



HAL
open science

Modélisation statistique de la distribution des grands carnivores en Europe

Julie Louvrier

► **To cite this version:**

Julie Louvrier. Modélisation statistique de la distribution des grands carnivores en Europe. Biodiversité et Ecologie. Université Montpellier, 2018. Français. NNT : 2018MONTG036 . tel-01954659

HAL Id: tel-01954659

<https://theses.hal.science/tel-01954659>

Submitted on 13 Dec 2018

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 L'UNIVERSITÉ DE MONTPELLIER

En Écologie et Biodiversité

École doctorale GAIA

Unité de recherche Centre d'Écologie Fonctionnelle et Évolutive

Modélisation statistique de la distribution des grands carnivores en Europe

Présentée par Julie LOUVRIER
Le 27 novembre 2018

Sous la direction de Olivier GIMENEZ

Devant le jury composé de

Stephanie KRAMER-SCHADT, directrice de recherche, Leibniz IZW

Wilfried THUILLER, directeur de recherche, LECA

Christophe DUCHAMP, chargé de recherche, ONCFS

Gurutzeta GUILLERA-ARROITA, Maitresse de conférences, University of Melbourne

Marc KERY, directeur de recherche, Vogelwarte

Anne CHARMANTIER, directrice de recherche, CEFE

Nolwenn DROUET-HOGUET, chargée de recherche, ONCFS

Olivier GIMENEZ, directeur de recherche, CEFE

Rapportrice

Rapporteur

Invité

Examinatrice

Examineur

Examinatrice

Examinatrice

Directeur



UNIVERSITÉ
DE MONTPELLIER

Remerciements

Trois ans de thèse, avec 6 mois de stage de M2, c'est une petite portion de vie mais qui fut tellement bien remplie par des rencontres autant professionnelles que personnelles que toutes ensemble elles m'ont fait grandir et mûrir. Je suis très reconnaissante de tous ces petits gestes et grands gestes qui, m'ont procuré la sensation d'avoir été toujours soutenue et aidée, jamais rabaissée, jamais seule. Cette thèse n'aurait pas été ce qu'elle est sans toutes ces personnes et je souhaitais les remercier chaleureusement.

Ma thèse s'est réalisée grâce à un co-financement de l'ONCFS, et plus particulièrement j'ai eu la grande chance d'avoir été guidée par Christophe Duchamp, Murielle Guinot-Ghestem et Eric Marboutin (et son schnaps) qui fut ensuite relayé par Nolwenn Drouet-Hoguet. **Merci** à vous pour votre aide, j'ai passé de super réunions techniques du CNERA-PAD avec un accueil chaleureux, qui m'ont aussi permis de remettre mes pieds sur terre, dans la réalité du terrain ainsi que des problématiques de l'ONCFS.

De plus, le travail de ma thèse a été basé sur de nombreuses collaborations pour chacun des chapitres. Je tenais donc particulièrement à remercier ces personnes avec qui j'ai travaillé, pour leur gentillesse, leur compréhension, et leur soutien. En particulier, je tenais à **remercier** Ian Renner pour son extrême gentillesse et son aide pour le projet sur le lynx en Norvège. Mais plus particulièrement, **merci** pour ces soirées mémorables ensemble, pour les karaokés enflammés à 4 heures du matin à Barcelone, et **merci** pour le soutien lorsque j'avais des questionnements. Enfin, **merci** pour toutes les références musicales que tu m'as envoyées, cette musique a été un énorme support pour moi vers la fin de ma thèse.

Un **merci** tout particulier à Julien Papaïx, pour ta réactivité et ton aide. Le projet mécastat n'aurait pas eu lieu sans ta précieuse aide et ta gentillesse.

Merci à Thierry Chambert, David Miller, Marc Kéry, Anja Molinari-jobin, John Linnell, Erlend Nielsen, Jenny Mattisson, Diana Bowler et Rémi Choquet pour m'avoir apporté leurs précieuses expertises afin de rendre les projets tels qu'ils sont aujourd'hui.

Un **merci** tout particulier à Stephanie Kramer-Schadt et Wilfried Thuiller d'avoir accepté d'être les examinateurs de ce manuscrit. **Merci** à Marc Kéry, Gurutzeta Guillera-Arroita, Nolwenn Drouet-Hoguet, Anne Charmantier et Christophe Duchamp d'avoir accepté de participer au jury.

Merci à ma petite Mélanie d'avoir apporté le soutien dont j'avais besoin et d'avoir toujours cru en moi. Tu es mon exemple de persévérance et de travail ! J'espère venir mouflonner très prochainement avec toi !

Un **merci** à toutes les personnes au CEFE, permanents, techniciens, administration, thésards, stagiaires, césuriens, postdocs, pour rendre ce lieu aussi agréable. **Merci** à PAC et Fred, Charlotte et Julien, **merci** d'être vous, le CEFE serait un tout autre endroit sans vous, croyez-moi. **Merci** à Gaia pour cette belle rencontre. **Merci** à Simon Benhamou de me rappeler de ne pas travailler trop tard (BONJOUUUUR). **Merci** à Florence de m'avoir tenue compagnie vers la fin lorsque j'étais dans ma période dark emo.

Merci à Hélène Terzian, pour m'avoir fait confiance et aidée lors des cours à l'université de Nîmes. Et **merci** aux étudiants que j'ai eu en cours pour m'avoir fait prendre conscience du plaisir d'enseigner.

Merci aux collègues et coaches de « ma thèse en 180 secondes », on s'est bien marrés. **Merci** à Manon et Marjo, mes collègues dans l'adversité. Un **merci** particulier à Iago pour ton aide précieuse lors de mon stage de M2. **Merci** à Coline et Justine pour avoir partagé le bureau avec moi et s'être serrées les coudes pour le stage de master.

Merci à mes grimpeurs favoris, André, Jérôme, Hugo pour m'aider à évacuer la pression en m'arrachant les doigts sur des falaises ou des murs de blocs. Et à ma petite Jess' pour les papotages ensemble.

Merci à Rowan, pour toutes nos soirées sur le Lez, pour l'amour des chats et pour nos voyages (oui oui j'ai mis au pluriel), pour les concerts, la culture et d'avoir été présente lorsque j'avais besoin de me vider la tête. Et **merci** pour la relecture de la thèse ;-).

Enfin, **merci** à la troupe d'artistes que j'ai eu la chance de rencontrer sur mon chemin. Merci à toi Arnaud, la vie en colloc avec toi restera un souvenir impérissable ! **Merci** pour toutes ces soirées, ces fou-rires, les soirées musicales, et pour tout l'amour qu'il y avait dans cet appartement. **Merci** à toi ma biche Aude, tu es chérie dans mon cœur, merci à Bastien, Vinc' Aviva et tout le reste de la bande, vous m'êtes précieux, **merci** de m'avoir fait voir autre chose dans la vie, et de m'avoir fait me sentir intégrée dans votre famille.

Merci à ma petite Lisae pour nos discussions et confessions, ainsi que les piercings ;) et **merci** à Hugo, mon coiffeur préféré pour ta gentillesse.

Merci à Agathe pour nos après-midi thé et d'être venue me rendre visite en Norvège afin de partager les bières sous les aurores boréales et se faire des frayeurs dans les chemins de rando « easy ».

Merci à Magali, mon amie d'enfance, celle qui me connaît par cœur et sait décoder toutes mes pensées avant même que je sois capable de les formuler. **Merci** d'être toujours là, depuis le début.

Merci à mes deux amours, Dragan et Eva, de faire retomber la pression, ensemble nous formons le citron vert, le sucre et la cachaça qui composent une Caipirinha. Vous-mêmes vous savez comment je vous aime.

Un **merci** très spécial à mes parents. **Merci** pour tout ce que vous avez fait et ce que vous continuez de faire, j'ai beaucoup de chance d'avoir une précieuse famille comme vous. **Merci** en particulier de me faire toujours cultiver la curiosité et de m'avoir fait aimer la connaissance. **Merci** d'avoir toujours été prêts à m'accueillir afin de m'aider et me soutenir et de prendre soin de moi. **Merci** à mon grand Aurélien, pour me montrer l'exemple du travail persévérant et **merci** à mon ptit Titouan pour la musique et les blagounettes nullardes que seuls nous pouvons comprendre. Une petite dédicace à mon neveu Aurèle, tu ne dois pas trop encore comprendre ce qu'est une thèse mais je suis heureuse et honorée d'être ta tata. **Merci** à mes tatas, Annie et Martine pour la culture que vous m'avez apportée. Enfin **merci** à Daniel, Cathy, Etienne et Simon d'être présents depuis que je suis toute petite.

Enfin, **merci** à toi, Borja, il me faudrait écrire un deuxième manuscrit afin de lister toutes les choses pour lesquelles je te suis reconnaissante. Je suis une amoureuse chanceuse et le projet dont je suis le plus fière au court de ces trois années, c'est notre relation. We're a team.

Merci à toutes les autres personnes que j'ai oubliées ☺

En bref, **merci** à tout ce petit monde, sauf Gilles Maurer (**pouet**).

Résumé

Les grands carnivores recolonisent l'Europe grâce à une augmentation de la superficie forestière et des populations d'ongulés sauvages ainsi que des mesures de conservation. Ce retour pose toutefois des défis, les carnivores entrant en interactions avec les activités humaines telles que l'élevage. Quantifier leur distribution peut aider à situer et prédire les impacts sur ces activités. Ces espèces sont toutefois très mobiles, difficiles à observer et vivent à de très faibles densités. Par conséquent, la modélisation de leur distribution présente plusieurs défis en raison 1) de leur détectabilité imparfaite, 2) de leur distribution dynamique dans le temps et 3) du suivi à grande échelle basé sur la collecte de données opportunistes sans mesure formelle de l'effort d'échantillonnage. Dans cette thèse, nous nous sommes concentrés sur deux espèces de grands carnivores, le loup (*Canis lupus*) et le lynx boréal (*Lynx lynx*), pour évaluer et développer les méthodologies liées à la modélisation de la distribution d'espèces (SDMs). Plus particulièrement, nous avons exploré l'application des modèles d'occupancy dans le contexte du suivi des grands carnivores en Europe. Ces modèles permettent d'établir le lien entre la présence d'une espèce et l'environnement dans le but d'établir la proportion d'une zone d'étude que l'espèce occupe, tout en prenant en compte une détectabilité imparfaite.

Plus précisément, nous avons d'abord évalué la dynamique de la distribution des loups en France de 1994 à 2016, tout en prenant en compte leur détection imparfaite. Nous avons montré l'importance de prendre en compte l'effort d'échantillonnage variant dans le temps et dans l'espace à l'aide de modèles d'occupancy dynamique.

Deuxièmement, comme des faux positifs peuvent être présents lors de la surveillance d'espèces rares, nous avons développé un modèle dynamique d'occupancy qui tenait compte simultanément des faux négatifs et des faux positifs pour analyser conjointement des données qui contenaient à la fois des détections certaines et des détections incertaines. L'analyse des données sur le lynx boréal dans les pays alpins a suggéré que l'incorporation de détections incertaines produisait des estimations des paramètres écologiques plus précises.

Troisièmement, nous avons développé un modèle qui prenait en compte l'hétérogénéité de la détection tout en traitant les faux positifs. En appliquant notre nouvelle approche au loup en France, nous avons démontré que l'hétérogénéité de la détection du loup était principalement due à un effort d'échantillonnage hétérogène dans l'espace.

Quatrièmement, pour traiter des sources de données multiples, nous avons développé un modèle de processus ponctuel de Poisson qui permettait l'inclusion de différentes sources de données lors de la construction des SDMs. Nous avons montré comment la combinaison des données sur la distribution permettait d'optimiser un suivi en répondant à la question de savoir quelle(s) source(s) d'information apporterait l'essentiel de l'information lors du suivi du lynx en Norvège.

Cinquièmement, pour comprendre les mécanismes sous-jacents de la colonisation des loups en France, nous avons développé un cadre statistique pour estimer l'occupation spatio-temporelle et la dynamique des effectifs en utilisant le cadre de diffusion écologique. Nous avons montré le potentiel de notre approche pour prédire la distribution future potentielle du loup à court terme, un élément qui pourrait contribuer à cibler des zones de gestion ou se concentrer sur des zones de conflit potentiel.

Dans l'ensemble, nos travaux montrent que les données opportunistes peuvent être analysées à l'aide de modèles de distribution d'espèces qui prennent en compte les contraintes liées au type de suivi utilisé pour produire les données. Nos approches peuvent être utilisées

par les gestionnaires pour optimiser la surveillance des grands carnivores, cibler des zones de présence potentielles et contribuer à proposer des mesures destinées à atténuer les conflits.

Mots-clés : Canis lupus, détection imparfaite, distribution d'espèce, dynamique des populations, écologie statistique, grands carnivores, hétérogénéité de détection, Lynx lynx, modèles d'occupancy, sciences participatives

Abstract

Large carnivores are recovering in Europe, as a consequence of increasing forest cover, ungulate populations and conservation measures. However, this return poses challenges as carnivores can interact with human activities such as livestock farming. Assessing their distributions can help to predict and mitigate conflicts with human activities. Because large carnivores are highly mobile, elusive and live at very low density, modeling their distributions presents several challenges due to 1) their imperfect detectability, 2) their dynamic ranges over time and 3) their monitoring at large scales consisting of opportunistic data without a formal measure of the sampling effort. In this thesis, we focused on two carnivore species, wolves (*Canis lupus*) and Eurasian lynx (*Lynx lynx*), to explore and develop the methodological aspects related to the modelling of species distributions. We considered the application of occupancy models in the context of monitoring large carnivores in Europe. These models allow the establishment of a link between the species' presence and environmental covariates while accounting for imperfect detectability, in order to establish the proportion of a study area occupied by the species.

Specifically, we first assessed wolf range dynamics in France from 1994 to 2016, while accounting for species imperfect detection and showed the importance of accounting for time- and space-varying sampling effort using dynamic site-occupancy models.

Second, acknowledging that false positives may occur when monitoring rare species, we showcased a dynamic occupancy model that simultaneously accounts for false negatives and positives to jointly analyze data that include both unambiguous detections and ambiguous detections. The analysis of data on the Eurasian lynx in Alpine countries suggested that incorporating ambiguous detections produced more precise estimates of the ecological parameters.

Third, we developed a model accounting for heterogeneity in detection while dealing with false positives. Applying our new approach to a case study with grey wolves in France, we demonstrated that heterogeneity in wolf detection was mainly due to a heterogeneous sampling effort across space.

Fourth, to deal with multiple data sources, we developed a Poisson point process approach which allows the inclusion of different data sources when building SDMs. By doing so, we also answered the question about which source(s) of information would provide most of the information when monitoring the lynx in Norway.

Fifth and finally, to understand the underlying mechanisms of the colonization of wolves in France, we developed a statistical framework for estimating spatiotemporal occupancy and abundance dynamics using the ecological diffusion framework. We demonstrated the potential of our approach to predict the potential future distribution of wolves in the short term, an element that could contribute to target management areas or focus on areas of potential conflict.

Overall our work shows that opportunistic data can be analyzed with species distribution models that control for issues linked to the type of monitoring used to produce the data. Our approaches have the potential for being used by decision-makers to optimize the monitoring of large carnivores and to target sites where carnivores are likely to occur and mitigate conflicts.

Keywords : Canis lupus, citizen sciences, detection heterogeneity, imperfect detection, large carnivores, , Lynx lynx, , occupancy models, population dynamic, species distribution, statistical ecology

1.	Contexte général	13
1.a.	Grands carnivores en Europe : pourquoi modéliser leur répartition spatiale ?	13
Encadré No 1 : La situation du loup en France et du lynx dans les pays alpins et en Norvège15		
1. b.	Quels outils utiliser pour modéliser la répartition spatiale des grands carnivores ? ..	18
1.c.	Sciences citoyennes et données opportunistes	19
2.	Les cas du loup en France et du lynx dans les pays alpins et en Norvège	20
2.a.	Cas du Loup en France	20
2.a.1.	Contexte du suivi	20
2.a.2.	Les verrous méthodologiques	22
2.a.3.	Les données	22
2.b.	Cas du lynx dans les pays alpins	28
2.b.1.	Contexte du suivi	28
2.b.2.	Les verrous méthodologiques	29
2.b.3.	Les données	29
2.c.	Cas du lynx en Norvège	34
2.c.1.	Contexte du suivi	34
2.c.2.	Les verrous méthodologiques	36
2.c.3.	Les données	36
3.	Les modèles d’occupancy pour quantifier la distribution d’une espèce	41
3.a.	Le modèle d’occupancy statique	41
3.b.	Le modèle d’occupancy dynamique	43
3.c.	Questionnements sur les hypothèses des modèles d’occupancy et leur application aux grands carnivores	44
4.	Organisation de la thèse.....	45
4.a.	Non-respect des hypothèses des modèles d’occupancy	45
4.b.	Gestion et suivi des populations de grands carnivores.....	46
5.	Article 1. Cartographier et expliquer la recolonisation du loup en France à l’aide de modèles d’occupancy dynamique et de données opportunistes	47
6.	Article 2. Utilisation de détections ambiguës pour améliorer les estimations des modèles de distribution d’espèces	61
Encadré No 2 : Intégrer les données incertaines pour modéliser la distribution du loup en France		
		73
		73

7. Article 3. Prise en compte des erreurs d'identification et de l'hétérogénéité dans les études d'occupancy en utilisant des modèles de Markov cachés	77
8. Article 4. Modélisation de la distribution du lynx (<i>Lynx lynx</i>) en Norvège en combinant des données d'occupancy et des données de presence-background	87
9. Article 5. Un modèle mécanistico-statistique de distribution d'espèces pour expliquer et prévoir la colonisation du loup (<i>Canis lupus</i>) dans le sud-est de la France.	121
10. Discussion	151
10.a. Synthèse de mon travail et intérêt.....	151
10.b. Interprétation et utilisation des résultats	152
10.b.1. Interprétation de l'occupancy.....	152
10.b.2. Interprétation des variables écologiques.....	154
10.b.3. Utilisation des résultats en termes de suivi.....	155
10.c. Vers une gestion adaptative des grands carnivores, sur la base de l'occupancy.....	157
Annexes	160
Annexe 1 : Article vulgarisé sur l'article 1, diffusé sur le site du CNRS.....	160
Annexes Article 1.....	163
Annexes Article 2.....	174
Annexes Article 3.....	180
Annexes Article 4.....	193
Annexes Article 5.....	200
Références.....	208
Table des figures.....	215

1. Contexte général

1.a. Grands carnivores en Europe : pourquoi modéliser leur répartition spatiale ?

Les grands carnivores sont souvent considérés comme des éléments clés du maintien des écosystèmes. L'étude de l'écologie des grands carnivores apparaît dès les premières études de la conservation de la nature avec Aldo Leopold qui faisait déjà référence à l'impact du déclin des prédateurs sur la croissance des populations d'herbivores et de plantes (Leopold et al. 1947), décrivant un processus qui fut intégré au concept de « cascade trophique » (Ripple & Beschta 2005; Layman et al. 2015; Ford 2017). En effet, en raison de leur position élevée dans la chaîne trophique, leur extinction peut entraîner des cascades trophiques et des changements dans la dynamique des populations et le fonctionnement des écosystèmes (Ripple et al. 2014). Au cours de la dernière décennie, la recherche sur les effets écologiques des grands carnivores s'est fortement développée (Ripple et al. 2014; Kuiper et al. 2016). Cette recherche est activement mobilisée pour promouvoir les efforts de conservation et de restauration des carnivores, y compris en Europe, où les carnivores tels que le loup (*Canis lupus*), l'ours brun (*Ursus arctos*), le lynx boréal (*Lynx lynx*) et le glouton (*Gulo gulo*) recolonisent rapidement leurs anciens territoires dans des écosystèmes qui sont intensément modifiés ou façonnés par l'humain (Linnell et al. 2010; Chapron et al. 2014).

Autrefois répandues en Europe, de nombreuses populations de grands carnivores ont disparu au cours du siècle dernier, principalement en raison des interférences avec les activités humaines comme l'élevage ou la chasse (Breitenmoser 1998; Ripple et al. 2014). Bien que ces espèces aient survécu, leur répartition et leurs effectifs ont été affectés de manière conséquente par l'humain, à la fois directement par les prélèvements et indirectement par l'exploitation industrielle des forêts et des ongulés sauvages (Ripple et al. 2014). Depuis les années 1970, les populations de grands carnivores ont commencé à se rétablir, en partie grâce à une augmentation de la superficie forestière, mais aussi à une augmentation de populations d'ongulés sauvages ayant retrouvé des niveaux records dans de nombreuses régions (Milner et al. 2007; Basille et al. 2009), ainsi que des mesures de conservation (comme par exemple la Directive habitats faune flore de l'Union européenne) qui ont attribué à ces prédateurs des statuts stricts de protection, interdisant à la fois les prélèvements et les nuisances à leur égard. De nombreuses populations de grands carnivores ont ainsi augmenté, soit par expansion naturelle (Wabakken et al. 2001; Valière et al. 2003), soit par réintroduction (Breitenmoser & al 2000) ou les deux à la fois (Linnell et al. 2009).

Souvent utilisé comme exemple de réussite en matière de conservation, le rétablissement de populations de grands carnivores dans des zones dominées par l'humain pose toutefois des défis, par exemple celui de savoir s'il reste encore des zones suffisamment vastes et fonctionnelles pour des populations viables de carnivores où il serait également possible d'atténuer les conflits potentiels avec les activités humaines (Packer et al. 2013). En effet, leur retour a été associé dans de nombreuses régions avec des impacts, tant économiques (Kaczensky 1999) que sociaux (Skogen & Haaland 2001). En effet, d'une part dans ces territoires majoritairement anthropisés, les grands carnivores ont besoin de larges surfaces pour survivre, rivalisant ainsi indirectement avec les humains pour l'espace et les ressources (Treves 2009). D'autre part, des impacts plus directs existent aussi, consistant principalement en des attaques sur les élevages domestiques et semi-domestiques par les grands carnivores

Contexte général

(Linnell & Cretois 2018). Au cours du temps, les humains ont développé de nombreuses méthodes pour protéger les troupeaux des attaques par les grands carnivores (Lescureux & Linnell 2014), et mis en place des programmes de contrôle et de protection des populations de prédateurs pour réduire leurs impacts (Boitani 1995). Par ailleurs, les grands carnivores étant protégés en Europe dans le cadre de la convention de Berne et la directive habitats (Trouwborst 2010), leur statut de conservation doit être maintenu comme « favorable » (Favourable Conservation Status FCS- Habitat directive 1992). Cet état favorable énoncé dans la directive habitat est basé sur le système utilisé par l'Union Internationale pour la Conservation de la Nature (UICN) afin d'établir la liste rouge des espèces menacées (IUCN 2012). Lorsque l'application des critères de la liste rouge, basés en partie sur l'évolution de la répartition spatiale des espèces, aboutit aux qualifications « en danger critique d'extinction », « en danger » ou « vulnérable », alors l'état de conservation est par définition défavorable (Trouwborst 2010). Les grands carnivores vivant à de faibles densités et sur de larges territoires, la conservation de populations viables doit donc être planifiée et coordonnée sur des échelles très larges, couvrant souvent plusieurs pays. Il est donc nécessaire de savoir comment coordonner la gestion de ces espèces à grande échelle, éventuellement au-delà des frontières (Linnell & Boitani 2012; Bischof et al. 2016), en particulier dans le contexte des traités internationaux.

Dans ce contexte, décrire (cartographier) et comprendre la dynamique de la répartition spatiale des grands carnivores est une étape clé de la gestion de ces espèces (Chapron et al. 2014). Cela peut aider à cibler des zones de présence potentielles et à prévoir des mesures destinées à atténuer les conflits potentiels souvent associés à la recolonisation des grands carnivores. Enfin, une meilleure connaissance des dynamiques de la répartition spatiale permet de renforcer l'évaluation du statut de conservation de l'espèce en Europe par l'UICN, il est donc crucial d'avoir accès à cette information dans le cadre de la conservation des grands carnivores en Europe, y compris à large échelle.

Encadré No 1 : La situation du loup en France et du lynx dans les pays alpins et en Norvège

Loup (*Canis lupus*) en France

Statut de conservation : « En danger d'extinction » en France (UICN, 2012).

Historique de la population : Après avoir été exterminé au début du siècle précédent, les loups sont réapparus dans les Alpes françaises au début des années 1990, suite à l'expansion de l'aire de répartition de la population italienne (Ciucci et al., 2009 ; Lucchini et al., 2004). Des analyses génétiques suggèrent que tous les loups français partagent le même patrimoine génétique que les loups italiens (Fabbri et al. 2007).

Effectifs : Les loups ont établi une population permanente en France avec 49 zones de présence permanente dans les Alpes en sortie d'hiver 2015-2016 parmi lesquelles 35 étaient constituées en meutes (Duchamp et al. 2017). Les effectifs ont évolué de quelques individus en début des années 1990 à autour de 250 à 300 individus en 2016.

Densité : La densité de loups varie de 1 individu pour 12 km² à 1 individu pour 120 km² en Europe (UICN, 2012).

Régime et habitat : Le loup est une espèce opportuniste qui peut adapter son régime alimentaire en fonction des espèces de proies disponibles (Pouille et al. 1997, Imbert et al. 2016). La composition de l'alimentation dans l'ensemble de l'aire géographique et les variations saisonnières dépendent des effectifs des proies. Il peut s'attaquer aux moutons comme proies domestiques. En général, les grandes zones forestières sont particulièrement adaptées aux loups en Europe, bien que les loups ne soient pas principalement des espèces forestières.

Dispersion : Les loups vont disperser pour pouvoir se reproduire lors de deux périodes de l'année, au printemps et en automne (Mech et Boitani 2010). Les loups peuvent parcourir de très longues distances, jusqu'à plus de 900 km (Ciucci et al. 2009), et peuvent aussi disperser à courte distance de la meute d'où ils sont originaires (Mech et Boitani 2010).

Lynx boréal (*Lynx lynx*) dans les pays alpins et en Norvège

Statut de conservation : Dans les pays alpins : « En danger d'extinction » ; En Norvège : « Préoccupation mineure » (UICN 2012).

Historique de la population : Dans les pays alpins, après son éradication totale dans les Alpes vers 1930, le lynx a été réintroduit à plusieurs reprises entre 1970 et aujourd'hui en Suisse, en Italie, en Autriche et en Slovénie (Molinari-Jobin et al. 2017). En Norvège, la population n'a jamais disparu mais a toujours été chassée de manière importante avec une diminution importante de la population au cours du siècle précédent et une ré-augmentation au début des années 1980 suite à la prise de décision d'instaurer des quotas de chasse (Linnell et al. 2010).

Effectifs : Dans les pays alpins, le nombre de lynx dans l'ensemble des Alpes a été estimé à environ 90-120 individus (Molinari-Jobin et al. 2012). En Norvège, la population est passée d'environ entre 400 et 500 individus à environ 300 au début des années 2000, pour augmenter à nouveau autour de 450-500 individus en 2008 (Linnell et al. 2010).

Densités : Dans les pays alpins, de 1 à 3 individus pour 100 km² (Breitenmoser et al. 1998). En Norvège, 0.3 individus pour 100 km² (Sunde et al. 2000).

Régime et habitat : Le lynx chasse principalement des proies d'ongulés sauvages, bien qu'il soit capable de chasser des espèces plus petites dans des endroits où les ongulés sont moins abondants. Dans les pays alpins, il chasse aussi des proies domestiques comme le mouton (Breitenmoser 1998) tandis qu'en Norvège il s'attaque aux moutons et aux rennes, des proies semi-domestiques (Mattisson et al. 2011).

Dispersion : La dispersion du lynx est généralement associée aux jeunes adultes qui quittent leur région natale à l'âge de 9 à 11 mois. Dans les pays alpins, la dispersion d'individus est en moyenne de 42 km (Kramer-Schadt et al., 2004). En Norvège, les distances de dispersion sont en moyenne de 100 km, en partie en raison de la taille beaucoup plus grande des domaines vitaux (Samelius et al., 2012).

Références:

- Breitenmoser U. 1998. Large predators in the alps: the fall and rise of man's competitors. *Biological Conservation* 83:279–289.
- Ciucci P, Reggioni W, Maiorano L, Boitani L. 2009. Long-Distance Dispersal of a Rescued Wolf From the Northern Apennines to the Western Alps. *Journal of Wildlife Management* 73:1300–1306.
- Duchamp C, Chapron G, Gimenez O, Robert A, Sarrazin F, Beudels-Jamar R, Le Maho Y. 2017. Expertise collective scientifique sur la viabilité et le devenir de la population de loups en France à long terme sous la coordination ONCFS-MNHN de: Guinot-Ghestem M. Haffner P., Marboutin É., Rousset G., Savouret-Soubelet A., Sibley JP, Trudelle L.(par or. alph.).
- Fabbri E et al. 2007. From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. *Molecular Ecology* 16:1661–1671.
- Imbert C, Caniglia R, Fabbri E, Milanese P, Randi E, Serafini M, Torretta E, Meriggi A. 2016. Why do wolves eat livestock? *Biological Conservation* 195:156–168.
- IUCN. 2012. IUCN Red List Categories and Criteria: version 3.1, 2nd edn. Gland, Switzerland and Cambridge, UK.
- Kramer-Schadt S, Revilla E, Wiegand T, Breitenmoser U. 2004. Fragmented Landscapes, Road Mortality and Patch Connectivity: Modelling Influences on the Dispersal of Eurasian Lynx. *Journal of Applied Ecology* 41:711–723.
- Linnell JDC, Broseth H, Odden J, Nilsen EB. 2010. Sustainably Harvesting a Large Carnivore? Development of Eurasian Lynx Populations in Norway During 160 Years of Shifting Policy. *Environmental Management* 45:1142–1154.
- Mattisson J, Odden J, Nilsen EB, Linnell JDC, Persson J, Andrén H. 2011. Factors affecting Eurasian lynx kill rates on semi-domestic reindeer in northern Scandinavia: Can ecological research contribute to the development of a fair compensation system? *Biological Conservation* 144:3009–3017.
- Mech LD, Boitani L. 2010. *Wolves: behavior, ecology, and conservation*. University of Chicago Press.
- Molinari-Jobin A et al. 2012. Monitoring in the presence of species misidentification: The case of the Eurasian lynx in the Alps. *Animal Conservation* 15:266–273.
- Molinari-Jobin A et al. 2018. Mapping range dynamics from opportunistic data: Spatiotemporal modelling of the lynx distribution in the Alps over 21 years. *Animal Conservation* 21:168–180.
- Pouille M-L, Carles L, Lequetie B. 1997. Significance of ungulates in the diet of recently settled wolves in the Mercantour mountains (southeast France).
- Samelius G, Andrén H, Liberg O, Linnell JDC, Odden J, Ahlqvist P, Segerström P, Sköld K. 2012. Spatial and temporal variation in natal dispersal by Eurasian lynx in Scandinavia: Natal dispersal by Eurasian lynx. *Journal of Zoology* 286:120–130.
- Sunde P, Kvam T, Moa P, Negård A, Overskaug K. 2000. Space use by Eurasian lynxes *Lynx lynx* in central Norway. *Acta Theriologica* 45:507–524.

1. b. Quels outils utiliser pour modéliser la répartition spatiale des grands carnivores ?

Les modèles de distribution d'espèces (« Species Distribution Models », SDMs) sont devenus un élément fondamental dans les domaines de l'écologie, la biogéographie, la conservation ainsi que la gestion des ressources naturelles (Guisan et al. 2017). Les SDMs peuvent se diviser en deux catégories, (i) les SDMs corrélatifs qui établissent la corrélation entre les données spatiales de présence d'une ou plusieurs espèces avec des facteurs environnementaux ou météorologiques qui ont du sens biologiquement pour la ou les espèces concernées (Guillera-arroita et al. 2015) et (ii) les SDMs mécanistes qui consistent à incorporer explicitement les liens mécanistes entre les caractéristiques fonctionnelles des organismes telles que la croissance ou la dispersion et leur environnement (Kearney & Porter 2009).

Souvent, les seules données de présence disponibles pour la modélisation de la répartition d'une espèce sont un ensemble de localisations où l'espèce étudiée a été observée, mais il n'y a généralement pas d'information sur les absences, c'est-à-dire des zones qui ont été étudiées mais où l'espèce n'a pas été observée. Ce type de données est souvent appelé données de « presence-only », ou plus communément « presence-background » (PB). Dans le cadre d'études à large échelle, la majeure partie des données sont généralement de type PB (Koshkina et al. 2017), car elles proviennent d'enquêtes dites « opportunistes », c'est-à-dire un suivi sans protocole où les observateurs prospectent et signalent lorsqu'un signe de présence est détecté mais ne fournissent pas d'information sur les zones prospectées sans avoir détecté l'espèce ainsi que les zones non prospectées (Van Strien et al. 2013). Pour faire des estimations de distribution à partir de ce type de données, il faut des informations environnementales de fond pour les zones où l'espèce n'a pas été observées (Guillera-arroita et al. 2015). Les SDMs basés sur des PB permettent d'estimer des préférences en matière d'habitat en comparant les caractéristiques environnementales des localisations où l'espèce a été détectée avec les caractéristiques environnementales de l'ensemble de la zone d'étude, ce qui constitue le « fond ». Il existe de nombreuses méthodes pour modéliser la distribution d'espèces à partir des données PB, y compris le modèle MaxEnt (Phillips et al. 2006), des modèles de régression (Elith et al. 2006) et les modèles de processus ponctuels spatiaux qui sont équivalents au modèle MaxEnt (Renner & Warton 2013; Renner et al. 2015).

Un deuxième type de données est appelé les données de « présence-absence » (PA). Elles sont constituées par un ensemble de localisations où l'espèce a été détectée mais aussi par l'ensemble des localisations où l'espèce n'a pas été détectée. L'un des avantages des données de PA est qu'elles fournissent des informations sur les zones qui ont été prospectées et celles qui ne l'ont pas été (Elith & Leathwick 2009). Les régressions logistiques sont souvent utilisées afin de modéliser la répartition d'espèces à partir de données de PA (Guillera-arroita et al. 2015).

Enfin, il existe un troisième type de données, les données de « patch-occupancy » (PO), qui sont similaires aux données de PA, mais le protocole de récolte est fait de telle sorte que le processus d'observation peut être inclus dans le SDM (MacKenzie et al. 2002). En effet, les données d'absence peuvent signifier plusieurs choses : soit un habitat inapproprié pour l'espèce, soit un habitat convenable mais inoccupé car inaccessible ou pas encore colonisé, soit à une zone où l'espèce était présente mais n'a pas été détectée (Elith & Leathwick 2009). Les

Contexte général

visites sont répétées dans le temps sur des sites et les détections et non-détections de l'espèce sont notées pour chaque visite. Ce processus permet ainsi de prendre en compte explicitement la probabilité de détecter l'espèce sur un site et ainsi d'étudier les variations de cette probabilité en fonction des sites et des visites (MacKenzie et al. 2002).

Le choix du SDM dépend des objectifs et de contraintes du sujet d'étude. Il convient donc de mettre en regard les avantages et les inconvénients de chacune des approches avec les particularités dans le cadre de la mesure de la répartition des grands carnivores. Tout d'abord, les grands carnivores vivent à faible densité sur de très larges territoires (UICN 2012, voir aussi encadré No 1), ce qui rend faible la probabilité de les détecter. Ne pas prendre en compte cette faible probabilité de détection peut biaiser les résultats de différentes manières. Tout d'abord la distribution observée peut être sous-estimée. Ensuite les facteurs environnementaux favorisant la détection de l'espèce peuvent être pris pour ceux influençant la présence de l'espèce (Kéry & Schaub 2011; Lahoz-Monfort et al. 2014).

De plus, comme nous l'avons vu précédemment, les grands carnivores sont en recolonisation d'espaces précédemment non occupés, leur répartition est donc dynamique dans le temps. Or, la plupart des SDMs est basée sur le concept de la niche écologique (Grinnell 1917; Hutchinson 1957). Ce concept repose sur le fait que l'espèce est en équilibre avec son environnement, en d'autres termes qu'elle occupe les habitats qui lui sont le plus favorables et donc que la dispersion n'est pas un facteur limitant (Jeschke & Strayer 2006). Les SDMs statiques ne prennent donc pas en compte les processus dynamiques et peuvent produire des résultats biaisés sur les estimations de distribution (Zurell et al. 2009; Yackulic et al. 2015).

Face à ces problématiques, j'ai exploré dans ma thèse les modèles dynamiques d'occupancy afin de modéliser la répartition des grands carnivores. Je développerai ces modèles dans la section 3.

1.c. Sciences citoyennes et données opportunistes

Les grands carnivores tels que le lynx et le loup ont un comportement élusif, vivent à de faibles densités et sur de grands territoires (Encadré No 1). Ces particularités rendent leur étude difficile sur la base d'observations directes uniquement, c'est-à-dire des observations de l'animal vivant dans son milieu (Ripple et al. 2014). Face à ces difficultés, l'étude des grands carnivores en Europe repose sur plusieurs méthodes de suivi, réalisées non seulement par des professionnels de la faune sauvage mais aussi par des non-professionnels, ici appelés « citoyens ».

On parle de « sciences citoyennes » ou « sciences participatives » (Godet 2010) comme le fait d'engager le public dans un projet scientifique afin de produire des données et des informations fiables et utilisables par les scientifiques et/ou les gestionnaires (Mckinley et al. 2017). La participation des citoyens permet de couvrir une grande zone d'étude sur de longues périodes de temps, ce qui serait très coûteux voire infaisable si des professionnels uniquement étaient impliqués (Molinari-Jobin et al. 2018). Les sciences citoyennes sont devenues d'une grande importance dans le domaine de l'écologie au cours de la dernière décennie (Silvertown 2009). Les citoyens participent désormais de plus en plus à des projets portant sur le changement climatique, les espèces envahissantes, la biologie de la conservation, la

Contexte général

restauration écologique, la surveillance de la qualité de l'eau, et l'écologie de la population (Silvertown 2009). Cette explosion est en partie due à l'utilisation d'outils facilement accessibles pour diffuser l'information sur les projets et recueillir des données auprès du public. Un deuxième facteur favorisant l'augmentation des sciences citoyennes est le fait que les citoyens peuvent être une source gratuite de main-d'œuvre et de compétences (Cohn 2008; Silvertown 2009). L'implication d'un public plus large permet la prospection sur de plus larges surfaces contrairement à ce qui se passe lors des programmes de suivis scientifiques coûteux (Kosmala et al. 2016).

Les méthodes d'observation généralement utilisées pour suivre les grands carnivores sont dites « indirectes », ce qui signifie que la présence des carnivores est étudiée à travers des signes indirects de présence tels que des traces dans la neige ou la boue, des carcasses de proies, des photos de pièges caméras ou des fèces (Linnell et al. 2007; Duchamp et al. 2012; Molinari-Jobin et al. 2018). On distingue ici deux types de suivis selon qu'un protocole d'échantillonnage est mis en place ou pas. Les suivis avec protocole d'échantillonnage sont par exemple ceux basés sur des « captures » via des pièges photographiques ou des pièges à poils pour faire des analyses génétiques, ou le comptage via des transects (Linnell et al. 2007; Duchamp et al. 2012; Molinari-Jobin et al. 2018). La récolte de telles données nécessite souvent un effort considérable et le suivi est généralement fait de manière locale (Linnell et al. 2010). L'autre type de suivi, opportuniste, consiste pour des citoyens (au sens donné plus haut) à récolter tous les signes de présences trouvés sans protocole formel. Il est alors possible de couvrir de grandes zones géographiques mais cela implique de ne pas avoir de mesure réelle de l'effort d'échantillonnage.

2. Les cas du loup en France et du lynx dans les pays alpins et en Norvège

Dans le cadre de ma thèse, mon travail s'est focalisé sur deux espèces : le loup en France et le lynx dans les pays alpins ainsi qu'en Norvège. Pourquoi ces cas d'étude ? C'est l'objet de cette section dans lequel je décris le contexte dans lequel les données ont été récoltées, quels étaient les verrous méthodologiques et quels développements nous proposons pour aller au-delà de ces limites. Cette partie développe de manière plus approfondie les sections « données » des articles 3, 4 et 5 et comporte donc des répétitions.

2.a. Cas du Loup en France

2.a.1. Contexte du suivi

Les premiers signes de présence du loup furent trouvés en 1992 dans le Parc National du Mercantour, avec d'abord une observation visuelle par des agents du parc puis un loup retrouvé mort la même année dans les Hautes-Alpes (Duchamp et al. 2012). A partir de cette date, la population de loups s'est développée en France notamment dans les Alpes puis les Pyrénées et le Massif Central à la fin des années 1990, pour ensuite atteindre les Vosges au début des années 2010. A partir du début des années 1990, un suivi a été mis en place par l'Office de la Chasse et de la Faune Sauvage (ONCFS) (Duchamp et al. 2017 p. 2017). Le premier objectif des autorités françaises était l'obtention d'estimations fiables de la répartition spatiale des loups, des tendances des effectifs de populations ainsi qu'un suivi des attaques de loups

sur les élevages. Le plan d'action national pour le loup (MEDEEM 2008) vise à répondre aux exigences des directives européennes consistant à atteindre un état de conservation favorable (Habitat directive, 1992). Cet état de conservation favorable est défini par (i) une dynamique de population indiquant que l'espèce est un élément viable des habitats auxquels elle appartient, (ii) une aire de répartition ne diminuant ni ne risquant de diminuer dans un avenir prévisible, (iii) un habitat suffisamment étendu pour que la population se maintienne à long terme (Duchamp et al. 2017). Le suivi de l'espèce a été mis en place pour pouvoir fournir (i) des estimations robustes de la dynamique de la population (par exemple la dynamique des effectifs, les taux de survie et de mortalité) ainsi que (ii) des mises à jour annuelles de ces estimations afin de pouvoir définir l'état de conservation de l'espèce. Le but final du suivi est de pouvoir développer un plan de gestion à partir de ces estimations. Les données de détection des loups ont donc commencé à être récoltées dès 1992, tout au long de l'année, et se poursuivent jusqu'à nos jours. Cette récolte de signes de présence indirecte se fait grâce à un réseau d'observateurs professionnels et non professionnels, le réseau « loup-lynx », géré par l'ONCFS. La taille du réseau est passée de quelques centaines de personnes en 1994 à plus de 3000 experts actuellement. Ces observateurs font partie de catégories socio-professionnelles variées dont, par exemple, des agents de l'ONCFS, des bénévoles, des chasseurs ou encore des agents de parcs nationaux (Tableau 1). Chaque observateur participe à une formation animée par l'ONCFS (Duchamp et al. 2012). Lors de cette formation, la reconnaissance des signes de présence, la biologie de l'espèce ainsi que le mode opératoire en cas de détection sur le terrain sont abordés. Chaque détection de signe de présence du loup est accompagnée d'un rapport standardisé qui renseigne, entre autres, la position GPS et la date de la détection. Sur la base de ce rapport, les experts de l'ONCFS classent la détection selon trois catégories : (i) « certain » comme par exemple une observation directe, une photo prise par un piège caméra, une piste fraîche ou encore des fèces en supposant qu'il n'y ait pas de fausses détections positives dans ces données, (ii) « incertain » comme par exemple une photo floue, une piste incomplète ou encore un témoignage, et (iii) « rejeté » pour des données trop dégradées ou manifestation pas un loup et exclues de la présente étude (Duchamp et al. 2012). Nous avons choisi délibérément de travailler à partir de l'année 1994, année où plus d'un signe de présence a été trouvé.

De plus, le suivi de la dynamique de la répartition spatiale du loup en France constitue une analyse complémentaire aux analyses de tendance dans les effectifs. L'étude des effectifs de la population de loups en France a été étudiée depuis le début de la recolonisation du loup en France (Cubaynes et al. 2010; Duchamp et al. 2017). Basées sur des études de capture-recapture, ces études ont pour but principal d'analyser les sources de mortalité de la population de loups ainsi que d'estimer les effectifs et à partir de cette estimation en déterminer le seuil minimum de viabilité, défini comme « l'effectif le plus faible à partir duquel le risque d'extinction serait considéré comme acceptable » (Duchamp et al. 2017). Ces études permettent aussi d'analyser et de mettre en place les tirs de prélèvements sur la population de loups, mesure mise en place afin de diminuer les risques d'impact sur les activités humaines. Cependant, les modèles d'estimation d'effectifs reposent en grande partie sur des données récoltées au sein des Alpes, zone qui rassemble l'essentiel des effectifs (Duchamp et al. 2012). Dans un contexte où l'espèce est en train de s'étendre de plus en plus largement autour des Alpes, la capacité d'estimation de la dynamique des effectifs dans les autres zones de la France par les modèles mis en place reste incertaine (Duchamp et al. 2017).

2.a.2. Les verrous méthodologiques

Comme nous l'avons vu précédemment, la collecte des signes de présence du loup en France est faite de manière opportuniste, c'est-à-dire sans protocole ni mesure de l'effort d'échantillonnage. Or, nous savons que cet effort peut varier 1) dans le temps : une zone peut être prospectée par plusieurs personnes une année, et par une seule personne l'année suivante ; et 2) dans l'espace : imaginons deux sites où l'espèce est présente, si l'effort d'échantillonnage est très bas sur l'un il est fort possible que l'espèce n'y soit pas observée, tandis que si l'effort d'échantillonnage y est fort sur l'autre, l'espèce y sera détectée plus facilement. Or, cela signifie en premier lieu qu'il est difficile voire impossible de différencier les zones qui n'ont pas été prospectées de celles qui ont été prospectées et où l'espèce n'a pas été détectée. Un nombre de détections augmentant peut donc simplement refléter une augmentation du nombre d'observateurs présents dans une zone plutôt que l'augmentation de la distribution de l'espèce étudiée (Kéry 2011). Un effort d'échantillonnage hétérogène dans le temps et dans l'espace peut donc mener à une probabilité de détecter l'espèce aussi hétérogène dans le temps et dans l'espace. Ne pas prendre en compte ce facteur peut mener à une sous-estimation de la probabilité de détection et donc à des estimations biaisées de la répartition de l'espèce (Van Strien et al. 2013).

De plus, l'effort d'échantillonnage fournit une information cruciale sur les sites qui sont visités et sur ceux qui ne le sont pas. Or cette information permet de faire la différence entre des données de PB et les données de PA et plus particulièrement celles de PO. En effet, faire la différence entre des sites visités où l'espèce n'a pas été détectée et des sites non visités permet d'estimer la probabilité de détection (Guillera-arroita et al. 2015). Il est donc apparu important dans le cadre de ma thèse de commencer par avoir une estimation de l'effort d'échantillonnage dans le cadre du suivi du loup afin de pouvoir établir des modèles d'occupancy dynamique.

2.a.3. Les données

Nous décrivons maintenant les données que nous avons utilisées dans le cadre de ma thèse. Tout d'abord, nous n'avons utilisé que les données d'hiver (Décembre, Janvier, Février et Mars) afin de minimiser les mouvements (émigration et entrée dans un nouveau site) et de répondre aux exigences des modèles que nous avons utilisés (voir section b.3. *Les hypothèses des modèles d'occupancy*). Afin de générer des détections et surtout les non-détections, nous avons ramené les données à l'échelle d'un site de 10 x 10 km, l'unité spatiale choisie pour établir la distribution de l'espèce (Figure 1). Si plusieurs signes de présence sont trouvés au sein d'un site lors d'un mois d'échantillonnage, alors nous les comptabilisons comme une seule détection à l'échelle du site et du mois. Pour générer des non-détections, il a fallu connaître les sites échantillonnés et non-échantillonnés par les observateurs. Le nombre de personnes faisant partie du réseau loup-lynx ayant augmenté au cours du temps et dans l'espace, un premier travail a consisté à quantifier à posteriori l'effort d'échantillonnage, défini comme le nombre d'observateurs par site chaque année (voir article 1). Une fois cet effort calculé, nous avons écarté de l'analyse les sites non échantillonnés (c'est-à-dire ayant un effort d'échantillonnage nul) et nous avons considéré les sites ayant un effort d'échantillonnage non nul comme visités mais où l'espèce n'y a pas été détectée.

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

Depuis le début du suivi du loup, le nombre de signes de présence n'a cessé d'augmenter au cours des années. Au début du suivi du loup, il y avait très peu de sites sur lesquels des données de présence ont été récoltées (entre 1 et 9 sites avec un signe de présence), jusqu'à environ 1998 où la population a aussi commencé à s'étendre spatialement dans les Pyrénées et les Vosges (Figure 1 et Tableau 2). De même, les détections classées comme incertaines ne sont pas apparues avant l'année 1996. De 1994 à 2016, 3215 observations (86 %) ont été classées comme certaines et 503 (14 %) comme incertaines. D'autre part, une augmentation des détections ne signifie pas nécessairement une augmentation de l'aire de répartition du loup ou du nombre d'individus. Cela peut aussi refléter une augmentation du nombre d'observateurs sur les sites. En 1994, l'effort d'échantillonnage maximum était de 35 observateurs alors qu'en 2016, sur ces mêmes sites l'effort d'échantillonnage était de plus de 150 observateurs. Imaginons que l'espèce soit présente sur 4 sites et qu'initialement il y avait très peu d'observateurs sur trois sites et une trentaine sur le quatrième. Il y a plus de chances que l'espèce soit détectée sur le quatrième site que sur les trois autres. Maintenant, si au cours du temps le nombre d'observateurs augmente jusqu'à 150 sur tous les sites, nous aurions sûrement constaté une augmentation des détections dans les trois premiers sites, ne reflétant pas l'apparition de l'espèce sur les trois premiers sites mais simplement l'augmentation du nombre d'observateurs par site. Soulignons la présence de signes de présence classés comme incertains en dehors du cœur de la distribution du loup en France, celui-ci se trouvant au niveau des Alpes, démontrant ainsi que, dans un lieu qui n'a pas encore été colonisé, les premières détections sont des détections incertaines (Figure 1).

