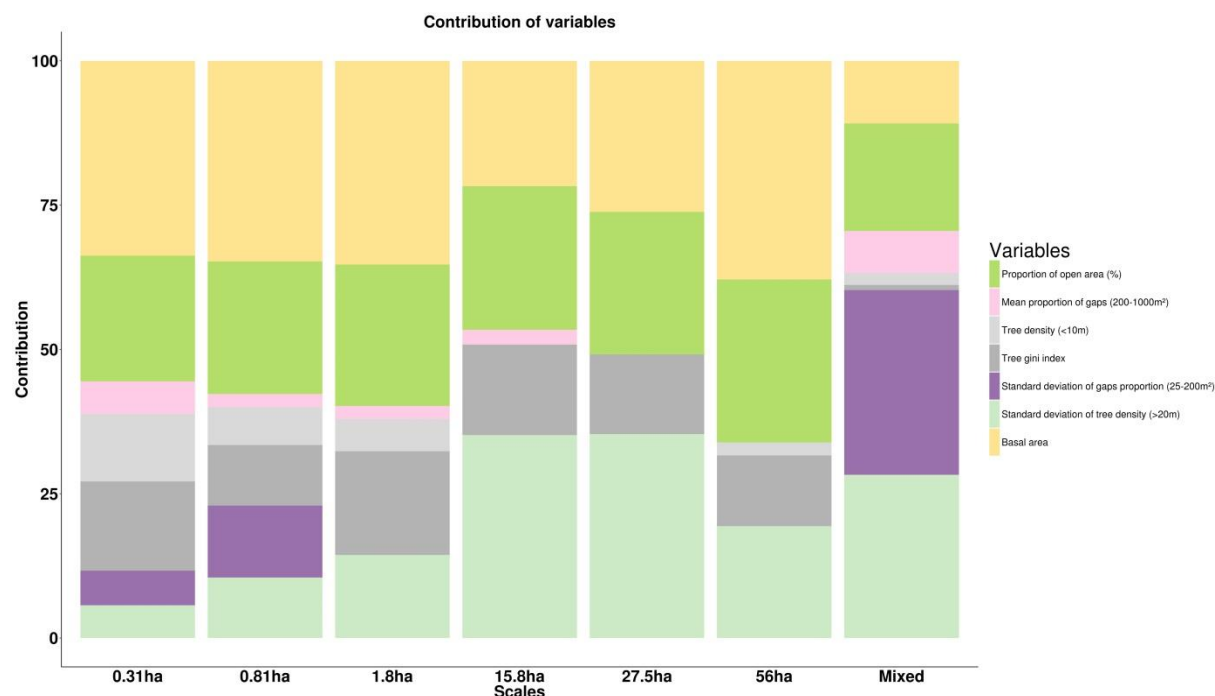
Scale	JURA	AIN
0,0625ha	0,75	0,8
0,31ha	0,75	0,81
0,81ha	0,75	0,81
1,8ha	0,74	0,81
15,8ha	0,75	0,83
27,5ha	0,76	0,83
56ha	0,76	0,82
Mixte	0,76	0,84

## 6. Contribution des variables

Dans la zone de l'Ain, les variables surface terrière et proportion de zone ouverte contribuent le plus aux modèles. La déviation standard de la densité des arbres (>20m) contribue le plus à large échelle (Figure 14). La variable déviation standard de la proportion de trouées (25-200m<sup>2</sup>) est la variable qui

contribue le plus au modèle mixte. Cette variable contribue modérément aux modèles à petite échelle. La variable indice de Gini contribue aux modèles aux échelles 0.31 ha et 1.8 ha.

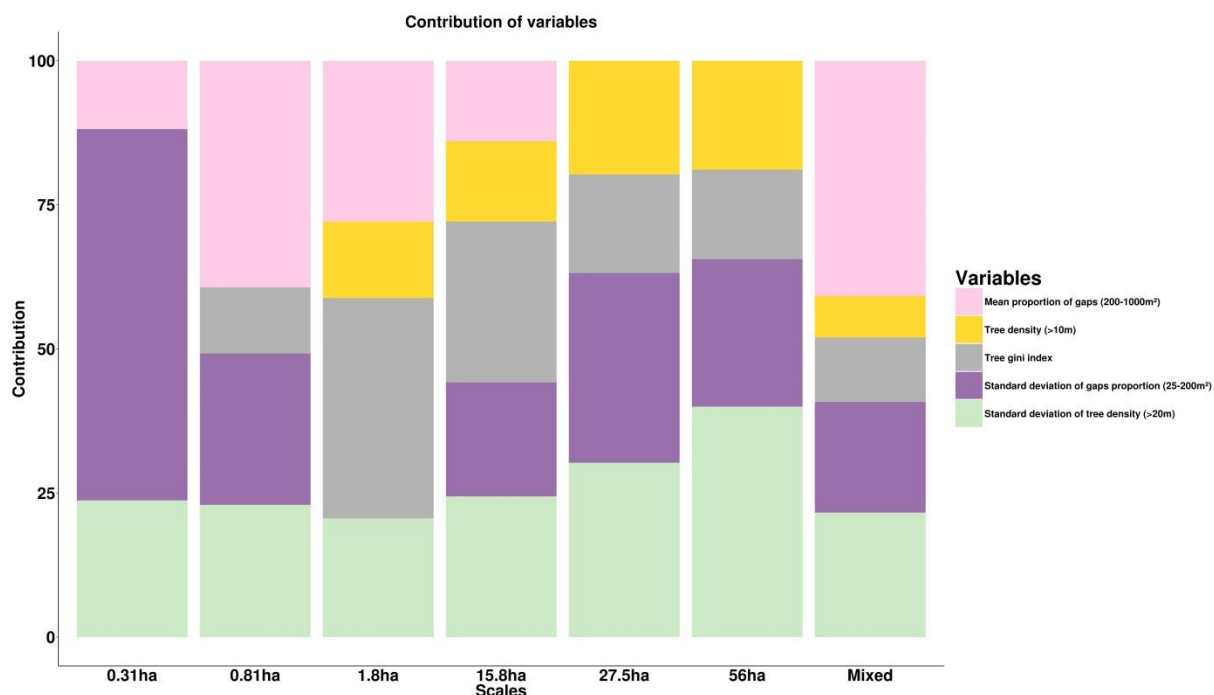
*Figure 14 Contribution des variables pour la zone de l'Ain.*



Dans la zone du Jura, les variables contribuant aux modèles diffèrent fortement entre les échelles (Figure 15). La variable déviation standard de la densité d'arbres (>20 m) participe à tous les modèles. La variable proportion de trouées (200-100 m²) participe aux modèles 0.81ha, 1.8ha et au modèle mixte. La déviation standard de la proportion de trouées (25-200 m²) et l'indice de Gini contribue à tous les modèles sauf à l'échelle de 1.8ha.



Figure 15 : Contribution des variables dans la zone du Jura pour les modèles été



## 7. Courbes de réponse

Les courbes de réponses sont similaires entre les échelles et le modèle mixte à l'exception des variables indice de Gini et déviation standard de la proportion de trouées (25-200m<sup>2</sup>) (Figures 16 et 17).

- Surface terrière

La courbe de réponse montre que les valeurs entre 20 et 25m<sup>2</sup>/ha sont plus favorables, avec des valeurs plus petites pour les plus petites échelles. Ce sont des valeurs plus faibles que celle observées pour les modèles hiver de chaque sexe. En été, la fréquentation par les poules de zones plus ouvertes, avec la présence d'une strate herbacée pour l'élevage des jeunes, peut être à l'origine d'une préférence pour des surfaces terrières plus faibles en été.

- Proportion de zones ouvertes

Les plus grandes proportions de zones ouvertes sont prédites comme étant défavorables pour les petites échelles et le modèle mixte. Cependant, les valeurs de proportion faibles et hautes (>25%) sont prédites comme étant favorable aux trois plus larges échelles. Ici, la présence à petite échelle de





zones ouvertes est défavorable et rejoint nos connaissances sur le Grand Tétrás. A plus large échelle, la réponse observée diffère et entre en contradiction avec nos connaissances. Comme précédemment, les limitations du modèle en termes de type de réponses peuvent engendrer des extrapolations aux valeurs extrêmes qui ne sont pas pertinentes.