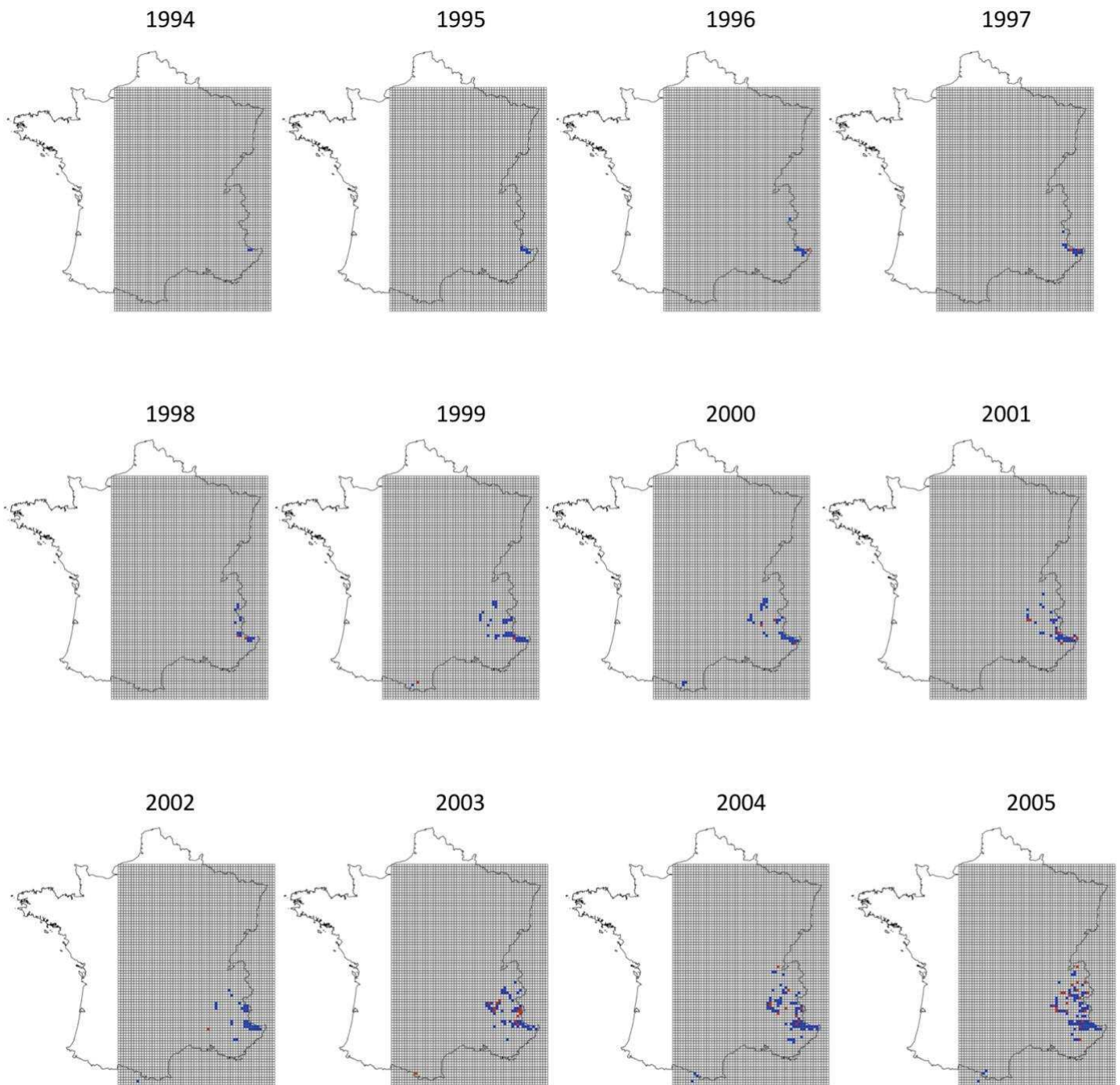
Tableau 1 : Catégories socio-professionnelles des observateurs présents dans le réseau loup-lynx depuis 1994 jusqu'à 2016.

Catégorie socio-professionnelle	Nombre d'observateurs
Administration	275
Agent de parc National	323
Agent de parc naturel régional	87
Agent de réserve naturelle	26
Scientifique	10
Profession agricole	91
Chasseur	423
Membre d'une association naturaliste	267
Bénévole	474
Agent de l'ONCFS	746
Agent de l'Office National des Forêts	354

Tableau 2 : Nombre de sites avec des détections certaines et incertaines au cours des mois d'échantillonnage (sélectionnés) du suivi du loup en France entre 1994 et 2016.

Année	Décembre (année n-1)		Janvier (année n)		Février (année n)		Mars (année n)	
	Certaines	Incertaines	Certaines	Incertaines	Certaines	Incertaines	Certaines	Incertaines
1994	0	0	1	0	1	0	2	0
1995	3	0	3	0	3	0	5	0
1996	4	0	5	0	6	1	7	0
1997	9	0	12	0	5	0	7	2
1998	7	1	9	2	8	0	9	1
1999	17	0	19	1	20	2	21	1
2000	13	0	17	1	24	1	20	1
2001	20	1	20	3	20	1	16	4
2002	17	0	20	0	23	0	18	1
2003	28	0	27	5	30	3	23	5
2004	36	4	39	3	43	3	41	2
2005	31	4	32	8	38	7	40	4
2006	33	4	46	9	37	6	44	5
2007	28	1	45	2	41	2	40	6
2008	35	4	42	2	53	1	50	8
2009	40	6	55	21	45	13	58	11
2010	39	8	49	16	34	13	48	13
2011	38	3	59	10	58	11	54	4
2012	45	9	64	10	62	5	58	8
2013	45	10	81	12	68	14	56	18
2014	69	8	83	11	69	18	80	9
2015	60	11	81	16	79	26	69	22
2016	40	13	70	7	75	14	71	11

Les cas du loup en France et du lynx dans les pays alpins et en Norvège



Les cas du loup en France et du lynx dans les pays alpins et en Norvège

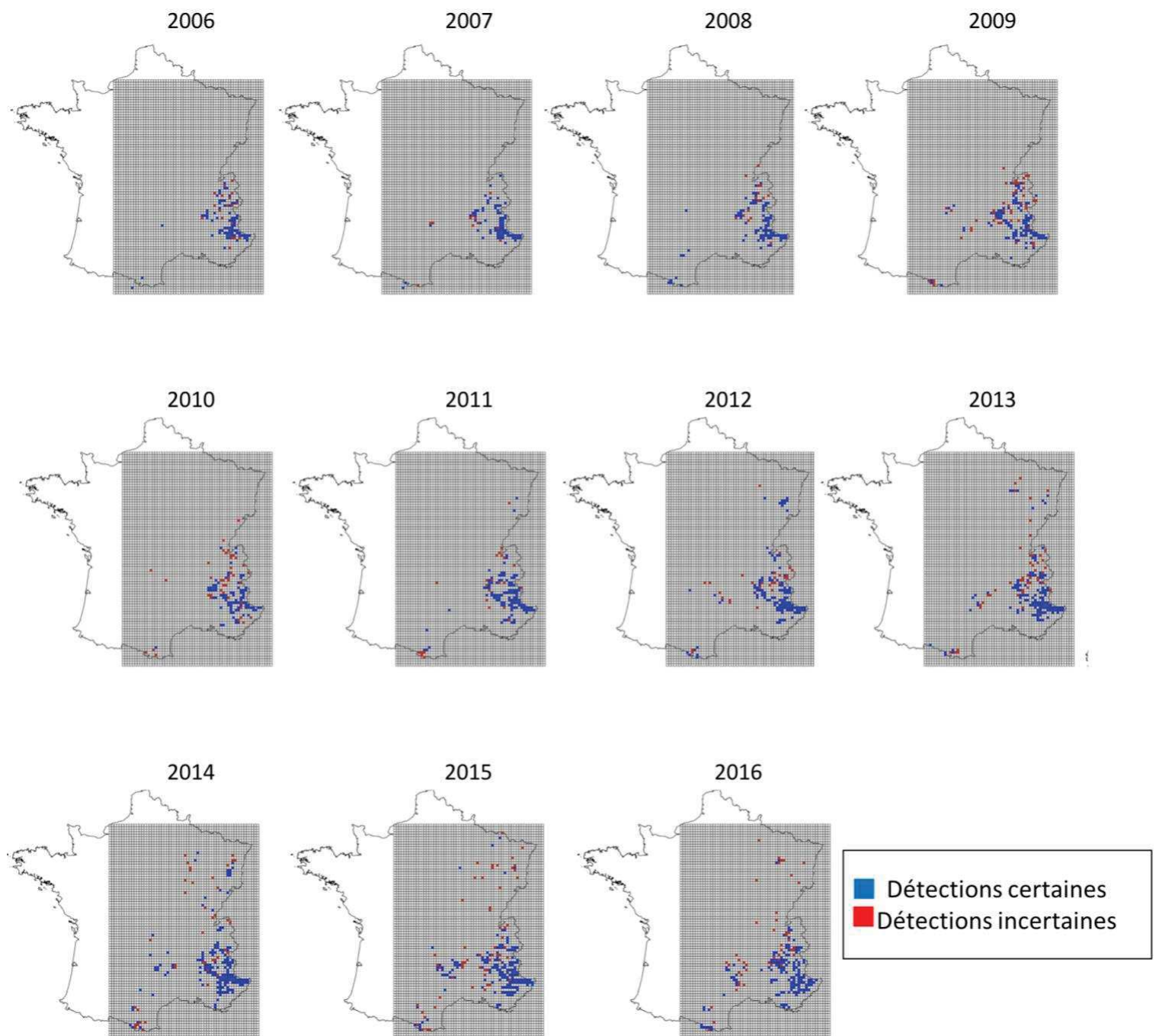


Figure 1: Cartes des détections certaines (en bleu) et incertaines (en rouge) du loup en France entre 1994 et 2016. La grille correspond aux sites de 10 x 10km considérés comme unités d'échantillonnage.

2.b. Cas du lynx dans les pays alpins

2.b.1. Contexte du suivi

Concernant la situation du lynx dans les pays alpins, celui-ci a été réintroduit dans les Alpes dans les années 1970, et plus de 40 ans plus tard, moins de 20% des Alpes ont été recolonisées (Molinari-Jobin et al. 2010). Les experts du lynx des sept pays alpins (France, Italie, Liechtenstein, Slovénie, Suisse et Allemagne) ont mis en place le projet visant à suivre l'état et la conservation de la population de lynx des Alpes (Status and Conservation of the Alpine Lynx Population -SCALP, Molinari-jobin et al. 2003), dont l'objectif est de rétablir et maintenir une population viable de lynx dans les pays alpins, en coexistence avec les populations humaines. Ce projet a pour but de coordonner et normaliser la surveillance et la conservation du lynx alpin à grande échelle, à travers les 7 pays concernés (Molinari-jobin et al. 2003). L'idée principale est de synchroniser la récolte de données à une échelle internationale, et de mettre en place une stratégie de conservation du lynx dans ces différents pays, la Stratégie pan-Alpine de Conservation du lynx (PanAlpine Conservation Strategy PACS) (Molinari-jobin et al. 2003).

Comme pour le suivi du loup en France, le suivi du lynx est donc réalisé grâce à un réseau de plus de 1300 observateurs tels que des gardes-chasses ou des agents des services forestiers. L'avantage d'un tel nombre est qu'il permet de couvrir largement l'ensemble du territoire du lynx dans les Alpes. De plus, des chasseurs, des naturalistes et d'autres personnes (i.e. des « citoyens ») peuvent faire partie du réseau, permettant une amélioration de l'acceptation sociale du suivi, tout en augmentant les chances de détecter la présence du lynx, étant donné que les membres ont des activités différentes sur le terrain (Molinari-jobin et al. 2003). Les membres du réseau recherchent activement les signes de présence toute l'année et les experts vérifient les signes qui leur sont signalés par les citoyens. Les signes de présence ont été classés en trois catégories de fiabilité : C1 correspond aux données indiscutables, par exemple un lynx mort, des photos de pièges photographiques ainsi que certains échantillons génétiques, C2 sont des détections qui ont été confirmées par un expert (les restes de proies sauvages et les traces) et C3 sont des données qui n'ont pas pu être vérifiées par des experts (traces non vérifiées et restes de proies sauvages trop dégradées). Ici, afin d'avoir la même classification que pour le loup en données certaines et incertaines, nous avons considéré les données C1 et C2 comme des détections certaines, en supposant qu'il n'y avait pas de fausses détections positives dans ces données, tandis que les données C3 ont été traitées comme des détections incertaines.

De plus, la proportion de la superficie occupée par l'espèce est actuellement considérée comme un possible paramètre pouvant quantifier l'état de conservation actuel de la population de lynx en Europe (Molinari-Jobin et al. 2018). En effet, l'état de conservation favorable a été défini pour le lynx en Europe comme un total de 1000 individus matures (ou 1300 individus indépendants) répartis dans les Alpes (Schnidrig et al. 2016; Molinari-Jobin et al. 2018). Toutefois, il est apparu difficile d'obtenir des estimations fiables d'effectifs de la population. A l'origine plusieurs études démographiques ont été menées pour appuyer la mise en place du réseau SCALP et la définition de normes communes depuis la réintroduction des lynx dans les années 1970 jusqu'en 1995. Ces mêmes études ont servies également à établir le statut de conservation du lynx pour chacun des pays (Molinari-Jobin et al. 2010). L'évaluation ultérieure de l'évolution démographique et de la répartition du lynx dans les Alpes s'est appuyée sur ces rapports d'état, dans lesquels le nombre et la répartition des signes du lynx étaient évalués

dans chaque pays. Cependant, la répartition spatiale de l'espèce n'était estimée qu'à travers les tendances des observations faites sans prendre en compte la probabilité de détection imparfaite du lynx (voir Molinari-Jobin et al. 2006, 2010), jusqu'à l'analyse de Molinari-Jobin et al. (2012) qui ont ajusté le premier modèle d'occupancy dynamique afin de comparer les répartitions obtenues avec les données certaines et les données incertaines séparément, sans prendre en compte la probabilité de faux positifs. Enfin, Molinari-Jobin et al. (2018) ont ajusté un modèle d'occupancy dynamique sur les données certaines seulement, avec des variables environnementales expliquant les variations des paramètres du modèle.

2.b.2. Les verrous méthodologiques

Comme nous l'avons vu précédemment, les méthodes utilisées pour suivre le lynx sont également menées à large échelle avec des observateurs venant de différentes catégories socio-professionnelles qui collectent des données indirectes. Or, dans ces données indirectes, il est possible que des erreurs d'identification du lynx se produisent. Pour pallier ce problème, un processus de filtrage a été mis au point afin de rejeter les détections qui ne sont évidemment pas du lynx, et garder les données certaines uniquement (les C1 et C2). Or, jusqu'à maintenant, la plupart des analyses de répartition des grands carnivores en Europe ne se font que sur ces données certaines (voir par exemple Molinari-Jobin et al. 2018), les données incertaines (les C3) étant écartées des analyses. Si les données incertaines contiennent des faux-positifs, elles contiennent également de vraies détections. Cela signifie donc qu'une partie des observations est mise de côté bien qu'elles puissent contenir des informations écologiques. Ici, je me suis attachée à développer un modèle d'occupancy dynamique prenant en compte les faux positifs dans les données incertaines.

2.b.3. Les données

Comme pour les données de suivi du loup en France, nous avons utilisé les données d'hiver, cette fois-ci par périodes de 2 mois (Novembre-Décembre, Janvier-Février et Mars-Avril). Afin de générer des détections et des non-détections, nous avons ramené les données de détection à l'échelle d'un site de 10 x 10 km (Figure 2). De même que pour le loup, si plusieurs signes de présence sont trouvés au sein d'un site lors d'un mois d'échantillonnage, alors nous comptabilisons une seule détection à l'échelle du site et du mois. Afin de générer les non-détections, nous avons considéré que l'effort n'était jamais nul durant toute la durée du suivi et sur toute l'aire d'étude (Molinari-Jobin et al. 2018). Nous avons donc considéré les sites n'ayant pas de détections comme des sites visités sur lesquels l'espèce n'avait pas été détectée. L'effort d'échantillonnage a été distingué en trois catégories : 0 signifie que l'effort est faible mais que gardes-chasse et gardes forestiers sont présents et rapportent des observations, 1 signifie que les agents du réseau sont présents et 2 signifie que des experts entraînés sont présents et cherchent activement des signes de présence du lynx.

Le nombre de détections du lynx dans les pays alpins a augmenté, avec la présence de données incertaines dès 1995. Le nombre de détections certaines est passé de 14 détections lors de la période Novembre-Décembre de l'année 1995, à 68 détections certaines en 2013 (Tableau 3). Les détections classées incertaines restent aussi en minorité, leur nombre varie de 8 par période de 2 mois à 33 lors de la période Mars-Avril de l'année 2005. A nouveau, nous

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

soulignons le fait qu'une augmentation des détections ne signifie pas nécessairement une augmentation de l'aire de répartition du lynx ou du nombre d'individus, cela peut aussi refléter une augmentation de la couverture spatiale du suivi et de l'effort de l'échantillonnage. Les signes de présence incertains apparaissent aussi régulièrement en dehors des cœurs de la distribution. Ces cœurs se trouvent en Suisse et en Slovénie (Molinari-Jobin et al. 2018). Nous pouvons aussi noter l'absence de détections certaines du lynx en Allemagne et seulement quelques détections incertaines dans le pays entre 2008 et 2013. Au total, de 1994 à 2013, 2188 observations (68 %) ont été classées comme étant des détections certaines et 1042 (32 %) comme étant incertaines.

Tableau 3 : Nombre de sites avec des détections certaines et incertaines au cours des mois d'échantillonnage du suivi du lynx dans les pays alpins entre 1995 et 2013.

Année	Novembre-Décembre (année n-1)		Janvier-Février (année n)		Mars-Avril (année n)	
	Certaines	Incertaines	Certaines	Incertaines	Certaines	Incertaines
1995	14	8	17	8	17	15
1996	18	17	35	14	27	19
1997	31	12	25	17	28	12
1998	30	17	18	17	35	19
1999	32	15	31	23	37	18
2000	21	32	35	20	42	18
2001	32	19	34	26	33	23
2002	28	20	28	22	43	16
2003	37	26	42	27	51	19
2004	32	26	36	24	52	23
2005	28	20	42	21	39	33
2006	32	17	38	17	28	32
2007	43	17	39	14	37	25
2008	47	21	50	17	50	16
2009	41	19	49	18	59	23
2010	37	14	49	10	52	14
2011	34	13	52	6	48	12
2012	36	13	56	11	71	12
2013	58	19	68	11	64	25

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

1995

1996

1997

1998

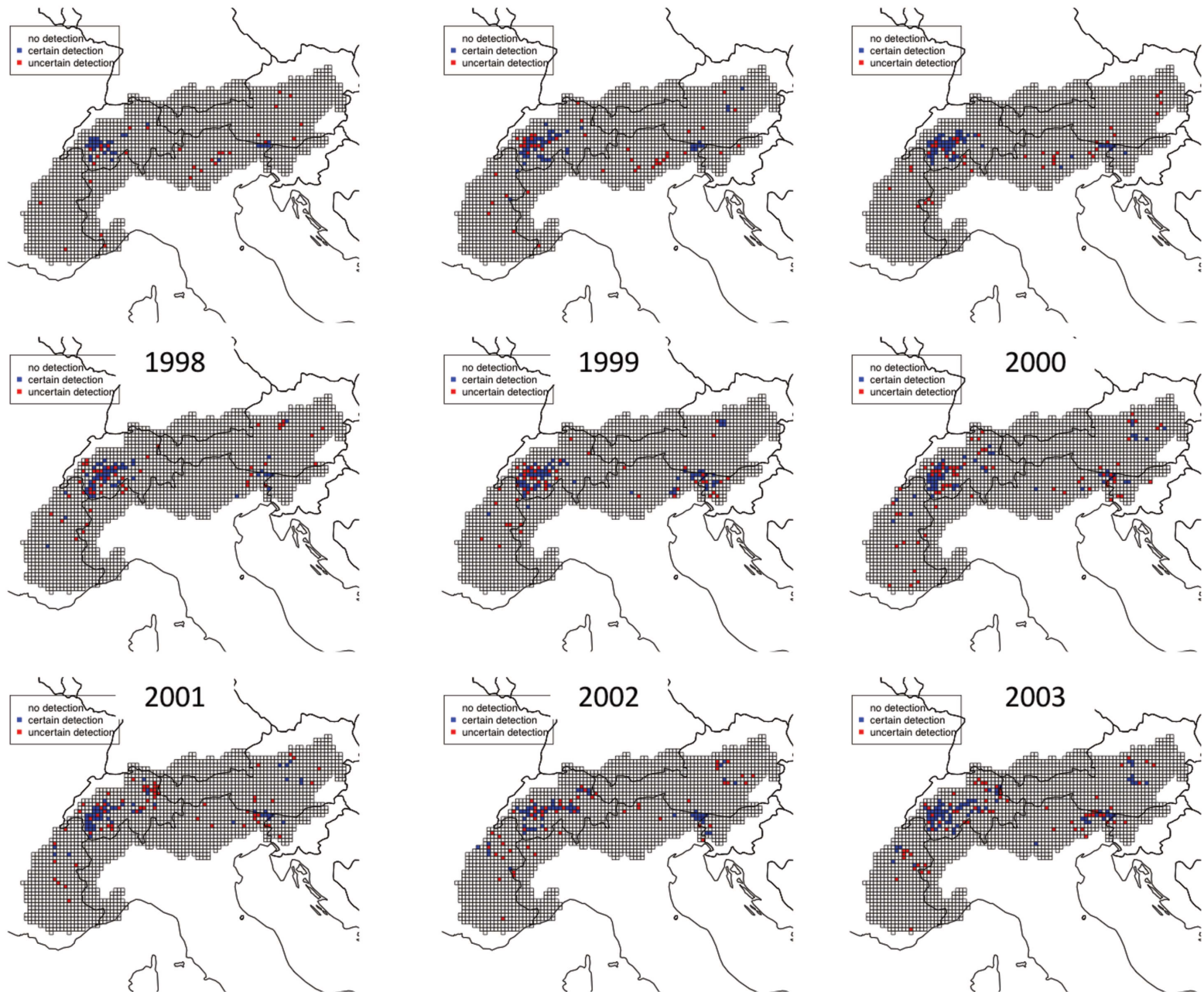
1999

2000

2001

2002

2003



Les cas du loup en France et du lynx dans les pays alpins et en Norvège

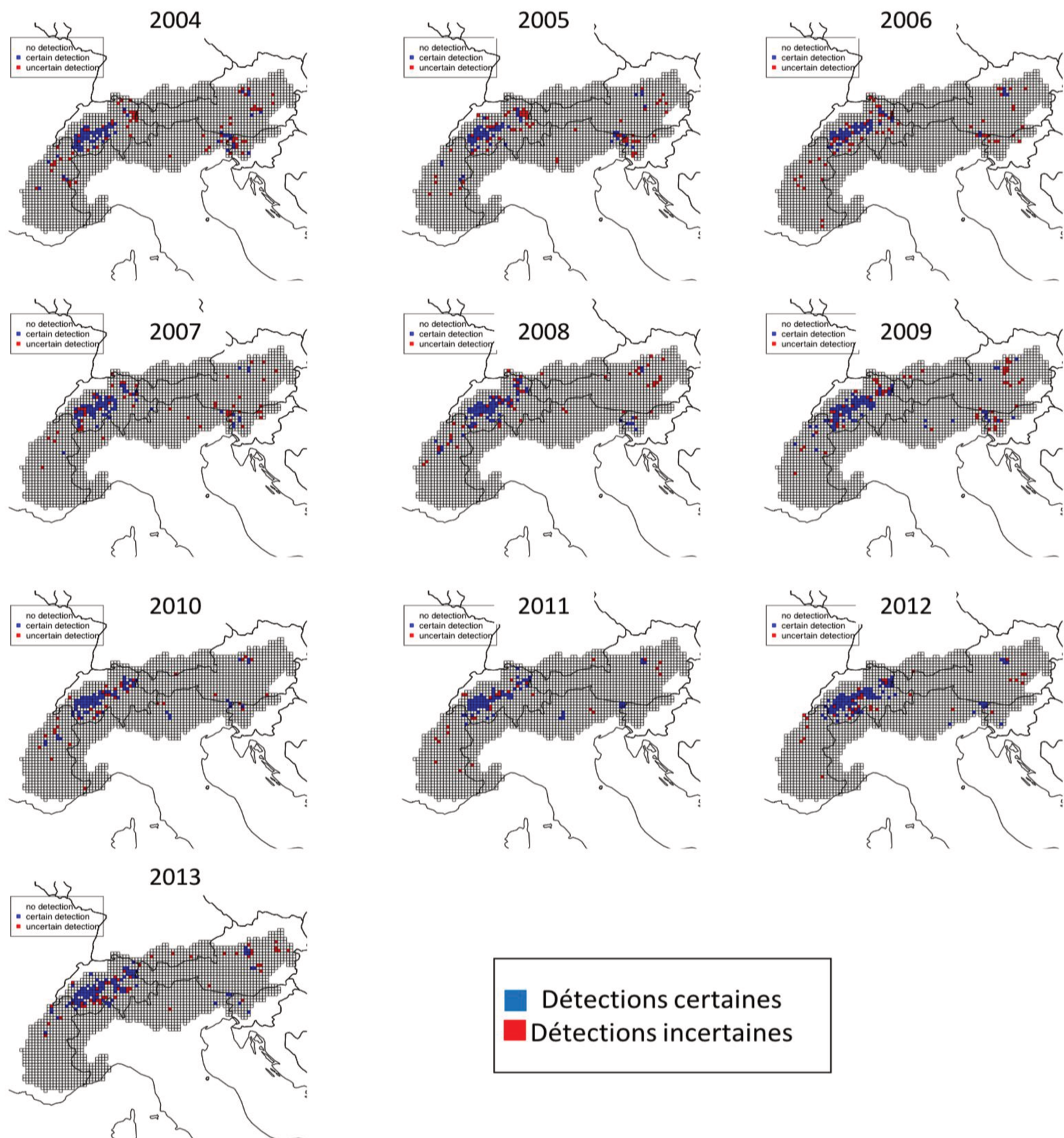


Figure 2: Cartes des détections certaines (bleu) et incertaines (rouge) du lynx dans les pays alpins entre 1995 et 2013. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage.

2.c. Cas du lynx en Norvège

2.c.1. Contexte du suivi

A la différence des deux cas précédents, le lynx en Norvège n'a jamais disparu du pays, cependant sa population a fortement diminué lors du siècle précédent due à une chasse non contrôlée (Linnell et al. 2010). De nos jours, le lynx en Norvège a un statut d'espèce chassée. La chasse est réalisée en partie pour limiter les dégâts sur les troupeaux et pour le loisir (Pedersen et al. 1999; Odden et al. 2002). L'humain exerce donc une forte pression sur les populations de lynx. Afin de préserver une population stable et une chasse durable, il est nécessaire d'estimer précisément les tendances de population afin de pouvoir fixer les quotas annuels de chasse.

La principale méthode de suivi du lynx en Norvège consiste à compter les groupes familiaux, définis comme une femelle adulte avec au moins un petit de l'année. Ce comptage est basé sur l'observation de traces de plus de deux individus dans la neige (Andrén et al. 2002; Linnell et al. 2007). Les petits restent auprès de leurs mères jusqu'à l'âge de 10 mois et la période de reproduction commençant fin mars, les mâles et les femelles vivent donc séparés avant cette date. Les traces de deux individus durant l'hiver correspondent à celles d'un groupe familial et sont signalées par les chasseurs, les garde-chasses ainsi que le public entre Octobre et Février. Les observations sont vérifiées par des agents d'état qui suivent alors les traces sur plus d'1 km. Toutes les traces vérifiées sont ensuite enregistrées dans une base de données nationale (ROVBASEN 3.0: www.rovbase.no) avec la date de l'observation ainsi que sa localisation. Dans ce jeu de données, nous n'avons donc pas eu accès aux données incertaines.

En plus du comptage de groupes familiaux, la mortalité de la population de lynx est étudiée via les individus chassés. Chaque lynx chassé est enregistré par le chasseur afin d'éviter de ne chasser plus d'individus que ce qui est autorisé. Tous les individus chassés sont enregistrés dans une base de données nationale, avec la date de la mort, la localisation, le sexe de l'individu et d'autres informations biologiques sur son état de santé général. La mortalité du lynx en Norvège est principalement due à l'humain, avec environ 90% de la mortalité reportée en Scandinavie due à la chasse (43 %) et au braconnage (47%) (Basille et al. 2013). Ces chiffres révèlent l'influence de l'humain comme principal facteur limitant la croissance de la

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

population de lynx. Chaque année, la saison de la chasse se déroule de début février jusqu'à fin mars (Bischof et al. 2012). Les quotas sont définis pour 8 régions du pays en fonction des estimations de population réalisées à partir du comptage des traces dans la neige.

Trois grands types d'impacts sont associés avec la présence du lynx en Norvège (Linnell et al. 2010). Tout d'abord, les moutons domestiques pâturent dans les forêts et les montagnes de tout le pays durant l'été. La plupart des troupeaux sont en pâturage libre, sans clôture ni supervision ou gardiennage. Ce pâturage libre mène donc à des attaques fréquentes des troupeaux par le lynx (Odden et al. 2002; Zimmermann et al. 2011). Le taux d'ovins chassés par le lynx est corrélé à la taille de la population locale de lynx (Herfindal et al. 2005). Les rennes semi-domestiques sont élevés par les Samis dans le centre et le nord de la Norvège. Les rennes sont également très vulnérables à la déprédation par le lynx (Sunde et al. 2000). Troisièmement, les chasseurs de chevreuils perçoivent le lynx comme un concurrent majeur pour la chasse de proies communes (Nilsen et al. 2009). Par conséquent, la gestion du lynx est considérée par les politiques publiques, les gestionnaires et le public comme un processus visant à trouver un équilibre entre la nécessité de maintenir des populations viables de lynx et la nécessité d'atténuer les impacts sur les élevages domestiques et semi-domestiques.

A ma connaissance, aucune étude n'avait encore été réalisée afin d'établir la répartition spatiale du lynx en Norvège à l'échelle du pays et son évolution au cours du temps. Des études locales ont été réalisées afin d'évaluer à fine échelle l'utilisation de l'espace (Bouyer et al. 2015), les facteurs influençant les paramètres démographiques (Basille et al. 2009) et de déterminer les comportements de chasse afin de comprendre les facteurs favorisant les attaques sur les proies domestiques, semi-domestiques et sauvages (Mattisson et al. 2011). Les données à larges échelles telles que le suivi des groupes familiaux et les suivis des lynx chassés annuellement étaient principalement utilisées afin d'estimer les effectifs annuels (Linnell et al. 2007, 2010). La modélisation de la répartition du lynx à l'échelle du pays apporte de nouvelles informations complétant les études démographiques. En prenant en compte le biais observateur et la détectabilité imparfaite de l'espèce nous avons estimé les sites où l'espèce était présente mais pourtant n'a pas été détectée. De plus, le lynx étant une espèce territoriale, il est possible d'évaluer la corrélation entre les estimations d'effectifs de population et le nombre de sites où l'espèce est présente et d'évaluer le lien entre la répartition de l'espèce et

les effectifs de population (Linden et al. 2017). Enfin, établir la répartition spatiale de l'espèce peut permettre d'identifier des populations isolées qui sont à risque d'extinction car séparées des autres populations.

2.c.2. Les verrous méthodologiques

Ici, nous pouvons voir que les données de groupes familiaux peuvent être transformées de la même manière que les données de suivi du loup en France et du lynx dans les pays alpins. En effet, grâce à la large couverture spatiale et temporelle que représente ce suivi au cours de l'hiver, et avec l'appui d'experts, il nous a été possible d'estimer que l'effort d'échantillonnage était non-nul partout. Ceci nous a permis de transformer les données de groupes familiaux en données de PO aisément. Cependant, il n'est pas possible de faire la même chose pour les données de chasse. Les données de chasse doivent donc être analysées comme des données de PB. Des efforts ont cependant été déployés pour élaborer des SDMs combinant à la fois des données de PB et des données de PO, ce qui permet d'avoir de meilleures estimations de la répartition que si les données étaient analysées séparément (Hefley & Hooten 2016). Ici, je me suis donc attachée tel un SDM afin de pouvoir combiner les deux types de données ensemble et comparer les informations apportées par chaque type de données.

2.c.3. Les données

Afin de générer les données de détections-non détections nécessaires pour ajuster un modèle de site occupancy, nous avons utilisé les données par mois d'hiver (c'est-à-dire d'octobre à février) que nous avons aussi ramenées à l'échelle d'un site de 10 x 10 km. Comme expliqué précédemment, après avis d'experts, nous avons considéré que l'effort d'échantillonnage était constant dans le temps et l'espace du fait de la large participation du public au suivi du lynx.

Octobre est le mois comprenant le moins de détections faites au cours des années avec un nombre de détections variant entre 0 en 1997, 2001 et 2005 et 10 détections en 2003 (Tableau 4). Janvier, à l'opposé, représente le mois avec le plus de sites sur lesquels l'espèce a été détectée avec 89 détections faites en 2008 (Table 3). Ces différences sont reliées aux

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

différences de chutes de neige qui ont lieu en Norvège entre le mois d'octobre, mois où l'épaisseur de la neige est la plus faible, et Janvier, mois au cours duquel l'épaisseur est à son maximum.

Nous avons utilisé ces données aussi comme des signes de présence du lynx mais nous les avons utilisées comme telles pour illustrer l'intérêt de combiner des données de différente nature, c'est-à-dire comme des données de type PB. Le nombre d'individus chassés a donc varié de 43 individus chassés en 2004 à 149 individus chassés en 2010 (Tableau 3). En observant les cartes de détections des groupes familiaux ainsi que des individus chassés, nous pouvons noter la présence du lynx notamment dans le Sud-Est du pays ainsi que dans le centre de la Norvège alors que le lynx est absent de la partie Sud-Ouest du pays, (Figures 3 et 4). Ceci reflète la décision prise par le gouvernement norvégien d'éliminer le lynx de régions où les conflits étaient trop nombreux, au cours des années 1990 (Linnell et al. 2007).

D'autres types de suivi se sont ajoutés afin de répondre à différentes problématiques de gestion et de recherche. Des comptages le long de transects ont été mis en place afin de compléter les estimations d'effectifs faites à partir des comptages de groupes familiaux (Linnell et al. 2007). Les mouvements de certains individus ont été enregistrés à l'aide de collier GPS afin de fournir de plus amples informations sur les mouvements qui se déroulent au sein d'un territoire ainsi que les comportements de chasse (Mattisson et al. 2011; Gervasi et al. 2014; Bouyer et al. 2015). Cependant, faisant face à des contraintes logistiques et financières, maintenir simultanément plusieurs suivis devient de plus en plus difficile pour le gouvernement norvégien. Il y a donc un besoin d'optimiser les suivis afin de concentrer les efforts sur le suivi qui apportera le plus d'informations sur la distribution du lynx. Je me suis intéressée à cette question en me focalisant sur les données des groupes familiaux et des individus prélevés à la chasse.

Tableau 4 : Nombre de sites avec des détections de groupes familiaux de lynx en Norvège au cours des mois d'échantillonnage ainsi que le nombre de lynx chassés entre 1997 et 2016.

	Groupes familiaux					Lynx chassés
	Octobre	Novembre	Decembre	Janvier	Février	
1997	0	0	38	0	0	93
1998	9	27	29	58	14	122
1999	8	32	25	82	14	97
2000	5	14	14	43	28	109
2001	0	6	14	39	24	86
2002	3	15	13	40	30	95
2003	10	12	16	23	17	70
2004	2	8	22	36	16	43
2005	0	17	43	54	24	50
2006	3	7	10	67	30	58
2007	6	20	44	70	55	77
2008	3	22	56	89	58	100
2009	7	18	52	67	52	119
2010	8	9	52	79	52	149
2011	7	51	21	50	39	145
2012	1	4	19	61	31	85
2013	9	10	36	40	40	86
2014	6	11	39	41	31	72
2015	6	16	19	50	47	80
2016	4	18	0	48	44	70

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

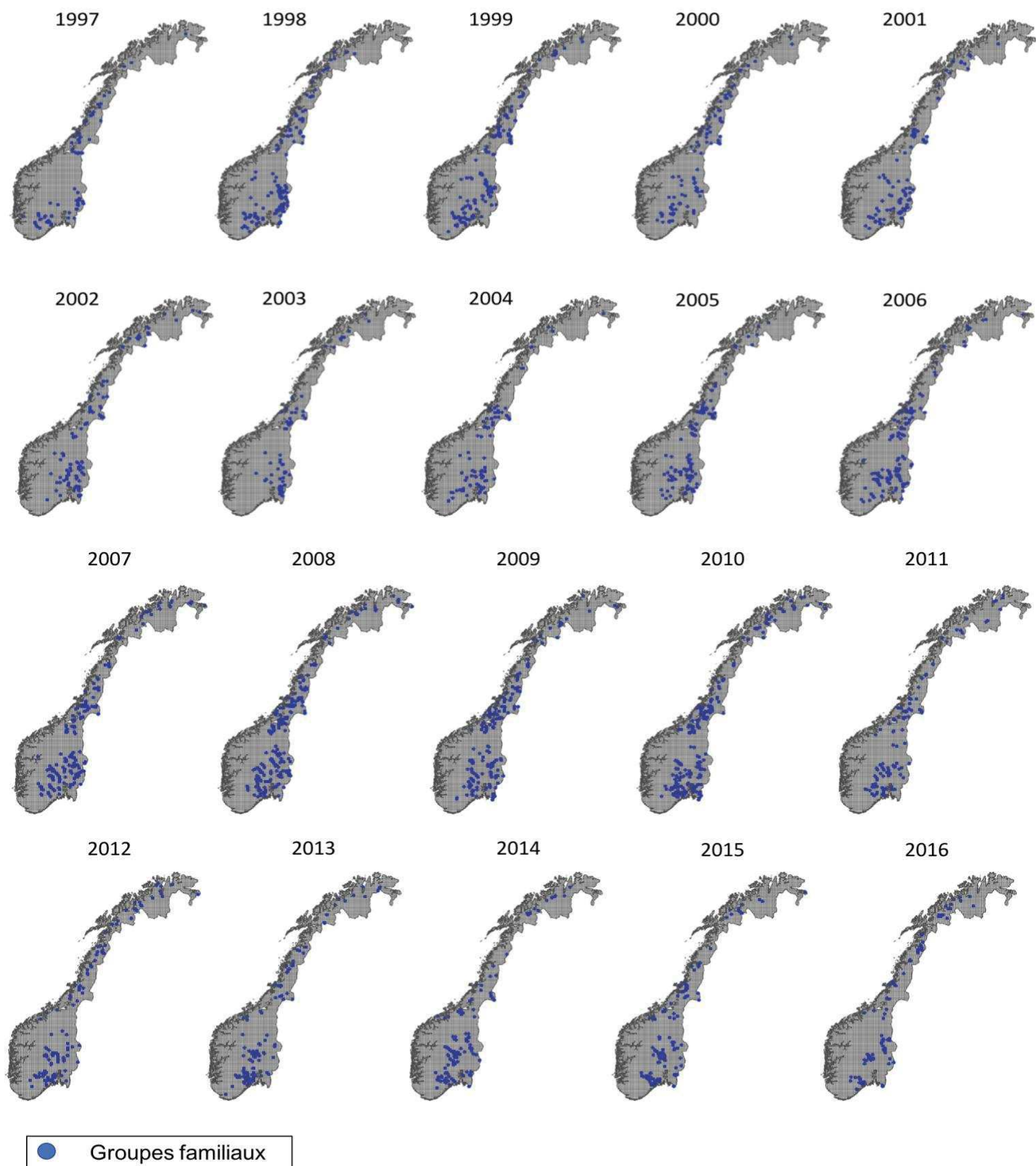


Figure 3 : Cartes des détections de groupes familiaux du lynx en Norvège entre 1997 et 2016. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage.

Les cas du loup en France et du lynx dans les pays alpins et en Norvège

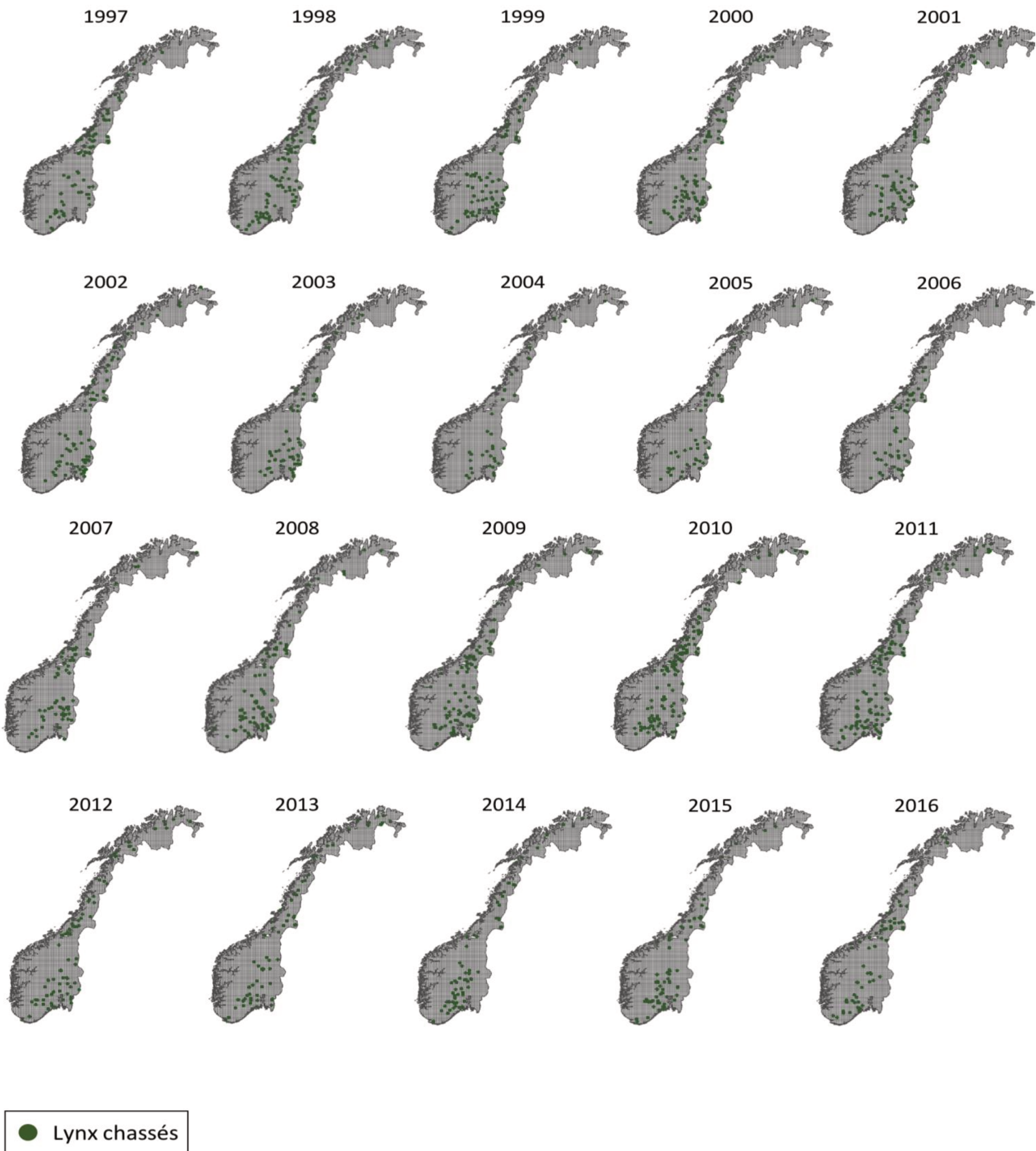


Figure 4 : Cartes des lynx chassés en Norvège entre 1997 et 2016. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage. Ici, les détections n'ont pas été ramenées à l'échelle d'un site de 10x10km mais les variables environnementales ont été calculées à l'échelle d'un site de 10x10 km

3. Les modèles d'occupancy pour quantifier la distribution d'une espèce

En écologie et en biologie de la conservation, comprendre la façon dont les espèces et les communautés sont réparties dans l'espace et comment cette répartition spatiale évolue dans le temps, est crucial. Afin de répondre à ces questions, les modèles d'occupancy ont été développés par MacKenzie et al. (2002). L'occupancy est définie comme la proportion d'une zone ou le nombre de sites occupés par l'espèce (MacKenzie et al. 2003; MacKenzie 2006).

Ces modèles prennent en compte explicitement le fait qu'une espèce peut être présente sur un site et peut ne pas avoir été observée due à une faible probabilité de détection, conséquence d'un comportement cryptique de l'espèce, des caractéristiques de l'habitat ou des capacités d'observation des observateurs (Kéry et al. 2009 ; Kéry 2011). Ils permettent donc d'établir le lien entre la présence d'une espèce et l'environnement dans le but d'établir la proportion d'une zone d'étude qui est occupée par une ou plusieurs espèces, tout en prenant en compte une détectabilité imparfaite. Nous nous plaçons donc à l'échelle d'un « site », dont la taille est définie en fonction de l'espèce étudiée (Steenweg et al. 2018). Le processus d'échantillonnage consiste à visiter plusieurs sites, dont la taille varie en fonction de l'espèce étudiée, et répéter ces visites dans un laps de temps défini de telle sorte que l'état du site (occupé ou non occupé) reste inchangé. Chaque visite répétée est appelée une « occasion secondaire ». Ces répétitions permettent, lorsque l'espèce est faiblement détectable, de répéter les chances de pouvoir la détecter.

Chaque site possède alors un historique de détection, c'est-à-dire le nombre d'occasions secondaires où l'espèce a été détectée. C'est à partir de cet historique de détection que le modèle d'occupancy fonctionne. L'approche nous permet aussi d'identifier les facteurs influençant la présence de l'espèce à un site ainsi que les facteurs influençant sa détection, ces deux types de facteurs environnementaux sont souvent différents car résultant de deux processus différents : le processus biologique gouvernant la présence de l'espèce et le processus de détection gouvernant l'observation de l'espèce sous condition que l'espèce soit présente.

3.a. Le modèle d'occupancy statique

Le modèle d'occupancy statique repose sur deux processus stochastiques. Le premier processus est le processus gouvernant l'état d'un site, c'est-à-dire s'il est occupé ou non. Nous définissons ici z_i , la variable qui décrit l'état latent écologique du site i . Cet état latent est partiellement observable, seulement lorsque l'espèce est détectée. Cette variable prend la valeur 1 si le site i est occupé et 0 s'il n'est pas occupé. Un site i est donc occupé par l'espèce avec la probabilité $P(z_i = 1) = \Psi_i$ et il est inoccupé avec la probabilité complémentaire ; $P(z_i = 0) = 1 - \Psi_i$. L'état latent z_i est le résultat d'un tirage dans une loi de Bernoulli dont la probabilité de succès est Ψ_i .

Le second processus gouverne l'observation de l'état du site, c'est-à-dire si l'espèce a été détectée ou pas détectée, conditionnel à sa présence. Nous définissons ici $y_{i,j}$, la variable qui décrit l'état observé du site i lors de l'occasion secondaire j . Cette variable prend la valeur 1 si l'espèce a été détectée au site i à l'occasion secondaire j et 0 si l'espèce n'a pas été détectée. Si l'espèce est présente à un site i , elle sera détectée durant l'occasion secondaire j avec la probabilité $P(y_{i,j} = 1 | z = 1) = p_{i,j}$, à l'inverse elle pourra ne pas être détectée avec la

Les modèles d'occupancy pour quantifier la distribution d'une espèce

probabilité complémentaire ; $P(y_{i,j} = 0 \mid z_i = 1) = 1 - p_{i,j}$. L'état observé $y_{i,j}$ est donc le résultat d'un tirage dans une loi de Bernoulli dont la probabilité de succès est $p_{i,j}$. Nous pouvons désormais relier les historiques de détections à la probabilité de l'évènement menant à ces historiques. Imaginons qu'un site i soit visité lors de trois occasions secondaires et l'historique de détection obtenu est le suivant : $H_i = '010'$, signifiant que l'espèce n'a pas été détectée lors de la première occasion secondaire, mais a été détectée lors de la deuxième occasion secondaire et à nouveau n'a pas été détectée lors de la troisième occasion secondaire. Nous sommes sûr que le site est occupé puisque l'espèce a été détectée lors de la deuxième occasion secondaire, nous pouvons donc traduire cet historique de détection en une probabilité comme suit :

$$P(H_i = '010') = P(z_i = 1)P(y_{i,1} = 0 \mid z_i = 1)P(y_{i,2} = 1 \mid z_i = 1)P(y_{i,3} = 0 \mid z_i = 1) = \Psi_i(1 - p_{i1})p_{i2}(1 - p_{i3})$$

Maintenant, considérons que l'espèce n'a jamais été détectée au cours des trois occasions secondaires. Nous avons donc un historique de détection comme suit : $H_i = '000'$. Nous ne savons pas si l'espèce est présente ou non. Soit l'espèce est présente et dans ce cas-là elle n'a pas été détectée au cours des trois occasions secondaires, soit l'espèce est absente et donc l'espèce est forcément non détectée. En traduisant cet historique de détection nous obtenons donc :

$$\begin{aligned} P(H_i = '000') &= [P(z_i = 1)P(y_{i,1} = 0 \mid z_i = 1)P(y_{i,2} = 0 \mid z_i = 1)P(y_{i,3} = 0 \mid z_i = 1)] + \\ &\quad [P(z_i = 0)] \\ &= \Psi_i(1 - p_{i1})(1 - p_{i2})(1 - p_{i3}) + (1 - \Psi_i) \end{aligned}$$

Pour chaque site, il est donc possible de traduire l'historique de détections en termes de probabilités et donc de définir une vraisemblance générale pour les données que nous avons avec j occasions secondaires et N sites comme suit :

$$\begin{aligned} \mathcal{L}(\Psi, p_1, p_2, \dots, p_j \mid h_1, h_2, \dots, h_N) &= \prod_{i=1}^N P(H_i) = \\ &= \left[\prod_{i=1}^m \Psi_i \prod_{j=1}^J p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{1 - y_{i,j}} \right] \left[\prod_{i=m+1}^N (\Psi_i \prod_{j=1}^J (1 - p_{i,j}) + (1 - \Psi_i)) \right] \end{aligned}$$

où m est le nombre de sites où l'espèce a été détectée au moins une fois et N le nombre total de sites.

Enfin, dans les analyses d'occupancy, nous sommes intéressés à expliquer les variations de la probabilité d'occupancy dans l'espace. Pour ce faire, nous avons recours à des variables environnementales. Afin de lier ces facteurs environnementaux aux paramètres du modèle, la fonction logistique est communément utilisée car elle permet de borner le paramètre d'occupancy entre 0 et 1 :

$$\text{logit}(\Psi_i) = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_C x_{i,C}$$

où $x_{i,1}, \dots, x_{i,C}$ représentent les valeurs des C facteurs environnementaux pour le site i et β_1, \dots, β_C les effets de ces facteurs environnementaux.

3.b. Le modèle d'occupancy dynamique

Les modèles que nous avons décrits précédemment permettent d'établir la distribution d'une espèce à un temps donné. Cependant dans des cas de recolonisation, d'invasions biologiques, ou de réduction d'une population, la distribution est dynamique au cours du temps, c'est-à-dire qu'elle peut changer entre le temps t et le temps $t+1$. Il est possible d'étendre les modèles précédemment décrits à une version dynamique, liant les états latents entre les temps t et $t+1$, autrement appelés les « saisons », ou occasions primaires. Ces modèles permettent de décrire et d'expliquer les processus influençant la dynamique de distribution et permettent d'identifier les paramètres environnementaux favorisant la présence d'une espèce (MacKenzie et al. 2003). Les occasions primaires sont donc choisies en fonction de la biologie de l'espèce, souvent entre les pics de migrations, de naissances ou de mortalité.

Pour tout processus dynamique, il faut définir l'état initial à partir duquel le processus va évoluer. Nous définissons donc $z_{i,1}$ comme l'état latent initial, résultant d'un tirage dans une loi de Bernoulli dont la probabilité de succès est $\Psi_{i,1}$.