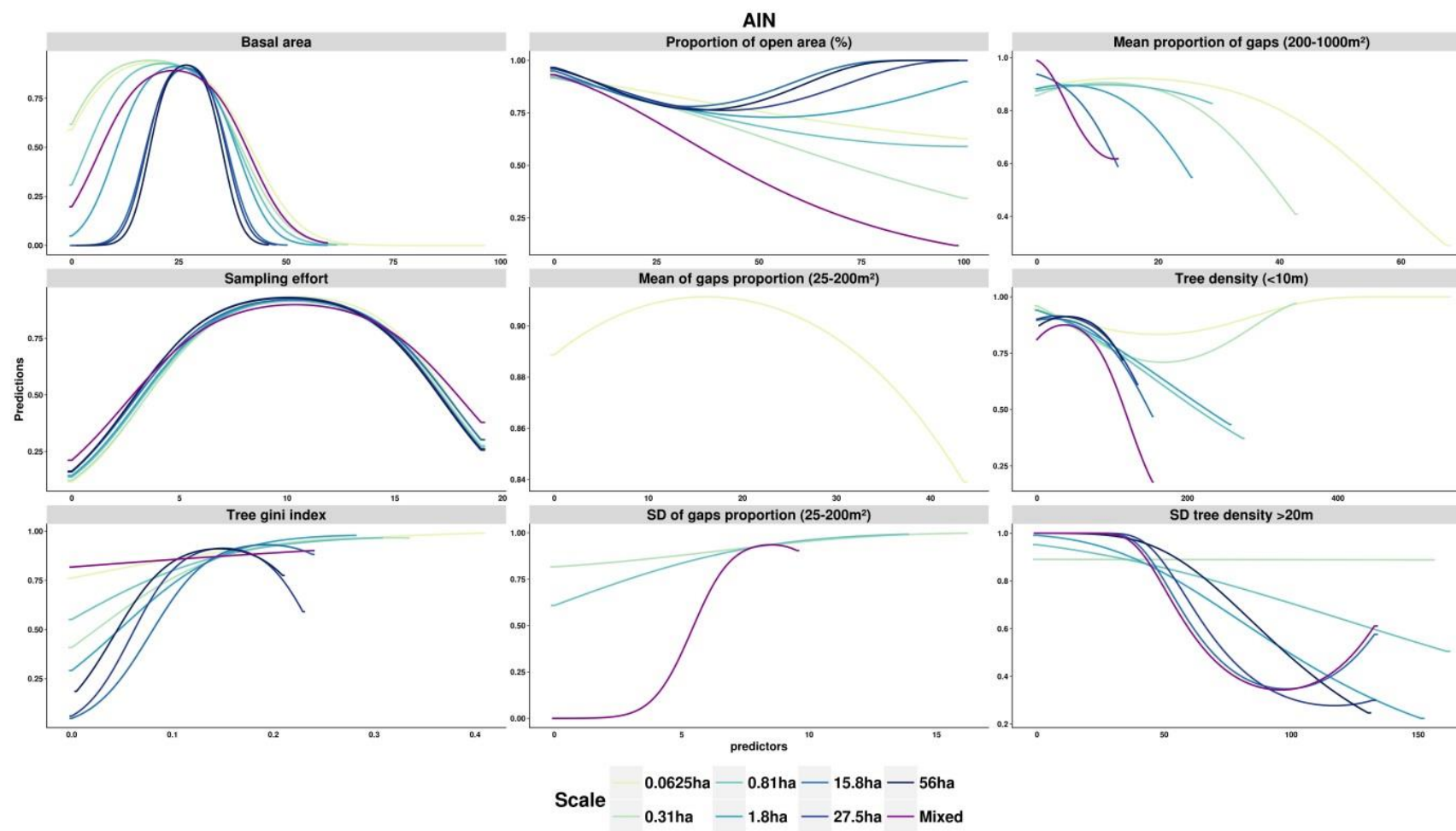
- Déviation standard de la densité des arbres (>20m).

A petite échelle et pour le modèle mixte, les valeurs les hautes sont défavorables. En revanche, à large échelle, une réponse quadratique est observée avec des valeurs défavorables autour de 100.

- Indice de Gini

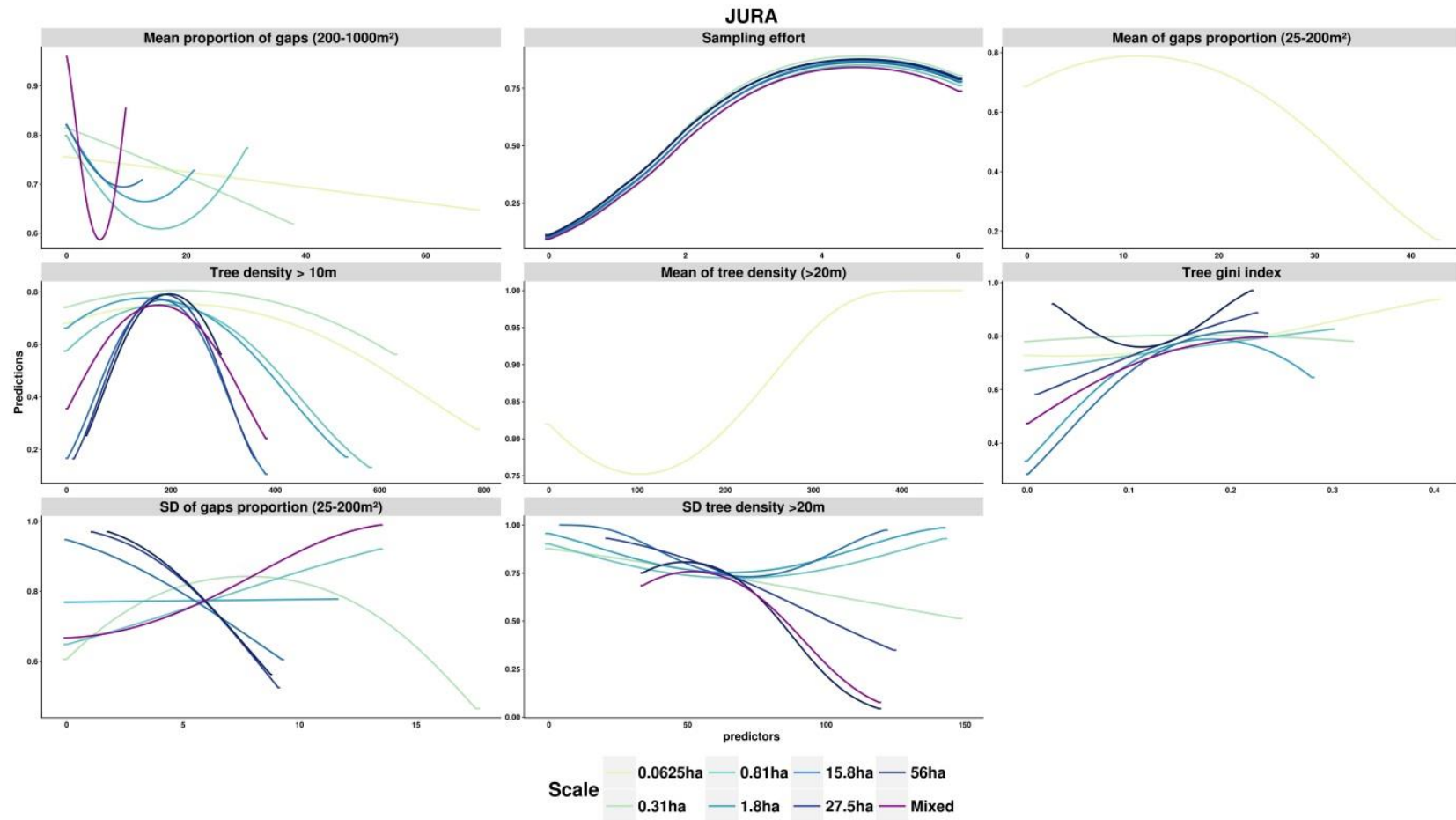
A petite échelle et pour le modèle mixte, un indice de Gini avec de hautes valeurs est prédit comme étant le plus favorable. En revanche, à large échelle une valeur entre 0.1 et 0.15 est le plus favorable. Cela rejoint nos observations des modèles d'hiver, où un habitat bien structuré est favorable pour l'espèce.

Figure 16 : Courbes de réponse pour la zone de l'Ain



Les courbes de réponse des modèles de la zone Jura diffèrent entre les échelles pour de nombreuses variables.

Figure 17 : Courbes de réponse pour la zone du Jura



- Déviation standard de la densité d'arbres (>20m)

Pour cette variable, aux échelles 0.31 ha et 27.5 ha, les valeurs de fortes hétérogénéités sont défavorables. En revanche, pour les échelles 0.81 ha à 15.8 ha, les valeurs de déviation standard intermédiaire ont été prédites défavorables. Au contraire, pour l'échelle 56 ha et le modèle mixte, l'hétérogénéité moyenne est plus favorable (valeur autour de 50).

- Proportion moyenne des trouées (200-1000m<sup>2</sup>)

Les réponses des deux plus petites échelles sont linéaires et montrent qu'une forte proportion de trouées est défavorable. Pour toutes les autres échelles et le modèle mixte, on observe une réponse quadratique, qui prédit un environnement défavorable pour des proportions de trouées de 10 à 20%. Les valeurs plus basses et plus hautes sont en revanche favorable en comparaison.

- Déviation standard de la proportion de trouées (25-200m<sup>2</sup>)

Dans le cas de cette variable, les réponses observées sont très différentes suivant les échelles. Ainsi, on observe à petite échelle (0.31ha) un optimum pour une valeur de 7. Le modèle à l'échelle 0.81ha et le modèle mixte montrent une réponse linéaire positive. Pour l'échelle 1.8ha, la réponse est monotone tandis que pour toutes les autres échelles, une réponse linéaire négative est observée.

- Indice de Gini

Pour l'indice de Gini, les réponses à petite échelle montrent que les plus hautes valeurs de cet indice sont favorables à l'espèce. En revanche, à large échelle, un optimum à une valeur de 0.2 est observé.

## V. INTERPRÉTATION DES CARTES DE PRÉDICTION

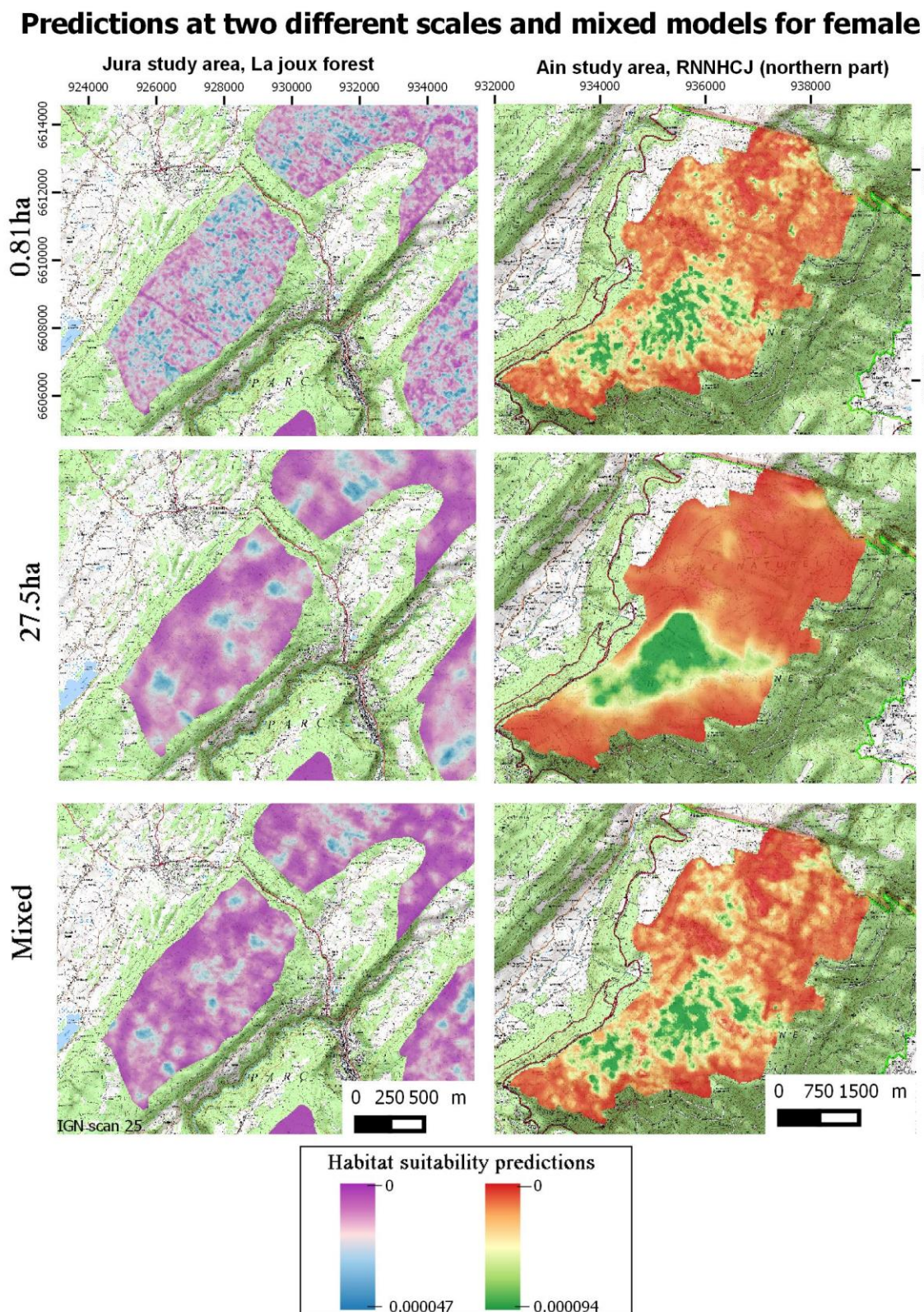
---

Des cartes de prédictions ont été réalisées pour chaque modèle. Ces cartes furent créées en utilisant le mode exponentiel de Maxent (figure 18). Les valeurs de chaque pixel représentent la valeur relative de la qualité de l'habitat par rapport à l'habitat présent dans le paysage. Ainsi, la comparaison des valeurs des pixels entre elles permet de hiérarchiser les habitats présents du meilleur au moins bon. Cependant, la meilleure valeur ne représente pas forcément la valeur du meilleur habitat de manière générale. Si dans le paysage, le meilleur habitat présent est un habitat de qualité moyenne du point de vue de l'espèce, il peut quand même avoir la valeur maximale parmi les prédictions du modèle. La valeur proposée en tant que telle est de plus dépendante du nombre d'observation et du nombre de point de fond donnés au modèle. Ainsi, une même valeur dans chacune des zones d'étude ne correspond pas à la même qualité de l'habitat. Il faut donc utiliser les résultats des deux zones d'études séparément.

Pour les cartes des modèles hiver, on remarque que les modèles à petite échelle mettent en lumière la présence de patchs très favorable au Grand Tétras et d'autres au contraire défavorables (hors zone ouvertes). A plus large échelle, de larges zones favorables apparaissent au sein des massifs forestiers. Pour les modèles mixtes, les résultats dépendent beaucoup de l'échelle de sélection des variables. Ainsi, si de nombreuses variables sont sélectionnées à petite échelle, on retrouve une carte avec l'apparition de petits patchs dans la zone d'étude.