Nous définissons à nouveau $z_{i,t}$, la variable qui décrit l'état latent écologique du site i mais qui est cette fois-ci lors de la saison t . Cette variable prend donc la valeur 1 si le site i est occupé lors de la saison t et 0 s'il n'est pas occupé. Pour chaque site i et chaque saison t , la probabilité $\Psi_{i,t}$ est définie et dépend de la probabilité d'occupancy de la saison précédente au site i ; $\Psi_{i,t-1}$. Nous définissons maintenant les paramètres permettant de faire le lien entre les saisons. Si un site i est inoccupé lors de la saison t il peut devenir occupé lors de la saison $t+1$ avec la probabilité de colonisation $P(z_{i,t+1} = 1 \mid z_{i,t} = 0) = \gamma_{i,t}$ et peut rester inoccupé à la saison $t+1$ avec la probabilité complémentaire $P(z_{i,t+1} = 0 \mid z_{i,t} = 0) = 1 - \gamma_{i,t}$. A l'inverse, si un site i est occupé lors de la saison t il peut devenir inoccupé lors de la saison $t+1$ avec la probabilité d'extinction $P(z_{i,t+1} = 0 \mid z_{i,t} = 1) = \varepsilon_{i,t}$ et peut rester inoccupé à la saison $t+1$ avec la probabilité complémentaire $P(z_{i,t+1} = 1 \mid z_{i,t} = 1) = 1 - \varepsilon_{i,t}$. Un site i est occupé lors de la saison $t+1$ soit car il était déjà occupé précédemment et l'espèce est restée, soit car il n'était pas occupé la saison précédente et a été colonisé par l'espèce. En termes de probabilités cela se traduit sous la forme suivante :

$$\begin{aligned} \Psi_{i,t+1} &= P(z_{i,t} = 1)P(z_{i,t+1} = 1 \mid z_{i,t} = 1) + P(z_{i,t} = 0)P(z_{i,t+1} = 1 \mid z_{i,t} = 0) \\ &= \Psi_{i,t} (1 - \varepsilon_{i,t}) + (1 - \Psi_{i,t}) \gamma_{i,t}. \end{aligned}$$

De la même manière, la vraisemblance peut être définie pour chaque historique de détection des sites. Imaginons cette fois-ci qu'un site i est visité lors de trois occasions secondaires au cours de 2 saisons et que l'historique de détection obtenu est le suivant : $H_i = '010\ 000'$, signifiant que l'espèce n'a pas été détectée lors de la première occasion secondaire, mais a été détectée lors de la deuxième occasion secondaire et à nouveau n'a pas été détectée lors de la troisième occasion secondaire et n'a jamais été détectée lors de la deuxième saison. Deux cas sont donc possibles : soit l'espèce n'a été détectée que lors de la saison 1 puis s'est éteinte du site lors de la saison 2, soit l'espèce était présente lors de la saison 2 mais n'a pas été détectée.

A nouveau, conditionnel à l'état latent d'un site, le processus d'observation permet de lier l'état latent écologique aux données observées. Dans ce cas-là, $y_{i,j,t}$ est donc la variable qui décrit l'état observé du site i lors de l'occasion secondaire j au sein de la saison t . L'état observé $y_{i,j,t}$, est donc le résultat d'un tirage dans une loi de Bernoulli dont la probabilité de succès est $p_{i,j,t}$. Comme décrit précédemment, chaque historique de détection peut alors être traduit sous forme de probabilités et il est donc à nouveau possible de définir la vraisemblance du modèle en fonction des données obtenues.

Les modèles statiques peuvent donc être vus comme un cas particulier des modèles dynamiques d'occupancy où les probabilités de colonisation et d'extinction sont nulles. La description que nous avons utilisée des modèles dynamiques d'occupancy suit une description dite à « espace d'états » où le modèle est divisé en deux composantes, une première décrivant le processus d'occupancy partiellement observable et l'autre décrivant les observations conditionnelles à ce processus (Royle & Kéry 2007). Au cours de ma thèse j'ai aussi utilisé une formulation en modèle de Markov caché des modèles d'occupancy (Gimenez et al. 2014), un cas particulier des modèles à espace d'états où l'on suppose que la dynamique des états latents est markovienne, c'est-à-dire que l'état à $t+1$ est dépendant de l'état au temps précédent t (Gimenez et al. 2012).

3.c. Questionnements sur les hypothèses des modèles d'occupancy et leur application aux grands carnivores

Les modèles d'occupancy reposent sur plusieurs hypothèses telles que l'indépendance des observations, la fermeture de la population ou encore l'absence de faux positifs (MacKenzie et al. 2003; MacKenzie 2006). Or ces hypothèses sont parfois difficiles à respecter dans le contexte des grands carnivores en Europe. Nous détaillons ici les hypothèses sous-jacentes des modèles d'occupancy en soulignant les limites et les verrous méthodologiques décrits aux paragraphes 2.a.2, 2.b.2, et 2.c.2.

Tout d'abord, ces modèles ne prennent pas en compte les cas de faux positif. Les estimations du modèle sont basées sur des observations dites certaines. Si cette hypothèse n'est pas respectée, ces observations erronées vont donc rajouter du bruit aux données de détection de l'espèce et donc biaiser les estimations de distribution (Miller et al. 2011). Il faut donc s'assurer d'un système de filtrage des données comme c'est le cas du suivi du loup en France et du lynx dans les pays alpins où le processus de filtrage permet d'éviter l'inclusion de faux-positifs. Cependant cette méthode de filtrage, si elle permet de convenir aux hypothèses des modèles d'occupancy, fait perdre une partie de l'information qui pourrait être utile par ailleurs. L'article 2 de ma thèse est axé sur ce point.

Deuxièmement, les historiques de détection des sites échantillonnés sont supposés être indépendants. Cela signifie que l'historique de détection d'un site ne doit pas dépendre de l'historique de détection d'un autre site. Si les historiques de détection ne sont pas indépendants, cela signifie qu'il y a de l'autocorrélation dans les détections due soit à la présence de l'espèce dans les sites voisins, soit à la présence d'observateurs sur plusieurs sites à la fois. Si l'espèce est présente sur plusieurs sites, cela signifie que l'occupancy pourra être surestimée (MacKenzie et al. 2002). Or, dans le cas d'une espèce mobile telle que le loup ou le lynx, la colonisation de nouveaux sites se fait via des individus qui vont disperser depuis leur lieu de naissance jusqu'à un nouveau site voisin à coloniser. Il est donc possible qu'il y ait de

Les modèles d'occupancy pour quantifier la distribution d'une espèce

l'autocorrélation dans les détections due aux individus qui dispersent. L'article 1 de ma thèse est axé en partie sur ce point.

Troisièmement, une fois les variables environnementales intégrées, il ne doit pas rester d'hétérogénéité non modélisée dans les paramètres du modèle. Cela signifie que la variation des paramètres doit pouvoir être expliquée par la variation des facteurs environnementaux sélectionnés. Ne pas prendre en compte cette hétérogénéité résiduelle peut mener à sous-estimer l'occupancy réelle si par exemple un facteur influençant positivement l'occupancy n'est pas pris en compte (MacKenzie et al. 2002). Or, comme nous l'avons vu, il est fort probable que la détection reste hétérogène si l'on ne prend pas en compte la variation de l'effort d'échantillonnage comme dans le cadre du suivi du loup en France. L'article 3 de ma thèse est axé sur ce point.

Quatrièmement, le statut d'un site (occupé ou pas) ne doit pas changer entre les occasions secondaires, c'est-à-dire entre les visites au cours d'une année par exemple (Rota et al. 2009). Cette hypothèse est appelée hypothèse de fermeture. Les observations doivent donc se dérouler à un moment du cycle biologique de l'espèce où les mouvements et les effectifs de la population sont les plus stables, lorsqu'il n'y a pas de mouvements entrants ou sortants, ni naissance ou mort. Or dans le cas des grands carnivores, la taille de leur territoire peut varier et dépasser la taille du quadrillage choisi. Cette hypothèse sera abordée en discussion de ma thèse.

4. Organisation de la thèse

C'est à partir d'une réflexion sur ces hypothèses, et sur le risque de ne pas les respecter dans le cas des grands carnivores, que j'ai axé une première partie de ma thèse. Dans un deuxième temps j'ai travaillé sur d'autres problématiques liées aux grands carnivores notamment i) l'analyse combinée de plusieurs jeux de données dans le but d'avoir une estimation plus complète de la répartition et ii) le problème de prédire la répartition du loup en France dans un contexte de recolonisation.

4.a. Non-respect des hypothèses des modèles d'occupancy

Dans les articles 1, 2 et 3 de ce manuscrit, j'ai développé des méthodes afin de répondre aux problèmes liés au non-respect des hypothèses des modèles d'occupancy.

Dans l'article 1, nous démontrons que prendre en compte l'effort d'échantillonnage est crucial dans le contexte de l'utilisation d'un modèle d'occupancy dynamique ajusté sur les données de suivi du loup en France entre 1994 et 2016. Notre but était d'établir une mesure de l'effort d'échantillonnage variant dans le temps et l'espace, et une fois cette variable prise en compte, d'estimer les paramètres écologiques influençant la colonisation et l'extinction du loup en France. De plus, nous montrons l'importance de prendre en compte l'autocorrélation spatiale en incluant le nombre de sites voisins observés occupés comme ayant une influence sur la probabilité qu'un site soit colonisé.

Dans l'article 2, nous montrons qu'il est possible d'inclure les données incertaines dans les modèles d'occupancy en prenant en compte la détection de faux-positifs. Le but ici est de

Organisation de la thèse

construire un modèle d'occupancy prenant en compte le fait qu'il y a des vrais positifs dans les données incertaines, c'est-à-dire des détections véridiques de lynx mais qui n'ont pas pu être validées. Nous montrons qu'inclure les données incertaines apporte donc de l'information quant à la distribution du lynx dans les pays alpins, une information peu ou pas utilisée jusqu'alors.

Dans l'article 3, nous montrons que même s'il est difficile d'avoir une mesure de l'effort d'échantillonnage, il est possible d'ajuster un modèle d'occupancy sur des données incertaines et certaines tout en prenant en compte l'hétérogénéité dans le processus de détection. Ici, le but est de démontrer, d'une part que les modèles d'occupancy prenant en compte les faux positifs (article 2) peuvent être formulés comme des modèles de Markov cachés. D'autre part, grâce à cette formulation, nous montrons qu'il est possible de prendre en compte l'hétérogénéité de détection.

4.b. Gestion et suivi des populations de grands carnivores

Dans les articles 4 et 5, nous explorons d'autres problématiques liées cette fois-ci à la gestion et aux suivis du loup en France et du lynx en Norvège.

Dans l'article 4, nous combinons un modèle d'occupancy ajusté sur les données des groupes familiaux avec un modèle spécifique appelé modèle à processus ponctuel ajusté sur des données de PB que sont les données de lynx chassés en Norvège. Le but est de montrer l'avantage de la combinaison de plusieurs types de données qui signifient des choses différentes tant du point de vue méthodologique (PO contre PB) que du point de vue de la biologie de l'espèce (groupes familiaux contre individus isolés des deux sexes) et du suivi (détection opportuniste de traces dans la neige contre recherche active). Le deuxième but dans cette étude est d'identifier quel suivi apporte le plus d'information dans l'optique d'optimiser le suivi du lynx en Norvège.

Enfin, dans l'article 5 nous montrons qu'il est possible de prédire la distribution du loup sur une courte échelle de temps en modélisant la diffusion écologique du loup tout en prenant en compte la détection imparfaite de l'espèce. Nous sommes partis d'une des conclusions de l'article 1 qui était que la probabilité de coloniser un site était majoritairement influencée par le nombre de sites voisins observés occupés à courte distance. La dispersion est donc apparue comme un facteur majeur de la colonisation de nouveaux sites par le loup.

Enfin, en discussion, je reviens sur l'interprétation des résultats que j'ai obtenus tant d'un point de vue méthodologique que biologique avec l'interprétation des variables écologiques et finalement ce qu'impliquent mes résultats en termes de suivis pour les espèces concernées. Je clos le manuscrit en replaçant mes résultats dans un contexte de gestion adaptative.

5. Article 1. Cartographier et expliquer la recolonisation du loup en France à l'aide de modèles d'occupancy dynamique et de données opportunistes

Résumé

Alors que les grands carnivores recolonisent l'Europe, l'évaluation de leur répartition spatiale peut aider à prévoir et à atténuer les impacts sur les activités humaines. Sachant qu'ils sont très mobiles, élusifs et vivent à faible densité, la modélisation de leur répartition spatiale présente plusieurs défis en raison de 1) leur détectabilité imparfaite, 2) leur répartition spatiale dynamique dans le temps et 3) leur suivi à grande échelle constitué principalement de données opportunistes sans mesure formelle de l'effort d'échantillonnage. Ici, nous nous sommes concentrés sur le loup *Canis lupus* qui a commencé à recoloniser la France dès le début des années 1990. Nous avons défini et évalué l'effort d'échantillonnage *a posteriori* comme le nombre d'observateurs présents par an dans un site basé sur les informations que nous avons sur leur adresse et leur activité professionnelle. A l'aide de modèles d'occupancy dynamique, nous avons ensuite évalué la dynamique de la répartition spatiale des loups de 1994 à 2016, tout en prenant en compte la détection imparfaite du loup et l'effort d'échantillonnage variant dans le temps et l'espace. Ignorer l'effet de l'effort d'échantillonnage sur la probabilité de détecter le loup conduit à sous-estimer le nombre de sites occupés de plus de 50% en moyenne. La colonisation du loup semble influencée négativement par la proportion du site dont l'altitude est supérieure à 2500 m et positivement influencée par le nombre de sites observés occupés à courtes et longues distances, le couvert forestier, le couvert agricole et l'altitude moyenne. Le taux d'expansion, défini comme le nombre de sites occupés au cours d'une année divisé par le nombre de sites occupés au cours de l'année précédente, a diminué au cours des premières années de l'étude, puis s'est stabilisé de 2000 à 2016. Nos travaux montrent que les données opportunistes peuvent être analysées avec des modèles de distribution d'espèces qui prennent en compte la détection imparfaite, à condition d'être dans la capacité de quantifier l'effort d'échantillonnage. Notre approche a le potentiel d'être utilisée par les gestionnaires de la faune sauvage pour cibler les sites où les grands carnivores sont susceptibles d'être présents et atténuer les conflits (Voir Annexe 1 pour un résumé de l'article vulgarisé).

Contributions

J'ai développé le modèle d'occupancy, transformé les données, calculé les variables environnementales et construit la méthode afin d'estimer l'effort d'échantillonnage. J'ai proposé une interprétation des résultats obtenus que nous avons discutée collectivement. J'ai rédigé un premier jet de l'article qui a été corrigé par mes co-auteurs. Les données sur les observateurs et les détections du loup m'ont été fournies par l'ONCFS via Christophe Duchamp, qui m'a aidée aussi à cibler les variables environnementales pouvant avoir une influence sur les paramètres du modèle de d'occupancy dynamique. Eric Marboutin m'a aidée pour la relecture du manuscrit tout comme Sarah Cubaynes. Valentin Lauret a permis de mettre à jour les données observateurs et de calculer l'effort d'échantillonnage final. L'encadré No 2 résume le travail que Valentin a réalisé dans le cadre de son stage que j'ai encadré avec Olivier Gimenez. La méthode de sélection de variables a été proposée par Olivier Gimenez.



Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data

Julie Louvrier, Christophe Duchamp, Valentin Lauret, Eric Marboutin, Sarah Cubaynes, Rémi Choquet, Christian Miquel and Olivier Gimenez

J. Louvrier (<http://orcid.org/0000-0003-1252-1746>) (julie.louvrier@cefe.cnrs.fr), V. Lauret, S. Cubaynes, R. Choquet and O. Gimenez (<http://orcid.org/0000-0001-7001-5142>), CEFÉ UMR 5175, CNRS, Univ. de Montpellier, Univ. Paul-Valéry Montpellier, EPHE, Montpellier, France. – C. Duchamp, Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux déprédateurs, Parc Micropolis, Gap, France. – E. Marboutin, Office National de la Chasse et de la Faune Sauvage, ZI Mayencin, Gières, France. – C. Miquel, Univ. Grenoble 1, CNRS, Laboratoire Ecologie Alpine, UMR 5553, Grenoble, France.

While large carnivores are recovering in Europe, assessing their distributions can help to predict and mitigate conflicts with human activities. Because they are highly mobile, elusive and live at very low density, modeling their distributions presents several challenges due to 1) their imperfect detectability, 2) their dynamic ranges over time and 3) their monitoring at large scales consisting mainly of opportunistic data without a formal measure of the sampling effort.

Here, we focused on wolves *Canis lupus* that have been recolonizing France since the early 1990s. We evaluated the sampling effort a posteriori as the number of observers present per year in a cell based on their location and professional activities. We then assessed wolf range dynamics from 1994 to 2016, while accounting for species imperfect detection and time- and space-varying sampling effort using dynamic site-occupancy models.

Ignoring the effect of sampling effort on species detectability led to underestimating the number of occupied sites by more than 50% on average. Colonization appeared to be negatively influenced by the proportion of a site with an altitude higher than 2500 m and positively influenced by the number of observed occupied sites at short and long-distances, forest cover, farmland cover and mean altitude. The expansion rate, defined as the number of occupied sites in a given year divided by the number of occupied sites in the previous year, decreased over the first years of the study, then remained stable from 2000 to 2016. Our work shows that opportunistic data can be analyzed with species distribution models that control for imperfect detection, pending a quantification of sampling effort. Our approach has the potential for being used by decision-makers to target sites where large carnivores are likely to occur and mitigate conflicts.

Large carnivores are often considered as key elements for maintaining ecosystems. Because of their high position in the trophic chain, their extinction can lead to trophic cascades and detrimental changes in species abundance and functioning of ecosystems (Ripple et al. 2014). Once widespread in Europe, many populations of large carnivores were extirpated over the last century, mainly due to interferences with human activities (Breitenmoser 1998, Ripple et al. 2014). Since the 1970s, all large carnivores have recovered, resulting in most of the European countries hosting at least one viable population of large predators (Chapron et al. 2014). Often used as a conservation success story, the recovery of large carnivores in human-dominated areas comes with challenges, including the question of whether there are any sufficiently large and functional areas left for viable populations (Packer et al. 2013). Another issue is how to coordinate management of these species at large scales, possibly across borders (Linnell and Boitani 2012, Bischof et al. 2015), in particular in the context of international treaties and directives (e.g. the European Habitats Directive).

In this context, mapping the distribution of a species can help to target potential area of presence and mitigate conflicts often associated with the recovery of large carnivores. Species distribution models (SDMs) have become important tools in the ecological, biogeographical and conservation fields (Guisan and Thuiller 2005). By correlating presence-only or presence–absence data of a species to environmental factors, SDMs provide an understanding of habitat preferences and predictions on future species distribution. This is especially relevant for species involved in conflicts, since predicting their future presence can help targeting contentious areas and guide management to reduce conflicts (Guillera-Arroita et al. 2015). However, the monitoring of large carnivores remains challenging to carry out in the field because these species live at low density and occupy wide areas (Woodroffe 2001). Therefore, assessing the distribution of these species comes with methodological challenges.

First, standard SDMs such as Maxent (Phillips et al. 2006) rely on the assumption that the focal species is detected everywhere it is present (Yackulic et al. 2013).

Going undetected at a given site does not necessarily mean that this species is absent from that site, but rather that it may simply be missed for various reasons related to observer abilities, habitat characteristics or species level of activity (Kéry et al. 2010, Kéry 2011). Ignoring the issue of imperfect detection can result in false absences that lead to flawed inference in two ways: 1) the distribution maps are biased by underestimating actual presences (Kéry and Schaub 2011, Lahoz-Monfort et al. 2014); 2) there may be confusion in identifying the drivers of the species distribution when detection depends on environmental explanatory variables that are independent from the variables influencing the species' actual presence (Lahoz-Monfort et al. 2014). To cope with this first issue, single-season or static site-occupancy models were developed (Mackenzie et al. 2006) and have been widely used for carnivores (Long et al. 2010, Thorn et al. 2011, Sunarto et al. 2012). Based on spatial and temporal replicated sampling of the target species, these models allow making the distinction between non-detections and true absences via the estimation of species detectability.

Second, most SDMs are implicitly based on the ecological niche concept (Grinnell 1917, Hutchinson 1957) and therefore rely on two main hypotheses: 1) the species is present in areas where environmental conditions are the most favorable and 2) dispersal is not a limiting factor (Jeschke and Strayer 2006). However, expanding species are often absent from an area not because conditions are unfavorable but because they have not yet dispersed to this area, or because of geographical barriers or dispersal constraints (Araújo and Guisan 2006). Hence, static SDMs ignore important dynamic processes, which may lead to bias in the resulting distributions and should therefore not be used for prediction (Zurell et al. 2009, Yackulic et al. 2015). To deal with this second issue, occupancy models have been extended (Mackenzie et al. 2003, Royle and Kéry 2007) to account for the influence of dynamic processes such as colonization and extinction on the species range dynamics (Mackenzie et al. 2003). So-called multi-season or dynamic site-occupancy models are increasingly used to assess the range dynamics of expanding or invasive species (Bled et al. 2011, Broms et al. 2016a), but remain rarely applied to carnivores (Marcelli and Fusillo 2012, Miller et al. 2013).

Third, data collection is particularly costly if not unfeasible for elusive species that need wide areas due to the large presence area required for sampling. In this context, citizen science is considered as an efficient source of information to assess changes in a species distribution by covering wide areas (Schmeller et al. 2009). However, data from citizen science are often collected with protocols that do not control for variation in the sampling effort 1) in time: a site can be sampled by several observers during a given year and not the following year and 2) in space: given two sites where the species is present, if the sampling effort is lower in one site, this might lead to recording a false absence in this site (Kéry et al. 2010). As a consequence, if sampling effort is not controlled for, detectability can be estimated low, for instance at sites with no sampling effort, leading to biased estimates of the distribution area (Van Strien et al. 2013).

Static and dynamic occupancy models hold promise to analyze population trends from opportunistic data because the data collection process is formally incorporated (Isaac

et al. 2014). However, to address the third issue and apply occupancy models to opportunistic data, one needs to differentiate between a site that was not sampled and a site that was sampled but the species was not detected. In the case of several species being monitored, the detection of a species in a site informs about the non-detection of other species because this site is known to have been sampled (Van Strien et al. 2013). This no longer holds for single-species settings, and the assumption is sometimes made that all sites where at least one detection occurred are sampled throughout the whole duration of the study (Molinari-Jobin et al. 2012, Rich et al. 2013).

Here, we considered grey wolves *Canis lupus* as a case study to illustrate the challenges in using opportunistic data and SDMs to infer the range dynamics of large carnivores. Wolves disappeared in most of the western European countries during the twentieth century (Promberger and Schröder 1993, Boitani 2010) except in Spain, Portugal and Italy (Boitani and Ciucci 1993). The species naturally recolonized the French Alps from the remaining Italian population (Valière et al. 2003, Fabbri et al. 2007). Because the species is protected by law while being a source of conflicts with sheepherding, its recolonization process needs to be carefully monitored.

Our main objective was to describe and determine the drivers of wolves' recolonization pattern in France between 1994 and 2016. To account for imperfect detection, we built a dynamic site-occupancy model (Mackenzie et al. 2006) and analyzed opportunistic data collected by a network of trained volunteers since 1992. To do so, we built a posteriori the sampling effort to account for biases in data collected through citizen science. To describe the recolonization process over time, we addressed two main questions: 1) what are the environmental and biological factors influencing colonization and extinction probabilities? 2) How can sampling effort be inferred a posteriori, i.e. after the data were collected, and to what extent does sampling effort correlate with detection probability?

Methods

Study species and area

The first wolf *Canis lupus* occurrence was detected in France in the early 1990s as a consequence of the Italian population's expansion (Valière et al. 2003, Ciucci et al. 2009). The species then spread outside the Alpine mountains to reach the Pyrenees and the Massif Central westward first in 1999, and the Vosges Mountains northward from 2011. The wolf is an opportunist species that can adapt its diet depending on available prey species (Pouille et al. 1997, Imbert et al. 2016). In areas with livestock farming, strong interactions between wolf presence and sheep breeding usually occur. The study area mostly covered eastern France and a major part of central France (Fig. 1).

Data collection

Wolf detection data were made of presence signs sampled all year long from 1992 to 2016 thanks to a network of

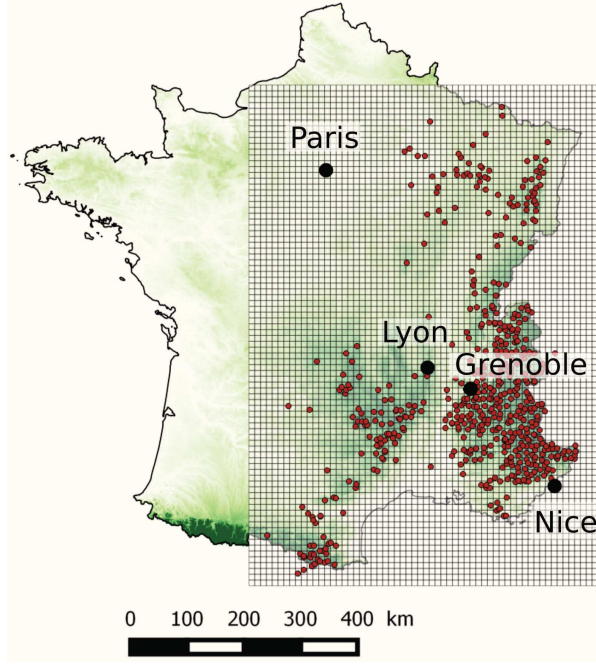


Figure 1. Maps of cumulated species detections (red dots) for the period 1994–2016. Sites were defined as 10×10 km cells within a grid covering all detections. Dark green areas represent mountainous areas with an altitude higher than 1500 m.

professional and non-professional observers. The network size has increased from a few hundred people in 1994, up to 3138 wolf experts in 2016. Every observer is trained during a 3-d teaching course led by the French National Game and Wildlife Agency (ONCFS) to document signs of the species' presence (Duchamp et al. 2012). Presence signs went through a standardized control process combining genetic identification tools, and validation standards to prevent misidentification (Duchamp et al. 2012). For every presence sign, the date and location of collection were stored in a geo-referenced database. These data are considered opportunistic in the sense that monitoring occurs all year long in an extensive manner without explicitly quantifying the sampling effort.

Dynamic site occupancy models

To model the colonization dynamics of wolf, we used dynamic site-occupancy models (Mackenzie et al. 2003) where sampling units were defined as 10×10 km cells (European Commission 2006). Site occupancy models rely on the closure assumption which states that the ecological state of a site (whether it is occupied or not) remains unchanged through occasions (or surveys) j within a year k . Sites were monitored mainly in winter from December to March, the most favorable period to detect the species between the two peaks of dispersal events in spring and fall (Mech and Boitani 2010). We defined the secondary occasions j as December, January, February and March and $y_{i,j,k}$, the observed state of site i equal to 1 if at least one sign of presence was found at site i during occasion j in the year k (and 0 otherwise).

We considered a state-space formulation of the dynamic occupancy model (Royle and Kéry 2007) in which the model is viewed as the combination of 1) the ecological process that involves the latent ecological state of a site, i.e. whether it is occupied or not; 2) the observation process that leads to the detections or non-detections by the observer conditional on the state of the system. The colonization probability $\gamma_{i,k}$ is the probability that an empty site i during year k becomes occupied during year $k + 1$, while the extinction probability $\epsilon_{i,k}$ is the probability that an occupied site i during year k becomes empty during year $k + 1$. We define $z_{i,1}$ as the initial latent state of site i as being drawn from a Bernoulli distribution with the success probability being $\Psi_{i,1}$,

$$z_{i,1} \sim \text{Bernoulli}(\Psi_{i,1})$$

All other latent states $z_{i,k}$ for $k > 1$ are drawn from a Bernoulli distribution as

$$z_{i,k+1} | z_{i,k} \sim \text{Bernoulli}(z_{i,k}(1 - \epsilon_{i,k}) + (1 - z_{i,k})\gamma_{i,k})$$

On top of the ecological process stands the observation process, in which the detections/non-detections are drawn from a Bernoulli distribution

$$y_{i,j,k} | z_{i,k} \sim \text{Bernoulli}(z_{i,k} p_{i,j,k})$$

where $p_{i,j,k}$ is the probability that the species is detected at site i for an occasion j during year k . The state-space formulation is appealing as it makes explicit the latent states $z_{i,k}$ that can be used to build distribution maps. We modelled detection probability with logistic regression using sampling effort, road density and months as covariates. We modelled colonization probability with logistic regression using forest cover, farming cover, rock cover, mean altitude, proportion of high altitude (> 2500 m), the number of observed occupied neighboring sites at short distance, the number of observed occupied neighboring sites at long distance and the distance to the nearest barrier (road or main river) as explanatory variables. Finally, we modelled extinction probability as a logistic function of 'year' as a continuous covariate. We describe these covariates below.

Sampling effort

Monitoring the range expansion of wolves at the country level prevented us from implementing any standardized experimental sampling design. Instead, the presence signs were sampled in an opportunistic way and the sites were defined a posteriori. We adopted an original approach to infer the non-detections based on the available qualitative information on the observers. When entering the network, observers attended a 3-d training session to learn how to identify the species and how it is monitored (Duchamp et al. 2012). During these training sessions, we recorded the observers' personal and professional address, socio-professional category and entry date into the network. The entry and exit dates (whenever known) were used to quantify how many observers were present in the network each year. If necessary, we updated their socio-professional category. We calculated a circular buffer for the prospection area for each observer based on a radius specific to his/her socio-professional category and a center located at his/her

address (Supplementary material Appendix 1 Table A1). For instance, for an observer belonging to the category 1 (departmental authority) whose address was located in the French Department number 39, his/her prospection area would be 4999 km², which is the size of the Department (Supplementary material Appendix 1 Fig. A1 and Table A2). For this observer, a circular buffer was built with a radius calculated as

$$Radius = \sqrt{\frac{prospection\ area}{\pi}}$$

For each 10 × 10 km cell, we then calculated the number of observers monitoring the species per year, i.e. the sampling effort, by summing the number of prospection areas overlapping the cell (Supplementary material Appendix 1 Fig. A2). Sites with a sampling effort equal to zero were not prospected by observers. To avoid estimating a detection probability at sites that were not prospected, we set the detection probability to zero when the sampling effort was null at these sites. When at least one observer was found in a cell in a given year, we considered that sampling occurred. We expected that the sampling effort had a positive effect on the detection parameter. We performed a sensitivity analysis to assess how a change in the construction of the sampling effort influenced the model parameter estimates (Supplementary material Appendix 1 Fig. A3).

Other covariates

Keeping in mind that wolves can adapt to a large range of different habitats, we incorporated proxies of variables that might shape the wolf distribution (Table 1). Using the CORINE Land Cover® database (U.E – SOeS, Corine Land Cover 2006), we defined three covariates to characterize the landscape of the study area: forest cover, farming cover and rock cover. Forest cover may structure the ungulate distribution (i.e. prey species). As a consequence, we expected that forest cover would have a positive effect on colonization,

and rock cover would have a negative effect on colonization. Farmland cover was defined as a covariate including pastures areas which can be used by livestock, a possible prey to wolves and a proxy for rural landscape under human influence. Altitude may also have an influence on colonization (Llaneza et al. 2012, Falcucci et al. 2013). We used the IGN BD_ALTI® database (250 m resolution) to calculate the mean altitude of each site as well as the proportion of altitude higher than 2500 m. We predicted a site with a high proportion of high-altitude (> 2500 m high) would be less attractive for the species as ungulate species might be less abundant above this limit.

Dispersal capacity is a key factor to explain the dynamic of wolf colonization (Boyd and Pletscher 1999, Kojola et al. 2006, Ciucci et al. 2009). Because cells occupied by established packs may act as a source of dispersers at short and long-distance (Yackulic et al. 2012), the neighborhood of an unoccupied cell may influence its colonization probability (Veran et al. 2016). In that spirit, the presence of individuals at short and long-distance could be accounted for by using conditional autoregressive models and auto-logistic models (Bled et al. 2013). However, due to the computational burden and convergence issues, we could not implement this approach here. We therefore defined two covariates that consisted of the observed number of contiguous occupied cells at both short and long-distances around the focal cell. The short-distance covariate was defined as the number of observed occupied cells directly contiguous to the focal cell i.e. situated within a distance of 10 km. The limit for the long-distance parameter was set to avoid a dilution effect due to the small number of observed occupied cells at very long-distances but large enough to account for most long-distance observed occupied cells that could play a role in the colonization probability. Based on observations of wolf dispersal in the western Italian Alps (Marucco and McIntire 2010), we set this limit at 150 km around the focal cell. We expected a positive effect of these two covariates on the probability of a site to be colonized.

Table 1. Description and expected effects of covariates used to describe the occupancy dynamics of wolf in France.

Covariate	Abbreviation	Parameter	Description	Expected effect	Reference
Forest cover	Forest	Colonisation (γ)	Percentage of mixt, coniferous or deciduous forests cover	+	Oakleaf et al. 2006, Fechter and Storch 2014
Farmland cover	Agr	Colonisation (γ)	Percentage of pasture lands and other farming activities cover	+/-	Glenz et al. 2001
Rock cover	Rock	Colonisation (γ)	Percentage of rock cover	-	
High altitude	Halt	Colonisation (γ)	Proportion of altitude higher than 2500 m	-	Glenz et al. 2001
Altitude	Alt	Colonisation (γ)	Mean altitude	+/-	Llaneza et al. 2012 Falcucci et al. 2013
Distance to the closest barrier	Dbarr	Colonisation (γ)	Minimal distance between a highway or one of the five main rivers in France	-	Falcucci et al. 2013
Short distance occupied neighboring cells	SDAC	Colonisation (γ)	Proportion of observed occupied contiguous cells	+	Bled et al. 2011
Long distance occupied neighboring cells	LDAC	Colonisation (γ)	Proportion of observed occupied cells within a 150 km radius without the contiguous cells	+	
Year (continuous)	Trend-year	Extinction (ϵ)	Year as a linear effect	-	Marucco 2009
Sampling effort	SEff	Detection (p)	Number of observers per site per year	+	
Road density	Rdens	Detection (p)	Percentage of site covered by roads	+	
Month-survey	survey	Detection (p)	Occasion of survey (categorical)	+/-	Marucco 2009

Because dispersal could be driven by the presence of physical barriers (Wabakken et al. 2001, Blanco et al. 2005), we defined a landscape covariate depicting the distance from the center of the site to the closest barrier defined as highways or rivers (U.E – SOeS, Corine Land Cover 2006). We expected this covariate to impact colonization negatively.

In the first few years after sites become newly colonized, extinction probability is expected to be high as long as only isolated individuals use them. Once a pack has settled, pack persistence is the rule for wolves when other packs are present in the surrounding areas (Mech and Boitani 2010). Pack splitting may rise from various sources including harvest or poaching of alpha pairs (Gehring et al. 2003, Brainerd et al. 2008) leading to a locally extinct site. Within the distribution of an actively expanding population, extinct sites might be recovered by surrounding individuals, either by dispersers or by neighboring packs. We therefore expected extinction probability to decrease over time.

Finally, in addition to sampling effort, we considered the potential effect of road densities on the species detectability, first through facilitation of site accessibility for the observers and second, because cross roads can be used as marking sites (Barja et al. 2004), which can lead to a higher detectability. Because presence signs rely partly on track records in the snow, we considered month as a categorical variable to account for the variation in detection conditions due to weather variations across the survey months (Marucco 2009).

Last, we considered the initial occupancy probability as constant since only two sites were occupied in the first year of the study, which was not enough to assess the effects of covariates on this parameter.

Model fitting, selection and validation

We performed covariate selection using stochastic search variable selection (SSVS; George and McCulloch 1993, O'Hara and Sillanpää 2009). In brief, SSVS builds a model that includes all covariate combinations as special cases. In practice, this is achieved by adding binary indicator variables, α_p equals 1 or 0, which allows the estimation of the regression parameter β_p or excludes it by setting it to a constant (Supplementary material Appendix 1 Table C1). In a Bayesian framework, we explored the model space generated by excluding or including covariates. The priors for regression parameters β_p were written as $(1 - w)$ Normal(0,0.0001) + w Normal(0,1) with $w \sim$ Bernoulli(0.5) therefore assuming a priori that each covariate had a 50-50 chance of being present in the model. We checked that the model space was well sampled by the SSVS and that we did not get stuck in a particular set of models. We used three different initial model configurations (with all covariates vs without any of the covariates vs a few covariates picked at random in the set of all covariates). We did not explore different priors as mixing and convergence were satisfying. Prior to model selection, we ran a Spearman test to check for correlations among covariates.

We used Markov chain Monte Carlo (MCMC) simulations and parameter estimation. We ran three MCMC chains with a burn-in period of 2500 iterations followed by

10 000 iterations on which we based our inference. We used posterior medians and 95% credible intervals to summarize parameter posterior distributions. To assess the effect of a covariate on a parameter, we set the other covariates to their mean value. We checked convergence visually by inspecting the chains and by checking that the R-hat statistic was below 1.1 (Gelman and Shirley 2011). We finally produced distribution maps of the latent states by using a posteriori means of the $z_{i,k}$ from the best model. To assess the fit of our final model, we used the posterior predictive checking approach (Gelman et al. 1996) that has recently been applied to occupancy models (Broms et al. 2016b) (Supplementary material Appendix 1 Fig. B1).

Data deposition

Data available from the Dryad Digital Repository: <<http://dx.doi.org/10.5061/dryad.g9s1d>> (Louvrier et al. 2017).

Results

The effect of covariates on detectability and the dynamic of occupancy

The model best supported by the data had detection as a function of sampling effort, road density and occasion (month) and colonization as a function of forest cover, farmland cover, mean altitude, proportion of high-altitude and the number of observed occupied cells at a short and long-distance neighborhood (Supplementary material Appendix 1 Table C1). This model appeared to fit the data adequately well (Supplementary material Appendix 1 Fig. B1).

Initial occupancy probability was low, as expected since few sites were detected as occupied at the beginning of the study (Supplementary material Appendix 1 Table C2).

As predicted, forest cover had a positive influence on the probability that a site became colonized. The proportion of farmland cover within a cell also appeared to have a positive influence on this probability. Below 1500 m of mean altitude, the probability that a site became colonized was close to zero, whereas above this limit the probability reached up to 0.07 (0.05; 0.11) (Fig. 2). This probability decreased with the high-altitude proportion in a site. Over time, the number of observed occupied neighboring cells increased at both short and long-distance (Supplementary material Appendix 1 Fig. D1). If all of the 8 neighboring cells were observed as occupied, the probability that the target cell became colonized was 0.48 (0.32; 0.58) compared to a colonization probability of 0.11 (0.08; 0.15) if the target site had only 0 to 2 contiguous neighboring cells observed occupied. As this number increased, the probability that a site became colonized increased accordingly (Fig. 2).

Sites located within the Alps had the highest number of observed occupied sites at both short and long-distance. Colonization probability was the highest in this area (Fig. 3). The highest part of the Alps (i.e. sites with the greatest proportions of high-altitude) remained with a low colonization probability (Supplementary material Appendix 1

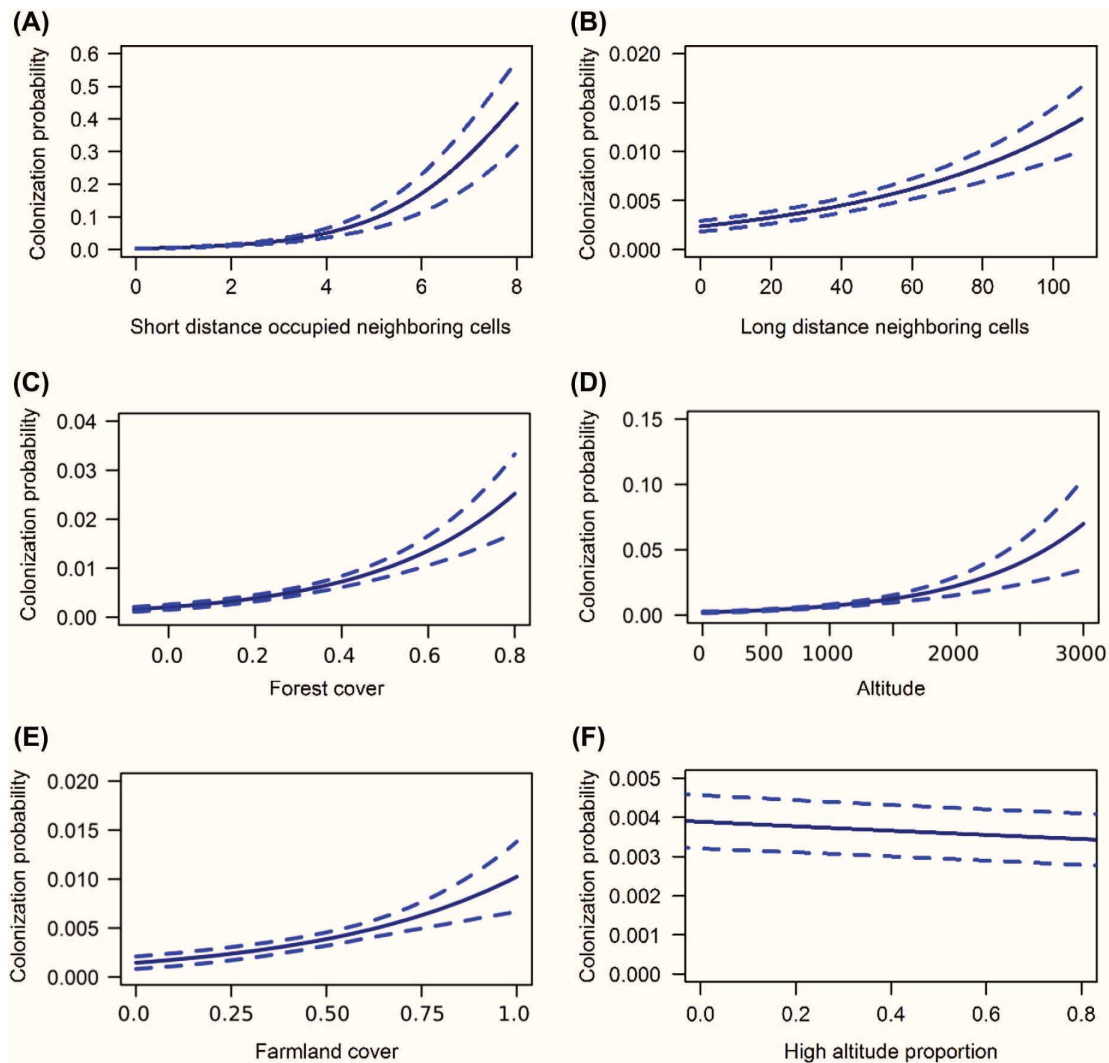


Figure 2. Relationship between the estimated colonization probability and (A) short-distance occupied neighboring cells, (B) long-distance occupied neighboring cells, (C) proportion of forest cover, (D) altitude, (E) proportion of farmland cover, and (F) site proportion of altitude higher than 2500 m.

Fig. D2). Overall, this probability remained higher than zero in mountainous areas and increased with time as the number of occupied sites increased (Fig. 3).

Finally, and as expected, detection probability varied according to the survey month with the lowest mean value of 0.17 (0.16; 0.18) in December and the highest value of 0.25 (0.24; 0.26) in January (Fig. 4). As expected, detection probability increased when the number of observers per site increased but, in contrast with what we expected, decreased with increasing road density. The sensitivity analysis showed weak effects of variations in the prospection areas used to build the sampling effort, except for the number of observed occupied sites at long distance (Supplementary material Appendix 1 Fig. A3).

Distribution map

From 1994 to 2016, 10 918 presence signs were recorded by the network and used in our analysis. The species was initially spotted in 2 cells in 1994 and was detected in 188 cells

in 2016 (around 70-fold increase, see top panel in Fig. 5). This led to an apparent occupancy (proportion of occupied sites on the total number of sites in the study area) varying from 0.001 in 1994 to 0.047 in 2016.

Accounting for both sampling effort and imperfect detection, we estimated the number of occupied sites as up to 10 (1; 19) in 1994 and up to 211 (195; 227) in 2016 (top panel in Fig. 5). Overall, the estimates were higher than the naïve estimates of occupancy. When we ignored the sampling effort in the detection process, we found an estimated number of occupied sites equal to 2 (1; 4) in 1994 and up to 192 (180; 204) in 2016. Most discrepancies between the two models (accounting for vs. ignoring the sampling effort) were found at the early stage of the colonization process when the network of observers was implemented mainly in eastern France (compare bottom left and right panels in Fig. 5; see also Supplementary material Appendix 1 Fig. D3). Accounting for the sampling effort allowed us to infer the species presence on sites that were not prospected or prospected with a low sampling effort (top panel in Fig. 5).

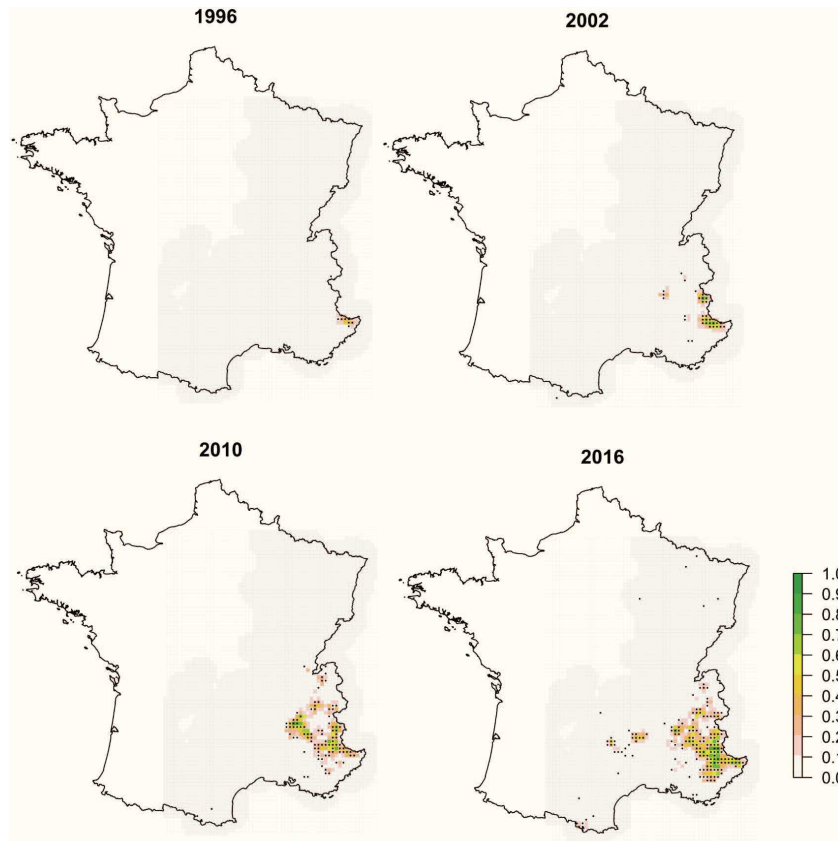


Figure 3. Maps of estimated colonization probability between 1995 and 1996, 2001 and 2002, 2009 and 2010, 2015 and 2016 from the best model (Table 2). Black dots represent detections made in 1995, 2001, 2009, and 2015.

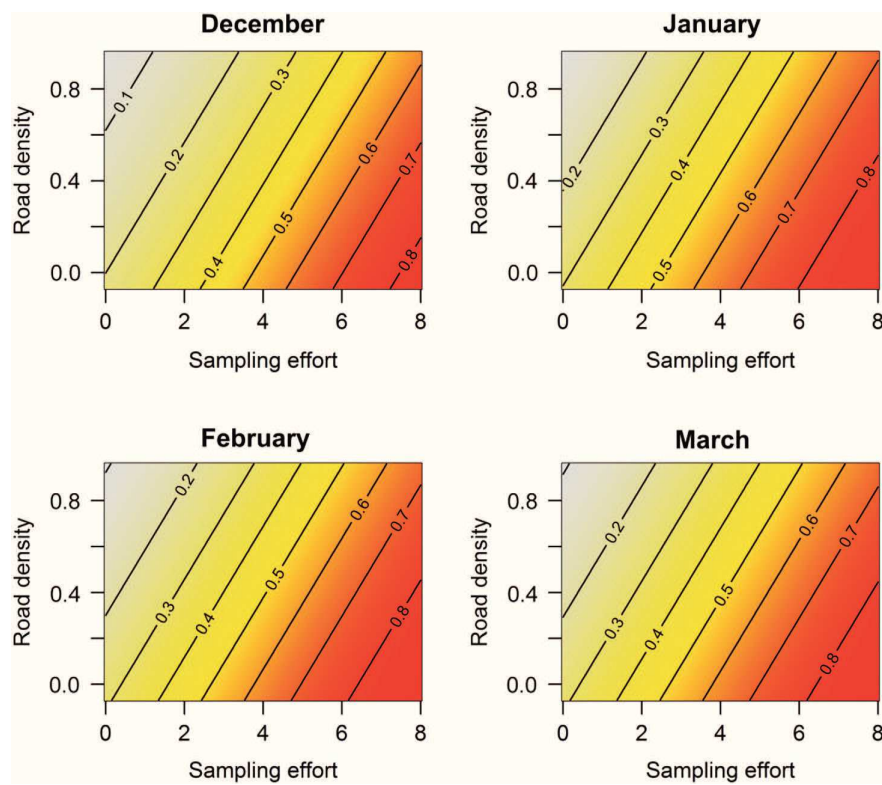


Figure 4. Joint effects of road density, standardized sampling effort and occasion (month) on the species detection probability.

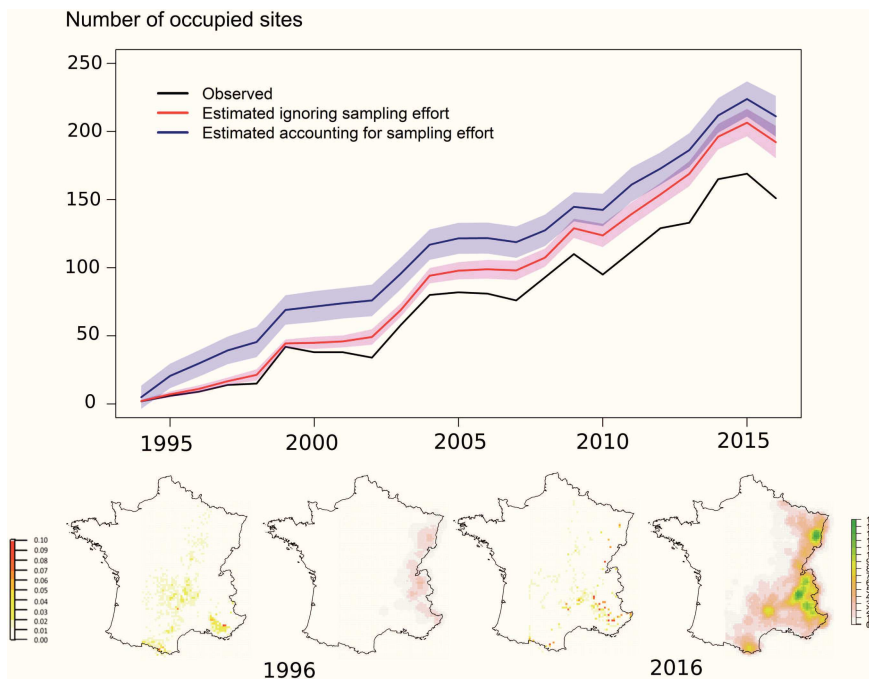


Figure 5. Up: number of 10×10 km cells observed (black), estimated occupied ignoring sampling effort (red) and estimated occupied accounting for sampling effort (blue) for each year from 1994 to 2016. Also displayed is the 95% credible interval for both estimates of the sampling effort. Down: maps of differences between estimates of occupancy from the model accounting for sampling effort and the one ignoring sampling effort. Dark red sites are sites that appeared estimated occupied by the model accounting for sampling effort but did not appear occupied once ignoring sampling effort. Both maps are associated with maps of the sampling effort on their right, for the years 1996 and 2016.