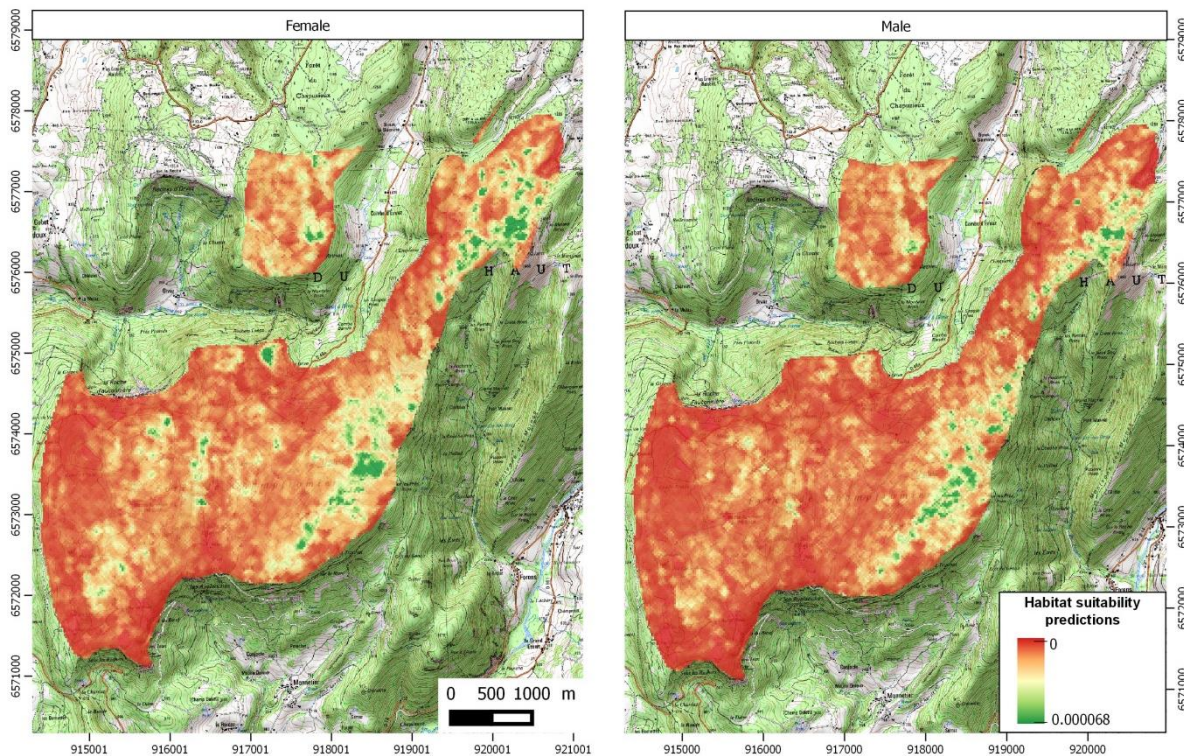
Figure 18: Cartes de prédiction à plusieurs échelles chacune des zones d'études des modèles hiver. Pour le Jura les couleurs roses indiquent les zones les moins favorables, tandis que pour la zone d'Ain ces zones sont montrées en rouge.





Quand on compare les résultats des modèles entre les poules et les coqs, on remarque tout d'abord que les mêmes zones sont prédites favorables pour les deux (Figure 19). Cependant, on remarque que ces patchs sont plus restreints pour le coq que pour la poule. Ainsi les poules sélectionnent des habitats similaires aux coqs mais fréquentent aussi d'autres types d'habitat moins favorable pour les mâles.

*Figure 19 : Comparaison entre les valeurs prédites pour les Coqs et les Poules à l'échelle 0.81 ha (Forêt de Champfromier, Ain)*



## 8. Comment les variables participent-elles aux valeurs prédites localement?

Si l'étude de la contribution des variables et des courbes de réponses montre à l'échelle du modèle les effets des variables sur la distribution du Grand Tétras, l'attribution d'une valeur de prédiction basse où haute peut être dû localement à l'effet d'une ou plusieurs variables. L'utilisation de « Local Interpretable Model-Agnostic Explanation » (LIME) permet, pour des valeurs spécifiques de prédictions, de savoir quelles variables et dans quelle mesure elles ont contribué à cette valeur. C'est particulièrement intéressant pour comprendre l'attribution de valeurs prédites extrêmes. Ainsi, le package R LIME a été utilisé pour observer cet effet local des variables pour 20 valeurs prédites parmi

les plus hautes (quantile 0.999) et plus basse (quantile 0.001) du paysage. Dans la Figure 20, sont présentés les résultats obtenus pour 20 valeurs parmi les plus hautes à l'échelle de 0.31ha dans la zone du Jura. Ainsi, on remarque que pour certaines valeurs (exemples des cas 4, 6, 10, 11), toutes les variables ont un effet positif, expliquant ainsi la valeur haute de la prédiction. Dans d'autres cas (1, 5, 7, 8 par exemple), une ou deux variables ont un effet négatif, sans pour autant que la valeur prédite ne soit basse. Ainsi, les variables telles que la densité d'arbre (>10m), l'indice de Gini ou la proportion de zones ouvertes ont une influence positive très forte si leur valeur locale est favorable.

Au contraire, pour des valeurs prédites parmi les plus basses du paysage (figure 21), on observe dans tous les cas une contribution négative de ces mêmes variables expliquant ainsi la valeur basse de la prédiction. Dans certains cas on observe une contribution positive de la valeur de la proportion de zones ouvertes, de la densité d'arbre (<10m) ou de la déviation standard de la proportion de trouées (25-200m<sup>2</sup>) (cas 2, 8, 15 par exemple), mais celles-ci ne compensent pas une valeur non favorable des autres variables.



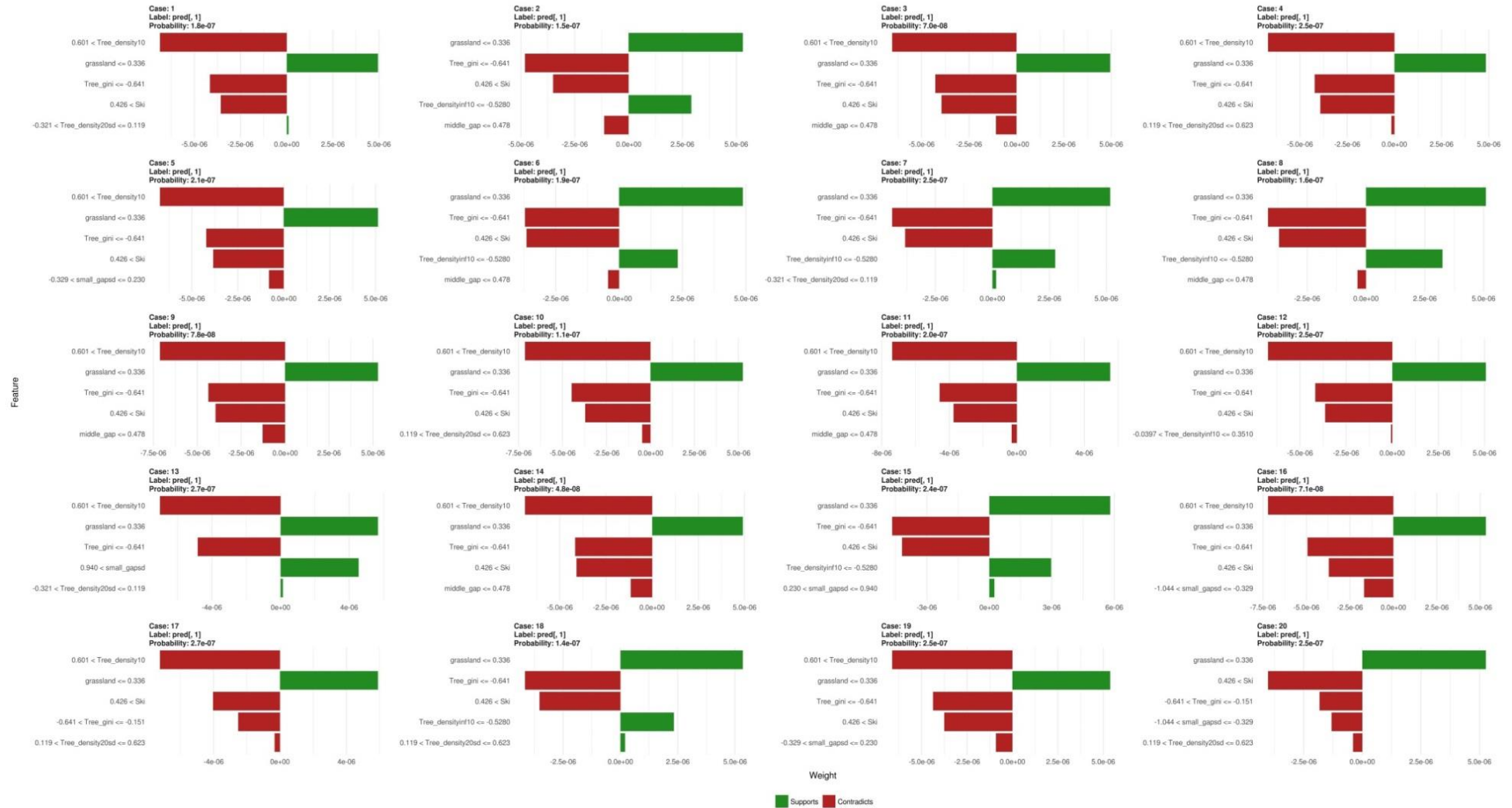
Figure 20 : 20 valeurs prédites parmi les 0.01% meilleures dans la zone du Jura (0.0625ha)

JURA, scale 0.31ha, 20 random values (Quantile = 0.999)



Figure 21: 20 valeurs prédites parmi les 0.01% plus mauvaise dans la zone du Jura (0.0625ha)

JURA, scale 0.31ha, 20 random values (Quantile = 0.001)



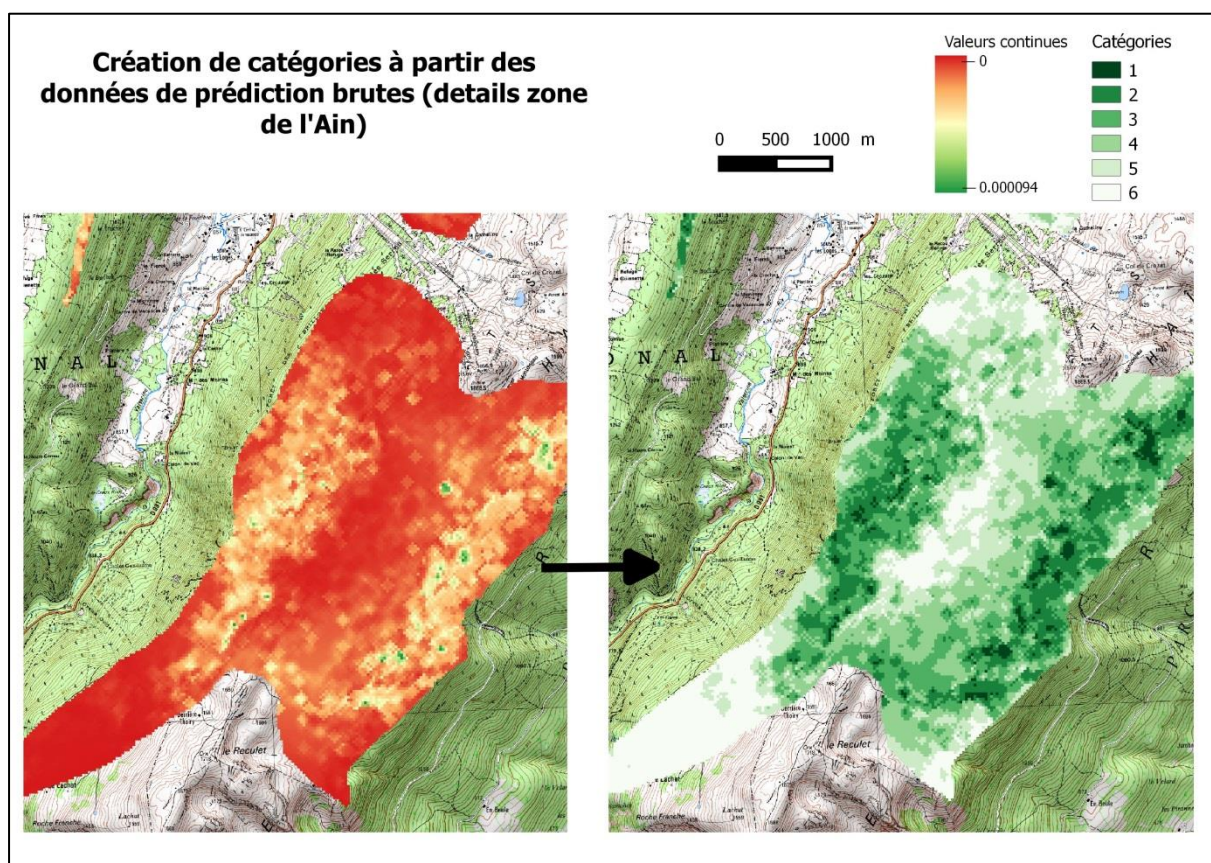
## 9. Création de catégories pour les cartes de prédictions

Dans le but de proposer des cartes de prédictions plus facilement interprétables, des cartes simplifiées ont été créées à partir des cartes de prédiction des modèles. Ainsi, les valeurs de prédictions ont été catégorisées en 6 classes définies par les valeurs seuil des quantiles 0.10, 0.25, 0.50, 0.75, 0.90 (Figure 22, Tableau 8). De même que pour les valeurs de prédictions, les catégories ne peuvent s'interpréter que dans le contexte définit de la zone d'étude. Ainsi, les valeurs de catégorie 1 dans la zone de l'Ain et la zone de Jura ne peuvent pas être comparées, chacune représentant les meilleures valeurs dans la zone d'étude uniquement.

*Tableau 8 : Catégories créées pour les cartes de prédiction*

Seuils des quantiles	Catégorie	Signification
0-0,10	6	Les 10% des valeurs les plus basses du paysage
0,1-0,25	5	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,1
0,25-0,50	4	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,25
0,50-0,75	3	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,50
0,75-0,90	2	Les 15% de valeurs au-dessus de la valeur seuil au quantile 0,75
0,90-1	1	Les 10% des valeurs les plus hautes du paysage

Figure 22 : Exemple de catégorisation des prédictions pour une échelle de 0.31ha.



## 10. Validation en forêt de Champfromier

Le lien entre les résultats des modèles et la qualité de l'habitat à dire d'expert sur le terrain a été évalué. Neuf placettes représentant les différentes catégories prédites par les modèles ont été sélectionnées et visitées en Forêt de Champfromier. Une bonne adéquation entre les valeurs des modèles a été observée, malgré des modèles plutôt optimistes concernant les habitats d'été. Les placettes P11 et P12 sont classées par les modèles en 1 ou 2 suivant les échelles et le sexe. En suivant le protocole du Groupe Tétras Jura, une note de qualité de l'habitat de B a été attribuée à ces placettes, indiquant un habitat de bonne qualité. La placette P10 a été classée 1 à petite échelle pour les deux sexes (puis 3 et 2 aux plus larges échelles) et correspond bien à une note d'habitat de B pour l'hiver. En revanche, cette placette ne représente qu'un habitat de qualité C pour l'été principalement à cause d'une haute strate composée d'herbacées et de mégaphorbiaies. Les placettes 17 et 14 ont été classées par les modèles en 4 ou 3 suivant les échelles et le sexe. La note D

obtenue par ces placettes sur le terrain indique en effet une mauvaise qualité de l'habitat pour le Grand Tétrás. La classification en catégorie 6 ou 5 est souvent observée pour les zones ouvertes et pâturages, ainsi les placettes sélectionnées à ces endroits n'ont pas été visitées. Ceci peut expliquer aussi, que des valeurs de classe 4 soit attribuées à des zones forestières défavorables, elles sont perçues comme plus favorable par le modèle que les zones ouvertes, obtenant ainsi une meilleure note. Ainsi, il semble raisonnable de considérer les catégories au-dessus de 4 comme défavorables pour l'espèce. Ceci a été confirmé par les placettes P3 et P8 classé en catégories 5 par le modèle et en qualité C sur le terrain. Nous avons aussi vu que la catégorie 1 représentant les meilleurs habitats semble comprendre des habitats de bonne qualité mais non optimaux. On peut supposer que la rareté des habitats optimaux pour l'espèce conduit à une classe des 10% des meilleures valeurs hétérogène.

Pour les modèles été, la discrimination entre les zones de bonne et mauvaise qualité est plus difficile, car cette qualité dépend pour beaucoup de la composition de la strate au sol (herbacées, myrtilles ...), qui sont difficilement détectables avec le LiDAR. Ainsi à l'exclusion des zones de chablis dont nous avons vu plus haut qu'elles posaient problème pour les modèles d'été, les catégories prédites sont similaires à celles prédites par le modèle hiver aux différentes placettes visitées. En particulier, la placette P10 identifiée comme étant un habitat de qualité B pour l'hiver mais C pour l'été, à cause d'une végétation au sol trop fournie, est classée en catégorie 1 par les deux modèles.