Our results showed that in 1994 the species was found only in the southern Alps, and then actively colonized towards the northern Alps at the beginning of the 2000s. The colonization process started to reach the Pyrenees and Massif Central area in early 2000, and the Vosges area in the very north-eastern part of France, at the beginning of the 2010s, indicating that the French wolf population is still in a phase of expansion west and northward from the alpine range. This led to an average expansion rate (i.e. number of occupied sites divided by the number of occupied sites the previous year) of 112% (100%; 128%) (Fig. 6). This expansion rate first decreased over time, from 225% (118%; 600%) at the early stage of the wolf colonization in 1994 to 103% (91%; 117%) in 2000 due to low number of occupied cells, then stabilized at 107% (98%; 117%) on average per year demonstrating that the population is still in an expanding phase mainly thanks to the colonization outside of the alpine range.

The model did not predict absence in places where presence signs were found (Fig. 7). Sites with high occupancy probability were mainly close to the sites where the species had been previously detected, mostly due to the effect of short-distance neighbors. Some sites had a high probability of being occupied (> 0.75), however the uncertainty associated with those predictions was also high (standard deviation [SD] > 0.30). We found sites with high probability of occupancy (> 0.75) with low uncertainty ($SD < 0.20$), and some of those sites were observed as occupied in the following year because the model propagates information backwards in time and so z_k is informed directly by z_{k+1} .

Discussion

Determining favorable areas is often accomplished by building distribution maps using habitat suitability models (Mladenoff et al. 1999) or occupancy models (Marucco 2009). However, these studies often rely on a static relationship between the species of interest and its environment

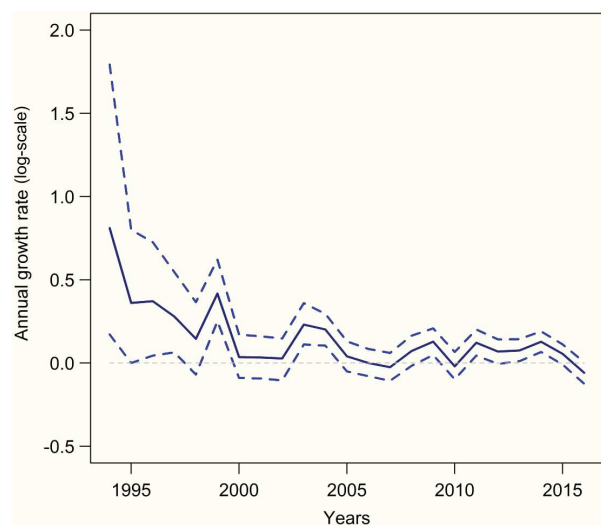


Figure 6. Growth rate (i.e. number of sites divided by the total number of sites the previous year) given for each year from 1994 to 2016, on a log scale.

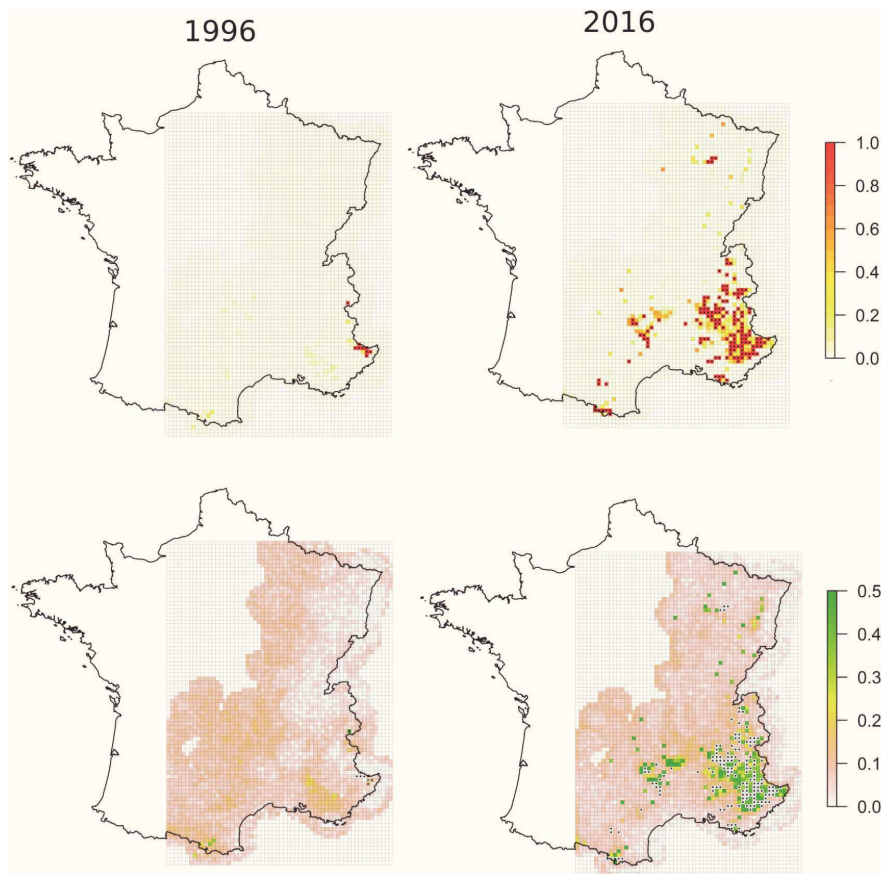


Figure 7. Maps of estimated occupancy (top) and associated standard deviation (bottom) for years 1996 and 2016. Black dots represent detections made in 1996 and 2016.

(Jedrzejewski et al. 2008). Here, we used dynamic site-occupancy models and brought new insights on the processes governing the dynamic of recolonization of a keystone carnivore species. By controlling for species detectability and heterogeneous sampling effort, our approach can be used to assess the distribution dynamics of any species based on opportunistic data, pending relevant information is gathered on the people collecting the data.

Model assumptions

Site occupancy models rely on several assumptions that need to be discussed (Mackenzie et al. 2003, 2006). First, the species should not be detected when absent from a site (i.e. no false positives). This is unlikely to happen in our case since we did not account for presence signs that were rejected because they did not fulfill the standardized criteria used to avoid species misidentification (Duchamp et al. 2012). If doubts persist about the occurrence of false positives, this assumption could be relaxed by using site-occupancy models that account for misidentifications (Miller et al. 2011, Rich et al. 2013).

Second, detection histories of all sampling units are assumed to be independent. However, detection histories were likely dependent in space because of a non-homogeneous spatial sampling effort inherent to opportunistic data. We partly accounted for this non-independence by quantifying

the sampling effort. Furthermore, by accounting for the number of observed occupied neighboring cells, we made the detection history of a focal cell dependent partly on the detection histories of the neighboring cells. If the source of dependence is unknown, spatial autocorrelation can be modeled using geostatistical tools on occupancy or extinction/colonization parameters and also on detection (Bled et al. 2013).

Third, the status of a site should not change during primary occasions – the closure assumption (Rota et al. 2009). We used the data provided within the winter period from November to March as a primary occasion because it corresponds to the most stable period in the social organization of the packs. If movements or mortality occurred inside or outside of the sampling sites, it is likely, in our study, that the probability of occupancy in a given time interval did not depend on the occupancy status of a site in the previous time interval (Mackenzie and Royle 2005). In this situation of so-called random temporary emigration, the bias in parameter estimates is minimal, but occupancy should be interpreted as use of the sampling area rather than the proportion of area occupied by the species (Mackenzie et al. 2004).

Fourth, there should be no unmodelled heterogeneity in the model parameters. Regarding the detection probability, some heterogeneity might remain due to a difference of detection in the presence signs, e.g. tracks vs hair (Graves et al. 2011). This was unlikely to occur in our study because

the vast majority of presence signs are tracks. Regarding the colonization parameter, even though we had data on the number of killed preys during the hunting season, we did not have information on wild prey density at such a large scale. Therefore, we used characteristics of their habitats as a proxy for their presence (Jedrzejewski et al. 2008).

Besides the usual assumptions of occupancy models, we also had to deal with opportunistic data that are collected through non-standardized sampling protocols. To cope with opportunistic data, we defined a grid of spatial units that was overlaid on the map of detections/non-detections. We used 10×10 km cells as sampling units, a choice we made in agreement with what was recommended by the European Union (European Commission 2006) and also shown to be the best tradeoff between the species territory size and sensitivity of the distribution to the size and shape of the unit cell (Marboutin et al. 2010). The average wolves' territory size vary between 100 and 400 km² in western and central Europe (Ciucci et al. 1997, Mech and Boitani 2010, Duchamp et al. 2012). Although these cells might not entirely cover wolves territories, Latham et al. (2014) studied the effect of grid size to assess wolf's occupancy and found that taking a large grid size may not be appropriate for areas with moderate to high wolf density as it can overestimate occupancy rate. On the other hand, if the size of the sampling unit is too small, then there is a risk of having very few detections within a year, which would make the estimation of the detection probability difficult.

Last, we assumed that observers were prospecting homogeneously inside the prospection area we assigned to them. This assumption may have been violated for two reasons. First, an observer might prospect more intensively near the center of the prospecting area, because it was defined as a home or work location, or near places where she/he already found presence signs (Duchamp et al. 2012). We also assumed that observers were prospecting homogeneously in time. However, observers may show different patterns in sampling frequency and some might not be prospecting during the months of winter. Finally, we assumed that once entered in the network, observers did not leave it unless we had information indicating the contrary such as a change of job or social status. Consequently, we might have overestimated the number of observers actually prospecting in the network. We therefore recommend recording carefully the activity of observers within the network to get a realistic picture of the actual sampling effort (Beirne and Lambin 2013).

Effects of environmental covariates

We used road density as a proxy of human presence and found a negative influence on the detection probability. When defining the road density covariate, we accounted for all types of roads (except highways). Because many observers from the network are wildlife professionals (Duchamp et al. 2012), main roads may not be used and accessibility to a site may consist mostly in dirt and forest roads or pathways. The negative influence could be explained by the fact that wolves tend to avoid roads (Whittington et al. 2005), therefore there might be fewer presence marks at sites with

high road density. As expected, we found that detection probability increased when sampling effort increased, therefore highlighting the importance to account for imperfect detection when it is likely to be inhomogeneous in time and space. Finally, detection varied according to the month of the survey, which can be explained by the variability in snow conditions in the study area.

We found that colonization was mainly influenced by the number of observed occupied neighbors at short and long-distances, showing that dispersal and competition for space with other packs is a key factor of the dynamic of occupancy. These results corroborate those of Adams et al. (2008) who showed that dispersal was the main component explaining wolf population dynamics. Several long-distance dispersal events have been documented across the alpine area (Wolf Alpine Group 2014) and in France (Duchamp et al. unpubl.). Further studies explicitly modeling dispersal processes could help to better predict wolves colonization by accounting for factors that could enhance or slow down the dispersal rate for instance (Broms et al. 2016a).

We found that mean altitude had a positive effect on colonization probability. Wolves are highly flexible and can live in various areas from maize cultures to high mountains (Kaczensky et al. 2013). Starting from central Italy (Lucchini et al. 2002, Fabbri et al. 2007), wolves reached the alpine range via the natural Apennine mountain corridor. Therefore, the effect of mean altitude may be related to the history of the wolves' natural recovery process. However, we also found a negative effect of the proportion of altitude higher than 2500 m, i.e. the higher the proportion of high-altitude, the less likely a site was to become colonized. Above 2500 m, vegetation turns to sparse vegetation with rocky covers and snow. In contrast, more forest cover associated with lower altitudes (< 2500 m) increased the probability that a site become colonized mainly because these habitats' structure and composition are much more suitable to the presence of key prey species (Darmon et al. 2012). To a lesser extent, the effect of farmland cover was also found to have a positive influence on the colonization probability. Although pasture areas host domestic preys (Meriggi and Lovari 1996) and may influence wolf colonization, the farmland covariate refers to rural exploited landscapes usually located down the valleys or lowlands. As wolf recovery is increasing over time, dispersers do not have other choices than to fulfill free available space to colonize. The overlap between human range activities and wolf settlement then increases as the recovering process is going on. The inclusion of more explicit covariates related to pastoral activity, such as the number of sheep in space, may provide a better understanding of the interaction between domestic prey and wolf presence, but these were not available to us.

Trends in wolf recolonization

Colonization patterns have been studied during recent decades (Wabakken et al. 2001). It appears that in Scandinavia, wolves were showing a colonization process that is typical of species with high dispersal capacities and pre-saturation dispersal (Swenson et al. 1998). This process is characterized by single long leaps forward and as a consequence, the

colonization front is less well defined (Hartman 1994) compared to a stepping stone dispersal strategy. Wolves seem to follow a similar pattern in France (Fig. 6). This biological trait used by wolves is mainly known as a mechanism to avoid competition with other packs (Hayes and Harestad 2000). Once the area becomes saturated, dispersers may settle at unoccupied sites at long distance with higher risks of mortality due to an Allee effect (Hurford et al. 2006, Sanderson et al. 2013) or demographic stochasticity (Vucetich et al. 1997). In line with Marescot et al. (2011) who estimated a positive rate of increase in abundance, we demonstrated that the spatial dynamic mechanism of the wolves' natural recovery is still going on, particularly outside the alpine range both northward and westward. However, this recovery appeared to slow down, mainly due to areas becoming saturated within the alpine range associated with the natural barrier along the Rhône valley slowing down dispersal and/or a recent increase in official wolf controls. We may expect an increase in occupancy once few new packs have settled apart the alpine range.

We used dynamic occupancy models to assess the current and dynamic distribution of a species that is expanding since it returned; there is a temptation to aim at forecasting its future distribution. However, we emphasize the difficulty of achieving this objective because we could not incorporate the drivers that may appear relevant to explain future colonization events. For instance, now that wolves have settled in the alpine range and continue to expand, they are likely to encounter new environments such as lowlands in the next few years, a landscape that may drive future colonization. Consequently, use of our model as a predictive tool should be considered in an adaptive framework, i.e. by updating the management rules and the distribution maps every year during the active colonization phase.

The outcomes of our analyses have important consequences for managing animal species because their conservation status must be assessed partly through trends in their distributions (see art. 1 of the Habitats Fauna Flora European Directive). Dynamic occupancy models are therefore relevant tools to the decision-making process by providing maps and spatio-temporal trends. In the case of the wolf, these models can help in focusing the prevention of damage to livestock (Miller 2015). The identification of areas where the species may or may not occur along with the surrounding uncertainty may be used to target specific sites and determine priorities for implementing mitigation measures.

Acknowledgements – The authors are sincerely thankful to all the volunteers and wolf regional referees belonging to the Large Carnivore Network for data collection and validation and local investment in the fieldwork. We warmly thank Ian Renner for editing our manuscript.

Funding – JL is thankful to the GDR Ecologie Statistique, Univ. of Montpellier and ONCFS for grants she received to conduct her work.

References

Adams, L. G. et al. 2008. Population dynamics and harvest characteristics of wolves in the Central Brooks Range, Alaska. – *Wildl. Monogr.* 170: 1–25.

- Araújo, M. B. and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. – *J. Biogeogr.* 33: 1677–1688.
- Barja, I. et al. 2004. The importance of crossroads in faecal marking behaviour of the wolves (*Canis lupus*). – *Naturwissenschaften* 91: 489–492.
- Beirne, C. and Lambin, X. 2013. Understanding the determinants of volunteer retention through capture–recapture analysis: answering social science questions using a wildlife ecology toolkit. – *Conserv. Lett.* 6: 391–401.
- Bischof, R. et al. 2015. Wildlife in a politically divided world: insularism inflates estimates of brown bear abundance. – *Conserv. Lett.* 9: 122–130.
- Blanco, J. C. et al. 2005. Wolf response to two kinds of barriers in an agricultural habitat in Spain. – *Can. J. Zool.* 83: 312–323.
- Bled, F. et al. 2011. Assessing hypotheses about site occupancy dynamics. – *Ecology* 92: 938–951.
- Bled, F. et al. 2013. Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. – *Ecol. Evol.* 3: 4896–4909.
- Boitani, L. 2010. Wolf conservation and recovery. – In: Mech, L. D. and Boitani, L. (eds), *Wolves: behavior, ecology, and conservation*. Univ. of Chicago Press, pp. 317–340.
- Boitani, L. and Ciucci, P. 1993. Wolves in Italy: critical issues for their conservation. – In: Promberger, C. and Schröder, W. (eds), *Wolves in Europe: status and perspectives*. Munich Wildlife Society, pp. 74–90.
- Boyd, D. K. and Pletscher, D. H. 1999. Characteristics of dispersal in a colonizing wolf population in the Central Rocky Mountains. – *J. Wildl. Manage.* 63: 1094–1108.
- Brainerd, S. M. et al. 2008. The effects of breeder loss on wolves. – *J. Wildl. Manage.* 72: 89–98.
- Breitenmoser, U. 1998. Large predators in the Alps: the fall and rise of man's competitor. – *Biol. Conserv.* 83: 279–289.
- Broms, K. M. et al. 2016a. Dynamic occupancy models for explicit colonization processes. – *Ecology* 97: 194–204.
- Broms, K. M. et al. 2016b. Model selection and assessment for multi-species occupancy models. – *Ecology* 97: 1759–1770.
- Chapron, G. et al. 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. – *Science* 346: 1517–1519.
- Ciucci, P. et al. 1997. Home range, activity and movements of a wolf pack in central Italy. – *J. Zool.* 243: 803–819.
- Ciucci, P. et al. 2009. Long-distance dispersal of a rescued wolf from the northern Apennines to the western Alps. – *J. Wildl. Manage.* 73: 1300–1306.
- Darmon, G. et al. 2012. Spatial distribution and habitat selection in coexisting species of mountain ungulates. – *Ecography* 35: 44–53.
- Duchamp, C. et al. 2012. A dual frame survey to assess time- and space-related changes of the colonizing wolf population in France. – *Hystrix* 23: 1–12.
- European Commission 2006. Assessment, monitoring and reporting under article 17 of the habitats directive: explanatory notes and guidelines. – European Commission.
- Fabbri, E. et al. 2007. From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. – *Mol. Ecol.* 16: 1661–1671.
- Falcucci, A. et al. 2013. Modeling the potential distribution for a range-expanding species: wolf recolonization of the Alpine range. – *Biol. Conserv.* 158: 63–72.
- Fechter, D. and Storch, I. 2014. How many wolves (*Canis lupus*) fit into Germany? The role of assumptions in predictive rule-based habitat models for habitat generalists. – *PLoS One* 9: e101798.
- Gehring, T. M. et al. 2003. Limits to plasticity in gray wolf, *Canis lupus*, pack structure: conservation implications for recovering populations. – *Can. Field-Nat.* 117: 419–423.

- Gelman, A. and Shirley, K. 2011. Inference from simulations and monitoring convergence. – In: Brooks, S. P. et al. (eds), Handbook of Markov Chain Monte Carlo. Chapman Hall, pp. 163–174.
- Gelman, A. et al. 1996. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). – Stat. Sinica 7: 733–807.
- George, E. and McCulloch, R. E. 1993. Variable selection via Gibbs sampling. – J. Am. Stat. Assoc. 88: 881–889.
- Glenz, C. et al. 2001. A wolf habitat suitability prediction study in Valais (Switzerland). – Landscape Urban Plan. 55: 55–65.
- Graves, T. A. et al. 2011. Linking landscape characteristics to local grizzly bear abundance using multiple detection methods in a hierarchical model. – Anim. Conserv. 14: 652–664.
- Grinnell, J. 1917. The niche-relationships of the California thrasher. – Auk 34: 427–433.
- Guillera-Arroita, G. et al. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. – Global Ecol. Biogeogr. 24: 276–292.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – Ecol. Lett. 8: 993–1009.
- Hartman, G. 1994. Long-term population development of a reintroduced beaver (*Castor fiber*) population in Sweden. – Conserv. Biol. 8: 713–717.
- Hayes, R. D. and Harestad, A. S. 2000. Demography of a recovering wolf population in the Yukon. – Can. J. Zool. 78: 36–48.
- Hurford, A. et al. 2006. A spatially explicit model for an Allee effect: why wolves recolonize so slowly in Greater Yellowstone. – Theor. Popul. Biol. 70: 244–254.
- Hutchinson, G. E. 1957. Concluding remarks. – Cold Springs Harbor Symp. Quant. Biol. 22: 415–427.
- Imbert, C. et al. 2016. Why do wolves eat livestock? Factors influencing wolf diet in northern Italy. – Biol. Conserv. 195: 156–168.
- Isaac, N. J. B. et al. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. – Methods Ecol. Evol. 5: 1052–1060.
- Jedrzejewski, W. et al. 2008. Habitat suitability model for Polish wolves based on long-term national census. – Anim. Conserv. 11: 377–390.
- Jeschke, J. M. and Strayer, D. L. 2006. Determinants of vertebrate invasion success in Europe and North America. – Global Change Biol. 12: 1608–1619.
- Kaczensky, P. et al. 2013. Status, management and distribution of large carnivores – bear, lynx, wolf & wolverine – in Europe. – Report to the EU Commission.
- Kéry, M. 2011. Towards the modelling of true species distributions. – J. Biogeogr. 38: 617–618.
- Kéry, M. and Schaub, M. 2011. Bayesian population analysis using WinBUGS – a hierarchical perspective. – Academic Press.
- Kéry, M. et al. 2010. Predicting species distributions from checklist data using site-occupancy models. – J. Biogeogr. 37: 1851–1862.
- Kojola, I. et al. 2006. Dispersal in an expanding wolf population in Finland. – J. Mammal. 87: 281–286.
- Lahoz-Monfort, J. J. et al. 2014. Imperfect detection impacts the performance of species distribution models. – Global Ecol. Biogeogr. 23: 504–515.
- Latham, M. C. et al. 2014. Can occupancy–abundance models be used to monitor wolf abundance? – PLoS One 9: e102982.
- Linnell, J. D. C. and Boitani, L. 2012. Building biological realism into wolf management policy: the development of the population approach in Europe. – Hystrix 23: 80–91.
- Llaneza, L. et al. 2012. Insights into wolf presence in human-dominated landscapes: the relative role of food availability, humans and landscape attributes. – Divers. Distrib. 18: 459–469.
- Long, R. A. et al. 2010. Predicting carnivore occurrence with noninvasive surveys and occupancy modeling. – Landscape Ecol. 26: 327–340.
- Louvrier, J. et al. 2017. Data from: Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data. – Dryad Digital Repository, <http://dx.doi.org/10.5061/dryad.g9s1d>.
- Lucchini, V. et al. 2002. Noninvasive molecular tracking of colonizing wolf (*Canis lupus*) packs in the western Italian Alps. – Mol. Ecol. 11: 857–868.
- Mackenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies : general advice and allocating survey effort. – J. Appl. Ecol. 42: 1105–1114.
- Mackenzie, D. I. et al. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. – Ecology 84: 2200–2207.
- Mackenzie, D. I. et al. 2004. Investigating species co-occurrence patterns when species are detected imperfectly. – J. Anim. Ecol. 73: 546–555.
- Mackenzie, D. I. et al. 2006. Occupancy estimation and modelling. Inferring patterns and dynamics of species occurrence. – Academic Press.
- Marboutin, E. et al. 2010. On the effects of grid size and shape when mapping the distribution range of a recolonising wolf (*Canis lupus*) population. – Eur. J. Wildl. Res. 57: 457–465.
- Marcelli, M. and Fusillo, R. 2012. Land use drivers of species re-expansion: inferring colonization dynamics in Eurasian otters. – Divers. Distrib. 18: 1001–1012.
- Marescot, L. et al. 2011. Capture–recapture population growth rate as a robust tool against detection heterogeneity for population management. – Ecol. Appl. 21: 2898–2907.
- Marucco, F. 2009. Spatial population dynamics of a recolonizing wolf population in the western Alps. – Univ. of Montana, Missoula.
- Marucco, F. and McIntire, E. J. B. 2010. Predicting spatio-temporal recolonization of large carnivore populations and livestock depredation risk: wolves in the Italian Alps. – J. Appl. Ecol. 47: 789–798.
- Mech, L. D. and Boitani, L. 2010. Wolves: behavior, ecology, and conservation. – Univ. of Chicago Press.
- Meriggi, A. and Lovari, S. 1996. A review of wolf predation in southern Europe: does the wolf prefer wild prey to livestock? – J. Appl. Ecol. 33: 1561–1571.
- Miller, D. A. W. et al. 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. – Ecology 92: 1422–1428.
- Miller, D. A. W. et al. 2013. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. – PLoS One 8: e65808.
- Miller, J. R. B. 2015. Mapping attack hotspots to mitigate human–carnivore conflict: approaches and applications of spatial predation risk modeling. – Biodivers. Conserv. 24: 2887–2911.
- Mladenoff, D. J. et al. 1999. Predicting gray wolf landscape recolonization: logistic regression vs. new field data. – Ecol. Appl. 9: 37–44.
- Molinari-Jobin, A. et al. 2012. Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. – Anim. Conserv. 15: 266–273.
- Oakleaf, J. K. et al. 2006. Habitat selection by recolonizing wolves in the northern Rocky Mountains of the United States. – J. Wildl. Manage. 70: 554–563.
- O’Hara, R. B. and Sillanpää, M. J. 2009. A review of Bayesian variable selection methods: what, how and which. – Bayesian Anal. 4: 85–118.
- Packer, C. et al. 2013. Conserving large carnivores: dollars and fence. – Ecol. Lett. 16: 635–641.

- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Pouille, M.-L. et al. 1997. Significance of ungulates in the diet of recently settled wolves in the Mercantour mountains (southeastern France). – *Rev. d'Ecologie (Terre vie)* 52: 357–358.
- Promberger, C. and Schröder, W. 1993. Wolves in Europe: status and perspectives. – Munich Wildlife Society.
- Rich, L. N. et al. 2013. Estimating occupancy and predicting numbers of gray wolf packs in Montana using hunter surveys. – *J. Wildl. Manage.* 77: 1280–1289.
- Ripple, W. J. et al. 2014. Trophic cascades from wolves to grizzly bears in Yellowstone. – *J. Anim. Ecol.* 83: 223–233.
- Rota, C. T. et al. 2009. Occupancy estimation and the closure assumption. – *J. Appl. Ecol.* 46: 1173–1181.
- Royle, J. A. and Kéry, M. 2007. A Bayesian state-space formulation of dynamic occupancy models. – *Ecology* 88: 1813–1823.
- Sanderson, C. E. et al. 2013. With Allee effects, life for the social carnivore is complicated. – *Popul. Ecol.* 56: 417–425.
- Schmeller, D. S. et al. 2009. Advantages of volunteer-based biodiversity monitoring in Europe. – *Conserv. Biol.* 23: 307–316.
- Sunarto, S. et al. 2012. Tigers need cover: multi-scale occupancy study of the big cat in Sumatran forest and plantation landscapes. – *PLoS One* 7: e30859.
- Swenson, J. et al. 1998. Geographic expansion of an increasing brown bear population, evidence for presaturation dispersal. – *J. Appl. Ecol.* 67: 708–715.
- Thorn, M. et al. 2011. Brown hyaenas on roads: estimating carnivore occupancy and abundance using spatially auto-correlated sign survey replicates. – *Biol. Conserv.* 144: 1799–1807.
- Valière, N. et al. 2003. Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. – *Anim. Conserv.* 6: 83–92.
- Van Strien, A. J. et al. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. – *J. Appl. Ecol.* 50: 1450–1458.
- Veran, S. et al. 2016. Modeling spatial expansion of invasive alien species: relative contributions of environmental and anthropogenic factors to the spreading of the harlequin ladybird in France. – *Ecography* 39: 665–675.
- Vucetich, J. A. et al. 1997. Effects of social structure and prey dynamics on extinction risk in gray wolves. – *Conserv. Biol.* 11: 957–965.
- Wabakken, P. et al. 2001. The recovery, distribution, and population dynamics of wolves on the Scandinavian peninsula, 1978–1998. – *Can. J. Zool.* 79: 710–725.
- Whittington, J. et al. 2005. Spatial responses of wolves to roads and trails in mountain valleys. – *Ecol. Appl.* 15: 543–553.
- Wolf Alpine Group 2014. Wolf population status in the Alps: pack distribution and trends up to 2012. – 7th “Wolf Alpine Group” Workshop “wolf monitoring over the Alps – towards a unique approach” 2013 March 19–20th – Jausiers, France.
- Woodroffe, R. 2001. Strategies for carnivore conservation: lessons from contemporary extinctions. – In: Gittleman, J. L. et al. (eds), *Carnivore conservation*. Cambridge Univ. Press, pp. 61–92.
- Yackulic, C. B. et al. 2012. Neighborhood and habitat effects on vital rates: expansion of the barred owl in the Oregon Coast Ranges. – *Ecology* 93: 1953–1966.
- Yackulic, C. B. et al. 2013. Presence-only modelling using MAXENT: when can we trust the inferences? – *Methods Ecol. Evol.* 4: 236–243.
- Yackulic, C. B. et al. 2015. To predict the niche, model colonization and extinction. – *Ecology* 96: 16–23.
- Zurell, D. et al. 2009. Static species distribution models in dynamically changing systems: how good can predictions really be? – *Ecography* 32: 733–744.

Supplementary material (Appendix ECOG-02874 at <www.ecography.org/appendix/ecog-02874>). Appendix 1.

6. Article 2. Utilisation de détections ambiguës pour améliorer les estimations des modèles de distribution d'espèces

Résumé


Alors que les grands carnivores recolonisent l'Europe, leur répartition spatiale doit être étudiée pour déterminer d'une part leur état de conservation et d'autre part évaluer les potentiels impacts avec les activités humaines. Cependant, le suivi efficace des grands carnivores est difficile en raison de leur rareté, de leur comportement élusif et de leurs grands domaines vitaux. Leur suivi peut inclure des observations opportunistes par des citoyens, en plus des enquêtes conçues à partir de protocoles scientifiques. Deux types d'erreurs d'observation peuvent se produire au sein de ces programmes de surveillance : les faux négatifs et les faux positifs. Les faux positifs peuvent être pris en compte dans les modèles de distribution d'espèces (SDMs) qui prennent en compte la détection imparfaite. Les faux positifs qui sont dus à une erreur d'identification de l'espèce ciblée ont rarement été pris en compte dans les SDMs. En général, les gestionnaires et les chercheurs utilisent des méthodes de filtrage des données pour écarter les observations ambiguës des analyses. Cependant, cette méthode de filtrage peut rejeter des informations écologiques précieuses sur la répartition d'une espèce. Nous avons utilisé un modèle d'occupation dynamique qui prend en compte à la fois des faux négatifs et des faux positifs pour analyser conjointement des données comportant des détections non ambiguës et des détections ambiguës. Nous avons utilisé des simulations pour comparer les performances de notre modèle avec un modèle basé uniquement sur des données non ambiguës. Nous avons testé les deux modèles dans 4 scénarios selon lesquels les paramètres qui contrôlent les détections faussement positives et les vraies détections variaient. L'ajout de détections ambiguës a augmenté la précision des estimations des paramètres. Nous avons appliqué notre modèle aux données de suivi du lynx d'Europe (*Lynx lynx*) dans les Alpes européennes. L'incorporation de détections ambiguës a permis d'obtenir des estimations plus précises des paramètres écologiques et de mettre en évidence d'autres sites occupés dans des zones où l'espèce est susceptible d'être présente. Dans l'ensemble, nous avons constaté que les données ambiguës devraient être prises en compte dans l'étude de la distribution des grands carnivores grâce à l'utilisation de modèles dynamiques d'occupancy qui prennent compte les erreurs d'identification.

Contributions

J'ai développé le modèle d'occupancy dynamique prenant en compte les incertains et je l'ai ajusté sur les données. J'ai proposé une interprétation des résultats obtenus que nous avons discutée collectivement. J'ai rédigé un premier jet de l'article qui a été corrigé par mes co-auteurs. Les données sur le lynx ont été fournies par le SCALP via Anja Molinari-Jobin. J'ai repris le modèle d'occupancy dynamique initialement développé par Marc Kéry pour y intégrer les données incertaines. Thierry Chambert et David Miller m'ont grandement aidé pour la relecture du manuscrit. J'ai adapté des codes écrits par Olivier Gimenez pour les simulations.



Use of ambiguous detections to improve estimates from species distribution models

Julie Louvrier ^{1,2*}, Anja Molinari-Jobin,³ Marc Kéry,⁴ Thierry Chambert,¹ David Miller,⁵ Fridolin Zimmermann,³ Eric Marboutin,⁶ Paolo Molinari,⁷ Oliver Müller,⁸ Rok Černe,⁹ and Olivier Gimenez¹

¹CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

²Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux déprédateurs, Parc Micropolis, 05000, Gap, France

³KORA, Thunstrasse 31, 3074, Muri, Switzerland

⁴Swiss Ornithological Institute, 6204, Sempach, Switzerland

⁵Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, U.S.A.

⁶ONCFS, Gières, France

⁷Italian Lynx Project, 33018, Tarvisio, Italy

⁸National Office of Forests, Vaduz, Liechtenstein

⁹Slovenia Forest Service, Ljubljana, Slovenia

Abstract: *As large carnivores recover throughout Europe, their distribution needs to be studied to determine their conservation status and assess the potential for human-carnivore conflicts. However, efficient monitoring of many large carnivore species is challenging due to their rarity, elusive behavior, and large home ranges. Their monitoring can include opportunistic sightings from citizens in addition to designed surveys. Two types of detection errors may occur in such monitoring schemes: false negatives and false positives. False-negative detections can be accounted for in species distribution models (SDMs) that deal with imperfect detection. False-positive detections, due to species misidentification, have rarely been accounted for in SDMs. Generally, researchers use ad hoc data-filtering methods to discard ambiguous observations prior to analysis. These practices may discard valuable ecological information on the distribution of a species. We investigated the costs and benefits of including data types that may include false positives rather than discarding them for SDMs of large carnivores. We used a dynamic occupancy model that simultaneously accounts for false negatives and positives to jointly analyze data that included both unambiguous detections and ambiguous detections. We used simulations to compare the performances of our model with a model fitted on unambiguous data only. We tested the 2 models in 4 scenarios in which parameters that control false-positive detections and true detections varied. We applied our model to data from the monitoring of the Eurasian lynx (*Lynx lynx*) in the European Alps. The addition of ambiguous detections increased the precision of parameter estimates. For the Eurasian lynx, incorporating ambiguous detections produced more precise estimates of the ecological parameters and revealed additional occupied sites in areas where the species is likely expanding. Overall, we found that ambiguous data should be considered when studying the distribution of large carnivores through the use of dynamic occupancy models that account for misidentification.*

Keywords: false positives, large carnivores, lynx, occupancy models, species imperfect detection

Uso de Detecciones Ambiguas para Mejorar las Estimaciones a partir de Modelos de Distribución de Especies

Resumen: *Conforme los carnívoros mayores se recuperan en toda Europa, su distribución requiere ser estudiada para determinar su estado de conservación y para evaluar el potencial de conflictos entre humanos y carnívoros. Sin embargo, el monitoreo eficiente de muchas especies de carnívoros mayores es complicada debido a su rareza, comportamiento elusivo y las grandes extensiones de su ámbito de hogar. Su monitoreo*

*email julie.louvrier@cefe.cnrs.fr

Article impact statement: *Use of ambiguous detections can improve understanding of large-carnivore distribution dynamics. Paper submitted December 29, 2017; revised manuscript accepted July 10, 2018.*

*puede incluir avistamientos oportunistas por parte de los ciudadanos, además de los censos diseñados. Pueden ocurrir dos tipos de errores de detección en dichos métodos de monitoreo: negativos falsos y negativos positivos. La detección de los falsos negativos puede justificarse en los modelos de distribución de especies (MDE) que manejan la detección imperfecta. La detección de falsos positivos por causa de la identificación errónea rara vez se justifica en los MDE. Los investigadores usan generalmente métodos con filtración de datos ad hoc para descartar las observaciones ambiguas previo al análisis. Estas prácticas pueden descartar información ecológica variable sobre la distribución de una especie. Investigamos los costos y beneficios de la inclusión de tipos de datos que podrían contener falsos positivos en lugar de descartarlos de los MDE para carnívoros mayores. Usamos un modelo dinámico de ocupación que justificó simultáneamente los falsos positivos y falsos negativos para analizar en conjunto los datos que incluían tanto las detecciones no ambiguas como las ambiguas. Usamos simulaciones para comparar el desempeño de nuestro modelo con el de un modelo ajustado solamente para datos no ambiguos. Probamos los dos modelos en cuatro escenarios en los que variaron los parámetros que controlan la detección de falsos positivos y de detecciones verdaderas. Aplicamos nuestro modelo a datos del monitoreo del lince euroasiático (*Lynx lynx*) en los Alpes. La suma de las detecciones ambiguas incrementó la precisión de las estimaciones de los parámetros. Para el lince euroasiático, la incorporación de las detecciones ambiguas produjo estimaciones más precisas de los parámetros ecológicos y reveló sitios ocupados adicionales en áreas en donde la especie probablemente se esté expandiendo. En general, encontramos que los datos ambiguos deberían ser considerados cuando se estudia la distribución de carnívoros mayores por medio del uso de modelos dinámicos de ocupación que justifican la identificación errónea.*

Palabras Clave: carnívoros mayores, detección imperfecta de especies, lince, modelos de ocupación, positivos falsos

摘要: 随着欧洲大型食肉动物的恢复, 我们需要研究这些物种的分布以确定它们的保护现状, 并评估人兽冲突的可能性。然而, 由于大型食肉动物的稀有性、回避行为和较大的活动范围, 对它们进行有效监测困难重重。大型食肉动物的监测除了设计调查外, 还可以包含公民的偶然目击。这样的监测方案可能出现两类检测误差: 假阴性和假阳性。假阴性的监测结果可用于物种分布模型分析, 该模型可处理有缺陷的监测。而因物种鉴定错误产生的假阳性结果在物种分布模型中却很少被考虑在内。通常情况下, 研究者在分析前会用特定的数据过滤方法来去除不确定的观察数据。这种做法也可能丢掉对物种分布有价值的生态学信息。我们研究了保留可能含有假阳性发现的数据类型 (而不是丢弃这些数据) 对大型食肉动物物种分布模型的利弊。我们利用同时考虑假阴性和阳性的动态占有模型, 共同分析了包含确定发现和不确定发现的数据。通过模拟, 我们比较了这个模型和仅用真实数据拟合的模型的表现。我们选择了不同的控制假阳性和真阳性的参数, 在四种情况下检验了这两个模型。我们还将这一模型应用于欧洲阿尔卑斯山脉的欧亚猞猁 (*Lynx lynx*) 的监测数据, 加入不确定的发现数据可以提高参数估计的精确度。对欧亚猞猁来说, 加入不确定的发现可以对生态学参数作出更精确的估计, 揭示这个物种的分布范围可能正在扩张。总之, 我们的研究表明, 在研究大型食肉动物的分布时, 可以用考虑到物种鉴定错误的动态占有模型来利用不确定的数据。 **翻译:** 胡怡思; **审校:** 聂永刚

关键词: 假阳性, 大型食肉动物, 猞猁, 占据模型, 物种不充分检测

Introduction

The distribution and abundance of large carnivores in many parts of the world has been declining for centuries because of habitat loss and human persecution (Ripple et al. 2014). Thanks to active conservation measures, several species of large carnivores have been recently expanding their ranges substantially in Europe. As a result, most European countries currently host at least 1 viable population of a large predator (Chapron et al. 2014). This expansion has led to the emergence of conflicts with humans (Ripple et al. 2014). In this context, accurate distribution mapping (i.e., species distribution models [SDMs]) (Elith & Leathwick 2009) is essential for determining conservation status and recovery success (IUCN 2012); targeting potential areas of occurrence; understanding large carnivore range dynamics; identifying possible areas for future population establishment

(Chapron et al. 2014); and mitigating conflicts associated with the recovery of large carnivores (Guillera-Aroita et al. 2015), for example, depredation of livestock by wolves (Marucco & Mcintire 2010). However, carnivore rarity, elusive behavior, and low densities render efficient monitoring difficult (Ripple et al. 2014).

The monitoring of large carnivores in Europe relies on several survey methods implemented by professionals and members of the public (citizens). In particular, citizen participation increases the ability to survey large areas over extended periods, an effort that would be costly if done by professionals only (Molinari-Jobin et al. 2018). A primary goal of citizen science (CS) is to produce reliable data and information that scientists or decision makers can use (McKinley et al. 2017), and it is becoming an important tool with which to study the distribution, abundance, and species richness of plants and animals (Silvertown 2009; Dickinson et al. 2012). However,

CS-generated data present potential quality problems, especially when the goal is to build SDMs.

Difficulty detecting large carnivores means animals can be missed at sites where they are present (i.e., producing false-negative observations). Occupancy models were developed to deal with false-negative errors (Guillera-Arroita 2017) and are recommended for analyzing CS data (Isaac et al. 2014). Although data sets produced by CS are valuable (Kosmala et al. 2016), professionals are better able to detect or identify species of interest, which diminishes the risk of identification errors (Fitzpatrick et al. 2009). False positives can occur when a species is detected through misidentification at a site where it does not occur (Miller et al. 2011). Recent studies have demonstrated the importance of accounting for misidentification for SDMs (Miller et al. 2011, 2013; Chambert et al. 2015). Ignoring misidentification may lead to overestimating a species range (Royle & Link 2006; McClintock et al. 2010).

Usually large carnivores are surveyed with indirect methods, for example, observations of tracks, scat, and prey remains or use of camera traps (Molinari-Jobin et al. 2018). Observations are then filtered by experts to assess the reliability of evidence of the observed presence. Recent studies of the distribution of European large carnivores have been based on only reliable observations, that is, those remaining after discarding ambiguous detections and validation by experts (Molinari-Jobin et al. 2018). This means that some observations are discarded, even though they may contain relevant ecological information on the species distribution. This raises the question of whether this information can somehow be extracted and made useful in the context of SDMs.

We investigated the pros and cons of removing ambiguous detections in SDMs of large carnivores versus keeping all records and formally accounting for misidentification. We used a dynamic occupancy model that accounts for both false-negative and false-positive errors (Miller et al. 2011, 2013) to jointly analyze unambiguous and ambiguous detections. To assess the performance of this approach, we performed a simulation study in which we compared the analysis of unambiguous and ambiguous detections with the use of unambiguous detections only.

We applied these methods to an SDM of the Eurasian lynx (*Lynx lynx*) in the European Alps (Molinari-Jobin et al. 2018). We expected improved precision in ecological parameter estimates when all data were included in an analysis, despite having to accommodate additional nuisance parameters to deal with misidentification.

Methods

Occupancy Model Accounting for Misidentification

Dynamic occupancy models allow estimation of occupancy and its temporal dynamics as a function of local ex-

inction and colonization probabilities while accounting for imperfect species detection (MacKenzie et al. 2003). These models can be formulated as state-space models to separate the state process (i.e., whether or not a species is present at a site and how presence changes over time) from the observational process (i.e., whether a species is observed at a site during a given period depending on whether or not it was actually present) (Royle & Kéry 2007). We defined $z_{i,1}$ as the initial latent occurrence state of site i ($z = 1$, presence; $z = 0$, absence) and $z_{i,t}$ as the latent state for site i at time t . The state process is initiated by the initial occupancy probability ($\psi_{i,1}$) for site i and governed by colonization probability ($\gamma_{i,t}$) (the probability that a site i that is not occupied at time t will become occupied at time $t + 1$) and extinction probability ($\varepsilon_{i,t}$) (the probability that an occupied site i at time t will become unoccupied at time $t + 1$). We modeled $z_{i,1}$ as a draw from a Bernoulli distribution with probability $\psi_{i,1}$. All subsequent latent states $z_{i,t}$ for $t > 1$ were draws from another Bernoulli distribution that combines both possible extinction and colonization events:

$$z_{i,t+1} | z_{i,t} \sim \text{Bernoulli}(z_{i,t}(1 - \varepsilon_{i,t}) + (1 - z_{i,t}) \gamma_{i,t}). \quad (1)$$

If a site is occupied in year (or season) t , it will still be occupied with probability $1 - \varepsilon_{i,t}$ or if it is unoccupied, it will become occupied with probability $\gamma_{i,t}$. Each year (or season), a site is surveyed several times (j) within a year or season. Site occupancy models rely on satisfaction of the site-closure assumption, whereby the latent occurrence state of a site does not change within a sampling season, whereas occupancy dynamics (colonization and extinction) happen between years (or seasons).

In addition to the state process, the observation process leads to the data $y_{i,j,t}$: the observed state of site i during a secondary survey j within year (or season) t . Hereafter, we do not use the indices when possible to ease reading of the model parameters. Here, $y = 0$ denotes no detection, $y = 1$ an unambiguous detection, and $y = 2$ an ambiguous detection. To account for unambiguous and ambiguous detections, we followed the formulation of Miller et al. (2013). We defined an additional parameter $d_{i,j,t}$ that had a value of 1 if any detection (ambiguous or unambiguous) was made at site i during survey j within year t and a value of 0 if there were no detections. For occupied sites, by definition, $d = 1$ and denoted a true detection. For unoccupied sites, $d = 1$ was a false-positive detection. For both occupied and unoccupied sites, $d = 0$ meant no detection and thus $y = 0$. At an occupied site, the possible observations were no detection ($y = 0$), unambiguous detection ($y = 1 | d = 1$), or ambiguous detection ($y = 2 | d = 1$). For occupied sites, the probability of a true detection (i.e., $d = 1$) during a secondary sampling occasion (or survey) was defined as $P(d = 1 | z = 1)$ (hereafter p_{11}). The probability that a true detection will be classified as unambiguous was

$P(y = 1 | d = 1)$ (hereafter b). The probability of an unambiguous detection was $p_{11}b$, and the probability for an ambiguous detection (i.e., $y = 2$) for an occupied site was $p_{11}(1 - b)$. For unoccupied sites (i.e., $z = 0$), by definition, unambiguous detections ($y = 1 | d = 1$) did not occur; thus, the only possible observations were an ambiguous detection ($y = 2 | d = 1$), which in this case is a false positive, or no detection ($y = 0$). The probability of a false positive detection (i.e., $d = 1$) occurring at an unoccupied site i during a secondary sampling occasion (or survey) j was $P(d = 1 | z = 0)$ (hereafter p_{10}). Then the probabilities, unconditional on state z of a site, of recording the 3 possible observed states (y) were

$$\begin{aligned} P(y = 0) &= P(z = 1)P(d = 0|z = 1) \\ &+ P(z = 0)P(d = 0|z = 0) = \psi(1 - p_{11}) \\ &+ (1 - \psi)(1 - p_{10}) \text{ for no detection;} \\ P(y = 1) &= P(z = 1)P(d = 1|z = 1)P(y = 1|d = 1) \\ &= \psi p_{11}b \text{ for unambiguous detection; and} \\ P(y = 2) &= P(z = 1)P(d = 1|z = 1)P(y = 2|d = 1) \\ &+ P(z = 0)P(d = 1|z = 0) = \psi p_{11}(1 - b) \\ &+ (1 - \psi)p_{10} \text{ for ambiguous detection.} \end{aligned}$$

Simulations

We conducted a simulation to examine the performance of a dynamic occupancy model that accounted for possible false positives (model unambiguous or ambiguous [MUA]) relative to a dynamic occupancy model that accounted only for false negatives (fitted with unambiguous data only) (model unambiguous [MU]). To assess the ability of both models to estimate ecological parameters, we tested 4 scenarios in which parameters that control false-positive detections and true detections varied (Table 1).

Because the ecological parameters have an influence on the amount of detections produced, occupancy probability was either high or low. In the high occupancy scenario, we set the initial occupancy probability ψ_1 at 0.8, the colonization probability γ at 0.4, and extinction probability at 0.1 to maintain a high occupancy probability. This scenario corresponds to a fairly well-established species reflected by its high occupancy probability across time. In the low occupancy scenario, we set the initial occupancy probability ψ_1 at 0.1, the colonization probability γ at 0.1, and extinction probability at 0.1 to maintain a low occupancy probability. This scenario corresponds to a rare species with a low occupancy probability across time.

Detection parameters also influence the amount of false-positive and true-positive detections. True detections are controlled by p_{11} and b . Therefore, in both high and low occupancy scenarios, we considered 2 situations

in which b was either high (i.e., set at 0.8) or low (i.e., set at 0.5), for a total of 4 scenarios. For all scenarios, we set p_{11} at 0.4. When $b = 0.8$, most of the true detections were classified as unambiguous. This scenario corresponds to monitoring of a species that is not easily mistaken for another or monitoring conducted by people trained to recognize accurately the presence signs of the species. When $b = 0.5$, a larger part of the true detections was classified as ambiguous. This scenario corresponds to the monitoring of a species that can easily be mistaken or done by untrained people, for instance from the general public. The amount of false-positive detections was controlled by p_{10} . In all 4 scenarios, we looked at how the models performed under 7 different values of p_{10} (range 0.01–0.3), for a total of 28 different simulation scenarios. Finally, because our main objective was to assess the effect of accounting for ambiguous data, environmental variation was not included in our simulations. The high occupancy high b scenario is referred to as HH; the high occupancy low b scenario is HL; the low occupancy high b scenario is LH; and the low occupancy low b is LL.

In our simulations, we generated data for 100 sites over 5 years and 3 surveys. To remain realistic in the simulations, the number of surveys mimicked the case-study characteristics. For each scenario, we simulated 500 data sets and fitted both models to each data set. For the initial occupancy probability, the colonization probability, and the extinction probability in both models in each scenario, we calculated the relative bias and mean squared error (MSE).