*Figure 23 : Placettes visitées en forêt de Champfromier*

**P12**

**Note modèle 1, Note terrain B**



**P3**

**Note modèle 5, Note terrain C**



**P14**

**Note modèle 4, Note terrain D**



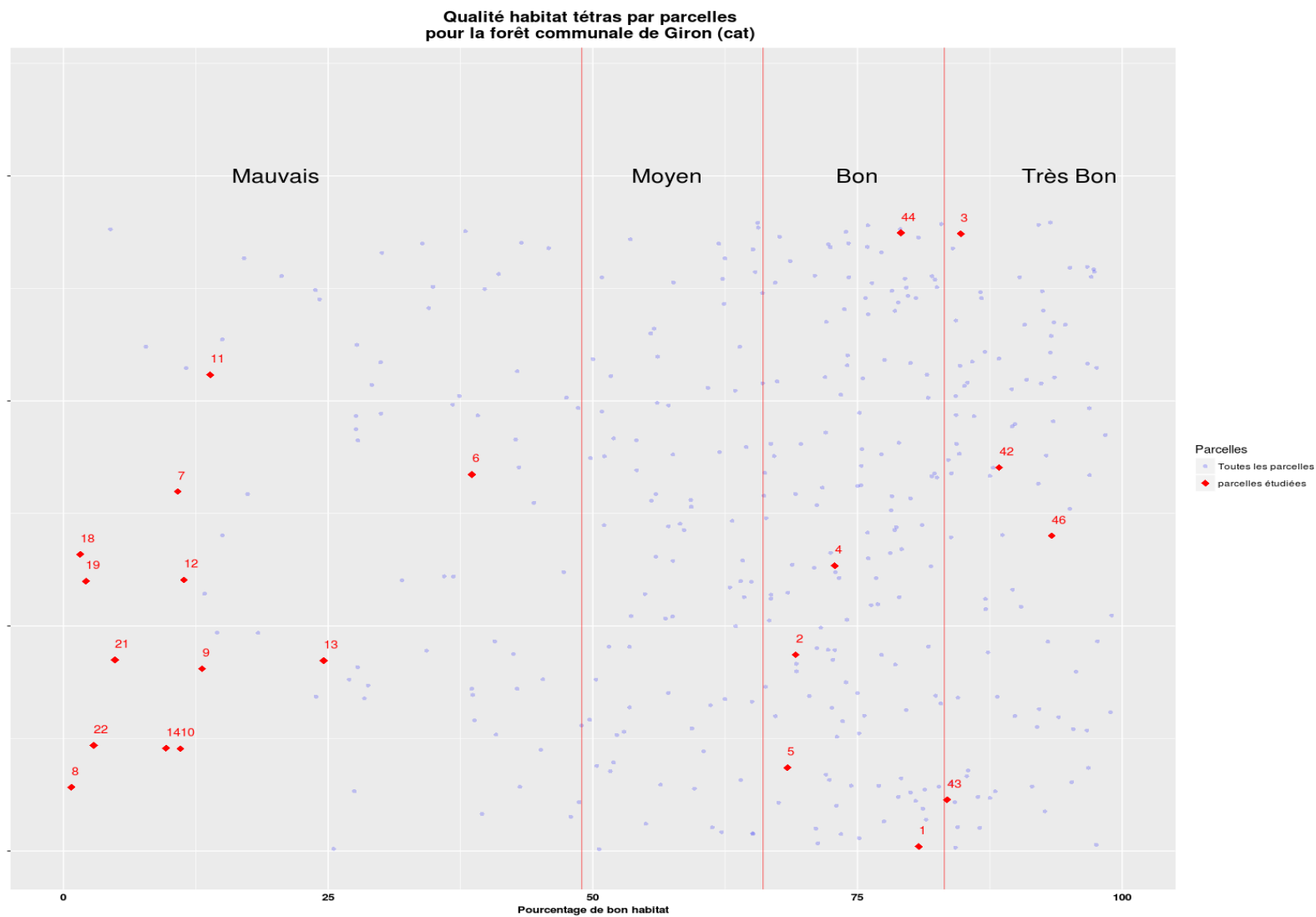
Cette validation en forêt de Champfromier nous a aussi permis de constater une limite du modèle concernant la discrimination entre les zones ouvertes, les routes et les chemins forestiers. En effet, avec la définition mise en place pour la détection des objets trouées, les routes et chemin de par leurs longueurs ont des surfaces très grande, et entrent de ce fait dans la catégorie zones ouvertes. Ces zones ouvertes sont prédites défavorables par les modèles. Hors les chemins forestiers, en particulier ceux présentant une strate herbacées fournie, sont en réalité des habitats favorables pour les nichées. Ainsi, une attention particulière doit être portée sur ces zones lors de l'interprétation du modèle.

## 11. Classer et prioriser les parcelles

Pour chaque parcelle communale, le pourcentage de pixel appartenant à chaque catégorie (de 1 à 6) a été calculé pour les modèles été et hiver. Ensuite, les catégories de 1 à 3 ont été classées en bon habitat, et le pourcentage a été calculé sur chaque parcelle pour l'été et l'hiver (P\_GT, P\_Coq, P\_Poule).

En utilisant les modèle été uniquement, une note bonne, moyenne ou mauvaise a ensuite été attribuée à chaque parcelle (NOTE). Pour cela, le pourcentage de bon habitat (P\_GT) a été extrait pour chacune des observations de Grand Tétras de la base de données. La moyenne et la déviation standard ont été calculées, donnant ainsi les limites de nos catégories : limite minimal = Moyenne - déviation standard = 48% (été) ; 45% (hiver), Moyenne=66% (été) ; 68% (hiver) et limite maximal = Moyenne + déviation standard = 83% (été) ; 92% (hiver). Pour une commune, il est possible de visualiser où se situent ces parcelles par rapport au reste des parcelles forestières (Figure 24).

Figure 24 : Qualité d'habitat pour le Grand Tétrás en été. En rouge les parcelles de la commune étudiée avec leur numéro et en bleu les parcelles de la zone d'étude.





On voit par exemple pour la forêt communale de Giron que de nombreuses parcelles contiennent peu d'habitat favorable, mais qu'il existe aussi des parcelles très favorables (42 et 46).

Deux pages en ligne permettent de choisir la commune désirée et ainsi obtenir le graphique:

[https://gladtj.shinyapps.io/Parcelles\\_Jura/](https://gladtj.shinyapps.io/Parcelles_Jura/)

[https://gladtj.shinyapps.io/Parcelles\\_Ain/](https://gladtj.shinyapps.io/Parcelles_Ain/)

Enfin une estimation des parcelles qui peuvent faire l'objet de travaux a été faite en utilisant deux méthodes.

#### v. Méthode de distance simple

Pour chaque parcelle, le nombre de parcelles adjacentes de bonne et moyenne a été calculés.

Pour les parcelles en bordure, une note pour une surface tampon de 100m a été calculée. Si une parcelle de note mauvaise est adjacente à une bonne parcelle ou un bon buffer elle est classée comme priorité 1. Les parcelles entourées de mauvais habitat sont classées en priorité 2. Enfin les parcelles avec une note bonne ou moyenne sont classées en priorité 0.

#### vi. Méthode utilisant la surface terrière (Ain seulement)

Ici, la surface terrière moyenne et totale a été calculée. Une surface terrière possible après une diminution de 1/3 de la surface totale a ensuite été calculée le but étant d'atteindre une surface terrière moyenne située entre 18m<sup>2</sup>/ha et 27m<sup>2</sup>/ha.

- Priorité 0 : Habitat bon ou moyen
- Priorité 1 : Parcelle proche d'un bon habitat (parcelle ou buffer) ayant déjà une surface terrière moyenne située entre 18m<sup>2</sup>/ha et 27m<sup>2</sup>/ha
- Priorité 2 : Parcelle proche d'un bon habitat (parcelle ou buffer) ayant une surface terrière moyenne possible après coupe inférieure à 27m<sup>2</sup>/ha

- Priorité 3 : Parcelle loin d'un bon habitat (parcelle ou buffer) ayant déjà une surface terrière moyenne située entre 18m<sup>2</sup>/ha et 27m<sup>2</sup>/ha ou une surface terrière moyenne possible après coupe située inférieure à 27m<sup>2</sup>/ha
- Priorité 4 : Parcelle loin d'un bon habitat (parcelle ou buffer) n'ayant pas une surface terrière moyenne située entre 18m<sup>2</sup>/ha et 27m<sup>2</sup>/ha ou ni une surface terrière moyenne possible après coupe située inférieure à 27m<sup>2</sup>/ha.

Les parcelles ayant une valeur de surface terrière inférieure à 18m<sup>2</sup>/ha sont notés NA.