Eurasian Lynx Case Study

After its total eradication in the Alps by around 1930, the Eurasian lynx (*Lynx lynx*) has been reintroduced multiple times since 1970 in Switzerland, Italy, Austria, and Slovenia (Molinari-Jobin et al. 2018). In the 1990s, experts from the 7 Alpine countries set up the international lynx monitoring program (Status and Conservation of the Alpine Lynx Population). To ensure coverage of its entire territory, the monitoring of the elusive lynx relies on >1300 trained experts (official game wardens and forest service personnel) in 7 Alpine countries. Hunters, naturalists, and other citizens also may be part of the monitoring network. Professional network members search actively for signs and check signs that are reported to them by citizens. We classified signs of presence into 3 reliability categories: C1, hard facts (e.g., dead lynx, lynx removed from the wild as a young orphan and put into captivity, lynx photos, and genetic samples); C2, detections confirmed by a lynx expert (livestock killed by lynx, wild prey remains, and tracks); and C3, data that could not be verified by experts (unverified tracks and wild prey remains) and unverifiable data (e.g., sighting, scat, and vocalization). We treated C1 and C2 data as unambiguous detections, assuming there were no false-positive

Table 1. Parameter values for the 4 simulation scenarios in which parameters controlling the total amount of false-positive and true detections varied.

Scenarios	Initial occupancy probability (ψ_1)	Colonization probability (γ)	Probability of classifying a true detection as unambiguous (b)	False-positive detection probability (p_{10})
High occupancy high b (HH)	0.8	0.4	0.8	0.01
				0.5
				0.10
				0.15
				0.20
				0.25
High occupancy low b (HL)	0.8	0.4	0.5	0.30
				0.01
				0.5
				0.10
				0.15
				0.20
Low occupancy high b (LH)	0.1	0.1	0.8	0.25
				0.01
				0.30
				0.5
				0.10
				0.15
Low occupancy low b (LL)	0.1	0.1	0.5	0.20
				0.01
				0.30
				0.5
				0.10
				0.15

detections in these data, and C3 data as ambiguous detections. From 1995 to 2014, 8415 observations (67%) were classified as unambiguous detections and 3991 (33%) as ambiguous. If unambiguous and ambiguous detections occurred at a site, we accounted for the unambiguous detections only. Nondetections were obtained on sites that were sampled but where no lynx presence was reported during a survey within a year. Molinari-Jobin et al. (2018) fitted a dynamic occupancy model with unambiguous detections only (i.e., they used our model MU) to assess the effects of environmental covariates on different parameters of the model and to assess distribution-based population trends. A 10 × 10 km grid was used to define the distribution units that corresponded to the approximate home range size of a female lynx in the Alps (Molinari-Jobin et al. 2018). Surveys were defined as 3 replicated 2-month periods: November–December, January–February, and March–April. We used the same data set as Molinari-Jobin et al. (2018), but we added the C3 data and fitted a dynamic occupancy model that combined both unambiguous and ambiguous detections

(MUA). We used the same covariates for the parameters that were common to the models MU and MUA. We considered the effects of forest cover and distance to the release site on ψ_1 ; the effects of year, forest cover, and number of observed occupied contiguous neighbors on ε ; and the same effects plus that of human density and elevation on γ .

For the new parameters in MUA, p_{11} and p_{10} , we used the effect of elevation and forest cover and a random site-by-winter effect to accommodate unmodeled spatial heterogeneity in detection rates in every combination of site and winter. A network covariate was also included to account for heterogeneity in sampling effort in time and space. This covariate took the following values based on the amount of effort for the location and time period: 0, no information available regarding the sampling effort, in which case we assumed it was small but never null because of the presence of at least game wardens and forest service agents who have not been trained in the recognition of lynx signs but report suspect cases (Molinari-Jobin et al. 2012); 1, presence of trained members of the

lynx monitoring network on the site; and 2, members of the lynx monitoring network actively searching for lynx signs. We also considered a linear year effect (i.e., an annual trend) on p_{10} to investigate whether this probability decreased as observers gained experience over time. Finally, we kept the probability b to classify a true positive detection as an unambiguous constant. We considered the effect of a covariate significant if its 95% credible interval (CRI) did not overlap 0.

To evaluate the added value of incorporating the C3 data (ambiguous detections) into the analysis, we compared the maps of occupancy produced by the 2 models by calculating and mapping the difference in the site- and year-specific estimates of realized occurrence $\hat{z}_{i,t}$ (MU) - $\hat{z}_{i,t}$ (MUA).

Codes to run the simulations and fit the models described above are in Supporting Information.

Results

Simulations

For MSE, MUA performed better than MU in all 4 scenarios when $p_{10} \leq 0.15$ (Supporting Information). Above this value both models performed equally well except in 1 scenario and for 1 parameter when estimating the ecological parameters: MUA estimated the colonization probability γ less precisely than MU only in the HL scenario for values of p_{10} between 0.20 and 0.30. The MSE was highest (range 0.04–0.25) in the HL scenario and ranged from 0.04 to 0.14 in the HH scenario. The MSE was lowest in the LH scenario and ranged from 0.02 to 0.06 and 0.02 to 0.11 in the LL scenario.

Both models estimated ψ_1 and γ with biases below or equal to 5% in the HH, HL, and LH scenarios (Supporting Information). In the LL scenario, MU estimated ψ_1 with a bias above 5% (up to 8%) and MUA had a lower bias than MU. For ε , MUA performed better or equivalently above 5% in terms of bias in the scenarios HH and HL and worse or equivalently above 5% in the LH and LL scenarios.

Lynx Case Study

When we fitted the MUA with both unambiguous and ambiguous detections (i.e., for C1, C2, and C3 data), p_{11} was higher on sites with a high forest cover and appeared to vary according to the season and network (Table 2). Elevation had no effect on p_{11} . The p_{10} was higher on sites with a high forest cover and varied according to network (Table 2). Although elevation and season had no significant effect on p_{10} , this probability decreased over time (Table 2). Both models gave similar estimates for ψ_1 , ε , and γ , but MUA produced more precise estimates than MU (Supporting Information).

The b was estimated at 0.81 with high precision (CRI 0.79–0.83). At the beginning of the study period, in the winter 1995 and 1996, we estimated the mean ψ over all sites at 0.04 (CRI 0.03–0.07), p_{11} at 0.11 (CRI 0.10–0.25), and p_{10} at 0.006 (CRI 0.004–0.01). For the end of the study period, winter 2013 and 2014, we estimated the mean ψ at 0.1 (CRI 0.0899–0.11), p_{11} at 0.17 (CRI 0.09–0.24), and p_{10} at 0.007 (CRI 0.003–0.010). MUA estimated a few more occupied sites than MU for both winters 1995 and 1996 and 2013 and 2014 (4 [1995 and 1996] to 13 [2013 and 2014]) (Fig. 1) and estimated occupied sites that were estimated occupied by MU too. The additional sites that were estimated occupied from MUA were sites where ambiguous detections occurred (Fig. 1).

Discussion

Assessing the distribution of large carnivores at large scales is vital information for assessing their conservation status and abundance (IUCN 2012; Jedrzejewski et al. 2018), identifying potential conflict areas (Marucco & Mcintire 2010), and, for successful management, understanding the mechanism behind distribution dynamics (Eriksson & Dalerum 2018). Producing more precise and less biased estimates by adding ambiguous data with a model accounting for false-positive detections can bring new insights to the distribution of species where getting unambiguous data is challenging. Due to the large areas involved, the monitoring of large carnivores in Europe relies on a large network of both professional and nonprofessional observers (Louvrier et al. 2018; Molinari-Jobin et al. 2018). Although false-negative detections have received much attention in the literature of species distribution modeling with the rise of occupancy models (MacKenzie et al. 2003; Bailey et al. 2014), dealing with ambiguous detections has been studied much less (Miller et al. 2011; Chambert et al. 2015). Here, our simulations demonstrated that jointly analyzing unambiguous and ambiguous detections with the appropriate dynamic occupancy models can lead to increased precision in the estimates of ecological parameters when p_{10} is low. When this probability was >0.20 , both models estimated ecological parameters with almost equivalent precision, which varied between its highest values in the high occupancy scenarios and its lowest values in the low occupancy scenarios. Both models produced estimates of ecological parameters with low bias except for one ecological parameter in one specific scenario. When looking at the results of the lynx analysis, adding ambiguous data helped produce more precise estimates and provided additional spatial information that improved inference in areas where the species likely occurred at very low density (e.g., at a colonization front).

Table 2. Parameters estimates for the detection probabilities from the dynamic occupancy model accounting for unambiguous data only and accounting for unambiguous and ambiguous data. Both models were fitted on Alpine lynx data from 1995 to 2014.^a

	<i>Model with unambiguous data only (MU)</i>				<i>Model with ambiguous and unambiguous data only (MUA)</i>				<i>Model with ambiguous and unambiguous data only (MUA)</i>			
	<i>Detection probability (p)</i>				<i>True detection probability (p₁₁)</i>				<i>False-positive detection probability (p₁₀)</i>			
	<i>mean</i>	<i>SD</i>	<i>x</i>	<i>y</i>	<i>mean</i>	<i>SD</i>	<i>x</i>	<i>y</i>	<i>mean</i>	<i>SD</i>	<i>x</i>	<i>y</i>
Intercept	-3.88	0.46	-4.88	-3.04	-3.14	0.47	-4.15	-2.33	-5.37	0.29	-5.96	-4.80
Effect of elevation	-0.11 ^b	0.05 ^b	-0.20 ^b	-0.01 ^b	-0.07	0.04	-0.16	0.01	-0.02	0.06	-0.13	0.09
Effect of forest	0.63 ^b	0.07 ^b	0.50 ^b	0.75 ^b	0.67 ^b	0.06 ^b	0.55 ^b	0.79 ^b	0.37 ^b	0.06 ^b	0.26 ^b	0.49 ^b
Effect of season 2	0.26 ^b	0.07 ^b	0.12 ^b	0.40 ^b	0.22 ^b	0.07 ^b	0.08 ^b	0.35 ^b	-0.14	0.12	-0.38	0.09
Effect of season 3	0.42 ^b	0.07 ^b	0.28 ^b	0.57 ^b	0.42 ^b	0.07 ^b	0.29 ^b	0.56 ^b	-0.04	0.12	-0.25	0.19
Effect of network 1	1.21 ^b	0.21 ^b	0.77 ^b	1.60 ^b	0.92 ^b	0.20 ^b	0.54 ^b	1.31 ^b	0.37 ^b	0.14 ^b	0.09 ^b	0.66 ^b
Effect of network 2	2.37 ^b	0.23 ^b	1.93 ^b	2.84 ^b	1.95 ^b	0.22 ^b	1.53 ^b	2.40 ^b	2.63 ^b	0.30 ^b	2.02 ^b	3.20 ^b
Residual effect	0.86	0.10	0.65	1.04	0.90	0.06	0.77	1.01	0.42	0.21	0.18	0.92
Effect of country France	-3.49 ^b	0.31 ^b	-4.09 ^b	-2.87 ^b	-2.76 ^b	0.28 ^b	-3.32 ^b	-2.21 ^b	-5.55 ^b	0.25 ^b	-6.05 ^b	-5.06 ^b
Effect of country Italy	-3.40 ^b	0.24 ^b	-3.86 ^b	-2.94 ^b	-2.85 ^b	0.23 ^b	-3.31 ^b	-2.40 ^b	-5.55 ^b	0.24 ^b	-6.01 ^b	-5.07 ^b
Effect of country Switzerland	-2.87 ^b	0.23 ^b	-3.32 ^b	-2.42 ^b	-2.23 ^b	0.22 ^b	-2.68 ^b	-1.81 ^b	-4.75 ^b	0.26 ^b	-5.25 ^b	-4.25 ^b
Effect of country Austria	-4.00 ^b	0.25 ^b	-4.48 ^b	-3.51 ^b	-3.25 ^b	0.21 ^b	-3.67 ^b	-2.84 ^b	-5.62 ^b	0.23 ^b	-6.07 ^b	-5.16 ^b
Effect of country Slovenia	-4.54 ^b	0.35 ^b	-5.23 ^b	-3.86 ^b	-3.83 ^b	0.34 ^b	-4.51 ^b	-3.19 ^b	-4.62 ^b	0.28 ^b	-5.17 ^b	-4.05 ^b
Effect of country Germany	-4.99 ^b	1.39 ^b	-8.28 ^b	-2.65 ^b	-3.93 ^b	1.52 ^b	-7.41 ^b	-1.51 ^b	-6.11 ^b	0.49 ^b	-7.21 ^b	-5.23 ^b
Effect of time in years	-	-	-	-	-	-	-	-	-0.04 ^b	0.02 ^b	-0.09 ^b	-0.01 ^b

^aBoth models were fitted on Alpine lynx data collected from 1995 to 2014. Means are posterior means, *x* and *y* are lower and upper bounds of the 95% Bayesian CI.

^bEffects with 95% Bayesian CI that do not contain zero.

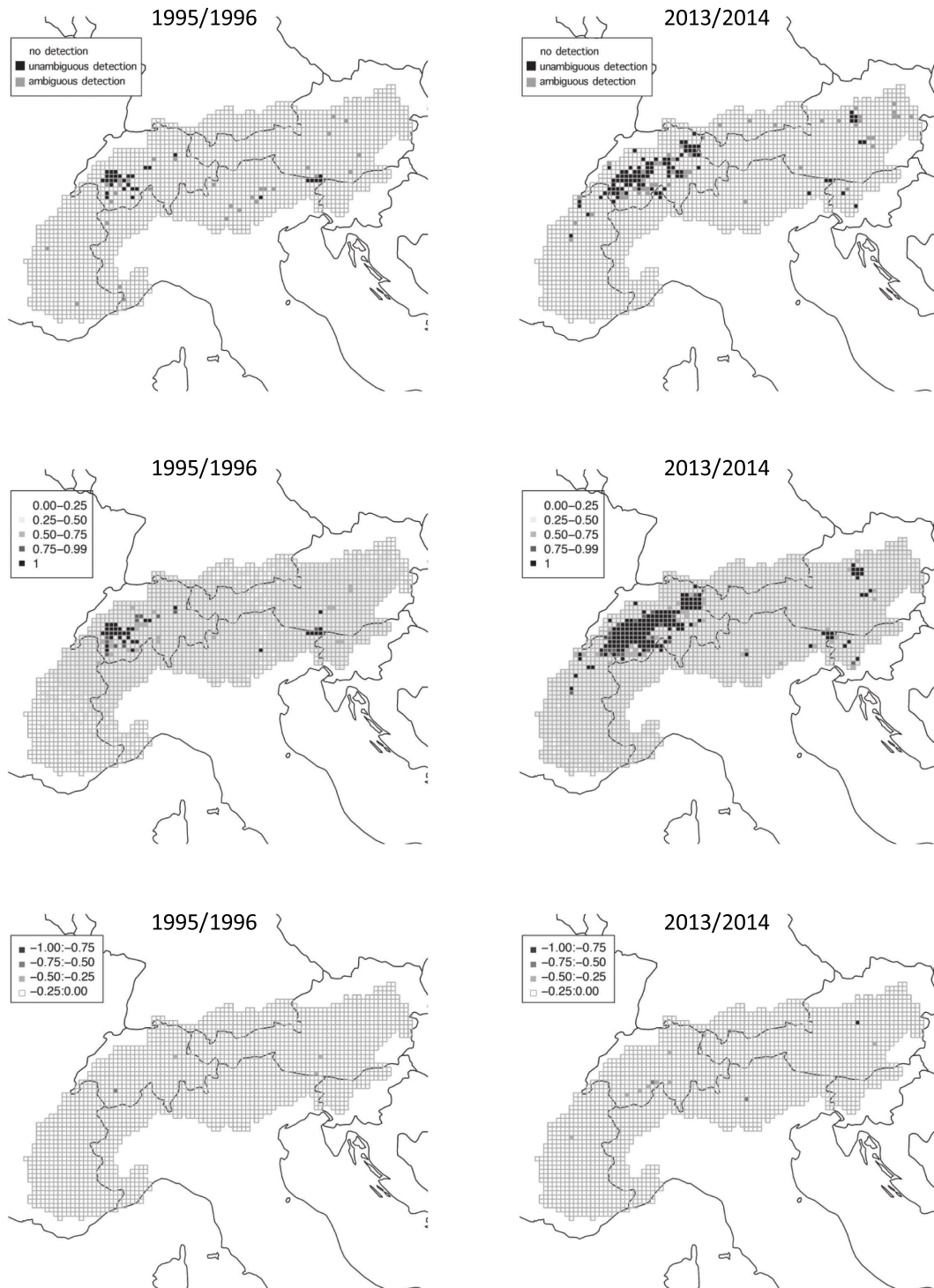


Figure 1. (Top row) Locations of Lynx observation (i.e., raw data) in the winters of 1994 and 1995 and 2013 and 2014 (black sites, unambiguous detections occurred; gray sites, ambiguous detections occurred); (middle row) estimated lynx distribution (probability of occurrence 0–1) in the Alps in the winters of 1994 and 1995 and 2013 and 2014 for the model with ambiguous data (not used on figure) (black sites, sites with a probability of occupancy of 1; white sites, sites with a probability of occupancy of 0); (bottom row) and mapped differences in occupancy estimates between the model with unambiguous data only (value on the right in key) and the model with unambiguous and ambiguous data for the winters 1994 and 1995 (left) and 2013 and 2014 (right) (value on left) (gray sites, model with ambiguous data predicted a higher occupancy probability than the model with unambiguous data only).

Simulation Lessons

MUA performed better than MU in most of the scenarios. Two factors seemed to have an influence on models' performances: the false-positive probability and the occupancy probability. In terms of precision, MUA performed better when p_{10} was low and performed equivalently when p_{10} was high. In the case of a low occupancy probability, the estimates of extinction probability were more biased positively under the MUA than the MU, leading to an overestimation of ε and the distribution. For the other parameters and the other scenarios, MUA produced estimates with low biases. Whether a species occurs at high or low occupancy probability can often be evaluated prior to the analyses based on the knowledge of the species ecology or on previous studies. We recommend always including ambiguous data because in most of the scenarios MUA performed better than or equivalently to MU in terms of both precision and bias for the ecological parameter estimates.

Accounting for Ambiguous Data When Studying Distribution of Large Carnivores

Our SDM exercise with lynx allowed assessment of trends in the distribution of the species, and these trends can be used to inform their conservation status (Guisan & Thuiller 2005). Our covariate effects were similar in direction and magnitude to those estimated by Molinari-Jobin et al. (2018), who fitted the simpler MU to the lynx data with unambiguous detections only (Table 2). (See Molinari-Jobin et al. [2018] for a detailed description of these effects and their possible biological interpretation.) Our results showed that the probability of a false-positive detection decreased over time. This could be due to observers remaining in the network becoming less likely to make false-positive detections over time as they became more experienced in recognizing the species (Jordan et al. 2012). This was corroborated by the fact that the number of ambiguous detections decreased over the study period (Molinari-Jobin et al. 2012). Additionally, the use of camera trapping has increased over time and has led to an increased number of C1 detections and therefore diminished the proportion of C3 in the data sets (Molinari-Jobin et al. 2018). The learning process of citizens in scientific monitoring programs has been studied (Dickinson et al. 2012; Jordan et al. 2012). The general public not only learns through participation, but also becomes more aware of the general ecological issues and more prone to understand scientific research (Bonney et al. 2009). We found that the probability of making a true detection was similar to the probability of detecting the species in MU fitted by Molinari-Jobin et al. (2018). This makes sense because the probability of detecting the species in MU was equal to the probability of making a true detection multiplied by the probability

of classifying a detection as unambiguous. We also found that there was a probability of 0.8 of classifying a true detection as unambiguous. This may be due to the fact that observers in the network were highly competent at detecting the species and producing reliable data. This finding may also reflect that it is relatively easy to identify the signs of presence of lynx because there is almost no confusion possible with other species present in the area. Whenever the focus species can be mistaken for another species, if the quality of data is not sufficient (e.g., dog tracks in the snow mistaken for wolf tracks), true detections can be classified as ambiguous. There can also be false-positive detections from misidentification when b is low. In this case, the amount of true detections in ambiguous data will not be negligible. In a case where b is low and only unambiguous data are used, a large part of true presences can be missed and the resulting distribution will be underestimated (Miller et al. 2011).

The occupancy estimates under both models suggest the lynx case study corresponds to the LH simulation scenario (compare Table 1 with Supporting Information). For the distribution maps produced by MUA, adding ambiguous detections brought new and useful information. Some sites were estimated as occupied by MUA, whereas these same sites were estimated as unoccupied by MU (Fig. 1). Because of the low occupancy of the lynx and its elusive behavior, the number of times the species was detected was very low. Because the probability to classify a detection as unambiguous b was high, only a few true detections were classified as ambiguous, which may explain why adding them did not change the parameter estimates but helped produce more precise estimates. In turn, adding ambiguous detections provided new insights related to management of a protected species (Guillera-Arroita et al. 2015). The sites we found to be occupied because of the incorporation of ambiguous detections could likely represent areas where the species is currently expanding. These same sites also may be places where lynx have not occurred before and negative interactions may occur due to the novelty of lynx presence. Sites that appeared occupied after including ambiguous data can inform the prediction of location of potential conflicts. Finally, if the objective is to map the colonization front to, for example, mitigate conflicts, ambiguous data should be included.

Recommendations

Dynamic occupancy models in general provide a powerful and natural analytical framework for changing species distributions (Kéry et al. 2013). More specifically, dynamic occupancy models accounting for misidentification represent a powerful method to deal with detections that cannot be categorized as certain in species distribution modeling. We recommend careful categorization of field observations into unambiguous or ambiguous

detections, for instance by using several experts to classify the detections and a standardized filtering classification process to avoid false-positive detections mistakenly classified as reliable data. This filtering process also allows rejecting detections that can be identified easily as false positives. If some detections cannot be checked by experts for instance and cannot be classified as unambiguous, observers may need to visit the sites where these detections were made to get more reliable detections. Even though occupancy models can deal with ambiguity, survey should be designed and data collected so as to avoid the production of false-positive detections or at least reduce their proportion. In the case of data from CS projects, models accounting for false-positive detections are a relevant tool to assess species distribution if a classification of detections is made (e.g., unambiguous vs. ambiguous). In the case of a species occurring at low density, such as the Eurasian lynx, additional information can provide new insights into the species distribution and help target specific sites where the species is likely to occur in the future.

Acknowledgment

We gratefully acknowledge the help of people who have collaborated with the Status and Conservation of the Alpine Lynx Population program to monitor the lynx in the Alps. This work was partly funded by a grant from CNRS and “Mission pour l’Interdisciplinarité” through its “Osez l’interdisciplinarité” call.

Supporting Information

The scripts used for running the simulations (Appendix S1) and to fit our model to the lynx data (Appendix S2), the resulting biases and MSE from 4 scenarios of the simulation study (Appendix S3), and the estimates of ecological parameters in the MU and MUA models (Appendix 4) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

- Bailey LL, Mackenzie DI, Nichols JD. 2014. Advances and applications of occupancy models. *Methods in Ecology and Evolution* **5**:1269–1279.
- Bonney R, Cooper CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, Shirk J. 2009. Citizen Science: a developing tool for expanding science knowledge and scientific literacy. *Bioscience* **59**:977–984.
- Chambert TC, Miller DAW, Nichols JD. 2015. Modeling false positive detections in species occurrence data under different study designs. *Ecology* **96**:332–339.
- Chapron G, et al. 2014. Recovery of large carnivores in Europe’s modern human-dominated landscapes. *Science* **346**:1517–1519.
- Dickinson JL, et al. 2012. The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment* **10**:291–297.
- Elith J, Leathwick JR. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**:677–697.
- Eriksson T, Dalerum F. 2018. Identifying potential areas for an expanding wolf population in Sweden. *Biological Conservation* **220**:170–181.
- Fitzpatrick MC, Preisser EL, Ellison AM, Elkinton JS. 2009. Observer bias and the detection of low-density populations. *Ecological Applications* **19**:1673–1679.
- Guillera-Arroita G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* **40**:281–295.
- Guillera-Arroita G, et al. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* **24**:276–292.
- Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**:993–1009.
- Isaac NJB, Strien AJVan, August TA, Zeeuw MPDe, Roy DB. 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution* **5**:1052–1060.
- IUCN. 2012. IUCN red list categories and criteria: version 3.1. 2nd edition. IUCN, Gland, Switzerland and Cambridge, United Kingdom.
- Jedrzejewski W, et al. 2018. Estimating large carnivore populations at global scale based on spatial predictions of density and distribution—application to the jaguar (*Panthera onca*). *PLOS ONE* **13**(e0194719) <https://doi.org/10.1371/journal.pone.0194719>.
- Jordan RC, Ballard HL, Phillips TB. 2012. Key issues and new approaches for evaluating citizen-science learning outcomes. *Frontiers in Ecology and the Environment* **10**:307–309.
- Kéry M, Guillera-arroita G, Lahoz-monfort JJ. 2013. Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography* **40**:1463–1474.
- Kosmala M, Wiggins A, Swanson A, Simmons B. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* **14**:551–560.
- Louvrier J, Duchamp C, Lauret V, Marboutin E, Cubaynes S, Choquet R, Miquel C, Gimenez O. 2018. Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data. *Ecography* **41**:647–660.
- MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* **84**:2200–2207.
- Marucco F, Mcintire EJB. 2010. Predicting spatio-temporal recolonization of large carnivore populations and livestock depredation risk: wolves in the Italian Alps. *Journal of Applied Ecology* **47**:789–798.
- McClintock BT, Bailey LL, Pollock KH, Simons TR. 2010. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology* **91**:2446–2454.
- McKinley DC, et al. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation* **208**:15–28.
- Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA. 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* **92**:1422–1428.
- Miller DAW, Nichols JD, Gude JA, Rich LN, Podruzny KM, Hines JE, Mitchell MS. 2013. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from

- public survey data. PLOS ONE **8**(e65808) <https://doi.org/10.1371/journal.pone.0065808>.
- Molinari-Jobin A, et al. 2018. Mapping range dynamics from opportunistic data: spatiotemporal modelling of the lynx distribution in the Alps over 21 years. *Animal Conservation* **21**:168–180.
- Molinari-Jobin A, et al. 2012. Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. *Animal Conservation* **15**:266–273.
- Ripple WJ, Beschta RL, Fortin JK, Robbins CT. 2014. Trophic cascades from wolves to grizzly bears in Yellowstone. *Journal of Animal Ecology* **83**:223–233.
- Royle JA, Kéry M. 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* **88**:1813–1823.
- Royle JA, Link AW. 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* **87**:835–841.
- Silvertown J. 2009. A new dawn for citizen science. *Trends in Ecology & Evolution* **24**:467–471.

Encadré No 2 : Intégrer les données incertaines pour modéliser la distribution du loup en France

Dans le cadre de ma thèse, j'ai eu l'opportunité d'encadrer Valentin Lauret lors de son stage de Master 2 avec Olivier Gimenez et Christophe Duchamp. Ce stage a porté sur l'intégration des données incertaines afin de modéliser la distribution du loup en France en appliquant un modèle prenant en compte les erreurs d'identification dans les données incertaines.

En effet, des faux positifs dus à des erreurs d'identification d'espèce peuvent se produire au cours du suivi du loup en France. Ces faux-positifs se trouvent dans les détections incertaines, qui peuvent contenir des informations écologiques sur l'espèce mais ne remplissaient pas suffisamment de critères lors du filtrage des données afin qu'elles soient classées comme des données incertaines.

Valentin a donc utilisé un modèle dynamique d'occupancy prenant en compte la détection imparfaite du loup et a illustré l'utilisation des données incertaines en prenant en compte les faux positifs. Il a comparé les résultats avec le modèle d'occupancy dynamique classique que j'ai développé dans l'article 1.

En plus d'avoir pris en compte les faux positifs, il a aussi testé deux variables basées sur des études de sciences citoyennes afin de caractériser le processus d'observation (Silvertown 2009, McKinley et al. 2016, Kosmala et al. 2016, Schmeller et al. 2009, Crall et al. 2011). Il a calculé la proportion de professionnels vs non-professionnels par site par année, afin de voir si cette proportion avait une influence sur les paramètres d'observation du modèle (p_{10} , p_{11} , et b : voir article 2 pour plus de détails sur la définition de ces paramètres). De plus, il a testé si la présence du loup dans un département de France pouvait influencer la probabilité que l'espèce soit détectée dans tout le département. L'échelle du département a été choisie en relation avec les impacts du loup qui sont gérés à l'échelle du département.

Il a montré que les données incertaines fournissaient des informations sur la colonisation à des endroits où aucune autre information n'était disponible. Il a également montré que les observateurs professionnels ne détectaient pas le loup plus souvent que les amateurs, mais que lorsqu'ils le faisaient, ils fournissaient des détections plus fiables. Et enfin, il a montré que la présence du loup dans le département avait une influence sur la probabilité de détecter l'espèce.

En particulier, la probabilité de faux positifs a augmenté lorsque le loup a été détecté présent de manière certaine dans un département, ce qui laisse à penser que la volonté des observateurs de détecter un loup pourrait être un facteur déterminant de détectabilité et donc conduire à des erreurs d'identification. Dans l'ensemble, cette approche permet de tirer le meilleur parti des données de sciences citoyennes en incorporant des données incertaines pour estimer la distribution d'espèces.

Références :

Crall AW, Newman GJ, Stohlgren TJ, Holfelder KA, Graham J, Waller DM. 2011. Assessing citizen science data quality: an invasive species case study: Assessing citizen science data quality. *Conservation Letters* 4:433–442.

Kosmala M, Wiggins A, Swanson A, Simmons B. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14:551–560.

Mckinley DC et al. 2017. Citizen science can improve conservation science , natural resource management , and environmental protection. *Biological Conservation* 208:15–28.

Schmeller DS et al. 2009. Ventajas del monitoreo de biodiversidad basado en voluntarios en Europa. *Conservation Biology* 23:307–316.

Silvertown J. 2009. A new dawn for citizen science. *Trends in Ecology and Evolution* 24:467–471.

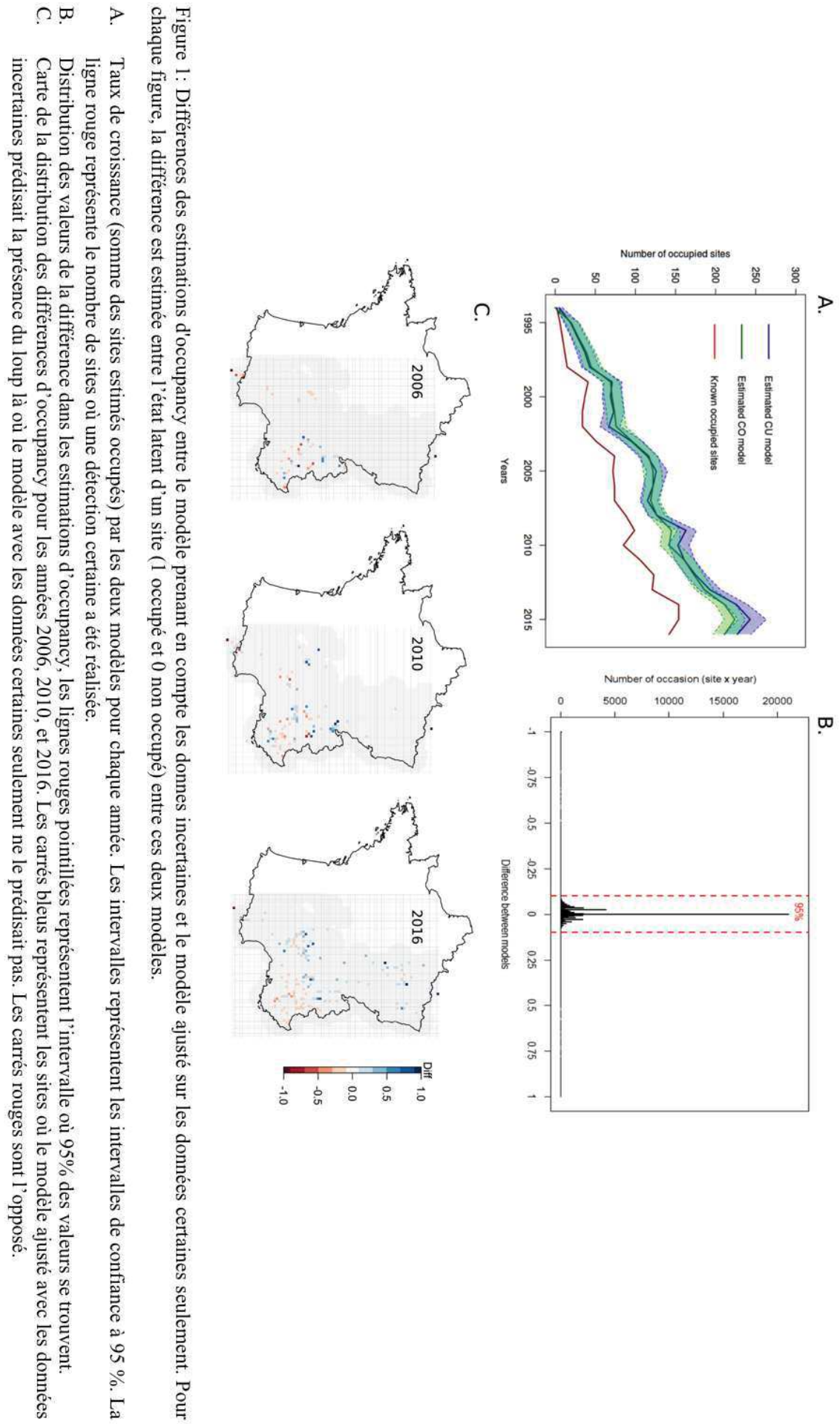


Figure 1 : Différences des estimations d'occupancy entre le modèle prenant en compte les données incertaines et le modèle ajusté sur les données certaines seulement. Pour chaque figure, la différence est estimée entre l'état latent d'un site (1 occupé et 0 non occupé) entre ces deux modèles.

- A. Taux de croissance (somme des sites estimés occupés) par les deux modèles pour chaque année. Les intervalles représentent les intervalles de confiance à 95 %. La ligne rouge représente le nombre de sites où une détection certaine a été réalisée.
- B. Distribution des valeurs de la différence dans les estimations d'occupancy, les lignes rouges pointillées représentent l'intervalle où 95% des valeurs se trouvent.
- C. Carte de la distribution des différences d'occupancy pour les années 2006, 2010, et 2016. Les carrés bleus représentent les sites où le modèle ajusté avec les données incertaines prédisait la présence du loup là où le modèle avec les données certaines seulement ne le prédisait pas. Les carrés rouges sont l'opposé.

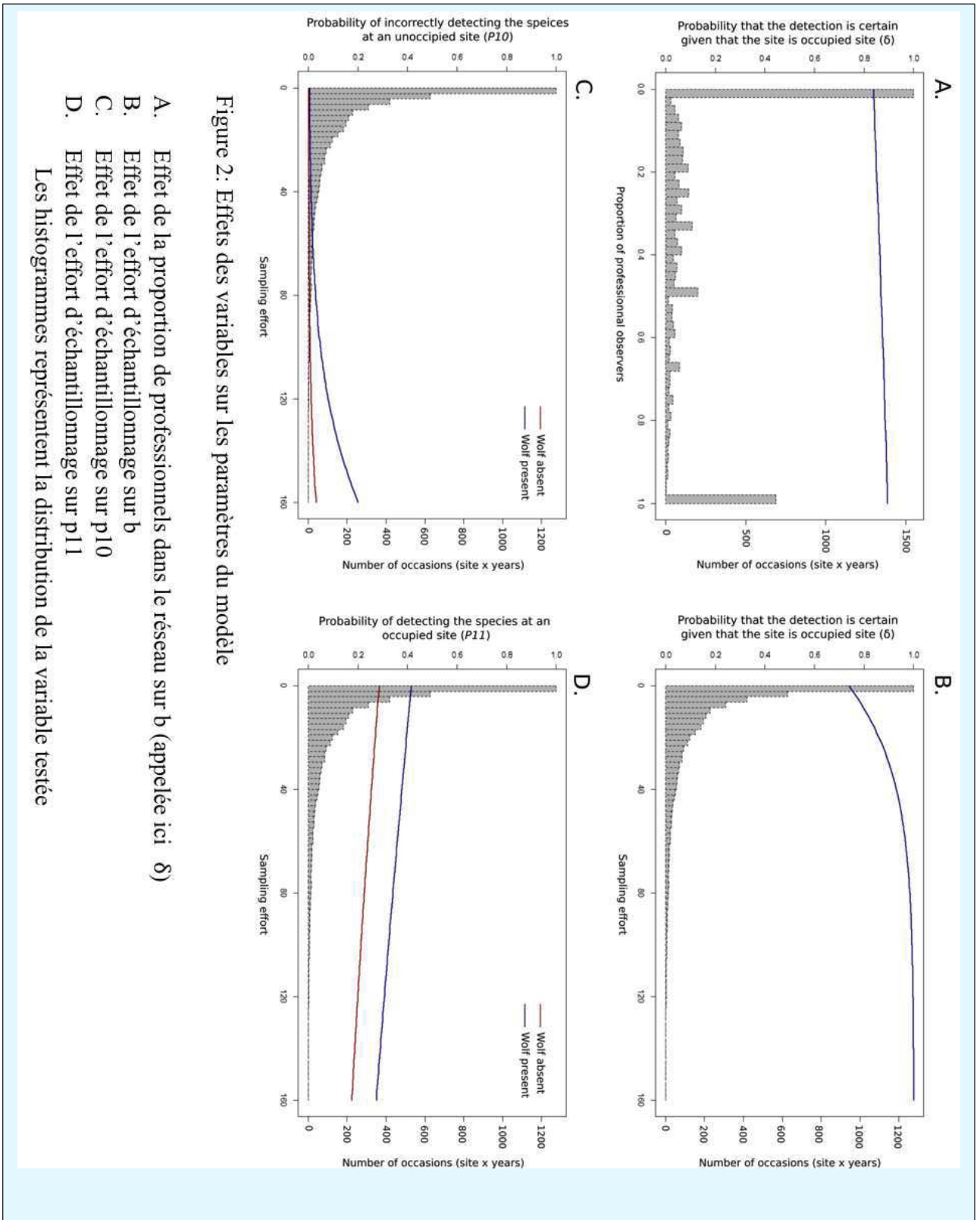


Figure 2: Effets des variables sur les paramètres du modèle

- A. Effet de la proportion de professionnels dans le réseau sur b (appelée ici δ)
 - B. Effet de l'effort d'échantillonnage sur b
 - C. Effet de l'effort d'échantillonnage sur p_{10}
 - D. Effet de l'effort d'échantillonnage sur p_{11}
- Les histogrammes représentent la distribution de la variable testée

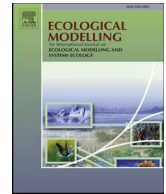
7. Article 3. Prise en compte des erreurs d'identification et de l'hétérogénéité dans les études d'occupancy en utilisant des modèles de Markov cachés

Résumé

Les modèles d'occupancy permettent d'évaluer l'occupation d'une espèce tout en prenant en compte la détection imparfaite. Comme pour tout modèle statistique, les modèles d'occupancy reposent sur plusieurs hypothèses parmi lesquelles (i) il ne doit pas y avoir d'hétérogénéité non modélisée dans la probabilité de détection et (ii) l'espèce ne devrait pas être détectée lorsqu'elle est absente d'un site, en d'autres termes il ne devrait pas y avoir de faux positif (par exemple, en raison d'une erreur d'identification). En pratique, ces deux hypothèses ne sont souvent pas respectées. Jusqu'à présent, élaborer des modèles qui prennent en compte à la fois l'hétérogénéité dans la détection et des faux positifs restait encore à faire. Ici, nous montrons d'abord comment les modèles d'occupancy prenant en compte les faux positifs peuvent être formulés comme des modèles de Markov cachés (HMM). Deuxièmement, profitant de la flexibilité des HMM, nous étendons les modèles à faux positifs afin de prendre en compte l'hétérogénéité dans la détection en utilisant des modèles de mélanges finis. À l'aide de simulations, nous avons démontré qu'à mesure que le niveau d'hétérogénéité augmente, les modèles d'occupancy prenant en compte à la fois l'hétérogénéité dans la détection et les erreurs d'identification donnent de meilleurs résultats en termes de biais et de précision que les modèles qui prennent en compte seulement les erreurs d'identification. Ensuite, nous illustrons l'application de notre nouveau modèle par une étude de cas sur les loups gris (*Canis lupus*) en France. Nous démontrons que l'hétérogénéité dans la détection du loup est principalement due à une hétérogénéité dans l'effort d'échantillonnage. En plus de fournir une nouvelle formulation des modèles d'occupancy, ce travail illustre la souplesse du cadre HMM pour formuler des modèles écologiques complexes et assouplir des hypothèses importantes qui ne sont pas toujours susceptibles d'être respectées. En particulier, nous montrons comment décomposer la structure du modèle en plusieurs des composantes simples, ce qui permet une interprétation écologique beaucoup plus claire et intuitive.

Contributions

J'ai développé la méthode et ajusté le modèle en prenant en compte l'hétérogénéité dans le processus de détection sur les données de détection du loup en France. J'ai proposé une interprétation des résultats obtenus que nous avons discutée collectivement. J'ai rédigé un premier jet de l'article qui a été corrigé par mes co-auteurs. J'ai adapté des codes écrits par Olivier Gimenez pour les simulations.



Accounting for misidentification and heterogeneity in occupancy studies using hidden Markov models



Julie Louvrier^{a,b,*}, Thierry Chambert^a, Eric Marboutin^c, Olivier Gimenez^a

^a CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France

^b Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux déprédateurs, Parc Micropolis, 05000 Gap, France

^c ONCFS, Gières, France

ARTICLE INFO

Keywords:

Occupancy models
Detection heterogeneity
Species imperfect detection
False-positives
Finite-mixture models

ABSTRACT

Occupancy models allow assessing species occurrence while accounting for imperfect detection. As with any statistical models, occupancy models rely on several assumptions amongst which (i) there should be no unmodelled heterogeneity in the detection probability and (ii) the species should not be detected when absent from a site, in other words there should be no false positives (e.g., due to misidentification). In the real world, these two assumptions are often violated. To date, models accounting simultaneously for both detection heterogeneity and false positives are yet to be developed. Here, we first show how occupancy models with false positives can be formulated as hidden Markov models (HMM). Second, benefiting from the HMM framework flexibility, we extend models with false positives to account for heterogeneity with finite mixtures. First, using simulations, we demonstrate that, as the level of heterogeneity increases, occupancy models accounting for both heterogeneity and misidentification perform better in terms of bias and precision than models accounting for misidentification only. Next, we illustrate the implementation of our new model to a real case study with grey wolves (*Canis lupus*) in France. We demonstrate that heterogeneity in wolf detection (false negatives) is mainly due to a heterogeneous sampling effort across space. In addition to providing a novel modeling formulation, this work illustrates the flexibility of HMM framework to formulate complex ecological models and relax important assumptions that are not always likely to hold. In particular, we show how to decompose the model structure in several simple components, in a way that provides much clearer ecological interpretation.

1. Introduction

Occupancy models (Mackenzie et al., 2006) are commonly used to infer species occurrence while accounting for imperfect detection (Bailey et al., 2014; Guisera-Arroita, 2017). These models rely on species detections and non-detections recorded during surveys repeated across time and across several spatial sampling units (sites). As with any statistical models, inferences made from occupancy analyses heavily rely on several assumptions that should be checked and validated (Mackenzie et al., 2003, 2006), although in reality this is rarely done (see however, Mackenzie et al., 2004; Warton et al., 2017).

Here, we focus on two important assumptions. First, there should be no unmodelled heterogeneity in species detection. In other words, all heterogeneity should be accounted for with covariates. If ignored, heterogeneity in detection leads to underestimating occupancy (Royle

and Nichols, 2003; Royle, 2006). Detection heterogeneity can be due to a heterogeneous sampling effort in space (Louvrier et al., 2018), variation in animal abundance (Royle and Nichols, 2003) or variation in site characteristics (Mackenzie et al., 2011). Often, site-level covariates can be measured on the field and incorporated in occupancy models to account for detection heterogeneity. However, unexplained variation may remain or measuring the relevant covariates may simply be impossible in the field. When we suspect substantial unmodelled heterogeneity to occur, we should consider modeling it, either with continuous latent variables (through normally distributed site random effects, e.g. Gimenez et al., 2014). Modelling heterogeneity using normally distributed random effect has long been studied in the field of theoretical biology (e.g., Perc, 2011). Alternatively, modelling heterogeneity can be done using finite mixtures. In finite-mixture models, a latent variable is defined to assign sites to a mixture components (i.e.,

Abbreviations: MMO, model accounting for misidentification only; MMH, model accounting for misidentification and heterogeneity in the detection probabilities; MMS, model accounting for misidentification only and sampling effort as a covariate on the detection probability; MMHS, model accounting for misidentification and heterogeneity in the detection probabilities and sampling effort as a covariate on the detection probabilities

* Corresponding author at: Centre d'Écologie Fonctionnelle et Évolutive, CNRS, 1919 route de Mende, 34090 Montpellier, France.

E-mail address: julie.louvrier@cefe.cnrs.fr (J. Louvrier).

<https://doi.org/10.1016/j.ecolmodel.2018.09.002>

Received 4 January 2018; Received in revised form 6 August 2018; Accepted 3 September 2018

0304-3800/ © 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

groups of heterogeneity) characterized by specific parameters (Royle, 2006; Pledger and Phillpot, 2008). While heterogeneity in detection probability using mixture models has been long studied in the capture-recapture (CR) literature (review in Gimenez et al., 2017), less attention has been given to this issue in occupancy models (Gimenez et al., 2014; Miller et al., 2015).

A second important assumption of occupancy models is that the species should not be detected when absent from a site (i.e. no false positives). False positives occur when the species of interest is detected at sites where it did not occur, usually as a result of misidentification (Miller et al., 2013). Several studies have underlined the importance of accounting for false positives on species distribution (Royle and Link, 2006; Miller et al., 2011, 2013; Chambert et al., 2015). Ignoring false positives and counting them as true positives causes important biases, such as overestimating occupancy and colonization probabilities, and underestimating extinction probability (Royle and Link, 2006; McClintock et al., 2010). Miller et al. (2011, 2013) developed static and dynamic occupancy models that accommodate both false negatives and false positives. As example of applications, these models have been used to estimate occurrence of amphibians (Miller et al., 2011), bats (Clement et al., 2014), and several large vertebrates in India (Pillay et al., 2014), as well as occurrence dynamics of wolves in Montana (Miller et al., 2013).

While several studies have accounted for heterogeneity in occupancy models with false positives by using site-level covariates (McClintock et al., 2010; Ferguson et al., 2015; Miller, 2015), methods that simultaneously account for both unmodelled heterogeneity through finite mixtures and false positives have yet to be developed. Here, we fill this gap and illustrate the use of hidden Markov modelling (HMM) framework as a powerful tool for further developments aiming at relaxing occupancy models' assumptions.

Standard occupancy models can be formulated as HMMs describing two time-series running in parallel. The first time-series captures the dynamics of the latent states with the state process following a Markovian sequence (e.g. site occupied vs. unoccupied); the other time series models the observation process consisting in detections conditional on the underlying but possibly unknown states (Gimenez et al., 2014). The originality of our approach is twofold. First, we show how occupancy models with false positives can be formulated as HMMs. Second, benefiting from the HMM framework flexibility, we extend models with false positives to account for unmodelled heterogeneity using a finite-mixture approach.

To assess the performance of our approach, we performed a simulation study comparing parameter bias and precision in a model accounting for misidentification and heterogeneity vs. a model accounting for misidentification only. To do so, we considered scenarios with an increasing level of heterogeneity in the probability of false positive detection. We also used a case study on the grey wolves' (*Canis lupus*) distribution in France to illustrate implementation of the method in a real-world scenario. Our objectives were (i) to investigate how detection heterogeneity, when ignored, affects the accuracy of occupancy estimation and (ii) assess the extent at which this heterogeneity might be explained by sampling effort variability across space.

2. Methods

In the statistical literature, there are three main problems of interest when using HMM (Rabiner, 1989). In what follows, we review each of these problems in the context of occupancy models. In the *evaluation* problem, we ask what the probability that the observations are generated by our model is – see Section 2.1. In the *decoding* problem, we ask what the most likely state sequence in the model that produced the observations is – see Section 2.5. In the *learning* problem, we ask how we should adjust the model parameters to maximize the likelihood – see Section 2.3.

2.1. HMM formulation of occupancy models with misidentification

Occupancy models can be viewed as HMM whereby the ecological states are considered as partially hidden states, i.e. imperfectly observed (Gimenez et al., 2014). Occupancy models incorporating false positives can also be framed within this approach. The HMM formulation allows flexibility in the model formulation. By decomposing the occupancy approach into simpler steps, the HMM formulation allows better understanding of the ecological and observation processes. To account for false positives, we followed the multi-season dynamic model formulation of Miller et al. (2013). For occupied sites, three observations can be made: (i) an unambiguous detection which is a true detection that has been validated, (ii) an ambiguous detection which is also a true detection but that could not be validated and (iii) no detection. At unoccupied sites, by definition, unambiguous detections cannot occur, thus, only two possible observations can be made: an ambiguous detection, which in this case is a false positive detection due to species misidentification, or no detection. The parameters of interest are ψ_1 the probability of initial occupancy, the probability of local extinction ϵ and of colonization γ , the probability of correctly detecting the species at an occupied site p_{11} , the probability to falsely detect the species at an unoccupied site p_{10} , and the probability b to classify a true-positive detection as unambiguous (Miller et al., 2011). The specification of a HMM is divided in three steps: the vector of initial state probabilities, the matrix of transition probabilities linking states between successive sampling occasions and the matrix of observation probabilities linking observations and states at a given occasion (Gimenez et al., 2014). We define $z_{i,k}$ the latent state of a site i during the primary occasion (e.g., season or year) k . At the first primary occasion, $k = 1$, a site can only be in one of two states ('unoccupied' or 'occupied'), with probabilities in the vector of initial state probabilities:

$$\psi = \begin{matrix} \text{unoccupied} & \text{occupied} \\ [1-\psi_1 & \psi_1] \end{matrix}$$

Then, the states are distributed as a first-order Markov chain governed by a transition matrix of the form:

$$T = \begin{matrix} & \text{unoccupied at } k+1 & \text{occupied at } k+1 \\ \begin{matrix} \text{unoccupied at } k \\ \text{occupied at } k \end{matrix} & \begin{bmatrix} 1-\gamma & \gamma \\ \epsilon & 1-\epsilon \end{bmatrix} \end{matrix}$$

where rows describe states at occasion k in, and columns describe states at $k + 1$.

Next, we describe the observation process, which is conditional on occupancy states. We define $y_{i,j,k}$ the observation of a site i during the secondary occasion (e.g. visit or survey) j during the primary occasion k . For unoccupied sites, unambiguous detections ($y_{i,j,k} = 1$) do not occur while ambiguous detections ($y_{i,j,k} = 2$) or no detections ($y_{i,j,k} = 0$) may occur. For occupied sites, unambiguous detections, ambiguous detections and no detection can all occur. For the sake of clarity, it is more convenient to write the observation process as the product of two matrices. The first matrix summarizes the detection state process conditional on occupancy state (rows) 'unoccupied' and 'occupied' at k . Columns describe the following intermediate latent detection states: 'no detection', 'true positive detection' and 'false positive detection':

$$P = \begin{matrix} & \text{no detection} & \text{true positive detection} & \text{false positive detection} \\ \begin{matrix} \text{unoccupied} \\ \text{occupied} \end{matrix} & \begin{bmatrix} 1-p_{10} & 0 & p_{10} \\ 1-p_{11} & p_{11} & 0 \end{bmatrix} \end{matrix}$$

It is important to keep in mind that the true, underlying state (i.e., false or true positive) of the detections is unknown. At this stage of the modeling, we are still dealing with latent state, not with actual data. The second matrix then summarizes the classification of a true-positive detection as unambiguous or ambiguous, with probability b and $1-b$, respectively. In this matrix, rows represent the above intermediate

latent detection states ('no detection', 'true-positive detection', 'false positive detection') while columns correspond to actual observations (data), i.e., 'no detection' ($y_{i,j,k} = 0$), 'unambiguous detection' ($y_{i,j,k} = 1$) and 'ambiguous detection' ($y_{i,j,k} = 2$):

$$B = \begin{matrix} & \begin{matrix} \text{no detection} & \text{unambiguous detection} & \text{ambiguous detection} \end{matrix} \\ \begin{matrix} \text{no detection} \\ \text{true positive detection} \\ \text{false positive detection} \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & b & 1-b \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

We reemphasize that, by definition, false detections can only occur in the form of ambiguous detection. This last process only deals with the classification of true detection as certain or ambiguous, depending on the level of confidence in the observation. The overall observation process can then be written in the form of a simple matrix product, $O = PB$, which gives:

$$O = \begin{matrix} & \begin{matrix} \text{no detection} & \text{unambiguous detection} & \text{ambiguous detection} \end{matrix} \\ \begin{matrix} \text{unoccupied} \\ \text{occupied} \end{matrix} & \begin{bmatrix} 1-p_{10} & 0 & p_{10} \\ 1-p_{11} & bp_{11} & (1-b)p_{11} \end{bmatrix} \end{matrix}$$

Note that for simplicity we assume observation parameters to be constant, but these could be specified as survey-specific (j), season-specific (k), site-specific (i) or any combination of these effects. We can then write the general probability of any detection history h such as:

$$P(h) = \overbrace{\Psi D [O(\cdot, y_{1,1} + 1)] T_0 D [O(\cdot, y_{2,1} + 1)] \dots T_0 D [O(\cdot, y_{j,1} + 1)]}^{\text{first primary occasion}} \times \overbrace{TD [O(\cdot, y_{1,2} + 1)] T_0 D [O(\cdot, y_{2,2} + 1)] \dots T_0 D [O(\cdot, y_{j,2} + 1)]}^{\text{second primary occasion}} \times \dots \times \overbrace{TD [O(\cdot, y_{1,k} + 1)] T_0 D [O(\cdot, y_{2,k} + 1)] \dots T_0 D [O(\cdot, y_{j,k} + 1)]}^{\text{last primary occasion}} \times \mathbf{1}_N$$

where for the sake of simplicity we have dropped the index i for site, $D(\theta)$ is the diagonal matrix with diagonal elements equal to the elements of the arbitrary vector θ , $O(\cdot, y_{j,k})$ is the column vector corresponding to the observation y at the secondary occasion j during primary occasion k , A_0 is the transition matrix with $\epsilon = \gamma = 0$ and $\mathbf{1}_N$ is the column vector of N ones, with N the number of occupancy states. The likelihood is then the product of the probabilities of all the site histories (Zucchini et al., 2016).

$$T = \begin{matrix} & \begin{matrix} \text{unoccupied (A)}k + 1 & \text{unoccupied (B)}k + 1 & \text{occupied (A)}k + 1 & \text{occupied (B)}k + 1 \end{matrix} \\ \begin{matrix} \text{unoccupied (A)}k \\ \text{unoccupied (B)}k \\ \text{occupied (A)}k \\ \text{occupied (B)}k \end{matrix} & \begin{bmatrix} 1-\gamma_A & 0 & \gamma_A & 0 \\ 0 & 1-\gamma_B & 0 & \gamma_B \\ \epsilon_A & 0 & 1-\epsilon_A & 0 \\ 0 & \epsilon_B & 0 & 1-\epsilon_B \end{bmatrix} \end{matrix}$$

In our study on wolves, we focused on static (single-season) models, such that extinction and colonization events do not occur. The probability of any site history h can thus be simplified to:

$$P(h) = \Psi D [O(\cdot, y_1 + 1)] T_0 D [O(\cdot, y_2 + 1)] \dots T_0 D [O(\cdot, y_j + 1)]$$

2.2. Occupancy model with heterogeneity in the detection probability

We now show how to incorporate site-to-site (i.e., spatial) heterogeneity in the detection process through the addition of a probabilistic process that assigns any site to a finite number of latent classes (Royle, 2006; Miller et al., 2015). For the sake of clarity, here we only consider two classes of heterogeneity (class A and B), but more classes could easily be considered. Like group effects, membership to a heterogeneity class is a constant feature of a site (i.e., it does not change over time), so the assignment process occurs at the first modelling step, describing initial states:

$$\Pi = \begin{matrix} \text{class A} & \text{class B} \\ \begin{bmatrix} \pi & 1-\pi \end{bmatrix} \end{matrix}$$

where π (respectively $1-\pi$) is the probability for any site to be assigned to class A (resp. to class B). This is the main difference with the model described above. The following processes, and the associated parameters remain the same, except that parameters are now allowed to vary according to the two classes of sites. We now define four occupancy states: occupied (A), occupied (B), unoccupied (A) and unoccupied (B). In terms of model parameters, in addition to π , we now have: ψ_{1A} (resp. ψ_{1B}) the probability of initial occupancy for sites of class A (resp. B), ϵ_A (resp. ϵ_B) the probability of local extinction for sites of class A (resp. B) and γ_A (resp. γ_B) the probability of local colonization for sites of class A (resp. B); p_{A11} (resp. p_{B11}) the probability of correctly detecting the species at an occupied site of class A (resp. B), p_{A10} (resp. p_{B10}) the probability to falsely detect the species at an unoccupied site A and the probability b to classify a true-positive detection as unambiguous (Miller et al., 2011).

The next process describes whether a site is initially occupied or not, but now conditionally on the site's class membership (A or B). The pattern matrix that links class membership to the four class-specific occupancy states is:

$$\Phi = \begin{matrix} & \begin{matrix} \text{unoccupied (A)} & \text{unoccupied (B)} & \text{occupied (A)} & \text{occupied (B)} \end{matrix} \\ \begin{matrix} \text{class A} \\ \text{class B} \end{matrix} & \begin{bmatrix} 1-\psi_{1A} & 0 & \psi_{1A} & 0 \\ 0 & 1-\psi_{1B} & 0 & \psi_{1B} \end{bmatrix} \end{matrix}$$

Here, rows correspond to the two conditioning states ('class A' and 'class B') and columns represent all possible initial states: 'unoccupied (A)', 'unoccupied (B)', 'occupied (A)' and 'occupied (B)' in columns.

The final vector of initial state probabilities is therefore the product of the row vector of class-assignment proportions and the matrix of occupancy probabilities:

$$\Psi = \Pi \Phi$$

Because we were interested in incorporating heterogeneity in the detection process only, here we assume $\psi_{1A} = \psi_{1B} = \psi_1$. The state transi-

tion process remains the same but is distinguished according to the two heterogeneity classes A and B. This leads us to define the following matrix for state transition from k to $k+1$:

The conditional observation process then happens independently at A sites and B sites. As above, rows describe occupancy states while

columns represent intermediate latent detection states ‘not detection’, ‘true positive detection’, ‘false positive detection’:

$$P = \begin{matrix} & \begin{matrix} \text{no detection} & \text{true positive detection} & \text{false positive detection} \end{matrix} \\ \begin{matrix} \text{unoccupied (A)} \\ \text{unoccupied (B)} \\ \text{occupied (A)} \\ \text{occupied (B)} \end{matrix} & \begin{bmatrix} 1-p_{A10} & 0 & p_{A10} \\ 1-p_{B10} & 0 & p_{B10} \\ 1-p_{A11} & p_{A11} & 0 \\ 1-p_{B11} & p_{B11} & 0 \end{bmatrix} \end{matrix}$$

Finally, the matrix describing the classification process of true positives as unambiguous or ambiguous detections remains unchanged (see matrix B above). Overall, the observation process can be described by the matrix product $O = P B$ with the occupancy states in rows and the observations in columns:

$$O = \begin{matrix} & \begin{matrix} \text{no detection} & \text{unambiguous detection} & \text{ambiguous detection} \end{matrix} \\ \begin{matrix} \text{unoccupied (A)} \\ \text{unoccupied (B)} \\ \text{occupied (A)} \\ \text{occupied (B)} \end{matrix} & \begin{bmatrix} 1-p_{A10} & 0 & p_{A10} \\ 1-p_{B10} & 0 & p_{B10} \\ 1-p_{A11} & bp_{A11} & (1-b)p_{A11} \\ 1-p_{B11} & bp_{B11} & (1-b)p_{B11} \end{bmatrix} \end{matrix}$$

The likelihood, for both static and dynamic models, is written in the same way as for the model without heterogeneity, only the composition of the matrices changes. Again, in the rest of this paper, we focus on static (single-season) models, so that $\epsilon_A = \epsilon_B = \gamma_A = \gamma_B = 0$.

2.3. Implementation

Occupancy models formulated as HMMs can be implemented in the software E-SURGE (Choquet et al., 2009), which allows for the decomposition of the observation and state processes in multiple steps, providing more flexible, and often more easily interpretable parameterization (Gimenez et al., 2014). We provide in Appendix A and Appendix B two detailed step-by-step procedure to implement occupancy models accounting for false positives with and without heterogeneity in E-SURGE. We also refer to the E-SURGE manual (Choquet and Nogue, 2011) as well as Choquet (2008) and Choquet et al. (2009) for additional details. We also provide in Appendix C and Appendix D the R (R. Core Team, 2013) codes to fit the models described above.

2.4. Simulations

We conducted a simulation study to assess performance of two models: one accounting for misidentification only and one accounting for both misidentification and heterogeneity. First, we simulated occupancy data, including both unambiguous and ambiguous detections, assuming two classes of heterogeneity for false positive probability p_{10} . To characterize and compare heterogeneity among the scenarios considered, we define a heterogeneity coefficient as:

$$\eta = \frac{\sigma^2}{\mu(1-\mu)}$$

with the mean value of the heterogeneity parameter $\mu = \pi p_{A10} + (1-\pi) p_{B10}$ and the variance between components $\sigma^2 = \pi(p_{A10} - \mu)^2 + (1-\pi)(p_{B10} - \mu)^2$ (Dorazio and Royle, 2003; Cubaynes et al., 2012). We

Table 1

Results of the simulation study to assess the performance of the misidentification occupancy model accounting for heterogeneity (MMH) vs. the model without heterogeneity (MMO). The first column corresponds to the heterogeneity coefficient calculated with π the proportion of sites of class A and p_{A10} the probability of making false positive on sites of class A and p_{B10} the probability of making false positive on sites of class B. The differences of AIC (ΔAIC) between the two models, obtained from 200 simulations for each scenario, are provided. Estimation accuracy (RMSE) and relative bias for the occupancy probability are both provided as measures of model performance.

Heterogeneity coefficient	π	p_{A10}	p_{B10}	ΔAIC	Sd (ΔAIC)	RMSE(ψ_1) _{MMH}	RMSE(ψ_1) _{MMO}	Relative bias(ψ_1) _{MMH}	Relative bias(ψ_1) _{MMO}
0.24	0.2	0.1	0.7	6.66	8.62	0.09	0.05	6.35	4.28
0.49	0.5	0.1	0.8	28.15	13.61	0.05	0.10	3.04	12.00
0.53	0.8	0.1	0.9	45.59	19.58	0.05	0.12	2.43	12.23

considered three scenarios with increasing heterogeneity coefficient, by varying parameters π and p_{B10} (Table 1). Other parameters were held constant: p_{A10} was set at 0.1; the initial occupancy probability ψ_1 was set at 0.8; probabilities of true positive p_{A11} and p_{B11} were both set at 0.5 and b was set at 0.7. For all simulation runs, we used 100 sites with 3 and 10 occasions. For each scenario, we simulated $S = 200$ datasets, and for each dataset, fitted both models. Simulation and analyses were done in the software R (see Appendix C and Appendix D). We compared the two models’ performance using averaged AIC differences, calculated from the 200 repeated runs. We chose the AIC because it has been shown to be effective at selecting the number of classes in finite-mixture capture-recapture models (Cubaynes et al., 2012). For each model, we also calculated the relative bias defined as:

$$\frac{1}{S} \sum_{s=1}^S \frac{(\hat{\theta}^{(s)} - \theta)}{\theta}$$

and root mean square error (RMSE) for occupancy probability estimates:

$$\sqrt{\frac{1}{S} \sum_{s=1}^S (\hat{\theta}^{(s)} - \theta)^2}$$

We performed a second simulation study, considering the same form and levels of heterogeneity, but applied to the true detection parameter p_{11} instead of the false positive detections parameter p_{10} . We tested three scenarios with increasing heterogeneity coefficient by varying parameters π and p_{B11} . Every other aspect of this simulation study was the same as in the first simulation study (Appendix E).

2.5. Application: wolf data

To illustrate use of our model accounting for both misidentification and heterogeneity, we analyzed wolves’ detection/non-detection data collected in France during the year 2013 (Louvrier et al., 2018). Signs of presence of the species such as tracks, scats, prey remains, dead wolves, camera trap photos or sightings are being collected thanks to a network of professional and non-professional observers (Duchamp et al., 2012). The data consisted of 250 unambiguous detections, 54 ambiguous detections and 12540 non-detections spread over a grid of 3211 10×10 km cells (Fig. 1). To respect the closure assumption, sites were visited between December and March, which corresponds to a period between the two peaks of dispersal events, in spring and fall (Mech and Boitani, 2010). In a previous study, we found that variability in occupancy probabilities was mostly explained by site’s altitude, while detection probability was primarily driven by sampling effort, defined as the number of observers per site and per year (Louvrier et al., 2018). Here, we compared four different models, all of which included altitude as a covariate on occupancy parameter. A first model accounted for misidentification only without heterogeneity in the detection process (MMO); a second model accounted for misidentification with heterogeneity in both detection probabilities (MMH); in the last model, we accounted for some detection heterogeneity using the sampling effort covariate (on both p_{11} and p_{10}) (MMS), instead of including a finite-mixture heterogeneity process. Sampling effort was indeed quite

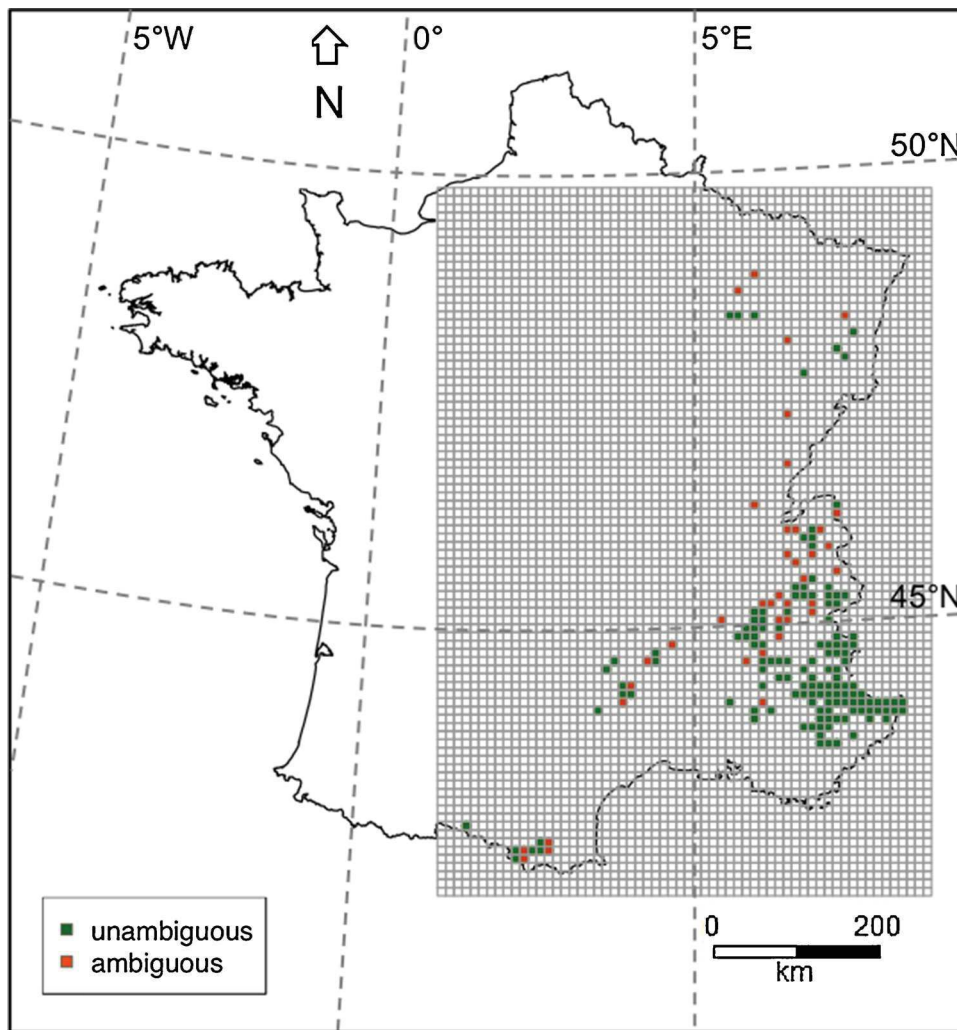


Fig. 1. Map of the unambiguous detections (green) and ambiguous detections (red) cumulated for the year 2013. Sites were defined as 10 × 10 km cells within a grid covering all detections (Louvrier et al., 2018). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

heterogeneous across space, with areas around the Alps subject to much higher sampling effort than the rest of the study area. We expected sampling effort to capture most of the heterogeneity in the detection parameters, which would translate in close AIC values for MMH and MMS and a lower AIC value for MMO. Finally, we fitted a fourth model, including both the finite-mixture process (for unobserved heterogeneity) and the sampling effort covariate (MMHS). This last model allowed assessing the relative fraction of heterogeneity due to sampling effort.

Using the parameter estimates from MMH, we finally built a map depicting the estimated assigned classes of heterogeneity for the 3211 sites of the study area. This was done with the Viterbi algorithm (Rouan et al., 2009), an approach that allows estimation of latent states from HMMs. Although the Bayes’ theorem could be used in this situation where no time series structure is involved, the Viterbi algorithm is a general approach that can be used in a dynamic occupancy framework.

3. Results

3.1. Simulations

Results from the simulations with 10 occasions showed that when heterogeneity in false positive probability increased (coefficient η going from 0.24 to 0.53), the ΔAIC between MMO and MMH ($\Delta AIC = AIC_{MMO}$

$- AIC_{MMH}$) increased from 6.66 (−10.23; 23.55) to 45.59 (7.21; 83.97). While both models showed low RMSE, in terms of occupancy estimation, MMH clearly outperformed MMO as heterogeneity increased. MMO produced increasingly biased parameters when heterogeneity increased while the bias in MMH always remained low (Table 1). We found similar results when we considered heterogeneity in the true detection probability p_{11} . MMH clearly outperformed MMO in terms of AIC, bias and RMSE as heterogeneity increased. The bias for MMO did not increase but remained important across the range of heterogeneity coefficients (Appendix E). Finally, with 3 occasions, MMH appeared to perform worse than MMO, highlighting the need to consider a minimal sampling effort to distinguish false positives and heterogeneity (Appendix F).

3.2. Wolf case study

We found evidence for heterogeneity in the detection process (Table 2), with MMO having a much larger AIC (2209.49) value than MMH (2084.14). Most of this heterogeneity was explained by spatial variation in sampling effort, as suggested by the fact that MMS had a lower AIC value (2071.51) than MMH, but, even after accounting for sampling effort, there remained some unobserved heterogeneity, as evidenced by the fact that MMHS still had a lower AIC value (1953.26) than MMS.

Table 2

Estimated parameters for the model with misidentification and heterogeneity and sampling effort on detection probabilities (p_{A11} , p_{B11} , p_{A10} and p_{B10}) (MMHS), the model with misidentification only with sampling effort on both detection probabilities (p_{11} and p_{10}) (MMS), for the model with misidentification and heterogeneity (MMH) and the model with misidentification only (MMO). The Akaike Information Criterion (AIC) value is given for each model as well as the lower (CI-) and upper (CI+) limits of the 95% confidence interval.

Model with misidentification and heterogeneity with sampling effort on detection probabilities (MMHS)	AIC	1953.26	
<i>Parameter</i>	<i>Value</i>	<i>CI-</i>	<i>CI+</i>
Proportion of sites in class A π	0.70	0.62	0.78
Probability of occupancy ψ	0.05	0.03	0.10
Probability to classify a true-positive detection as unambiguous b	0.89	0.84	0.92
Model with misidentification only with sampling effort on both detection probabilities (MMS)	AIC	2071.51	
<i>Parameter</i>	<i>Value</i>	<i>CI-</i>	<i>CI+</i>
Probability of occupancy ψ	0.06	0.04	0.08
Probability to classify a true-positive detection as unambiguous b	0.91	0.87	0.94
Model with misidentification and heterogeneity (MMH)	AIC	2084.14	
<i>Parameter</i>	<i>Value</i>	<i>CI-</i>	<i>CI+</i>
Proportion of sites in class A π	0.81	0.75	0.86
Probability of occupancy ψ	0.03	0.02	0.06
Probability of false-positive detection in sites A p_{A10}	0.00	0.00	0.00
Probability of false-positive detection in sites B p_{B10}	0.00	0.00	0.00
Probability of true-positive detection in sites A p_{A11}	0.04	0.02	0.06
Probability of true-positive detection in sites B p_{B11}	0.56	0.47	0.64
Probability to classify a true-positive detection as unambiguous b	0.86	0.81	0.90
Model with misidentification only (MMO)	AIC	2209.49	
<i>Parameter</i>	<i>Value</i>	<i>CI-</i>	<i>CI+</i>
Probability of occupancy ψ	0.03	0.02	0.03
Probability of false-positive detection p_{10}	0.00	0.00	0.00
Probability of true-positive detection p_{11}	0.42	0.37	0.47
Probability to classify a true-positive detection as unambiguous b	0.91	0.86	0.94

Overall, all four models estimated a low occupancy probability (Table 2). All four models estimated a very low probability of false-positive detection p_{10} (Fig. 2). MMHS estimated that for 70% of sites, p_{10} was equal to zero, while for the remaining 30% p_{10} strongly increased with sampling effort. MMS results also suggested that p_{10} strongly increased with sampling effort. According to MMH, 81% of the sites had a value of p_{10} similar to the one estimated from MMO. For the remaining 19% p_{10} was equal to zero, indicating that virtually no false-positive errors occurred at these sites. As we could expect, both MMS and MMHS results suggested that true detection probability p_{11} strongly increases with sampling effort for both classes of sites. MMH estimated p_{11} quite low, at around 0.04 (0.02; 0.06) for 81% of the sites and high in 19% of the sites (Fig. 3). Finally, the sites with high detectability corresponded to sites with high sampling effort (Fig. 4).

4. Discussion

While heterogeneity in detection probability has long been studied in the capture-recapture literature (review in Gimenez et al., 2017), less attention has been given to this issue in occupancy modeling (Miller et al., 2015). We developed single- (static) and multi-season (dynamic) occupancy models that account for both heterogeneity and false-positives, benefiting from the flexibility of HMM frameworks to decompose a complex likelihood structure in multiple simpler components more amenable to biological interpretation (Zucchini et al., 2016). Using simulations, we showed that, when ignored, detection heterogeneity induces biases in occupancy estimators. On a case study on wolves, we also found that finite-mixture models of detection could be used to

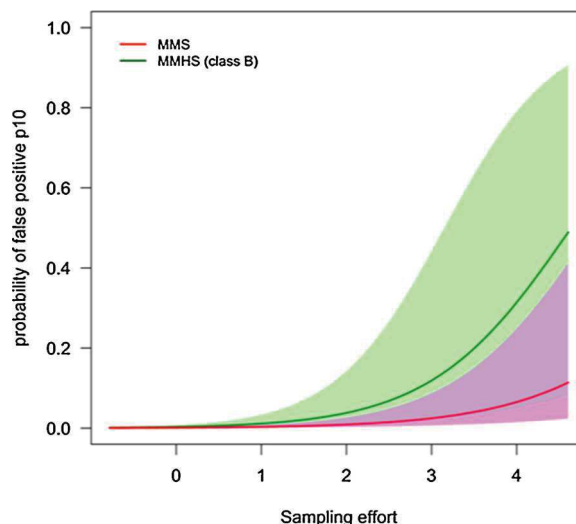


Fig. 2. Estimates of false positive detection from the model accounting for misidentification and sampling effort on the detection probability (MMS; red line) and the model accounting for misidentification and heterogeneity with sampling effort on detection probabilities (MMHS; green line); Lighter shades represent the estimated 95% confidence interval. The estimate of p_{10} for sites of class A was on average zero in the MMHS. In both models (MMHS and MMS) sampling effort had a positive effect on p_{10} in the MMS and on p_{B10} in the MMHS. When the sampling effort increased, p_{B10} increased but remained below 0.5. The probability p_{10} in the MMS model increased as well when sampling effort increased but p_{10} remained below p_{B10} . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

capture part of the heterogeneity due to variable sampling efforts. We note that, although such mixture models are very useful to account for unobserved detection heterogeneity, the associated parameters remain difficult to interpret (Gimenez et al., 2017).

4.1. Advantages of the HMM framework

The HMM formulation of occupancy models provides a great flexibility in the way detection heterogeneity and misidentification can be specified. First, the HMM formulation provides a straightforward mean to modelling the ecological and the observation processes separately (Pradel, 2005). Besides, each process can be split in multiple steps that match exactly the underlying, relevant ecological or observation events, making the modeling exercise and interpretation more intuitive to the biologist. This latter feature has been illustrated in several capture-recapture studies (Pradel, 2005; Sanz-Aguilar et al., 2011; Avril et al., 2012) but it is fairly new for occupancy models (Gimenez et al., 2014).

In this study, we focused on issues of heterogeneity in the detection process, but the framework provided can equally be used to model heterogeneity in occupancy probabilities (Gimenez et al., 2014). We may also allow a site to change heterogeneity status by introducing a transition parameter from class A to class B (Pradel, 2009; Gimenez et al., 2012). This model could be useful in the wolf case study to accommodate an increase in sampling effort over time and the possibility for some sites with low detectability to get a higher detectability. The use of finite mixtures allows capturing relevant levels of heterogeneity without the need to include a large number of explanatory covariates. It is however important to keep in mind that parameter redundancy can become an issue as we add more classes of heterogeneity and/or choose to apply finite mixtures on several parameters (Gimenez et al., 2014). For instance, if we wanted to consider classes of heterogeneity for all parameter (occupancy, true detection and false detection), our model could quickly become too complex (Gimenez et al., 2014). If parameter redundancy is suspected, it could be assessed using the methods that

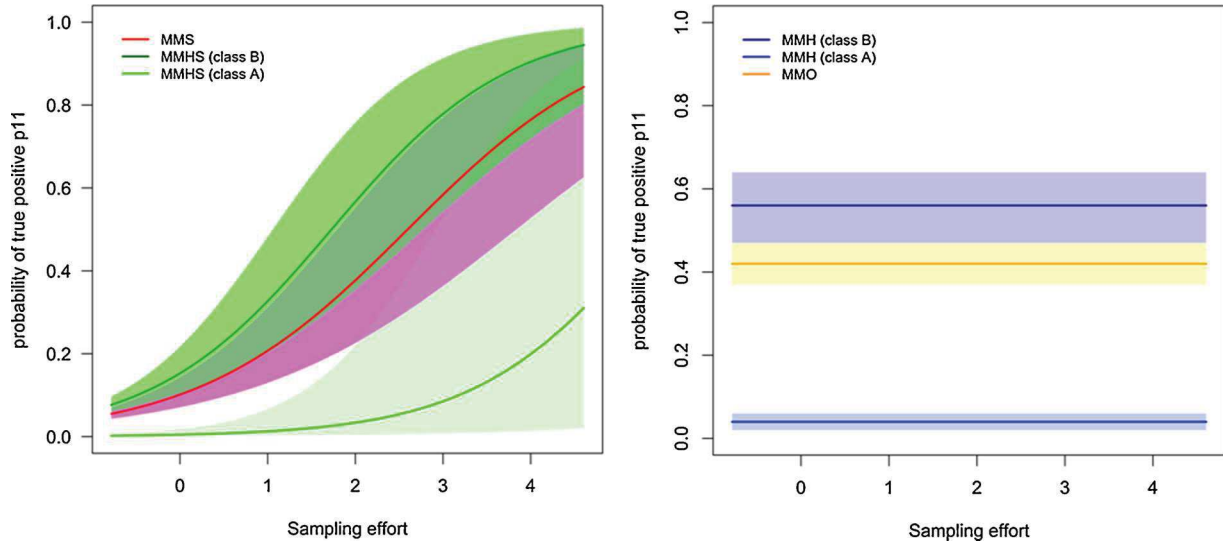


Fig. 3. Estimates of true positive detection from the model accounting for misidentification and the standardized sampling effort (number of observers per site per year) on the detection probability (MMS; red line) and the model accounting for misidentification and heterogeneity with sampling effort on detection probabilities (MMHS; green lines) (left panel), the model accounting for misidentification and heterogeneity with sampling effort on detection probabilities for misidentification only (MMO; yellow line). Lighter shades represent the estimated 95% confidence interval. Sampling effort was defined as the number of observers per cell per year. In both models (MMHS and MMS) sampling effort had a positive effect on p_{11} in the MMS and on p_{A11} and p_{B11} in the MMHS. When sampling effort increased, p_{B11} increased as well, before almost reaching a plateau for the maximum values of sampling effort. p_{A11} increased as well but remained below 0.4 when sampling effort is at the maximum value. With (left panel) or without (right panel) sampling effort as a covariate on this probability, p_{A11} was much lower than p_{B11} . On the right panel, no covariate was used in the detection probabilities, which explains the flat lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

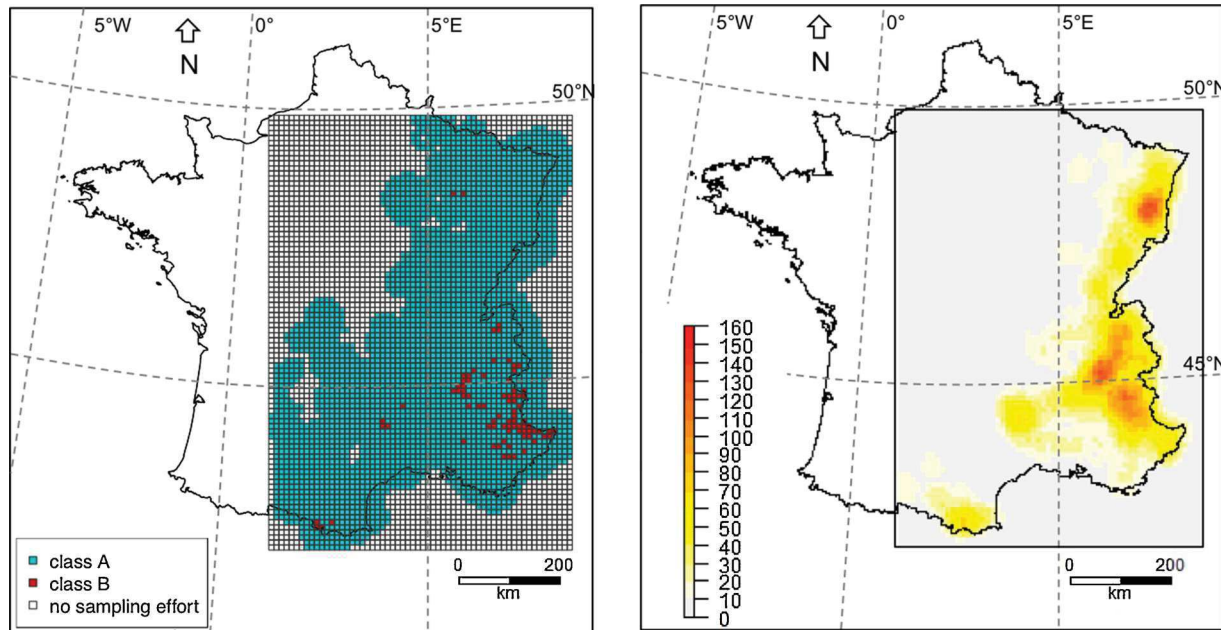


Fig. 4. Visualizing heterogeneity. Left: Map depicting the class of heterogeneity to which each site belongs to (obtained through the use of the Viterbi). Sites within the study area that are colored in white correspond to sites where sampling effort was null. Sites colored in blue are sites affiliated to class A, and sites colored in red are sites affiliated to class B. Right: Map of the sampling effort defined as the number of observers per site per year for the year 2013. The grey rectangle represents the study area where we defined our 10 x 10 km sites (Louvrier et al., 2018). Sites of class A, which correspond to the ones with a null false positive probability p_{A10} and a very low true detection probability p_{A11} are the sites represented in light blue where sampling effort is very low as well. In opposition, sites of class B seemed to be mainly where sampling effort was important. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

have been developed for this purpose (Choquet and Cole, 2012).

4.2. Simulations

We considered three scenarios in which the heterogeneity coefficient increased with the proportion of sites of class A and the probability of false-positive detections in sites B. As expected, models ignoring heterogeneity produced biased and imprecise occupancy estimates when heterogeneity was actually increasing. Those results are the same whether heterogeneity occurs in probabilities of false positive detections or true detections. Previous studies have emphasized the importance of incorporating covariates, when available, on the true positive probability to avoid underestimating occupancy parameters (Miller et al., 2015). In absence of measurable candidate covariates, we recommend using the finite-mixture approach described in this paper. An alternative to finite mixtures is to consider a site random effect (Gimenez et al., 2014). An advantage of finite-mixture models is that they can easily be fitted in a frequentist or a Bayesian framework. In contrast, site random effects are naturally fitted in a Bayesian framework, but at the cost of higher computational burden; in the frequentist framework, site random effects require a specific, non-trivial treatment of the likelihood (e.g., Gimenez and Choquet, 2010).

4.3. Wolf case study

We found strong evidence of among-site heterogeneity in the detection probability of wolves. This heterogeneity was mostly explained by spatial variation in sampling effort. As expected, the probability of true positive detection increased with sampling effort. In this study, more sampling effort occurs at the core of the species' distribution, around the Alps and in the North-Eastern part of France (Fig. 1). This variable effort is explained by the fact that in France, the first recolonizing wolves were detected in the Alps, before they started spreading out (Valière et al., 2003).

Average values of detection parameters estimated by the MMH were close to the average values estimated by the MMS and the MMHS, highlighting the similarities between these models. However, the MMH and MMS provided slightly different estimates of occupancy probability. AIC comparison between these two models revealed that sampling effort explains most of the variation in detection probabilities. A previous study (Louvrier et al., 2018) had also found strong support for this covariate, and the authors had shown that ignoring the effect of sampling effort leads to underestimating occupancy probabilities, especially at sites with low effort. Results from the MMHS confirm this hypothesis as the estimate of occupancy probability was close to the one provided by the MMS. Sampling effort is thus an important covariate to account for, but the fact that MMHS had the lowest AIC suggests that some unobserved detection heterogeneity remains. If sampling effort cannot easily be measured, models accounting for unobserved heterogeneity could be used as an alternative to control for detection variability and avoid estimation biases. Such an approach should prove especially useful for analyses of opportunistic monitoring data collected by large networks of citizens, which rely on protocols that rarely include explicit measures of sampling effort.

The fact that the estimated probability of false positive detection was low suggests that the continuous training, over the years, of the observers of the network (Duchamp et al., 2012) has been efficient. The low value of the false positive detection could also mean that the filtering process of observations was effective. Part of this filtering process consisted in rejecting detections that could be identified as false positives to avoid noise in the data. We can conclude that most false positives may very well have already been rejected during this filtering process. Based on this fact and the value of the probability to classify a true detection as unambiguous b we could conclude that most of the ambiguous data were actually true positive detections, which could be considered as unambiguous detections and analyzed with standard

occupancy models. However, running the model accounting for false positives would still be a necessary step to ensure that false positive probability is extremely low, before we decide to treat ambiguous data as true detections.

In our wolf study, winter observations consisted mostly of tracks found in the snow, scats, carcasses and camera trap photographs (Duchamp et al., 2012), which were all observations that could easily be verified. However, on the colonization front, a larger proportion of detections consisted of visual sightings, and because these types of data could not be verified they were therefore classified as ambiguous. Here, integrating these ambiguous data is especially interesting because it brings new information about the wolf distribution on the colonization front. These data could thus provide new insights in the context of managing an expanding protected species where new identified sites could help understanding the processes underlying the recolonization of wolves (Guillera-arroita et al., 2015).

We found from the Viterbi algorithm that the sites having the higher true positive probability were those where the sampling effort was high. However, the fact that the estimated number of sites of class B from the Viterbi algorithm is lower than what the MMH estimated can be due to the low value of the occupancy probability. This low occupancy estimate can lead to a greater uncertainty in assessing the class of a site (Rouan et al., 2009). Despite this fact, it is possible to see the link between sampling effort and the two classes of heterogeneity. Sites belonging to the heterogeneity class A, which were primarily found on the colonization front of the species (Louvrier et al., 2018), had a lower detection probability p_{11} than sites of class B. This strongly suggests that the species is less likely to be detected on the colonization front, where new conflicts with human activities can arise.

In conclusion, we recommend devoting efforts in the monitoring process to minimize heterogeneity across sites, and possibly identifying and measuring covariates that may affect detection probability such as the sampling effort. If not possible, we recommend using occupancy models accounting for detection heterogeneity if covariates possibly explaining site-to-site variation cannot be measured in the field. The HMM formulation we propose allows an easy implementation of these models.

Data accessibility

We provide the scripts to simulate the data in the Appendix B and Appendix C as well as the wolf data in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.g9s1d>

Acknowledgements

We gratefully acknowledge the help of people who have collaborated with the wolf monitoring network supervised by the French game and wildlife agency (ONCFS). We warmly thank Christophe Duchamp for his help and for making the data available to us. We thank the Montpellier University and the ONCFS for a PhD grant to the first author.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at [doi:https://doi.org/10.1016/j.ecolmodel.2018.09.002](https://doi.org/10.1016/j.ecolmodel.2018.09.002).

References

- Avril, A., Letty, J., Pradel, R., Léonard, Y., Santin-Janin, H., Pontier, D., 2012. A multi-event model to study stage-dependent dispersal in radio-collared hares: when hunting promotes costly transience. A multi-event model to study stage-dependent dispersal in radio-collared hares : when hunting promotes costly transience. *Ecology* 93, 1305–1316.
- Bailey, L.L., Mackenzie, D.I., Nichols, J.D., 2014. Advances and applications of occupancy

- models. *Methods Ecol. Evol.* 5, 1269–1279.
- Chambert, T., Miller, D.A.W., Nichols, J.D., 2015. Modeling false positive detections in species occurrence data under different study designs. *Ecology* 96, 332–339.
- Choquet, R., 2008. Automatic generation of multistate capture-recapture models. *Can. J. Stat.* 36, 43–57.
- Choquet, R., Cole, D.J., 2012. Mathematical Biosciences A hybrid symbolic-numerical method for determining model structure. *Math. Biosci.* 236, 117–125.
- Choquet, R., Nogue, E., 2011. E-SURGE 1.8 User's Manual. CEFÉ, Montpellier, France.
- Choquet, R., Rouan, L., Pradel, R., 2009. Program E-surge: a software application for fitting multievent models. In: Thomson, D.L., Cooch, E.G., Conroy, M.J. (Eds.), *Modeling Demographic Processes in Marked Populations*. Springer US, Boston, MA, pp. 845–865.
- Clement, M.J., Rodhouse, T.J., Ormsbee, P.C., Szwczak, J.M., Nichols, J.D., 2014. Accounting for false-positive acoustic detections of bats using occupancy models. *J. Appl. Ecol.* 51, 1460–1467.
- Cubaynes, S., Lavergne, C., Marboutin, E., Gimenez, O., 2012. Assessing individual heterogeneity using model selection criteria: how many mixture components in capture-recapture models? *Methods Ecol. Evol.* 3, 564–573.
- Dorazio, R.M., Royle, J.A., 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* 59, 351–364.
- Duchamp, C., Boyer, J., Briaudet, P.E., Leonard, Y., Moris, P., Bataille, A., Dahier, T., et al., 2012. A dual frame survey to assess time- and space-related changes of the colonizing wolf population in France. *Hystrix* 23, 14–28.
- Ferguson, P.F.B., Conroy, M.J., Hepinstall-Cymerman, J., 2015. Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys. *Methods Ecol. Evol.* 6, 1395–1406.
- Gimenez, O., Choquet, R., 2010. Individual heterogeneity in studies on marked animals using numerical integration : capture – recapture mixed models. *Ecology* 91, 951–957.
- Gimenez, O., Lebreton, J., Gaillard, J., Choquet, R., Pradel, R., 2012. Estimating demographic parameters using hidden process dynamic models. *Theor. Popul. Biol.* 82, 307–316.
- Gimenez, O., Blanc, L., Besnard, A., Pradel, R., Doherty, P.F.J., Marboutin, E., Choquet, R., 2014. Fitting occupancy models with E-SURGE : hidden Markov modelling of presence – absence data. *Methods Ecol. Evol.* 5, 592–597.
- Gimenez, O., Cam, E., Gaillard, J., 2017. Individual heterogeneity and capture-recapture models: what, why and how? *Oikos* 127 (5), 664–686.
- Guillera-arroita, G., Lahoz-monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., et al., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Glob. Ecol. Biogeogr.* 24, 276–292.
- Guillera-Arroita, G., 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40, 281–295.
- Louvrier, J., Duchamp, C., Lauret, V., Marboutin, E., Cubaynes, S., Choquet, R., Miquel, C., et al., 2018. Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data. *Ecography* 41, 647–660.
- Mackenzie, D.I., Nichols, J.D., Hines, J.E., Knutson, M.G., Franklin, A.B., 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84, 2200–2207.
- Mackenzie, D.I., Bailey, L.L., Nichols, J.D., 2004. Investigating species co-occurrence patterns when species are detected imperfectly. *J. Anim. Ecol.* 73, 546–555.
- Mackenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2006. *Occupancy Estimation and Modeling*. Academic Press.
- Mackenzie, D.I., Bailey, L.L., Hines, J.E., Nichols, J.D., 2011. An integrated model of habitat and species occurrence dynamics. *Methods Ecol. Evol.* 2, 612–622.
- Mcclintock, B.T., Bailey, L.L., Pollock, K.H., Simons, T.R., 2010. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology* 91, 2446–2454.
- Mech, L.D., Boitani, L., 2010. *Wolves: Behavior, Ecology, and Conservation*. University of Chicago Press.
- Miller, J.R.B., 2015. Mapping attack hotspots to mitigate human–carnivore conflict: approaches and applications of spatial predation risk modeling. *Biodivers. Conserv.* 24, 2887–2911.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Weir, L.A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology* 92, 1422–1428.
- Miller, D.A.W., Nichols, J.D., Gude, J.A., Rich, L.N., Podruzny, K.M., Hines, J.E., Mitchell, M.S., 2013. Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. *PLoS One* 8.
- Miller, D.A.W., Bailey, L.L., Grant, E.H.C., McClintock, B.T., Weir, L.A., Simons, T.R., 2015. Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods Ecol. Evol.* 6, 557–565.
- Perc, M., 2011. Does strong heterogeneity promote cooperation by group interactions? *New J. Phys.* 13, 123027.
- Pillay, R., Miller, D.A.W., Hines, J.E., Joshi, A.A., Madhusudan, M.D., 2014. Informant interviews. *Divers. Distrib.* 20, 223–235.
- Pledger, S., Phillpot, P., 2008. Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biometr. J.* 50, 1022–1034.
- Pradel, R., 2005. Multievent : an extension of multistate capture – recapture models to uncertain states. *Biometrics* 61, 442–447.
- Pradel, R., 2009. The stakes of capture–recapture models with state uncertainty. *Modeling Demographic Processes in Marked Populations*. Springer, Boston, MA, pp. 781–795.
- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286.
- Rouan, L., Gaillard, J.-M., Guédon, Y., Pradel, R., 2009. Estimation of lifetime reproductive success when reproductive status cannot always be assessed. *Modeling Demographic Processes in Marked Populations*. Springer, pp. 867–879.
- Royle, J.A., 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics* 62, 97–102.
- Royle, J.A., Link, A.W., 2006. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology* 87, 835–841.
- Royle, J.A., Nichols, J.D., 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84, 777–790.
- Sanz-Aguilar, A., Tavecchia, G., Genovart, M., Igual, J.M., Oro, D., Rouan, L., Pradel, R., 2011. Studying the reproductive skipping behavior in long-lived birds by adding nest inspection to individual-based data. *Ecol. Appl.* 21, 555–564.
- Team, R., 2013. R development core team. *RA Lang. Environ. Stat. Comput.* 55, 275–286.
- Valière, N., Fumagalli, L., Gielly, L., Miquel, C., Lequette, B., Pouille, M.-L., Weber, J.-M., et al., 2003. Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Anim. Conserv.* 6, 83–92.
- Warton, D.I., Stoklosa, J., Guillera-arroita, G., Mackenzie, D.I., Welsh, A.H., 2017. Graphical diagnostics for occupancy models with imperfect detection. *Methods Ecol. Evol.* 8, 408–419.
- Zucchini, W., MacDonald, I.L., Langrock, R., 2016. *Hidden Markov Models for Time Series: An Introduction Using R*. CRC press.

8. Article 4. Modélisation de la distribution du lynx (*Lynx lynx*) en Norvège en combinant des données d'occupancy et des données de presence-background

Résumé

Les modèles de distribution d'espèces (SDMs) sont devenus un outil de plus en plus utile pour comprendre les mécanismes qui sous-tendent la répartition des espèces. Récemment, des SDMs combinés ont été mis au point pour analyser conjointement les données de présence-background et les données de détection/non-détection, fournissant des estimations moins biaisées et plus précises des paramètres de distribution que les SDMs standards. Nous avons utilisé ici une nouvelle approche développée par Renner et Gimenez (2018) qui formule un SDM combiné à l'aide d'un cadre unifié de processus ponctuels, permettant ainsi l'utilisation des techniques existantes pour vérifier les résidus spatiaux et sélectionner efficacement les variables environnementales. Nous avons appliqué le nouveau SDM combiné au lynx d'Europe (*Lynx lynx*), un grand carnivore chassé en Norvège. Avec un modèle d'occupancy, nous avons analysé les données de groupes familiaux, c'est-à-dire la présence de femelles adultes avec un petit. Les données de lynx chassés ont été quant à eux analysés comme des données de présence-background avec des modèles type Maxent en utilisant l'équivalence avec les modèles de processus ponctuels. Les deux jeux de données ont été analysés ensemble à l'aide du SDM combiné. Nous avons sélectionné des facteurs influençant la distribution du lynx, comme la pente et la productivité primaire. Les données d'occupancy sur les groupes familiaux ont fourni des informations sur la détection imparfaite, tandis que les données de présence-background ont fourni des informations indépendantes de l'échelle à laquelle les données sont récoltées. Nos résultats apportent de nouvelles connaissances sur la gestion du lynx en Norvège. Ils mettent en évidence les endroits où le lynx peut être présent et aident donc à cibler des zones spécifiques afin de déterminer les priorités pour la mise en œuvre de mesures de protection et enfin à définir des quotas durables de chasse.

Contributions

J'ai transformé et trié les données, ajusté les modèles séparés et combinés et réalisé les analyses d'agrégation spatiale et les sélections de variables sur toutes les années. J'ai proposé une interprétation des résultats obtenus que nous avons discutée collectivement. J'ai rédigé un premier jet de l'article qui a été corrigé par mes co-auteurs. Les données ont été fournies par John Linnell. J'ai calculé les variables environnementales avec des contributions de Diana Bowler. John Linnell, Erlend Nielsen et Jenny Mattisson m'ont aidé à cibler les variables environnementales importantes. Ian Renner a fourni le package et les scripts permettant les analyses ainsi que le calcul des variables environnementales. Olivier Gimenez m'a aidé pour l'établissement de la méthode et la simplification du problème.

1 **Modelling the distribution of Eurasian lynx (*Lynx lynx*) in Norway using the**
2 **combination of occupancy data and presence-background data**

3 Julie Louvrier^{1,2}, Ian Renner², John Linnell³, Erlend Nilsen³, Diana Bowler³, Jenny Mattisson³,
4 Olivier Gimenez¹

5 ¹ CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier,
6 France

7 ² Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux
8 déprédateurs, Parc Micropolis, 05000 Gap, France

9 ³ School of Mathematical and Physical Sciences, The University of Newcastle, Australia

10 ⁴ Norwegian Institute for Nature Research, PO Box 5685 Sluppen, NO-7485 Trondheim,
11 Norway

12 Corresponding author: Julie Louvrier, CEFE UMR 5175, CNRS, Université de Montpellier,
13 Université Paul-Valéry Montpellier, EPHE, Montpellier Cedex 5, France,
14 julie.louvrier@cefe.cnrs.fr

15 ian.renner@newcastle.edu.au

16 John.Linnell@nina.no

17 Erlend.Nilsen@nina.no

18 diana.e.bowler@gmail.com

19 Jenny.Mattisson@nina.no

20 olivier.gimenez@cefe.cnrs.fr

21 **To be submitted to Journal of Applied Ecology**

22 **Abstract**

23 Species distribution models (SDMs) have become an increasingly useful tool to understand the
24 mechanisms underlying the distribution of species. Recently, combined SDMs have been

25 developed to analyze jointly presence-background and detection/non-detection data, providing
26 less biased and more precise estimates of distribution parameters than standard SDMs. Here,
27 we used a new approach developed by Renner and Gimenez (2018) that formulates combined
28 SDM in a unified point process modeling framework, therefore allowing the use of existing
29 techniques to check spatial residuals and efficiently select covariates. We applied the new
30 combined SDM to a hunted large carnivore in Norway, the Eurasian lynx (*Lynx lynx*). We
31 analyzed family groups data, i.e. adult females with a cub, with site-occupancy models and data
32 on harvested lynx with presence-background models, and both datasets together with the
33 combined SDM. We found important factors such as slope and primary productivity that
34 underlie lynx distribution across years. Occupancy data on family groups provided information
35 on imperfect detection, while PB data provided scale-independent information. Our results
36 bring new insights into the management of lynx in Norway by highlighting places where lynx
37 may occur and therefore help in targeting specific areas, determining priorities for
38 implementing protection measures and defining sustainable harvesting quotas.