Les résultats se présentent sous la forme d'un fichier shapefile avec la légende ci-dessous :

### ETE

Légende	Correspondance
IIDNT_F	Code commune
LLIB_FR	Nom commune
CCOD_PR	Numéro parcelle
ID	ID unique des parcelles
GT	Valeurs exponentielle moyenne
Surfacepar	Surface parcelle en m <sup>2</sup>
P_GT	Pourcentage de bon habitat été
NOTE	Note parcelle
N_voisins	Nombre de parcelles voisine
N_bon	Nombre de parcelles voisine de bonne qualité
P_buffer	Pourcentage de bon habitat pour le buffer de 100m
notebuffer	Note du buffer
Priorite	Priorité selon la distance
Prob_hetre	Probabilité moyenne de présence de régénération (Ain)
ST_tot	Surface terrière totale
ST_moy	Surface terrière moyenne
St_poss	Surface terrière possible après coup de 1/3
PrioriteST	Priorité selon la distance et la surface terrière
Couv_hetre	Pourcentage moyen de la couverture de régénération (Jura)

## HIVER

Légende	Correspondance
IIDNT_F	Code commune
LLIB_FR	Nom commune
CCOD_PR	Numéro parcelle
ID	ID unique des parcelles
Coq	Valeurs exponentielle moyenne pour le Coq
Poule	Valeurs exponentielle moyenne poule la Poule
Surfacepar	Surface parcelle en m <sup>2</sup>
P_Coq	Pourcentage de bon habitat hiver pour le Coq
NOTE_C	Note parcelle pour le Coq
N_voisins_C	Nombre de parcelles voisine pour le Coq
N_bon_C	Nombre de parcelles voisine de bonne qualité pour le Coq
P_buffer_C	Pourcentage de bon habitat pour le buffer de 100m pour le Coq
Notebuf_C	Note du buffer pour le Coq
Priorite_C	Priorité selon la distance pour le Coq
P_Poule	Pourcentage de bon habitat hiver pour la Poule
NOTE_P	Note parcelle pour la Poule
N_voisins_P	Nombre de parcelles voisine pour la Poule
N_bon_P	Nombre de parcelles voisine de bonne qualité la Poule
P_buffer_P	Pourcentage de bon habitat pour le buffer de 100m la Poule
Notebuf_P	Note du buffer pour la Poule
Priorite_P	Priorité selon la distance pour la Poule
Prob_hetre	Probabilité moyenne de présence de régénération (Ain)
ST_tot	Surface terrière totale
ST_moy	Surface terrière moyenne
St_poss	Surface terrière possible après coup de 1/3
PrioriteST	Priorité selon la distance et la surface terrière
Couv_hetre	Pourcentage moyen de la couverture de régénération (Jura)

## VI.DISCUSION

Les modèles de distribution pour le Grand Tétras créés à l'aide de données LiDAR comme variables environnementales ont permis de produire des cartes à plusieurs échelles. Les cartes produites représentent la distribution potentielle de l'espèce, où l'habitat est favorable en suivant les variables prises en compte dans chaque modèle. De nombreux autres processus tels que la capacité de dispersion, la dynamique des populations ou la compétition avec d'autres espèces peuvent

restreindre dans la réalité la présence de population même dans des habitats de structure et composition favorable. Ces effets sont encore très peu pris en compte dans la réalisation de modèles de distribution d'espèces (Guisan et Thuiller 2005). Les plus petites échelles donnent des résultats détaillés sur la qualité relative de l'habitat, il est cependant probable que la précision de ces modèles soit moins bonne que ceux à plus large échelle due à une diversité potentiellement plus importante localement. L'étude des variables contribuant à la valeur de prédiction montre d'ailleurs que des prédictions fortes peuvent être dues à l'association positive de différentes variables, décrivant donc des environnements différents. Les performances des modèles aux différentes échelles ont montré que les modèles mixtes sont, dans le cas du Grand Tétrás, plus performants. Cependant, les cartes des modèles mixtes présentent entre elles de larges différences, car l'aspect de la carte dépend en grande partie des échelles sélectionnées pour construire le modèle. Les modèles d'été ont été créés avec peu de données par rapport aux modèles hiver. De plus, l'effort d'échantillonnage n'était pas connu, ce qui peut être à l'origine des difficultés observées dans la zone du Jura. La collection de plus de données en été pourra permettre d'améliorer les modèles. Une autre limitation de ces modèles est la date de survol LiDAR, qui est déjà suffisamment ancienne pour la zone de l'Ain (2014) pour que des éléments de végétation aient changés (chablis, coupes, croissance et régénération de la végétation).

La déclinaison des cartes en 6 catégories permet une interprétation plus facile des prédictions et ainsi de cibler de futures zones de gestion. Les cartes non-catégorisées peuvent être utilisées pour avoir une vision plus fine de la qualité de l'habitat dans une zone précise. Concernant l'interprétation des courbes de réponses, la limitation principale provient du fait que deux types de réponses peuvent être modélisées, linéaire et quadratique. Ceci a pour effet de rendre certaines interprétations non compatibles avec nos connaissances de l'écologie de l'espèce, par exemple, des zones de densité d'arbres de 0 données favorables. Cependant, les prédictions finales proviennent des variations combinées des variables, ainsi la présence d'une variable pourcentage de milieu ouvert permet dans certains cas de compenser des estimations mauvaises aux valeurs extrêmes des autres variables. Le cas des modèles d'été dans la zone du Jura montre aussi que de trop nombreuses

variables peuvent mener à un modèle peu réaliste quand on le compare à nos connaissances du terrain.

La sélection initiale des variables ne couvre pas toutes les variations présentes dans l'habitat. En particulier le LiDAR ne donne que peu d'information sur les espèces végétales présentes dans le milieu. Ceci peut être amélioré avec l'ajout de données satellites qui, associées au LiDAR, peuvent permettre une meilleure caractérisation de la structure et la composition des zones forestières (Bae et al. 2014). Une autre limitation concerne en particulier la détection du sous-étage qu'il est difficile de détecter sous la canopée et qui dépend beaucoup des caractéristiques de la campagne LiDAR (densité des points) ou de la saison (feuillus avec ou sans feuilles). Ceci pose un problème en particulier pour le Grand Tétrás dont l'habitat se réduit dans le Jura avec une forte régénération des semis de hêtre. Ainsi, l'apport d'un modèle spécifique concernant le hêtre peut apporter de nouvelles informations sur la qualité de l'habitat.

## VII. BIBLIOGRAPHIE

---

Bae, Soyeon, Bjoern Reineking, Michael Ewald, et Joerg Mueller. 2014. « Comparison of airborne lidar, aerial photography, and field surveys to model the habitat suitability of a cryptic forest species—the hazel grouse ». *International Journal of Remote Sensing* 35 (17): 6469–89.

Bellon, B. 2011. « Évolution de la répartition communale du petit gibier de montagne en France Décennie 2000-2009 ». *Supplément Faune sauvage*, n° 290.

Bollmann, Kurt, Patrick Weibel, et Roland F Graf. 2005. « An analysis of central Alpine capercaillie spring habitat at the forest stand scale ». *Forest Ecology and Management* 215 (1): 307–18.

El-Gabbas, Ahmed, et Carsten F Dormann. s. d. « Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent ». *Ecography*.

Elith, Jane, Steven J Phillips, Trevor Hastie, Miroslav Dudík, Yung En Chee, et Colin J Yates. 2011. « A statistical explanation of MaxEnt for ecologists ». *Diversity and Distributions* 17 (1): 43–57.

Eysn, Lothar, Markus Hollaus, Eva Lindberg, Frédéric Berger, Jean-Matthieu Monnet, Michele Dalponte, Milan Kobal, Marco Pellegrini, Emanuele Lingua, et Domen Mongus. 2015. « A Benchmark of Lidar-Based Single Tree Detection Methods Using Heterogeneous Forest Data from the Alpine Space ». *Forests* 6 (5): 1721–47.

Franklin, Janet. 2009. « Moving beyond static species distribution models in support of conservation biogeography ». *Diversity and Distributions* 16 (3): 321–30.

Graf, Roland F, Kurt Bollmann, Harald Bugmann, et Werner Suter. 2007. « Forest and landscape structure as predictors of capercaillie occurrence ». *Journal of Wildlife Management* 71 (2): 356–65.

Graf, Roland F, Lukas Mathys, et Kurt Bollmann. 2009. « Habitat assessment for forest dwelling species using LiDAR remote sensing: Capercaillie in the Alps ». *Forest Ecology and Management* 257 (1): 160–67.