39

40 **Introduction**

41 Understanding the mechanisms underlying the distribution of species is essential in ecology,
42 evolution and conservation biology (Guisan & Thuiller 2005; Elith & Leathwick 2009). Species
43 Distribution Models (SDMs) have become an increasingly useful tool to address this issue.
44 They allow understanding and predicting the spatial distribution of species presence based on
45 the location of observed individuals. In some cases, SDMs can be used to study changes through
46 times in the context of management of threatened species, conservation planning as well as
47 predicting the future ranges of invasive species (Guillera-arroita et al. 2015).

48 SDMs may rely on locations where a species or signs of its presence were observed
49 (Dorazio 2014), known as presence-background (PB) data, collected during opportunistic
50 surveys, i.e. without measure of the sampling effort. PB data are an attractive source of
51 information for SDMs because collecting them is easier and less expensive than planned
52 surveys and allows covering a large spatial study site (Dorazio 2014). To provide an
53 understanding of the mechanisms underlying the distribution of a species, PB data are
54 correlated to potential environmental and climatic covariates (Guisan & Thuiller 2005). A
55 variety of SDMs have been developed to analyze PB data (Elith et al. 2006), including Maxent
56 (Phillips et al. 2006; Elith et al. 2011) and models of spatial point patterns (Warton & Shepherd
57 2010; Renner et al. 2015), with a close link between the two (Renner & Warton 2013).

58 In addition to PB data, other sources of data may be collected. Planned surveys lead to
59 detections and non-detections of the species at sites (Koshkina et al. 2017), generating so-called
60 occupancy data. Site-occupancy models have been developed to analyze occupancy data and
61 are attractive because they allow explicitly incorporating the imperfect detection of a species
62 based on several repeated surveys over each site (MacKenzie et al. 2002). The parameters of
63 site-occupancy models have, however, scale-dependent definitions and need to be cautiously
64 interpreted (Koshkina et al. 2017).

65 Recently, efforts have been put into the development of SDMs combining both PB and
66 occupancy data, therefore providing better estimates of distribution parameters than SDMs
67 based on PB data or occupancy data analyzed in isolation. (Dorazio 2014) first developed a
68 model combining PB data with data from point-count surveys in which imperfect detection was
69 accounted for. In parallel, Fithian et al. (2015) developed a similar approach to combine PB
70 data and occupancy data using information on several species to correct for the observer bias
71 in the PB data.

72 However, SDMs combining different sources of data have only recently been introduced
73 (Hefley & Hooten 2016), with no assessment of their limitations. For example, combining
74 models makes the process of selecting relevant explanatory covariates more complex because
75 of the increase in the number of potential model combinations. Additionally, combined models
76 are limited by the lack of methods for checking residuals and assessing model fit. Building on
77 (Dorazio 2014)'s work, Renner and Gimenez (2018) recently developed an SDM approach that
78 combines machine learning methods and point process tools for efficient covariate selection
79 and model validation.

80 Here, we showcase the Renner and Gimenez (2018)'s new approach using large
81 carnivores as a case study. Large carnivores' populations have increased for the past 40 years,
82 resulting in most of the European countries hosting at least one viable population of large
83 predators (Chapron et al. 2014). However, the presence of large carnivores in human-dominated
84 areas comes with challenges. Their presence can lead to conflictual situations with human-
85 related activities (Linnell et al. 2007). In addition, the question of finding optimal areas left for
86 viable populations needs to be addressed (Packer et al. 2013). In this context, mapping the
87 distribution of large carnivores can help management decisions by targeting factors affecting
88 their presence to mitigate conflicts and protect the species.

89 Monitoring large carnivore populations is also a challenge due to their low densities and
90 elusive behaviors. As a consequence, studies on large carnivores are mostly based on indirect
91 sampling methods, where scats, tracks and other opportunistic signs are collected and used as
92 presence signs (Thompson et al. 1998; Duchamp et al. 2012; Louvrier et al. 2018b; Molinari-
93 Jobin et al. 2018). Here we illustrate these constraints by using two different types of ecological
94 datasets, PB data and occupancy data on the Eurasian lynx, *Lynx lynx* in Norway. In Norway,
95 lynx have always been present but they are currently heavily harvested, to limit conflictual
96 situations and also to provide hunting opportunities (Pedersen et al. 1999; Odden et al. 2002;

97 Linnell et al. 2010). Monitoring their population and assessing their distribution is therefore
98 crucial to get population trends, target potential favorable areas, and harvest sustainably. The
99 main monitoring method for lynx in Norway was put in place to assess population trends and
100 set hunting quotas. This monitoring consists in counts of family groups (adult females with
101 dependent kittens) based on observations of tracks in the snow (Andrén et al. 2002; Linnell et
102 al. 2007). In addition to this monitoring, other types of monitoring have been considered for
103 various objectives: transects along which lynx tracks (in the snow) are recorded (Linnell et al.
104 2007); quota bags are collected with biological information about the hunted individuals and
105 the location of the hunt from hunters, which give yearly information about population trends
106 (Linnell et al. 2010); GPS collared individuals are followed to study fine scale habitat and
107 hunting preferences, territory sizes and individual movements (Mattisson et al. 2011; Bouyer
108 et al. 2015b, 2015a). However, with the logistical difficulties and financial constraints in
109 running several monitoring schemes in parallel, there is a need to optimize the surveys to target
110 surveys that would bring most of the ecological information about the lynx distribution. Here,
111 we tackled the issue by focusing on the family group monitoring (occupancy data) and the
112 harvested data (PB data).

113 We first describe the new SDM that combines PB data with occupancy data (Renner
114 and Gimenez, 2018). Then we demonstrate the advantage in combining information by
115 analyzing the Norwegian lynx data with family groups and harvested lynx together or by
116 considering them separately. We perform covariates selection in the combined model using the
117 LASSO technique to investigate the effects of environmental covariates on lynx distribution.
118 We show how to assess the underlying assumptions of the combined model by highlighting the
119 analysis of spatial residuals. Finally, by running SDM models for each year, we quantify the
120 distribution dynamics of the lynx in Norway.

121

122 **Material and methods**

123 1- Data collection

124 *Family groups*

125 Counting lynx family groups has been the main monitoring method used for lynx national
126 monitoring program in Norway since 1996 (Braa et al. 1999; Linnell et al. 2010). Family groups
127 are defined as groups of one reproductive female with one or more dependent kittens (Linnell
128 et al. 2007). Kittens stay with their mother until they are 10 months old and mating starts late
129 March, males and females remaining separated before this date. Tracks in the snow of more
130 than one individual during early winter indicated a family group. Records of lynx tracks were
131 collected by hunters, game wardens and the public from beginning of October until the end of
132 February. These observations were all verified by state wardens who followed the tracks for
133 more than 1 km. For every family group track, the date and location of collection were stored
134 in a national monitoring database (ROVBASEN 3.0: www.rovbase.no). Here, we used records
135 of snow-tracking observations from family groups from each winter between 1997 and 2016
136 (Fig. 1, Table 1).

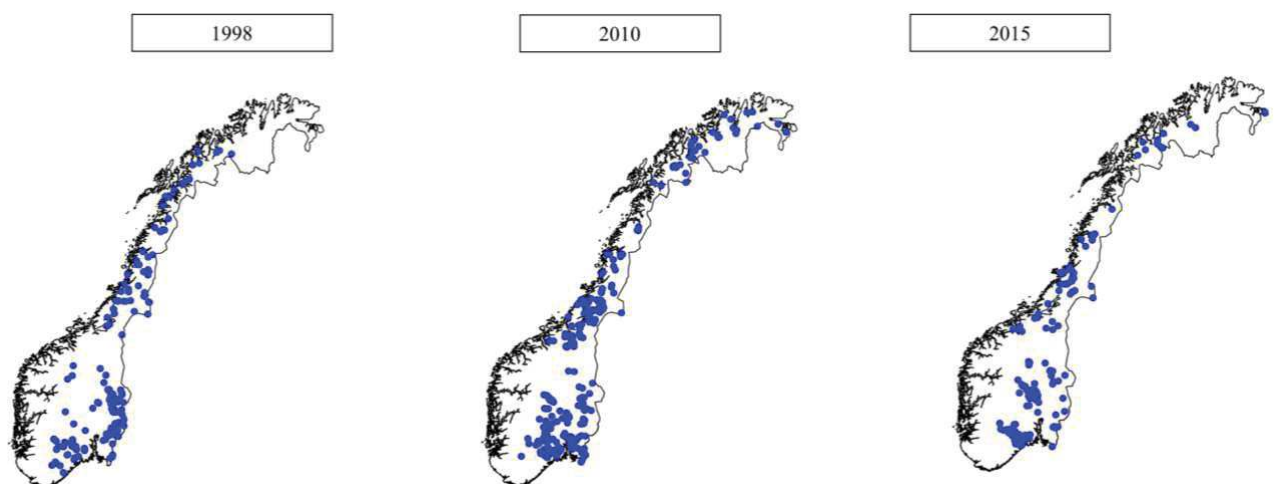


Figure 1: Maps of observation of family groups detected between October and February for the years 1998, 2010 and 2015.

139 Lynx are managed as a game species in most Scandinavia. Each year, a quota-limited hunt
140 season occurs from beginning of February until March 31st (Bischof et al. 2012). Quotas were
141 defined annually for 8 regions of Norway covering several counties based on the population
142 estimates made from family groups counts (Fig. 2) (Linnell et al. 2010). Records of harvested
143 lynx were reported by local hunters. For every harvested lynx, the date, location and sex of the
144 individual were reported. Here, we used records of harvested lynx from each year between 1997
145 and 2016 (Fig 3, Table 1).

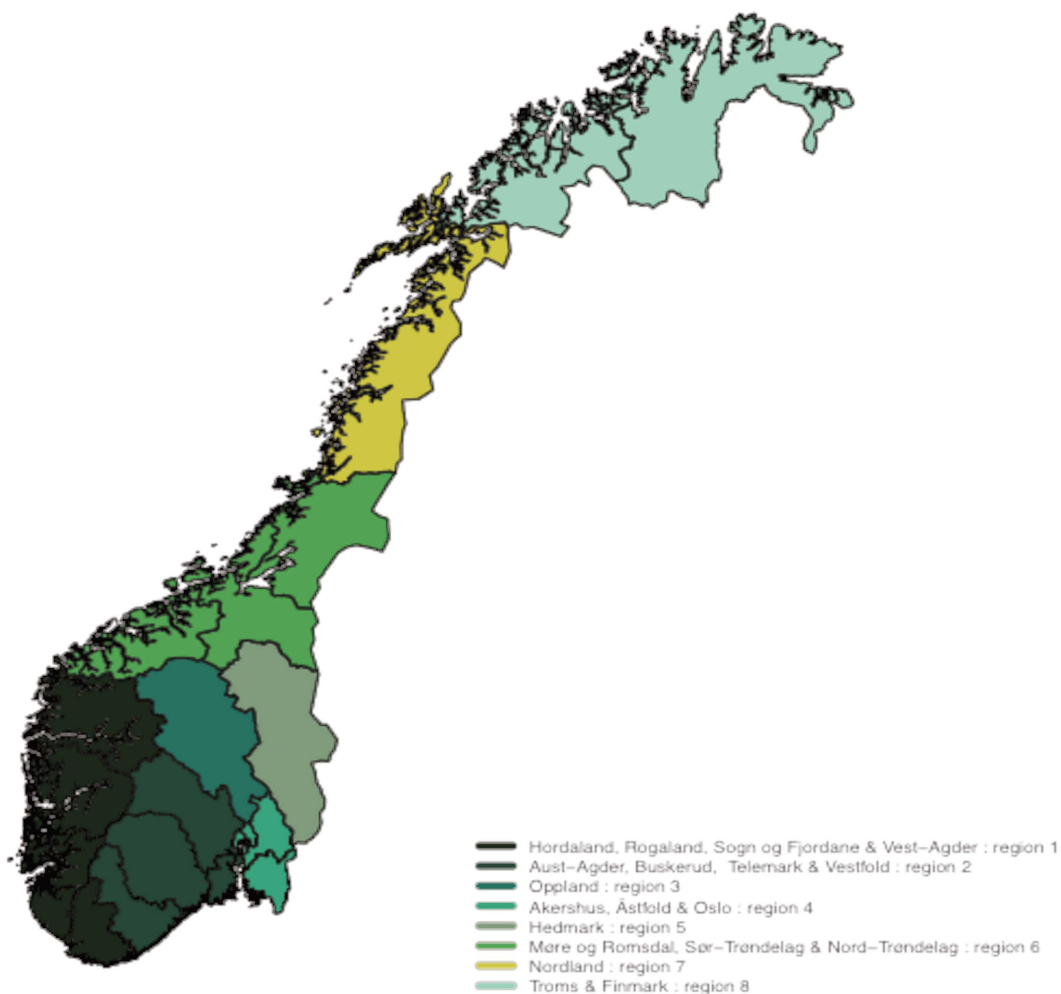
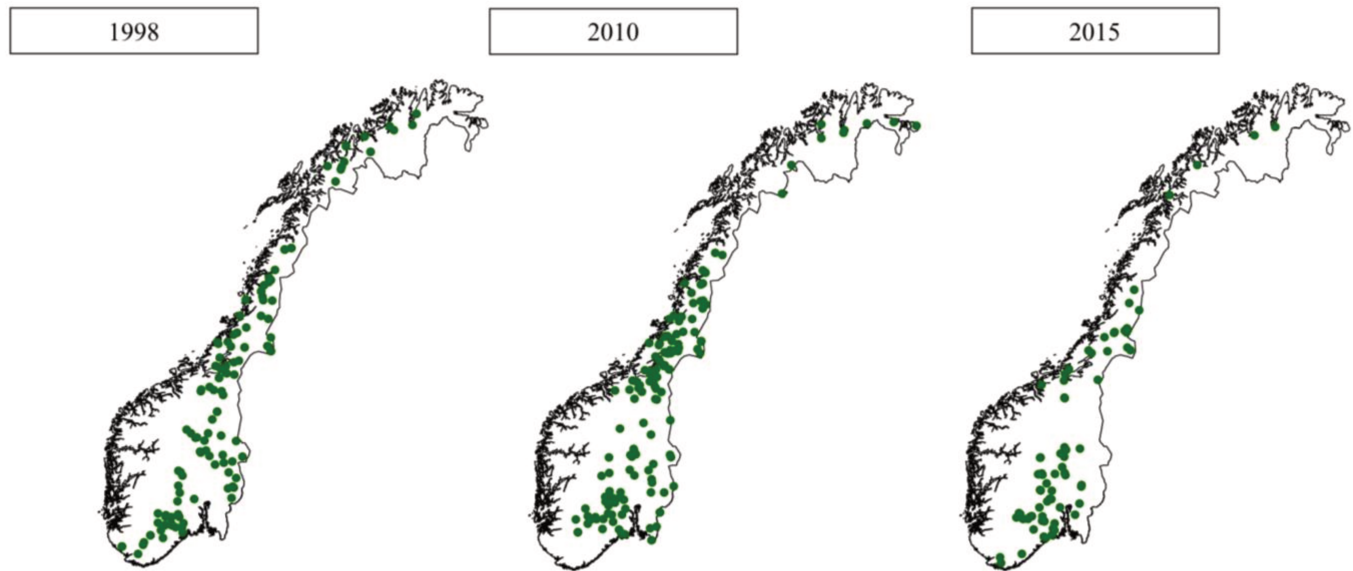


Figure 2: Map of the 8 regions of aggregated counties where quotas are defined each year for hunting lynx.



146

147 **Figure 3:** Maps of observation of harvested lynx for the years 1998, 2010 and 2015.

148

149 **Table 1:** Number of detections from the secondary surveys for the family group monitoring
 150 and number of harvested lynx for each year between 1997 and 2016

	Family groups					Harvested lynx
	October	November	December	January	February	
1997	0	0	38	0	0	93
1998	9	27	29	58	14	122
1999	8	32	25	82	14	97
2000	5	14	14	43	28	109
2001	0	6	14	39	24	86
2002	3	15	13	40	30	95
2003	10	12	16	23	17	70
2004	2	8	22	36	16	43
2005	0	17	43	54	24	50
2006	3	7	10	67	30	58
2007	6	20	44	70	55	77
2008	3	22	56	89	58	100
2009	7	18	52	67	52	119
2010	8	9	52	79	52	149
2011	7	51	21	50	39	145
2012	1	4	19	61	31	85
2013	9	10	36	40	40	86
2014	6	11	39	41	31	72
2015	6	16	19	50	47	80
2016	4	18	0	48	44	70

151 **Statistical modeling**

152 To model the distribution of lynx in Norway based on two different sources of data (family
153 groups and harvested lynx), we developed a combined SDM that fit both occupancy data and
154 presence-background data simultaneously. To compare the performance of the combined
155 model, we first analyzed separately family groups with site-occupancy models and harvested
156 lynx with point process models (PPM), then we fitted the combined SDM. We now describe in
157 more details the different models.

158 *Occupancy models for family groups*

159 Occupancy data are records of detection and non-detections at large scale through planned
160 surveys over the region of interest (Koshkina et al. 2017). To model the distribution of lynx
161 based on family groups monitoring, we used static (aka single-season) site-occupancy models
162 (MacKenzie et al. 2002). Site-occupancy models are particularly relevant to account for
163 imperfect detectability because the data collection process is formally incorporated (Isaac et al.
164 2014). We divided the country into 4044 sampling units, defined as 10x10 km cells. Site
165 occupancy models rely on the closure assumption which states that the ecological state of a site
166 (whether it is occupied or not) remains unchanged through sampling occasions j . The
167 monitoring occurred during the winter months, so we defined the sampling occasion j as
168 October, November, December, January or February. For each site i ($i = 1, \dots, 4044$) during
169 sampling occasion j ($j = 1, \dots, 5$) we defined the binary observations $y_{i,j}$ with $y_{i,j} = 1$ if the species
170 was detected at site i during sampling occasion j and $y_{i,j} = 0$ otherwise. The sampling effort
171 may vary in time and space but is non-null everywhere in the country (Linnell et al. 2010). This
172 assumption allowed us considering sites with no clues of presence as non-detections (i.e. $y_{i,j} =$
173 0) (Louvrier et al. 2018b; Molinari-Jobin et al. 2018). We defined Ψ_i the probability of
174 occupancy at site i and $p_{i,j}$ the probability to detect the species at site i during sampling occasion
175 j . Covariates were introduced using a logistic function on these parameters with β_x the vector

176 of regression coefficients on the occupancy probability and α the vector of regression
 177 coefficients on the detection probability to be estimated. The log-likelihood of a site-occupancy
 178 model can be written as follows (MacKenzie et al. 2002):

$$179 \quad l_{occ}(\alpha; \beta_x) = \log \left[\prod_{i=1}^k \Psi_i \prod_{j=1}^J p_{ij}^{y_{ij}} (1 - p_{ij})^{1 - y_{ij}} \right]$$

$$180 \quad + \log \left[\prod_{i=k+1}^I (\Psi_i \prod_{j=1}^J (1 - p_{ij}) + (1 - \Psi_i)) \right].$$

181 The first part of the log-likelihood formula corresponds to the k sites where at least one detection
 182 was made during the J occasions, and the second part to the $I - k$ sites with no detection.

183 *Point process models for harvested lynx*

184 PB data are made of a set of locations where signs of the species were observed, without any
 185 information on sites where the monitoring occurred but the species was not present, i.e.
 186 absences (Renner et al. 2015; Koshkina et al. 2017). Based on the analysis of PB data, point
 187 process models (PPMs) yield an intensity of species observations per unit area (Renner et al.
 188 2015). However, they usually suffer from observer bias, i.e., spatial patterns in the distribution
 189 of observers that may influence the distribution of observations. By correcting for observer
 190 bias, the output can be interpreted as an intensity of species observations per unit area (Renner
 191 et al. 2015). It is assumed that the locations of presence-background data are a realization of an
 192 inhomogeneous Poisson PPM (Cressie & Wikle 2015). Inhomogeneous Poisson PPMs are
 193 characterized by an intensity μ_i , corresponding to the mean number of individuals in a site i . We
 194 assumed that the point locations were independently distributed, conditional on environment.
 195 The intensity μ_i can be modelled as a log-linear function of covariates at location i .

$$196 \quad \log(\mu_i) = x_i \beta_x + v_i \beta_v$$

197 where x_i and v_i are vectors of environmental and observer bias variables at location s_i and β_x
 198 and β_v are the regression coefficients to be estimated. We chose the spatial resolution to be the

199 same as for the site-occupancy model, i.e., 10x10 km. To fit a Poisson PPM, the log-likelihood
 200 needs to be maximized and is written as follows (Renner et al. 2015):

$$201 \quad l_{PPM}(\beta_x; \beta_v) = \sum_{i=1}^m \log(\mu_i) - \int_{s \in A} \mu(s) ds$$

202 with m the number of locations, the first part of the log likelihood representing the realized
 203 number of presence points and the integral representing the expected number of presence
 204 points in the whole study region A .

205 *Combined model*

206 The combined SDM uses the continuous space in the Poisson PPM process to model the
 207 detection and non-detections used in a site-occupancy model framework. By doing so, we
 208 analyze both datasets within the same framework. Ψ_i can be formulated as the probability that
 209 at least one individual is on site s_i , that is, if we defined N_i the number of individuals present at
 210 site i :

$$211 \quad \Psi_i = P(N_i > 0) = 1 - P(N_i = 0).$$

212 We assume that N_i follows a Poisson distribution depending on the intensity function at site s_i :

$$213 \quad N_i \sim \text{Poisson}(\mu_i A_i)$$

214 where A_i is the surface of site i . Following (Dorazio 2014), we built the likelihood of the
 215 combined SDM by multiplying the likelihoods of the Poisson PPM model and the occupancy
 216 model, assuming that the presence-background and the occupancy datasets are independent of
 217 each other. Here, this assumption is likely to be valid because the data are collected during
 218 different time periods and for different goals. The log-likelihood for the combined SDM can
 219 therefore be expressed as:

$$220 \quad l_{COMB}(\beta_x; \beta_v; \alpha) = l_{PPM}(\beta_x; \beta_v) + l_{OCC}(\alpha; \beta_x)$$

221 Finally, to estimate the parameters of the combined SDM, we maximized this log-likelihood.

222 *Model validation and spatial dependence between observations*

223 PPMs assume that data points are independent in space. We ran a residual analysis (Baddeley
224 et al. 2015) to assess adequacy of the model for intensity, in particular by checking for a spatial
225 trend in residuals with a varying area of interaction radius between 1 km and 10 km (Fig. 4).
226 We found that data points were not independent in space because there was evidence of
227 additional clustering of observed locations beyond that which was explained by the ecological
228 and observer bias variables. We accounted for spatial dependence between data points by fitting
229 an area-interaction model for each year (Widom & Rowlinson 1970; Baddeley & Van Lieshout
230 1995). Area interaction models are point process models assuming interactions among all points
231 within a certain radius. The choice of the interaction radius was made by selecting the radius
232 that maximized the pseudo-likelihood of the spatial point process, using the *profilepl* function
233 in the *spatstat* R package (Baddeley et al. 2015). Finally, we fitted the PPM and combined SDM
234 both with and without point interactions for each year.

235 *Environmental covariates*

236 For the probability of occupancy, a range of environmental covariates were used including 5
237 habitat type covers, slope, environmental productivity, presence of wild preys and presence of
238 domestic preys. The variables were chosen according to previous established knowledge
239 regarding lynx habitat preferences in Scandinavia (May et al. 2008; Bouyer et al. 2015a).

240 Habitat type covers were obtained from a 30 x 30m Norway Land Cover (NORUT;
241 Johansen et al. 2009). We calculated the percentage of cover of 5 aggregated habitat classes in
242 each 10x10 km cell. The 5 classes were considered to be meaningful for lynx based on

243 knowledge of their habitat use and factors affecting their demography (Table 2) (May et al.
 244 2008; Basille et al. 2009; Bouyer et al. 2015a).

245

246 **Table 2:** Classification of the vegetation covers based on the aggregation of layers from the
 247 Norut database (Johansen et al. 2009).

Aggregated classes	Layers from the Norut map
Urban area (class 1)	Cities and built-up areas
Agricultural area (class 2)	Agricultural areas
Forest (class 3)	Coniferous Forest – dense canopy layer Coniferous forest and mixed forest - open canopy Lichen rich pine forest Low herb forest and broad leaved deciduous fore Tall herb - tall fern deciduous forest Bilberry- low fern birch forest Crowberry birch forest Lichen-rich birch forest
Open area below treeline (class 4)	Ombrotrophic bog and low-grown lawn vegetation Tall-grown lawn vegetation Wet mires, sedge swamps and reed beds
Open area above treeline (class 5)	Exposed alpine ridges, scree and rock complex Graminoid alpine ridge vegetation Heather-rich alpine ridge vegetation Lichen-rich heathland Heather- and grass-rich early snow patch com. Fresh heather and dwarf-shrub communities (up-/lowland) Herb-rich meadows (up-/lowland) Grass and dwarf willow snow-patch vegetation Bryophyte late snow patch vegetation Glacier, snow and wet snow-patch vegetation

248

249 Slope was averaged from a 50 x 50m Digital Elevation Model (DEM) raster (Norwegian
250 Mapping Authority). Environmental productivity was obtained from the normalized difference
251 vegetation index (NDVI; Gutman et al. 1997). For each cell, we determined whether wild preys
252 (roe deer *Capreolus capreolus* and red deer *Cervus elaphus*) or domestic preys (reindeer
253 *Rangifer tarandus*) were present based on maps realized during previous studies on distribution
254 of lynx preys (Odden et al. 2006; Mattisson et al. 2011; Gervasi et al. 2014).

255 For the detection probability in the site-occupancy model, we considered that two
256 factors could affect the parameter: the annual hunting quota and the population density. With
257 higher quotas, we expected the search effort to be high during the winter before the hunting
258 season hence increasing the detectability (Linnell et al. 2010). A previous study found that if
259 the number of people surveying the species increased, the probability to detect the species
260 increased as well (Louvrier et al. 2018b), therefore we expected human density to have a
261 positive effect on detectability. Human population was averaged in each 10x10 km from a 1x1
262 km raster from the Earth Observing System Data and Information System (EOSDIS). The
263 quotas information was given by the Norwegian Institute for Nature Research (NINA).

264 The ecological variables for the intensity parameter μ in the point process model were
265 chosen to be the same as the variables for the probability of occupancy. Finally, we set only
266 human density as a variable for the observer bias as it may have an influence on the number of
267 people going hunting.

268 *Variable selection*

269 We carried out variable selection using the LASSO technique (Renner et al. 2015). The LASSO
270 consists in augmenting the likelihood with a penalty function, which is the sum of the absolute
271 values of the α 's and β 's. The penalty is then maximized to control model complexity and
272 reduce the variance of parameter estimates. Several parameters are then estimated at 0 (the so-

273 called ‘sparsity’ property of the LASSO), which provides automatic variable selection. We ran
274 the LASSO for each year from 1997 to 2016. All the analyses were done in R using the
275 *ppmlasso* package (Renner et al. 2015). Below we display the results of the analyses of spatial
276 dependence for years 1997, 1998, 1999, 2002, 2003, 2004, 2006, 2010, 2011, 2012, and 2015.
277 We encountered convergence issues for years 2000, 2001, 2005, 2007, 2008, 2009, 2013, 2014
278 that precluded us from getting results. For convenience, we display the results from the variable
279 selection and further results for three years spread over the study period, namely 1998, 2010,
280 and 2015.

281 **Results**

282 *Accounting for spatial dependence*

283 We checked the residuals spatial trend with point interactions, to check for any aggregation left
284 for each year (Fig. 4). The radius of area interaction in order to account for spatial dependence
285 varied according to the years, between 1 km for the year 1998 and 10 km for the year 2015
286 (Appendix 1). Multiplication of the smoothed residuals by the square root of the area allowed
287 for comparisons to the standard ± 2 thresholds of Pearson residuals. Based on the analyses of
288 the Pearson residuals for the PPM only, accounting for spatial dependence appeared to be
289 significant for years 1998, 2002, 2003, 2006, 2010, and 2011 (Table 3).

290

291

292

293

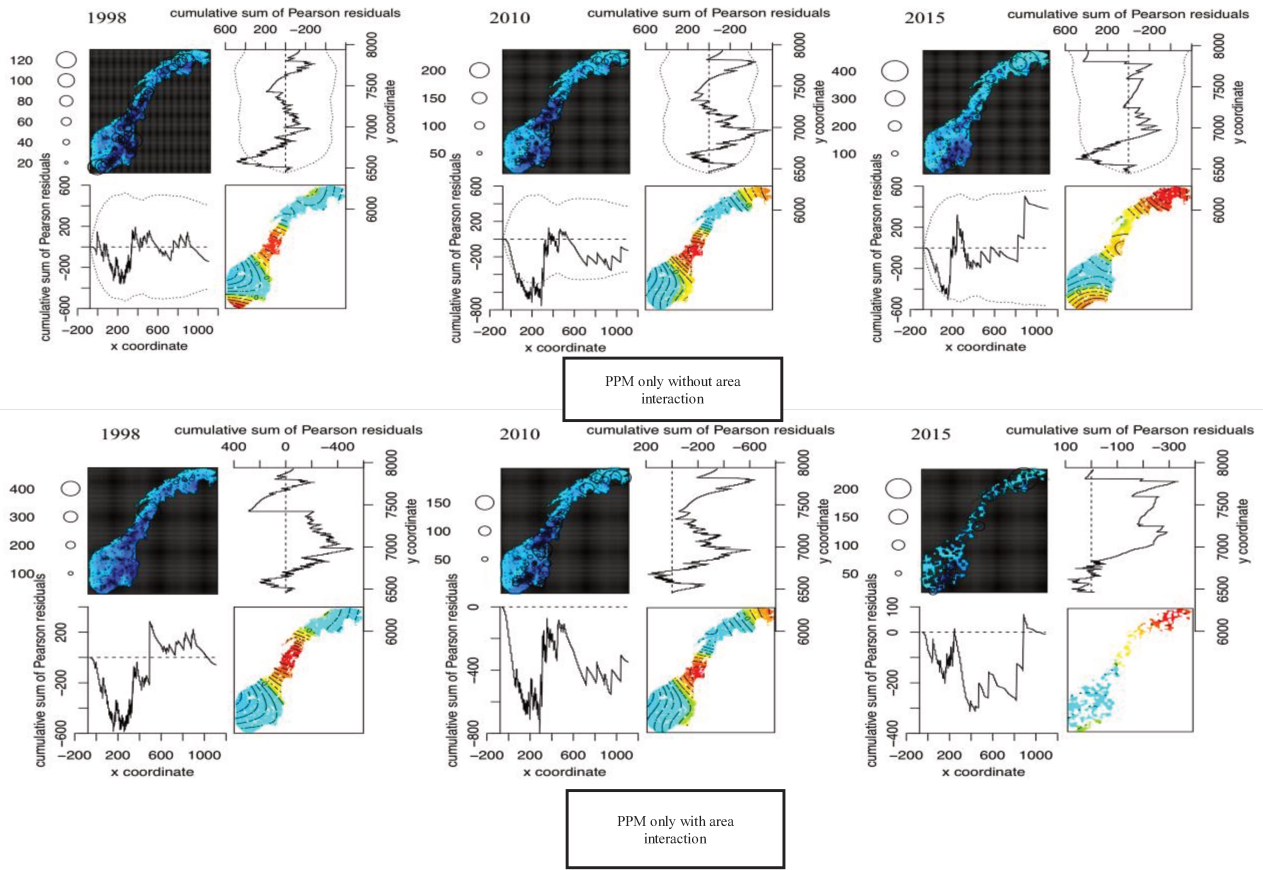
294 **Table 3:** Boundaries of residuals calculated as the multiplication of the range of smoothed field
 295 found in Appendix 1 with the area of entire window for the model not accounting for area
 296 interaction, and the range of smoothed field multiplied by the area of clipped window for the
 297 model accounting for area interaction. Accounting for area interaction appears significant when
 298 the difference of the resulting boundaries is higher than 2. The years during which accounting
 299 for spatial dependence appeared significant are displayed with a *.

300

Year	Without area interaction		With area interaction	
	Lower boundary of residuals	Upper boundary of residuals	Lower boundary of residuals	Upper boundary of residuals
1997	-0.8016501	0.834763763	-2.6511298	2.53473299
1998*	-0.9240023	0.797534272	-3.1527588	4.8201286
1999	-0.8201713	0.737293588	-2.6038621	2.37402264
2002*	-1.225954	1.241669004	-3.7347429	4.01775858
2003*	-0.9264344	1.261312705	-3.1663483	3.01154639
2004	-0.8574008	0.532999095	-2.4626497	1.69868469
2006*	-0.9372852	1.110523912	-3.4464097	3.75128665
2010*	-0.9715213	1.305464262	-3.140351	3.90963361
2011*	-0.950381	1.477019253	-3.352465	2.49987302
2012	-0.6922066	0.94214933	-2.3391627	3.39146092
2015	-0.6968837	1.514248744	-2.3403444	3.09958257

301

302



303 **Figure 4:** Analyses of the Pearson residuals without accounting for spatial dependence (up) and
 304 accounting for spatial dependence (bottom) for the years 1998, 2010 and 2015.

305
 306 *Environmental Covariates*

307 The variable selection using the LASSO technique according to the AIC criterion selected all
 308 the covariates we included for all the years (Fig. 5). The intensity of the PPM accounting for
 309 spatial dependence, the intensity of the combined model accounting for spatial dependence and
 310 the probability of occupancy increased when the slope, the primary productivity (NDVI), the
 311 proportion of agricultural area (cover 2), and the proportion of forest (cover 3) increased (table
 312 5) for years 1998, 2010 and 2015 (Table 5). For years 1998 and 2010, the intensities for both
 313 Combined and PPM only and the occupancy probability slightly increased when the proportion
 314 of urban area increased, and those ecological parameters decreased when the proportion

315 increased for the year 2015. The same ecological parameters decreased when proportion of
 316 open area below treeline (cover 4) increased for the years 1998, 2010 and 2015 except for the
 317 PPM only models in the years 1998 and 2015. In 1998, only in the occupancy model the
 318 probability of occupancy decreased when the proportion of open area above treeline (cover 5)
 319 decreased (Table 5). In 2010, for all models, the ecological parameters increased when this
 320 proportion increased. In 2015, only in the occupancy model the probability of occupancy
 321 increased when this proportion increased. Finally, for 1998 and 2010, the observer bias for the
 322 combine d and PPM only decreased when human density increased (Table 5).

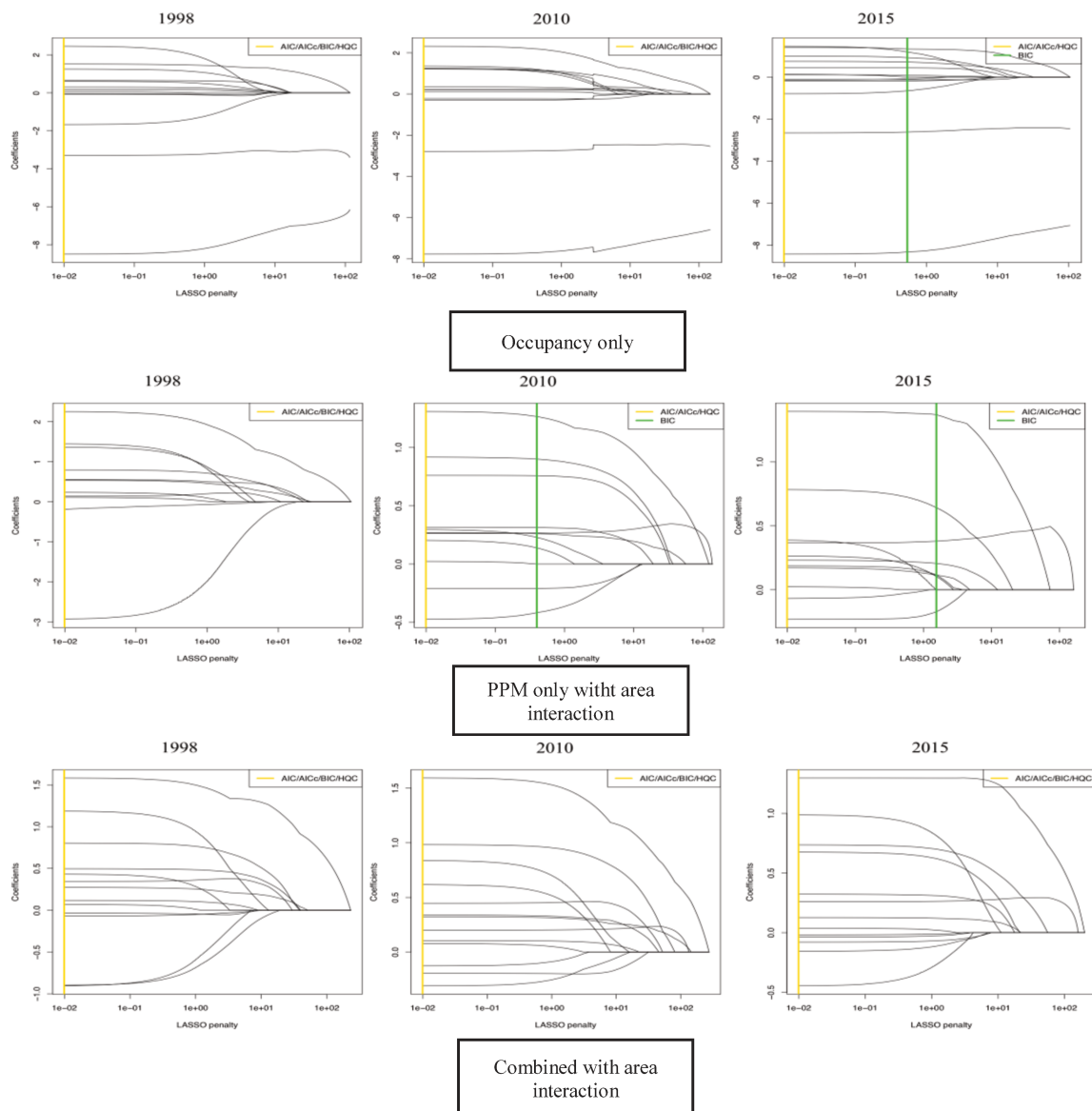


Figure 5: Variable selection using the LASSO technique (Renner et al. 2015) for the years 1998, 2010 and 2015. The Poisson PPM likelihood is augmented with a penalty function. On the plot, each curve represents a variable in the model. The x axis is a function of the regularization penalty parameter. The y axis gives the value of the variables. The vertical line shows, depending on the selection criteria, the optimal value of the penalty parameter and the values of the selected variables.

323 In 2015, the observer bias increased when human density increased. For 1998, 2010 and 2015,
 324 for both the occupancy and the combined models, the detection probability decreased when
 325 human density increased. The detection probability increased in both the combined and the
 326 occupancy models when the quota increased for the years 1998 and 2010 and decreased for
 327 year 2015.

328
 329 **Table 5:** Effects of the ecological and observer biases on the intensity of the combined model
 330 with area interaction, the PPM with area interaction fitted on one source of data (harvested
 331 lynx) and on the probability for a site to be occupied in the occupancy model fitted to family
 332 group data, for years 1998, 2010 and 2015. Environmental biases and observer biases for the
 333 PPM and combined models are given on a log scale and covariates for the detection and
 334 occupancy probability for the occupancy model are given on a logit scale

	Combined model			PPM Only model			Occupancy model		
	with area interaction			with area interaction					
1998	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
Intercept for									
environmental	-9.28	-9.13	-7.89	-10.12	-9.66	-8.20	-8.49	-8.17	-6.48
bias									
Slope	0	0.48	0.50	0	0.50	0.55	0	0.48	0.61
NDVI	0	0.83	1.19	0	0.81	1.45	0	1.75	2.46
Reindeer	0	0.75	0.80	0	0.71	0.80	0	1.05	1.25
Wild	0	0.35	0.38	0	0.15	0.23	0	0.63	0.65
Cover 1	0	0	0.07	0	0.10	0.24	0	0	0.06
Cover 2	0	0.24	0.27	0	0.46	0.54	0	0.25	0.30
Cover 3	0.25	1.46	1.58	0.22	1.92	2.25	0.28	1.44	1.51

Cover 4	-0.05	-0.04	0	0	0.01	0.11	-0.10	-0.04	0
Cover 5	0	0.21	0.43	0	0.86	1.36	-0.15	-0.07	0
Human density									
for observer	-0.90	-0.60	0	-2.92	-1.97	0			
bias									
Intercept for									
detection	-1.98	-1.95	-1.92				-3.30	-3.22	-3.01
probability									
Quota	0	0.10	0.12				0	0.15	0.17
Human density	-0.90	-0.48	0				-1.67	-1.21	0
2010									
Intercept for									
environmental	-8.81	-8.71	-7.64	-9.10	-9.01	-7.97	-7.77	-7.61	-6.66
bias									
Slope	0	0.32	0.34	0	0.31	0.32	0	0.29	0.34
NDVI	0	0.61	0.84	0	0.13	0.30	0	0.84	1.24
Reindeer	0	0.94	0.98	0	0.87	0.92	0	1.06	1.23
Wild	0	0.45	0.46	0	0.75	0.76	0	0.22	0.26
Cover 1	0	0.03	0.08	0	0	0.02	0	0.10	0.11
Cover 2	0	0.31	0.32	0	0.25	0.27	0	0.99	1.22
Cover 3	0.22	1.49	1.59	0.10	1.19	1.31	0.22	1.99	2.32
Cover 4	-0.20	-0.20	0	-0.21	-0.20	0	-0.24	-0.23	0
Cover 5	0	0.45	0.62	0	0.03	0.20	0	0.95	1.35

Human density									
for observer	-0.31	-0.22	0	-0.47	-0.35	0			
bias									
Intercept for									
detection	-1.76	-1.71	-1.71				-2.78	-2.73	-2.44
probability									
Quota	0	0.10	0.10				0	0.21	0.21
Human density	-0.13	-0.07	0				-0.30	-0.27	0
2015									
Intercept for									
environmental	-9.52	-9.42	-8.13	-10.36	-10.12	-8.60	-8.42	-8.25	-7.11
bias									
Slope	0	0.69	0.74	0	0.66	0.78	0	0.69	0.75
NDVI	0	0.78	0.99	0	0.04	0.39	0	1.04	1.47
Reindeer	0	0.61	0.67	0	0.13	0.26	0	0.88	1.00
Wild	0	0.31	0.32	0	0.13	0.19	0	0.44	0.45
Cover 1	-0.16	-0.11	0	-0.07	0	0	-0.13	-0.05	0
Cover 2	0	0.12	0.13	0	0.21	0.23	0	0.08	0.12
Cover 3	0.25	1.30	1.30	0	1.37	1.40	0.22	1.33	1.42
Cover 4	-0.04	-0.02	0	0	0.12	0.17	-0.18	-0.16	0
Cover 5	-0.02	-0.01	0	-0.23	-0.19	0	0	0	0.13
Human density									
for observer	0	0.02	0.04	0	0	0.02			
bias									

Intercept for								
detection	-1.71	-1.69	-1.63			-2.65	-2.58	-2.40
probability								
Quota	-0.08	-0.07	0			-0.12	-0.11	0
Human density	-0.45	-0.24	0			-0.79	-0.54	0

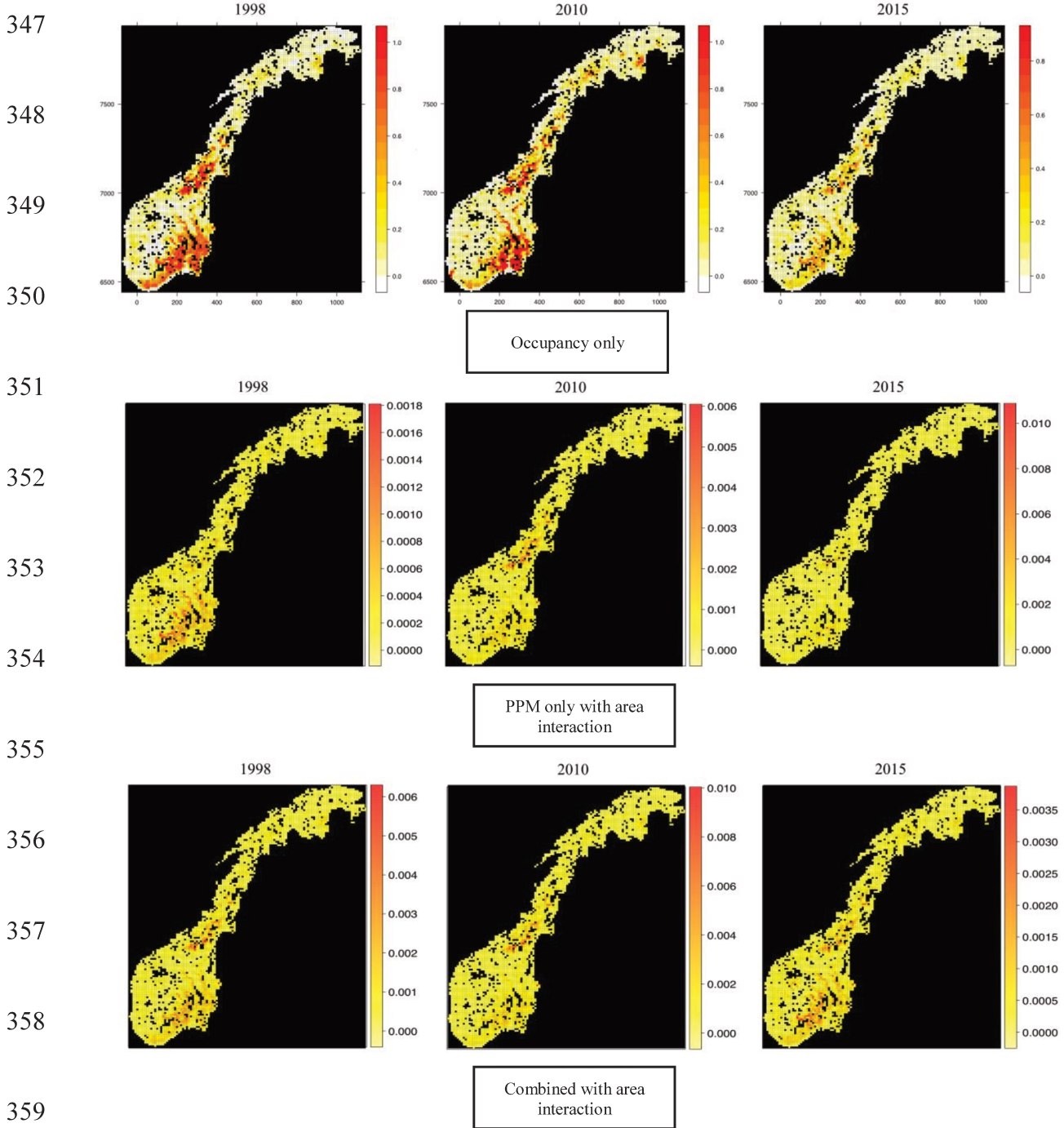
335

336 *Distribution dynamics of lynx in Scandinavia*

337 Overall, the distribution of lynx in Scandinavia decreased according to the occupancy model
338 (Fig. 6 first row), with two cores in central Norway and in the South-East in 1998 and 2010 and
339 a single core in central Norway in 2015. According to the PPM model accounting for spatial
340 dependence, the distribution shifted from South to central Norway as well, but with an
341 increasing average intensity (Fig. 6 second row). In the combined SDM, the average intensity
342 was higher than in the PPM with harvested data only, except for year 2015. The resulting
343 distribution according to the combined model shifted to central Norway between 1998 and 2010
344 but shifted South with a core distribution in central Norway in 2015.

345

346



360 **Figure 6:** Probability of occupancy maps for the years 1998, 2010 and 2015 (first row), intensity
 361 maps from the PPM model accounting for spatial dependence for the years 1998, 2010 and 2015
 362 (second row) and intensity maps from the combined model accounting for spatial dependence
 363 for the years 1998, 2010 and 2015 (last row).

364 **Discussion**
365

366 We demonstrated the use of a combined SDM, using PB data and occupancy data
367 simultaneously. By casting our combined SDM in a unifying point process modeling
368 framework, we used existing techniques to inspect spatial residuals and efficiently select
369 covariates. We applied the new combined SDM to the Norwegian lynx and found important
370 factors such as slope and primary productivity that underlie lynx distribution across years.

371 *Effects of environmental variables on lynx distribution in Norway*

372 Overall, we found that the effects of the environmental factors were similar across models, i.e.
373 the combined SDM and each SDM in isolation. Slope, primary productivity, forest cover,
374 agricultural cover and the presence of wild preys and reindeers explained lynx presence, in
375 agreement with previous studies (Basille et al. 2009, 2013; Bouyer et al. 2015b, 2015a). Human
376 presence is usually negatively correlated to slope (Basille et al. 2009), and here the positive
377 effect of slope highlighted the influence of human presence on lynx presence. We found that
378 the presence of lynx increased with the presence of reindeer and wild preys. Lynx are
379 considered a specialist species because they predate only on a small range of specific species
380 (Breitenmoser & al 2000). However, in the areas in which preys were present, we didn't have
381 information about variation in prey density. The primary productivity indirectly represented the
382 variation in prey density. We found that the presence of lynx increased with the primary
383 production, which confirms that lynx was more likely to be present wherever preys density
384 increased. We found that the presence of lynx increased with increased forest cover. We
385 explained this pattern by the fact that lynx live mainly in forest (Niedziałkowska et al. 2006),
386 and may hunt on roe deer and red deer, which are known to live in forests and farmland-forest
387 mosaics (Bunnefeld et al. 2006; Meisingset et al. 2013). We also found a positive effect of
388 farmland cover which confirms the presence of lynx influenced by the presence of roe deer and

389 red deer. Reindeer are more associated with alpine tundra and high-altitude areas (Mattisson et
390 al. 2011) however, depending on the years, the effects of the proportion of open area below and
391 above treeline, which corresponds to the habitat of reindeer was positive or negative, which
392 might be due to the sparsity of detections found in these areas, making this relationship
393 uncertain. We found a positive effect of urban cover on lynx presence, which might be due to
394 the detection of family groups near roads and habitations, leading to some bias in the effect of
395 urban cover on the presence of lynx (Linnell et al. 2010). Finally, we found a negative effect of
396 human density on the observer bias and the detection probability. This effect might be explained
397 by the fact that family groups and hunting data are based on detection of snow tracks, which
398 might become undetectable with too many people passing by.

399 *Trends in lynx distribution in Norway between 1997 and 2016*

400 Lynx got almost exterminated by the mid 20th century and persisted in one or two remaining
401 cores in southeastern and central Norway (Basille et al. 2009). Those two cores still remain
402 occupied by the species (Fig. 6). Linnell et al. (2010) found that population abundance varied
403 greatly from a peak in 1997 until its lowest point in 2004 and increased again, until it exceeded
404 the national conservation goals in several regions. Four years after the introduction of quota
405 hunting, the species was completely exterminated from South-western Norway (Region 1)
406 where an unlimited quota was set (Linnell et al. 2010). We found that in central Norway, lynx
407 remained present across the years in this core despite a fluctuating maximum intensity varying
408 between 0.0015 in 2002 and 0.010 in 2010 (Fig. 6). In South-East Norway, lynx remained but
409 with a lower intensity around 0.002 than in the core in Central Norway. In Northern Norway,
410 lynx was present in the area in a sporadic way, absent in 1999 and 2006. Finally, no lynx was
411 present in South-western Norway, which confirms what was found by Linnell et al. (2010).