Guisan, Antoine, et Wilfried Thuiller. 2005. « Predicting species distribution: offering more than simple habitat models ». *Ecology letters* 8 (9): 993–1009.

Hofstetter, Lea, Raphaël Arlettaz, Kurt Bollmann, et Veronika Braunisch. 2015. « Interchangeable sets of complementary habitat variables allow for flexible, site-adapted wildlife habitat management in forest ecosystems ». *Basic and Applied Ecology* 16 (5): 420–33.

Kramer-Schadt, Stephanie, Jürgen Niedballa, John D Pilgrim, Boris Schröder, Jana Lindenborn, Vanessa Reinfelder, Milena Stillfried, Ilja Heckmann, Anne K Scharf, et Dave M Augeri. 2013. « The importance of correcting for sampling bias in MaxEnt species distribution models ». *Diversity and Distributions* 19 (11): 1366–79.

McGarigal, Kevin, Ho Yi Wan, Kathy A Zeller, Brad C Timm, et Samuel A Cushman. 2016. « Multi-scale habitat selection modeling: a review and outlook ». *Landscape ecology* 31 (6): 1161-75.

Phillips, Steven J., Robert P. Anderson, Miroslav Dudík, Robert E. Schapire, et Mary E. Blair. 2017. « Opening the black box: an open-source release of Maxent ». *Ecography*, mai, n/a-n/a. <https://doi.org/10.1111/ecog.03049>.

Rechsteiner, Christian, Florian Zellweger, Anatole Gerber, Frank T. Breiner, et Kurt Bollmann. 2017. « Remotely sensed forest habitat structures improve regional species conservation ». *Remote Sensing in Ecology and Conservation*, n/a-n/a. <https://doi.org/10.1002/rse2.46>.

Roussel, Jean-Romain, et David Auty. 2016. *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications. R package version 1.0.0*. <https://github.com/Jean-Romain/lidR>.

Storch, Ilse. 1993. « Habitat selection by capercaillie in summer and autumn: Is bilberry important? » *Oecologia* 95 (2): 257-65.

Thiel, Dominik, Susanne Jenni-Eiermann, Veronika Braunisch, Rupert Palme, et Lukas Jenni. 2008. « Ski tourism affects habitat use and evokes a physiological stress response in capercaillie *Tetrao urogallus*: a new methodological approach ». *Journal of applied ecology* 45 (3): 845-53.

Thiel, Dominik, Emmanuel Ménoni, JEAN-FRANÇOIS BRENOT, et Lukas Jenni. 2007. « Effects of recreation and hunting on flushing distance of capercaillie ». *The Journal of Wildlife Management* 71 (6): 1784-92.

## ANNEX IX. FILES PRODUCED

- LiDAR metrics: All extracted LiDAR variables for both study area at each scales as raster files
- Hazel Grouse predicted maps: Predictions for all models and at each scale as raster files
- Capercaillie predicted maps: Predictions for all models and at each scale as raster files
- Beech regeneration predicted maps: Predictions from the best models in the Jura study area as raster files
- Communal plots: communal plots classification regarding Capercaillie habitat suitability as shapefiles
- Interactive graphics by communal plots:
  - [https://gladgtj.shinyapps.io/Parcelles\\_Ain/](https://gladgtj.shinyapps.io/Parcelles_Ain/)
  - [https://gladgtj.shinyapps.io/Parcelles\\_Jura/](https://gladgtj.shinyapps.io/Parcelles_Jura/)

## ANNEX X. PRÉSENTATIONS

- Presentation of a poster at the IUFRO conference Landscape Ecology Conference September 24-29 2017: “Remote Sensing contribution towards the understanding of ecosystem functioning and ecosystem accounting group”, Halle, Germany (September 2017)
- The work was presented at the *Séminaire Plan Départemental Tétrionidés* on January 19<sup>th</sup>, 2018.
- Upcoming workshop “LiDAR et Tétrionidés” on October 16<sup>th</sup>, 2018, Relais Nordique de Giron.

## ANNEX XI. OUTREACH

Article published in the Groupe Tétrás Jura newsletter on the use of LiDAR for the Capercaillie and Hazel Grouse conservation (“*Le LiDAR au secours du Grand-Tétrás et de la Gélinoche des bois !*”, issue 36, June 2016).





## Abstract

In the general context of biodiversity erosion, two forest bird species occurring in the French Jura massif, the Capercaillie (*Tetrao urogallus*) and the Hazel Grouse (*Bonasa bonasia*), are threatened by habitat loss and fragmentation at the regional scale. In particular, intensive beech regeneration patches extension in the Jura massif is leading to the transformation of the understory cover, once suitable with bilberry and herbaceous vegetation, to closed unfavorable habitat. The fate of those two emblematic species is depending for the first on future management actions and for the second on a better knowledge of the species population's dynamics and occurrences. In particular, the cutting of the beech regeneration patches is one of the efficient management actions undertaken to restore the habitat. However, management actions and surveys are money and time consuming due to the large area that need to be covered. The opportunity to use two Light Detection and Ranging (LiDAR) datasets covering a major part of the distribution of the two species in the Jura massif initiated the PhD project, with the objective to support managers in their decisions and actions by the creation of adapted distribution predicted maps using Species Distribution Models (SDMs) (Hazel Grouse, Capercaillie and beech regeneration). The realization of this objective is depending on the reliability of the models produced and on the capacity of the researcher to transfer the results to managers who are not familiar with modeling methods. In a first step, the choice of the appropriate modeling method regarding the datasets characteristics and the objectives was investigated (sampling bias correction, scales, and algorithms). In addition, the use of object-oriented LiDAR predictors (trees and gaps) pertinent from both species and managers point of view to facilitate the results transfer was tested. The results obtained were used to create appropriate multi-scale SDMs and to predict distribution maps for both target species, demonstrating the capacity of LiDAR to represent vegetation structures that influence the targeted forest bird species occurrences. Models at a fine scale were also created to map the beech regeneration distribution in the Jura massif.

**Keywords:** Species Distribution Models, LiDAR, Capercaillie, Hazel Grouse, conservation, management.



## Résumé

Dans le contexte général de l'érosion de la biodiversité, deux espèces d'oiseaux forestiers, le Grand Tétras (*Tetrao urogallus*) et la Gélinotte des bois (*Bonasa bonasia*), présentes dans le massif Jurassien sont menacées par la perte et la fragmentation de leur habitat à l'échelle régionale. En particulier, dans le massif Jurassien l'extension progressive des tâches de régénération du hêtre induit la transformation du couvert végétal constitué de myrtilles et d'herbacées favorable en un habitat fermé défavorable. Le destin de ces deux espèces emblématiques dépend pour la première d'actions de gestions et pour la seconde d'une meilleure connaissance de la distribution et de la dynamique des populations. La coupe des zones de régénération fait partie des principales actions envisagées pour restaurer l'habitat forestier. Cependant ces actions de gestion ou de suivi des populations sont couteuses en temps et en argent. Ainsi, l'opportunité d'utiliser deux jeux de données LiDAR (Light Detection and Ranging) couvrant la majorité de l'aire de distribution des deux espèces dans le massif Jurassien a initié le projet de cartographie des habitats de chaque espèce et de la présence des tâches de régénération du hêtre en utilisant des modèles de distribution d'espèces (SDMs). L'objectif est de soutenir les gestionnaires dans leurs décisions et actions grâce à la production de prédictions spatiales adaptées. La réalisation de cet objectif dépend de la fiabilité des modèles produits, mais aussi de la bonne transmission des résultats par le chercheur aux gestionnaires qui ne sont pas familiers avec les méthodes utilisées. Dans un premier temps, le choix d'une méthode de modélisation appropriée (correction du biais d'échantillonnage, échelles, algorithmes) par rapport aux caractéristiques des jeux de données et aux objectifs a été évalué. Dans un second temps, l'utilisation de variables environnementales LiDAR orienté-objet (arbres et trouées) pour faciliter l'appropriation des résultats par les gestionnaires a été testée. Enfin, les résultats obtenus ont permis la création de modèles multi-échelles et de carte de prédictions pour chacune des espèces démontrant la capacité du LiDAR de représenter la structure de la végétation qui influence la présence des espèces d'oiseaux forestières étudiées. Des modèles de distribution de la régénération du hêtre ont pu aussi être créés à une échelle fine.

**Mots-clefs :** Modèles de Distribution d'espèces, LiDAR, Grand Tétras, Gélinotte des bois, conservation, gestion.