412

413 *Advantages in combining presence-background data with occupancy data*

414 Modelling species distribution based on PB data is widely used because these data are
415 frequently available across broad scales (Fletcher et al. 2016). However, PB data often suffer
416 from sample selection bias, whereby presence signs are collected near roads or easily accessible
417 areas (Phillips et al. 2009). In contrast, site-occupancy data are detections and non-detections
418 and can account for imperfect and possibly heterogeneous detection. However, site-occupancy
419 models depend on the spatial resolution of the study site, with occupancy estimates being higher
420 in larger study sites even though abundance remains constant (Steenweg et al. 2018). In
421 contrast, the parameters of a point-process model do not depend on the spatial scale because
422 the outputs are expressed as an intensity per area in opposition to an intensity per grid cell (as
423 in MaxEnt, (Renner & Warton 2013). As a consequence, combined SDMs can be used to
424 predict occurrence of individuals for any sampling site size (Dorazio 2014; Koshkina et al.
425 2017).

426 When it comes to our case study, the combination had several advantages. First, the
427 results of the PB data analysis should be interpreted carefully. The PB data are based on hunted
428 lynx therefore an increase in the distribution intensity might actually mean an increase in lynx
429 mortality. The combination with occupancy data based on indirect signs should lower this
430 potential bias. Second, the occupancy monitoring relies on the detection of adult females with
431 cubs. By adding the harvested data in the analysis, which include both females and males, the
432 combined SDM produce the species distribution for both sexes. Third, we found an increased
433 precision in the estimates of each covariate's effect in comparison with the estimates from only
434 one data source, as expected (Dorazio 2014). Making optimal use of all the data we have is of
435 particular importance when dealing with a cryptic species such as the lynx. Fourth, occupancy
436 data were collected by people detecting tracks in the snow, hence in rather populated areas,

437 while harvest lynx data were collected by hunters going to places with low human density. The
438 combined analysis of both datasets therefore allows sharing information, as illustrated in the
439 combined SDM map (Fig. 6 and Appendix 3). For most of the years, the occupancy model
440 estimated a broader distribution of lynx in central and South East Norway while the point
441 process based on the PB data barely estimated a few sites with high intensity. In 2002, only the
442 occupancy model estimated occupied sites in North-East Norway.

443 *Methodological novelty*

444 The approach developed by Renner and Gimenez (2018) offers solutions to existing issues in
445 combined SDMs. First, covariate selection was difficult for combined models, because no
446 method was really developed to deal with variable selection. Here we used the LASSO
447 technique in a single step and automatically making the covariate selection process easier for
448 users.

449 Second, checking the residuals of combined SDM was previously not possible. In the
450 lynx case study, accounting for spatial dependence was however needed for most of the years
451 according to the residuals check. Spatial dependence in the data locations might have resulted
452 from both biological processes such as social aggregation or the observation process due to the
453 unmeasured sampling effort.

454 Third, current combined SDMs rely on the Poisson assumption, which no longer holds
455 when spatial clustering occurs. The new approach allows fitting combined SDMs that explicitly
456 account for spatial clustering across the years. As a result, we found that inference about the
457 species distribution was improved by reducing the magnitude of the residuals.

458 *Limits and perspectives*

459 An important hypothesis that needs to be discussed is that species that are present at a site do
460 not leave (or enter from another site) during the data collection (Rota et al. 2009). This
461 hypothesis applies especially for site-occupancy data. This hypothesis is likely to be satisfied
462 for species with limited movements but is likely to be violated for mobile species such as lynx.
463 Because lynx territory can vary from a few hundred km² up to 1000 km² (Sunde et al. 2000), it
464 is very likely that movements occurred inside or outside of the sampling sites. Assuming that
465 those movements are random, the bias in occupancy estimates is minimal, but occupancy should
466 be interpreted as habitat use rather than the proportion of area occupied by the species
467 (Mackenzie & Royle 2005).

468 In the Norwegian lynx monitoring, other datasets are available such as GPS data and
469 citizen science data collected by the public with a smartphone application. The challenge in
470 dealing with GPS data is that they occur at much finer time and spatial scales. Such fine-scale
471 data could help bring new insight into the distribution and territory use of lynx at a finer scale
472 and compare the outputs we have at a broader scale. On the other hand, combining site
473 occupancy data with citizen science data could bring a valuable source of information by
474 covering even wider areas (Schmeller et al. 2009). However, the lynx citizen science data are
475 not checked through a standardized protocol, which means that false positives might occur (i.e.
476 false detections of the species at sites where it is absent). Accounting for false positives has
477 been developed in the site occupancy modelling framework (Miller et al. 2011), in particular
478 for large carnivores (Louvrier et al. 2018a) but remains to be investigated in point process
479 models.

480 The family group monitoring is the most comprehensive monitoring of lynx in Norway
481 (Linnell et al. 2007). This monitoring was set up to get a minimum abundance index.
482 Interestingly, we found trends in the lynx distribution that coincide with trends in the estimated
483 abundance, therefore raising the question of using occupancy as a proxy for abundance

484 (Mackenzie 2006). While the occupancy-abundance relationship has been questioned when
485 occupancy is estimated in continuous habitats (Efford & Dawson 2012), Linden et al. (2017)
486 have showed that occupancy and abundance estimates were correlated when the area of the
487 study sites scale is similar to that of home ranges. In the lynx case study, the area of grid cells
488 might be too small to establish a reliable link between occupancy and abundance. The fine-
489 scale GPS data could help in comparing the variation in territory sizes and the territory use in
490 time with the occupancy data.

491 A widely used method for assessing the predictive ability of species distribution models
492 is to evaluate the area under the receiver operating characteristic curve (AUC – Elith et al.
493 2006). However, the AUC has been shown to be biased in occupancy models due to imperfect
494 detection (Zipkin et al. 2012). The development of an AUC criterion to assess the predictive
495 ability of combine SDMs remains an open issue.

496 Finally, we used static SDMs to assess trends in the lynx distribution in Norway, by
497 analyzing each year separately. To better understand the mechanisms underlying trends, our
498 approach should be extended to allow the estimation of local colonization and extinction
499 processes. While dynamic (aka multi-season) site-occupancy models have been developed
500 (MacKenzie et al. 2003), the point process modelling framework yet fails to acknowledge time.

501

502

503 **Conclusion**

504 Combining PB data with occupancy data in a combined model to understand species
505 distribution offers a number of advantages. Occupancy data on females provided useful insight
506 into the distribution by accounting for imperfect detection, while PB data provided a scale-
507 independent framework and information about the distribution of males. The combined SDM
508 formulated in a unifying point process framework offers the methodological flexibility to check

509 spatial residuals and look for spatial clustering, to fit area-interaction models if needed and to
510 use the LASSO technique for automatic covariate selection. Finally, our results bring new
511 insights into the management of lynx in Norway, by highlighting places where lynx may occur
512 and therefore help in targeting specific areas, determining priorities for implementing
513 protection measures and defining sustainable harvesting quotas.

514

515 **References**

- 516 Andrén H, Linnell JDC, Liberg O, Ahlqvist P, Andersen R, Danell A, Franzén R, Kvam T,
517 Odden J, Segerström P. 2002. Estimating total lynx *Lynx lynx* population size from censuses
518 of family groups. *Wildlife Biology* 8:299–306.
- 519 Baddeley A, Rubak E, Turner R. 2015. *Spatial point patterns: methodology and applications*
520 with R. Chapman and Hall/CRC.
- 521 Baddeley AJ, Van Lieshout M. 1995. Area-interaction point processes. *Annals of the Institute*
522 *of Statistical Mathematics* 47:601–619.
- 523 Basille M, Herfindal I, Santin-Janin H, Linnell JDC, Odden J, Andersen R, Arild Høgda K,
524 Gaillard J-M. 2009. What shapes Eurasian lynx distribution in human dominated landscapes:
525 selecting prey or avoiding people? *Ecography* 32:683–691.
- 526 Basille M, Van Moorter B, Herfindal I, Martin J, Linnell JDC, Odden J, Andersen R, Gaillard
527 J-M. 2013. Selecting Habitat to Survive: The Impact of Road Density on Survival in a Large
528 Carnivore. *PLoS ONE* 8:e65493.
- 529 Bischof R, Nilsen EB, Brøseth H, Männil P, Ozoliņš J, Linnell JDC. 2012. Implementation
530 uncertainty when using recreational hunting to manage carnivores: Carnivore hunting quotas.
531 *Journal of Applied Ecology* 49:824–832.
- 532 Bouyer Y, Gervasi V, Poncin P, Beudels-Jamar RC, Odden J, Linnell JDC. 2015a. Tolerance
533 to anthropogenic disturbance by a large carnivore: the case of Eurasian lynx in south-eastern
534 Norway: Lynx tolerance to human disturbance in Norway. *Animal Conservation* 18:271–278.
- 535 Bouyer Y, Rigot T, Panzacchi M, Moorter BV, Poncin P, Beudels-Jamar R, Odden J, Linnell
536 JDC. 2015b. Using Zero-Inflated Models to Predict the Relative Distribution and Abundance
537 of Roe Deer Over Very Large Spatial Scales. *Annales Zoologici Fennici* 52:66–76.
- 538 Braa JT, Brainerd SM, Brøseth H, Knutsen E, Linnell JD. 1999. Forslag til Nasjonalt
539 overvåkingsprogram for store rovdyr [Proposal for a National Program for Monitoring Large
540 Carnivores]. *Utredning for DN* 2000 1.
- 541 Breitenmoser U, al. 2000. Action plan for the conservation of the Eurasian lynx in Europe
542 (*Lynx lynx*). Council of Europe.
- 543 Bunnefeld N, Linnell JDC, Odden J, van Duijn MAJ, Andersen R. 2006. Risk taking by
544 Eurasian lynx (*Lynx lynx*) in a human-dominated landscape: effects of sex and reproductive
545 status. *Journal of Zoology* 270:31–39.

- 546 Chapron G et al. 2014. Recovery of large carnivores in Europe's modern human-dominated
547 landscapes. *Science* 346:1517–1519.
- 548 Cressie N, Wikle CK. 2015. *Statistics for spatio-temporal data*. John Wiley & Sons.
- 549 Dorazio RM. 2014. Accounting for imperfect detection and survey bias in statistical analysis
550 of presence-only data. *Global Ecology and Biogeography* 23:1472–1484.
- 551 Duchamp C et al. 2012. A dual frame survey to assess time- and space-related changes of the
552 colonizing wolf population in France. *Hystrix* 23:14–28.
- 553 Efford MG, Dawson DK. 2012. Occupancy in continuous habitat. *Ecosphere* 3:1–15.
- 554 Elith J et al. 2006. Novel methods improve prediction of species' distributions from
555 occurrence data. *Ecography* 29:129–151.
- 556 Elith J, Leathwick JR. 2009. *Species Distribution Models : Ecological Explanation and
557 Prediction Across Space and Time*. *Annual review of ecology, evolution, and systematics*
558 40:677–697.
- 559 Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011. A statistical explanation of
560 MaxEnt for ecologists. *Diversity and Distributions* 17:43–57.
- 561 Fithian W, Elith J, Hastie T, Keith DA. 2015. Bias correction in species distribution models:
562 pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*
563 6:424–438.
- 564 Fletcher RJ, McCleery RA, Greene DU, Tye CA. 2016. Integrated models that unite local and
565 regional data reveal larger-scale environmental relationships and improve predictions of
566 species distributions. *Landscape Ecology* 31:1369–1382.
- 567 Gervasi V, Nilsen EB, Odden J, Bouyer Y, Linnell JDC. 2014. The spatio-temporal
568 distribution of wild and domestic ungulates modulates lynx kill rates in a multi-use landscape:
569 Lynx prey selection in a multi-use landscape. *Journal of Zoology* 292:175–183.
- 570 Guillera-arroita G, Lahoz-monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, Mccarthy
571 MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose? Matching
572 data and models to applications. *Global Ecology and Biogeography* 24:276–292.
- 573 Guisan A, Thuiller W. 2005. Predicting species distribution : offering more than simple
574 habitat models. *Ecology Letters* 8:993–1009.
- 575 Gutman G, Tarpley D, Ignatov A, Olson S. 1997. *Global monthly AVHRR climatology over
576 land clear-sky top-of-the-atmosphere variables*. Boulder, Colorado.
- 577 Hefley TJ, Hooten MB. 2016. *Hierarchical Species Distribution Models*. *Current Landscape
578 Ecology Reports* 1:87–97.
- 579 Isaac NJB, Strien AJV, August TA, Zeeuw MPD, Roy DB. 2014. *Statistics for citizen
580 science : extracting signals of change from noisy ecological data*. *Methods in Ecology and
581 Evolution* 5:1052–1060.
- 582 Johansen B, Arrestad P, Øien D. 2009. *Vegetasjonskart for Norge basert på satellittdata;
583 Delprosjekt 1: Klasseinndeling og beskrivelse av utskilte vegetasjonstyper*. Rapport 3:2009.
- 584 Koshkina V, Wang Y, Gordon A, Dorazio RM, White M, Stone L. 2017. Integrated species
585 distribution models : combining presence-background data and site-occupancy data with
586 imperfect detection. *Methods in Ecology and Evolution* 8:420–430.
- 587 Linden DW, Fuller AK, Royle JA, Hare MP. 2017. Examining the occupancy-density

588 relationship for a low-density carnivore. *Journal of Applied Ecology* 54:2043–2052.

589 Linnell JDC, Broseth H, Odden J, Nilsen EB. 2010. Sustainably Harvesting a Large
590 Carnivore? Development of Eurasian Lynx Populations in Norway During 160 Years of
591 Shifting Policy. *Environmental Management* 45:1142–1154.

592 Linnell JDC, Fiske P, Odden J, Brøseth H, Herfindal I, Andersen R. 2007. An evaluation of
593 structured snow-track surveys to monitor Eurasian lynx *Lynx lynx* populations. *Wildlife*
594 *Biology* 13:456–466.

595 Louvrier J et al. 2018a. Use of ambiguous detections to improve estimates from species
596 distribution models. *Conservation Biology*.

597 Louvrier J, Duchamp C, Lauret V, Marboutin E, Cubaynes S, Choquet R, Miquel C, Gimenez
598 O. 2018b. Mapping and explaining wolf recolonization in France using dynamic occupancy
599 models and opportunistic data. *Ecography* 41:647–660.

600 Mackenzie DI. 2006. Modeling the probability of resource use: The effect of, and dealing
601 with, detecting a species imperfectly. *Journal of Wildlife Management* 70:367–374.

602 MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB. 2003. Estimating
603 occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*
604 84:2200–2207.

605 MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002.
606 Estimating site occupancy rates when detection probabilities are less than one. *Ecology*
607 83:2248.

608 Mackenzie DI, Royle JA. 2005. Designing occupancy studies: General advice and allocating
609 survey effort. *Journal of Applied Ecology* 42:1105–1114.

610 Mattisson J, Odden J, Nilsen EB, Linnell JDC, Persson J, Andrén H. 2011. Factors affecting
611 Eurasian lynx kill rates on semi-domestic reindeer in northern Scandinavia: Can ecological
612 research contribute to the development of a fair compensation system? *Biological*
613 *Conservation* 144:3009–3017.

614 May R, van Dijk J, Wabakken P, Swenson JE, Linnell JDC, Zimmermann B, Odden J,
615 Pedersen HC, Andersen R, Landa A. 2008. Habitat differentiation within the large-carnivore
616 community of Norway's multiple-use landscapes. *Journal of Applied Ecology* 45:1382–1391.

617 Meisingset EL, Loe LE, Brekkum Ø, Van Moorter B, Mysterud A. 2013. Red deer habitat
618 selection and movements in relation to roads. *The Journal of Wildlife Management* 77:181–
619 191.

620 Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA. 2011. Improving
621 occupancy estimation when two types of observational error occur: Non-detection and species
622 misidentification. *Ecology* 92:1422–1428.

623 Molinari-Jobin A et al. 2018. Mapping range dynamics from opportunistic data:
624 Spatiotemporal modelling of the lynx distribution in the Alps over 21 years. *Animal*
625 *Conservation* 21:168–180.

626 Niedziałkowska M, Jędrzejewski W, Mysłajek RW, Nowak S, Jędrzejewska B, Schmidt K.
627 2006. Environmental correlates of Eurasian lynx occurrence in Poland – Large scale census
628 and GIS mapping. *Biological Conservation* 133:63–69.

629 Odden J, Linnell JDC, Andersen R. 2006. Diet of Eurasian lynx, *Lynx lynx*, in the boreal
630 forest of southeastern Norway: the relative importance of livestock and hares at low roe deer
631 density. *European Journal of Wildlife Research* 52:237–244.

632 Odden J, Linnell JDC, Moa PF, Herfindal I, Kvam T, Andersen R. 2002. Lynx Depredation
633 on Domestic Sheep in Norway. *The Journal of Wildlife Management* 66:98–105.

634 Packer C et al. 2013. Conserving large carnivores: Dollars and fence. *Ecology Letters*
635 16:635–641.

636 Pedersen VA, Linnell JDC, Andersen R, Andrén H, Lindén M, Segerström P. 1999. Winter
637 lynx *Lynx lynx* predation on semi-domestic reindeer *Rangifer tarandus* in northern Sweden.
638 *Wildlife Biology* 5:203–211.

639 Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species
640 geographic distributions. *Ecological Modelling* 190:231–259.

641 Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample
642 selection bias and presence-only distribution models: implications for background and
643 pseudo-absence data. *Ecological Applications* 19:181–197.

644 Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI.
645 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution*
646 6:366–379.

647 Renner IW, Gimenez O. 2018. Species distribution modelling with combined data sources.
648 Unpublished

649 Renner IW, Warton DI. 2013. Equivalence of MAXENT and Poisson Point Process Models
650 for Species Distribution Modeling in Ecology: Equivalence of MAXENT and Poisson Point
651 Process Models. *Biometrics* 69:274–281.

652 Rota CT, Fletcher Jr RJ, Dorazio RM, Betts MG. 2009. Occupancy estimation and the closure
653 assumption. *Journal of Applied Ecology* 46:1173–1181.

654 Schmeller DS et al. 2009. Ventajas del monitoreo de biodiversidad basado en voluntarios en
655 Europa. *Conservation Biology* 23:307–316.

656 Steenweg R, Hebblewhite M, Whittington J, Lukacs P, McKelvey K. 2018. Sampling scales
657 define occupancy and underlying occupancy – abundance relationships in animals. *Ecology*
658 99:172–183.

659 Sunde P, Kvam T, Moa P, Negård A, Overskaug K. 2000. Space use by Eurasian lynxes *Lynx*
660 *lynx* in central Norway. *Acta Theriologica* 45:507–524.

661 Thompson WL, White GC, Gowan C. 1998. Monitoring vertebrate populations. Elsevier.

662 Warton DI, Shepherd LC. 2010. Poisson point process models solve the “pseudo-absence
663 problem” for presence-only data in ecology. *The Annals of Applied Statistics* 4:1383–1402.

664 Widom B, Rowlinson JS. 1970. New Model for the Study of Liquid–Vapor Phase Transitions.
665 *The Journal of Chemical Physics* 52:1670–1684.

666 Zipkin EF, Campbell Grant EH, Fagan WF. 2012. Evaluating the predictive abilities of
667 community occupancy models using AUC while accounting for imperfect detection.
668 *Ecological Applications* 22:1962–1972.

669

9. Article 5. Un modèle mécanistico-statistique de distribution d'espèces pour expliquer et prévoir la colonisation du loup (*Canis lupus*) dans le sud-est de la France.

Résumé

Les modèles de distribution d'espèces (SDMs) sont d'importants outils statistiques qui permettent aux écologues de comprendre et prévoir l'aire de répartition des espèces. Cependant, les SDMs standard n'intègrent pas explicitement des processus dynamiques comme la dispersion. Cette limitation peut conduire à un biais dans l'inférence de la distribution de l'espèce étudiée. Nous adoptons ici la théorie de la diffusion écologique récemment introduite en écologie statistique pour intégrer les processus spatio-temporels dans les modèles écologiques. Comme étude de cas, nous avons considéré la recolonisation naturelle du loup (*Canis lupus*) qui est réapparu dans l'Est de la France via les Apennins au début des années 1990. À l'aide d'équations différentielles partielles nous modélisons la diffusion et la croissance du loup dans un paysage fragmenté. Nous développons un modèle spatio-temporel mécanistico-statistique qui tient compte de la diffusion écologique, de la croissance logistique et de la détection imparfaite de l'espèce. Nous avons mené une étude à partir de simulations et démontré la capacité de notre modèle à (i) estimer les paramètres écologiques dans diverses situations avec une probabilité de détection et un nombre de sites surveillés variables et (ii) à prédire la distribution future. Nous avons constaté que le taux de croissance de la population de loups en France s'expliquait par la proportion de couvert forestier, que la diffusion était influencée par la densité humaine et que la détectabilité augmentait avec l'effort d'échantillonnage. À l'aide des paramètres que nous avons estimés pour la période 2007-2015, nous avons prédit la distribution du loup pour l'année 2016 que nous avons validée par les détections réelles faites cette année-là. Notre approche peut être utile pour gérer les espèces qui sont en interaction avec les activités humaines afin d'anticiper les conflits potentiels.

Contributions

J'ai proposé une interprétation des résultats obtenus que nous avons discutée collectivement. J'ai rédigé un premier jet de l'article qui a été corrigé par mes co-auteurs. J'ai réalisé les simulations. Les analyses ont été réalisées ensemble avec Olivier Gimenez et Julien Papaïx. J'ai mené à terme les prédictions après avoir été mise sur la voie par Olivier Gimenez. Un squelette des simulations m'a été proposé par Olivier Gimenez, que j'ai repris et finalisé.

1 **A mechanistic-statistical species distribution model to explain and forecast wolf (*Canis***
2 ***lupus*) colonization in South-Eastern France**

3 Julie Louvrier^{1,2}, Julien Papaix³, Christophe Duchamp², Olivier Gimenez¹

4

5 ¹ CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier,
6 France

7 ² Office National de la Chasse et de la Faune Sauvage, CNERA prédateurs et animaux
8 déprédateurs, Parc Micropolis, 05000 Gap, France

9 ³ BioSP, INRA, Avignon, France

10

11 Corresponding author: Julie Louvrier, CEFE UMR 5175, CNRS, Université de Montpellier,

12 12 Université Paul-Valéry Montpellier, EPHE, Montpellier Cedex 5, France, 13

13 julie.louvrier@cefe.cnrs.fr

14 julien.papaix@inra.fr

15 christophe.duchamp@oncfs.gouv.fr

16 olivier.gimenez@cefe.cnrs.fr

17 **To be submitted to The American Naturalist**

18

19 **Abstract**

20 Species distribution models (SDMs) are important statistical tools for ecologists to understand
21 and predict species range. However, standard SDMs do not explicitly incorporate dynamic
22 processes like dispersal. This limitation may lead to bias in inference about species distribution.

23 Here, we adopt the theory of ecological diffusion that has recently been introduced in statistical
24 ecology to incorporate spatio-temporal processes in ecological models. As a case study, we
25 considered the wolf (*Canis lupus*) that has been naturally recovering Eastern France through

26 dispersal from the Apennines since the early 90's. Using partial differential equations for
27 modelling species diffusion and growth in a fragmented landscape, we developed a
28 mechanistic-statistical spatio-temporal model accounting for ecological diffusion, logistic
29 growth and imperfect species detection. We conducted a simulation study and showed the
30 ability of our model to i) estimate ecological parameters in various situations with contrasted
31 species detection probability and number of surveyed sites and ii) forecast distribution in the
32 future. We found that the growth rate of the wolf population in France was explained by the
33 proportion of forest cover, that diffusion was influenced by human density and that species
34 detectability increased with increasing survey effort. Using the parameters, we estimated on the
35 2007-2015 period, we forecasted wolf distribution in 2016 and found good agreement with the
36 actual detections made that year. Our approach may be useful for managing species that interact
37 with human activities to anticipate potential conflicts.

38 **Introduction**

39 Assessing the dynamic of species distribution is a fundamental topic in ecology
40 (Guillera-Arroita 2017). Species distribution models (SDMs) are important tools in the fields
41 of ecology, biogeography and conservation biology to understand and predict species
42 distribution by correlating occurrence data to environmental covariates (Guisan & Thuiller
43 2005). SDMs can be used to study distribution dynamics through time (Elith & Leathwick 2009;
44 Hefley & Hooten 2016; Koshkina et al. 2017), which is especially relevant in conservation for
45 the management of threatened species, conservation planning, as well as predicting the likely
46 range of invasive species (Elith & Leathwick 2009; Guillera-arroita et al. 2015).

47 Despite being the most widely used methods in ecological applications, SDMs based on
48 regressing presence locations on environmental factors suffer from several limitations (Hefley
49 & Hooten 2016; Hefley et al. 2017). SDMs rely on the hypotheses that species will be present
50 in the most favorable areas and that dispersal is not a limiting factor (Jeschke & Strayer 2006).

51 However, expanding species may be absent from an area because they have not yet dispersed
52 to this area, or because of geographical barriers or dispersal constraints (Araújo & Guisan
53 2006), not necessarily because conditions are unfavorable.

54 Species may expand through colonization defined as the ecological process of
55 populations' establishment in unoccupied areas, in which populations can often face novel
56 environments (Koontz et al. 2017). Colonization is therefore a dynamic process, underlying the
57 past, present and future distribution of species (Clark et al. 2001; Wikle 2003; Wikle & Hooten
58 2010; Williams et al. 2017). Colonization can be a natural process, or the consequence of
59 anthropogenic pressures, for example biological invasions (Sakai et al. 2001; Ricciardi 2007).
60 Being able to understand the underlying mechanisms of the colonization has significant
61 implications for wildlife managers (Koontz et al. 2017). Ignoring the dynamic process involved
62 can lead to biased inferences and some authors have discouraged the use of SDMs for
63 predictions (Yackulic et al. 2015).

64 Mechanistic spatio-temporal models have been developed to offer an alternative to
65 regression-based SDMs that encounter difficulties associated with dispersal processes (Hefley
66 et al. 2017). Mechanistic models are based on biological processes, such as survival or
67 dispersal, describing processes through which environmental factors affect the targeted
68 biological system (Morin & Thuiller 2009; Mouquet et al. 2015; Gauthier et al. 2016). SDMs
69 accounting for dynamic mechanisms are relevant tools to assess ecological colonization,
70 because they improve our ability to get predictions in space and time and include reliable
71 measures of prediction errors (Williams et al. 2017).

72 The theory of ecological diffusion is an essential component of mechanistic models to
73 assess spatial distributions dynamics and population dynamics (Soubeyrand & Roques 2014;
74 Roques & Bonnefon 2016; Hefley et al. 2017). To model dynamic ecological processes,
75 mechanistic models are usually expressed as partial differential equations (PDEs) (Wikle &

76 Hooten 2010). By combining Bayesian estimation methods with mechanistic models, these
77 PDEs can be fitted to data, so-called mechanistic-statistical models, and facilitate probabilistic
78 forecasts of spatio-temporal processes (Wikle et al. 1998).

79 Here, we aimed at exploring the use of mechanistic-statistical models to gain insight
80 into the colonization of large carnivores, with a particular emphasis on the observation process.
81 Indeed, data collection is particularly costly for elusive species that need wide areas to live
82 and/or disperse. Monitoring large carnivores often requires sampling large areas. In this context,
83 citizen sciences are increasingly used as an efficient source of information to assess the
84 dynamics of such species (Schmeller et al. 2009; Louvrier et al. 2018). The monitoring system
85 then often relies on the only available opportunistic data, leading to a set of presence locations,
86 with no information about absences (Koshkina et al. 2017). These data need to be analyzed
87 cautiously as they are collected without any measure of time- and space-varying sampling
88 effort, possibly leading to biased estimates of the actual factors influencing the distribution
89 (Van Strien et al. 2013). Furthermore, large carnivores can go undetected at sites where they
90 are actually present, due to imperfect detection (Kéry 2011). Ignoring the issue of detectability
91 < 1 can lead to underestimating the actual distribution (Kéry & Schaub 2011; Lahoz-Monfort
92 et al. 2014) and confusion in the environmental factors driving the distribution dynamics
93 (Lahoz-Monfort et al. 2014).

94 We developed a mechanistic-statistical model accounting for ecological diffusion,
95 logistic growth and imperfect detection. To assess the performance of our approach, we
96 performed a simulation study to assess bias and precision of parameter estimates and evaluate
97 forecasting performance in contrasted scenarios of varying species-level detectability and
98 number of monitoring sites. Then, we fitted our model to opportunistic data on wolves in South-
99 Eastern France between 2007 and 2015. We considered grey wolves (*Canis lupus*) as a case
100 study to illustrate the challenges of using detections/non-detections data to infer the dynamics

101 of a recolonizing large carnivore. Wolves disappeared in western European countries during
102 the twentieth century (Mech & Boitani 2010; Chapron et al. 2014) except in Spain, Portugal
103 and Italy (Ciucci et al. 2009). The species naturally recolonized the French Alps from the
104 remaining Italian population (Valière et al. 2003). From the 1990s, the species then spread
105 outside the Alpine mountains to reach the Pyrenees and the Massif Central then later, the
106 Vosges Mountains northward. In areas with livestock farming, strong interactions between wolf
107 presence and sheep breeding usually occur. Because the species is protected by law while being
108 a source of conflicts with shepherding, its recolonization process needs to be carefully
109 monitored. Besides quantifying the wolf colonization process over the study period, we
110 explored the ability of our model to forecast wolf recolonization in the short term.

111 **Material and Methods**

112 1- Model

113 To model the spatial expansion of wolves, we developed an approach to infer the parameters
114 of a mathematical formulation of the temporal dynamics of species distribution from
115 detection/non-detection data collected in the field. We adopted the framework of ecological
116 diffusion (Hefley et al. 2017). We considered a state-space modelling approach in which the
117 model is formulated in two parts: 1) the observation process that handles the detections and
118 non-detections conditional on 2) the state process describing the hidden ecological dispersion
119 process.

120

121 *Observation process*

122 Let $y_{i,j,t}$ be a random variable that takes value 1 if at least one individual is detected at site $i =$
123 $1, \dots, K$ at site i within a study area S ($i \in S \subset \mathbb{R}^2$) during secondary occasion (or survey) $j =$

124 $1, \dots, J$ in year $t = 1, \dots, T$, and takes value 0 otherwise. Let $n_{i,t}$ be the true abundance at site i in
 125 year t . The probability $p_{i,t}$ for the species to be detected at site i in year t is likely to be
 126 influenced by $n_{i,t}$. To link the detection process to abundance, we used the Royle-Nichols
 127 approach (Royle & Nichols 2003) developed to deal with heterogeneity in the detection
 128 probability due to variation in abundance and/or occupancy surveys (Williams et al. 2017). If
 129 at a site i during year t there are $n_{i,t}$ individuals present, assuming that each individual within
 130 an occupied cell has an identical detection probability $q_{i,t}$, and that there is independence of
 131 detections among individuals, then the probability to detect the species is equal to the
 132 probability to detect at least one individual. This latter probability is the complementary
 133 probability of detecting no individuals which can be written as $(1 - q_{i,t})^{n_{i,t}}$. Then the
 134 probability to detect at least one individual at site i during year t can be written as follows:

$$135 \quad p_{i,t} = 1 - (1 - q_{i,t})^{n_{i,t}}$$

136 Conditioning the observation $y_{i,j,t}$ on the latent, true abundance $n_{i,t}$ through the species-level
 137 detection probability $p_{i,t}$, and assuming a binomial observation process, a constant survey
 138 effort, and that $q_{i,t}$ and $n_{i,t}$ remain unchanged across the J surveys, we then have

$$139 \quad y_{i,t} = \sum_{j=1}^J y_{i,j,t} \sim \text{Binomial}(J, p_{i,t})$$

140 The J repeated surveys within each year t are used to estimate the species-level detection
 141 probability. Note that if $n_{i,t} = 0$ then $p_{i,t} = 0$ and $y_{i,j,t} = 0$ for all j .

142 Covariates are incorporated in the individual-level detection probability using a logistic
 143 regression-like relationship. To avoid estimating the detection probability where sampling
 144 effort was null, we set the detection probability to zero when sampling effort was equal to zero.

145

146 *State process*

147 We assumed that the true abundance $n_{i,t}$ at site i during year t was Poisson distributed over a
148 site i

$$149 \quad n_{i,t} \sim \text{Poisson}(\lambda(i,t))$$

150 where $\lambda(i,t)$ is a spatiotemporal process that describes the dynamics of the number of individuals
151 in site i during year t . We then defined this parameter as follows:

$$152 \quad \lambda(i,t) = \int_{B_i} u(x,t) dx$$

153 where $u(x,t)$ is the density of individuals at the spatial location x at time t and B_i is the study
154 area in which counts occur.

155 We used Partial Differential Equations (PDE) known as ecological diffusion (Williams
156 et al. 2017) to describe diffusion and growth dynamics. The ecological diffusion PDE
157 describing the variation of density of individuals at location x at time t , $u(x,t)$ over time, in two
158 dimensions with logistic growth, can be written as follows:

$$159 \quad \frac{\partial u(x,t)}{\partial t} = \Delta(D(x) u(x,t)) + R(x) u(x,t) \left(1 - \frac{u(x,t)}{K(x)}\right)$$

160 where Δ is the Laplace 2D diffusion operator (i.e. the sum of the second derivatives with respect
161 to the coordinates). This operator describes uncorrelated random walk movements, with the
162 coefficient $D(x)$ measuring heterogeneous mobility. The term $R(x)$ is the intrinsic growth rate
163 at low density and $K(x)$ is the carrying capacity. In addition, we assumed reflecting boundary
164 conditions, meaning that there were no population flows going outside the boundaries of the
165 study area due to actual barriers (i.e. seas) or symmetric inward and outward flows. Last, we

166 assumed that the scale at which data were collected coincides with the numerical scale in which
 167 we solve $u(x,t)$, which prevents from integrating over B_i .

168 *Approximation*

169 Calculating the density $u(x,t)$ requires solving the PDE described above. We used finite
 170 differencing in space, which consists in discretizing the spatial domain into C_s grid cells of N
 171 rows and L columns, leading to an approximation of the PDE system by a system of Ordinary
 172 Differential Equations (ODE; Hooten et al. 2013; Williams et al. 2017). The corresponding
 173 densities were then taken at each time t as

174
$$U_t^T = [u(1,t), u(2,t), \dots, u(C_s,t)]$$

175 We defined M as a $C_s \times C_s$ diffusion matrix that describes how $u(i,t)$ varies through time as a
 176 function of direct neighboring cells in four cardinal directions and diffusion parameter $D_{k,l}$. The
 177 i^{th} row of M represents the link between the C_s sites to site i . The approximation of the
 178 differential operator is then:

179
$$[MU_t]_{s_{k,l}} = \frac{1}{h^2} (D(s_{k+1,l})u(s_{k+1,l}, t) + D(s_{k-1,l})u(s_{k-1,l}, t) + D(s_{k,l+1})u(s_{k,l+1}, t)$$

 180
$$+ D(s_{k,l-1})u(s_{k,l-1}, t) - 4D(s_{k,l})u(s_{k,l}, t)$$

181 With $s_{k,l}$ the coordinates of the site i and $i = l(k - 1) + l$. With $k = 1, \dots, R$ and $l = 1, \dots, L$ and
 182 $R L = C_s$. Exceptions are the cells bordering non-habitat cells as the later are excluded of the
 183 dynamics due to the reflecting boundary conditions. This leads to the approximation of the
 184 density dynamics using the following ODE system:

185
$$\dot{U}_t = R U_t \left(1 - \frac{U_t}{K} \right) + M U_t.$$

186 The term $\lambda(i,t)$ is then calculated over the surface A_i of site i as follows:

187
$$\lambda(i,t) = u(i,t)A_i$$

188 with $u(i, t)$ the discretized approximation of $u(x, t)$ at site i . The system was solved using the
189 lsoda method (Petzold 1983) through the R-package deSolve (Soetaert et al. 2010).

190 2- Simulations

191 We conducted a simulation study to assess the ability of the model to estimate ecological
192 parameters. We defined four scenarios in which we explored the effect of a variation in the grid
193 resolution (see section *Approximation* above) and in the individual-level detectability q . To
194 mimic the characteristics of the wolf case study (see below), we set the number of surveys to 4
195 and the number of years to 20, while we set the carrying capacity to 5 individuals per 100 km²,
196 the diffusion coefficient to 5 individuals per cell and the growth rate to 80%. Because we
197 discretize the spatial domain, we expected a lower bias and a better precision of the ecological
198 parameters estimates when increasing the grid cell resolution. We defined a low resolution (LR)
199 scenario in which the spatial domain was divided into 16 cells and a high resolution (HR)
200 scenario in which we divided the spatial domain into 100 cells. Under the Royle-Nichols model
201 hypothesis, individual-level detectability can have a positive effect on the species-level
202 detectability until a certain level of abundance, hence it can influence whether the species is
203 detected or not. We then defined a high detectability (HD) scenario in which the individual-
204 level detectability was set at 0.8, and a low detectability (LD) scenario in which this probability
205 was set at 0.2. For each scenario (LR-LD, LR-HD, HR-LD, HR-HD), we simulated 100 datasets
206 and we fitted the model to each dataset. We calculated the relative bias and mean squared error
207 (MSE) for the carrying capacity K , the growth rate R and the species-level detectability q . Note
208 that in the simulation study we assumed that K , R and q were constant over space and time. To
209 explore the ability of our model to forecast the abundance of individuals per site in the four
210 scenarios, we fitted our model to the first ten years and forecasted the distribution for the next
211 ten years.

212 3- Case study: Wolf colonization in France 2007-2015

213 Wolf detections and non-detections were made of presence signs sampled all year long thanks
214 to a network of spatially distributed professional and non-professional observers (Duchamp et
215 al. 2012). Presence signs went through a standardized control process to prevent
216 misidentification.

217 To define the observation process, we used a grid to cover the study area with 10x10
218 km cells that we used as sampling units ($C_s = 975$ in the notation above). Wolf monitoring
219 occurred mainly in winter from December to March, the most favorable period to detect the
220 species. Secondary occasions were defined as December, January, February and March ($J = 4$).
221 We focused on the south-eastern part of France and the period 2007-2015 ($T = 9$) (Fig. 1). We
222 used the sampling effort, defined as the number of observers at site i in year t ($Eff_{i,t}$) and the
223 road density at site i ($RoadD_i$) to explain variation in the individual-level detection probability
224 ($q_{i,t}$) as:

$$225 \text{logit}(q_{i,t}) = \beta_0 + \beta_1 \text{Eff}_{i,t} + \beta_2 \text{RoadD}_i$$

226 We expected that the sampling effort had a positive effect and road density had a
227 negative effect on the individual-level detection probability q . We also used environmental and
228 anthropogenic covariates to model spatial variation in parameters R and D . Using the CORINE
229 Land Cover[®] database (U.E – SOeS, Corine Land Cover 2006), we calculated forest cover as
230 the percentage of mixt, coniferous or deciduous forest cover. Because forests may structure the
231 ungulate distribution (i.e. prey species), we expected that forest cover would have a positive
232 effect on the logistic growth rate R (Louvrier et al. 2018).

233 Human density was found in previous studies to influence habitat choice and dispersal
234 of wolves in Italy (Corsi et al. 1999; Marucco & McIntire 2010). We therefore considered

235 human density as a candidate covariate possibly explaining spatial variation in the diffusion
236 parameter D . Human population was averaged in each 10x10 km from a 1x1 km raster from
237 the Earth Observing System Data and Information System (EOSDIS). For both parameters, we
238 tested a linear and a quadratic effect through a logistic regression-type relationship.

239 Finally, we initialized the model with $\lambda = 0.01$ for the sites with at least one wolf
240 detection during the period 1994-2006 preceding our study period, which corresponds to one
241 individual per 100 km² cell, and zero otherwise.

242 To explore the ability of our model to forecast wolf colonization, we used the parameter
243 estimates we obtained on the 2007-2015 period and forecasted the abundance in 2016. We
244 assessed qualitatively our prediction by confronting the estimated probability of a site being
245 occupied (forecasted abundance at that site > 0) in 2016 to the actual detections made in that
246 same year.

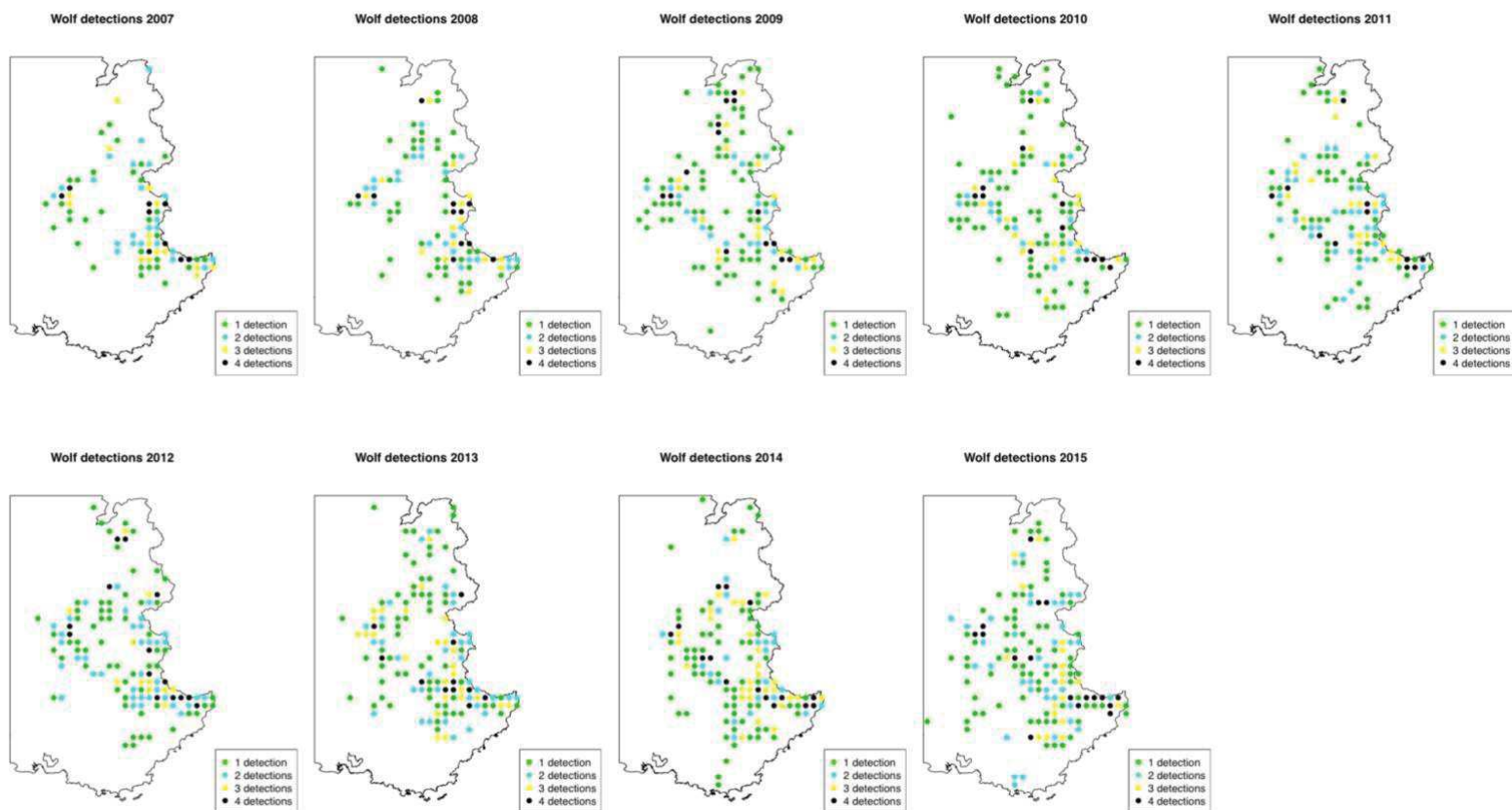


Figure 1: Maps of the yearly detections of wolf in the study area in France from 2007 to 2015.

248 4- Bayesian inference

249 To complete the Bayesian specification of our model, we specified normal priors with mean 0
250 and variance 1 for the parameters to be estimated, except for parameter K for which we used a
251 logistic distribution between 0 and 0.2. The simulations were implemented in OpenBUGS
252 (Lunn et al. 2010). The wolf analyses were implemented in JAGS using the JAGS package
253 mecastat (Rey et al. 2018). We used Markov chain Monte Carlo (MCMC) simulations for
254 parameter estimation and forecasting (Hobbs & Hooten 2015). We ran three MCMC chains
255 with 40,000 iterations preceded by 10,000 iterations as a burn-in. We used posterior medians
256 and 95% credible intervals to summarize parameter posterior distributions. We checked
257 convergence visually by inspecting the chains and by checking that the R-hat statistic was
258 below 1.1 (Gelman & Shirley 2011). We produced distribution maps of the latent states by
259 using a posteriori means of the $n_{i,t}$ from the model. We provide the scripts for running the
260 simulations and for fitting the model on the wolf data in Appendix 3.

261 5- Forecasting

262 To forecast the abundance of individuals per site, along with the associated prediction
263 uncertainty, we needed to assess the probability distribution of the true state in the future when
264 data will be collected, conditional on the collected data in the past (Williams et al. 2018). The
265 Bayesian formulation of our model allowed assessing the forecast densities by simulating
266 yearly data from $t = 2, \dots, T + 1$ and sampling $\lambda(i, T+1)$ on each iteration of the MCMC chains.
267 The posterior distribution was then assessed from the forecast distribution by sampling into the
268 forecast n_{T+1} . In the simulation study, we compared this posterior distribution with the
269 simulated data for the last year. For the wolf case study, we assessed the probability that the
270 site i was occupied, which boiled down to calculating $P(z_i = 1)$ where z_i is the latent status of

271 the site (occupied or not) as the number of MCMC iterations producing a strictly positive
272 abundance, i.e. $P(z_i = 1) = P(N_i > 0)$.

273

274 **Results**

275 1- Simulations

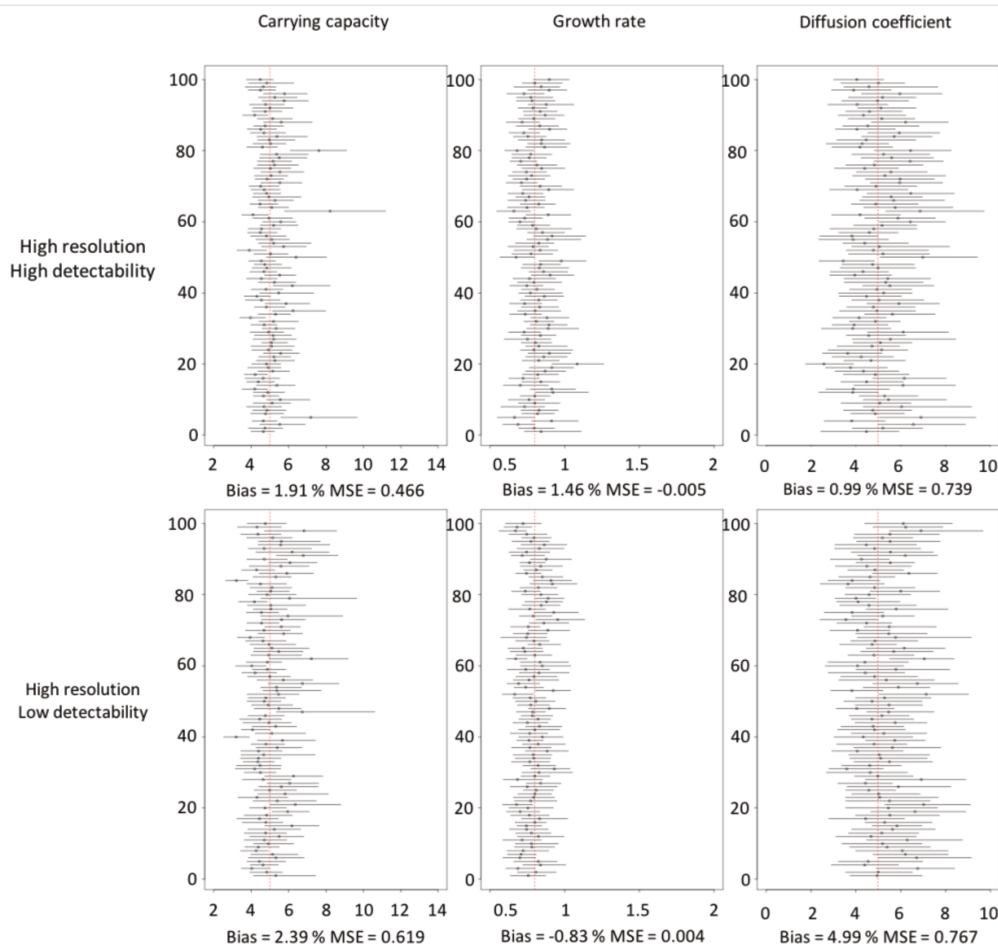
276 When the resolution in the simulation scenario increased from 16 cells to 100 cells, the model
277 produced less biased results for all parameters (Fig. 2 and 3). The largest decrease in bias was
278 found for the carrying capacity: from 8.94 % in LR-HD and 7.84 % in LR-LD to 1.91 % in HR-
279 HD and 2.39 % in HR-LD. The bias also decreased for the growth rate: 3.80 % in LR-HD and
280 4.71 % in LR-LD to 1.46 % in HR-HD and -0.83 % in HR-LD. For the diffusion coefficient,
281 the combination of high resolution and high individual-level detectability led a bias reduction
282 from 4.82 % in LR-HD, 5.02 % in LR-LD and 4.99 % in HR-LD to 0.99 % in HR-HD.

283 The model also produced more precise results for all parameters (Fig. 2 and 3). The
284 largest MSE reduction was found for the diffusion parameter. The MSE decreased for the
285 carrying capacity from 0.84 in LR-HD and 1.65 in LR-LD to 0.466 in HR-HD and 0.619 in
286 HR-LD. The MSE also decreased for the growth rate (0.02 in LR-HD and 0.017 in LR-LD to -
287 0.005 in HR-HD and 0.004 in HR-LD). For the diffusion coefficient the MSE decreased greatly
288 from 2.187 in LR-HD and 2.44 in LR-LD to 0.767 in HR-HD and 0.739 in HR-LD.

289 When the resolution was low, and the individual-level detectability increased from 0.2
290 to 0.8, leading to an increase in the species-level detectability, there was no significant change
291 in bias for the 3 parameters estimates (Fig. 3). For the carrying capacity the MSE was lower
292 when the detectability was high (from 1.65 to 0.84). For the other parameters there was no
293 significant change.

294 When the resolution was high, and the individual-level detectability increased from 0.2
295 to 0.8, there was no significant change in the parameters estimates in terms of bias and MSE
296 except for the diffusion coefficient as explained above.

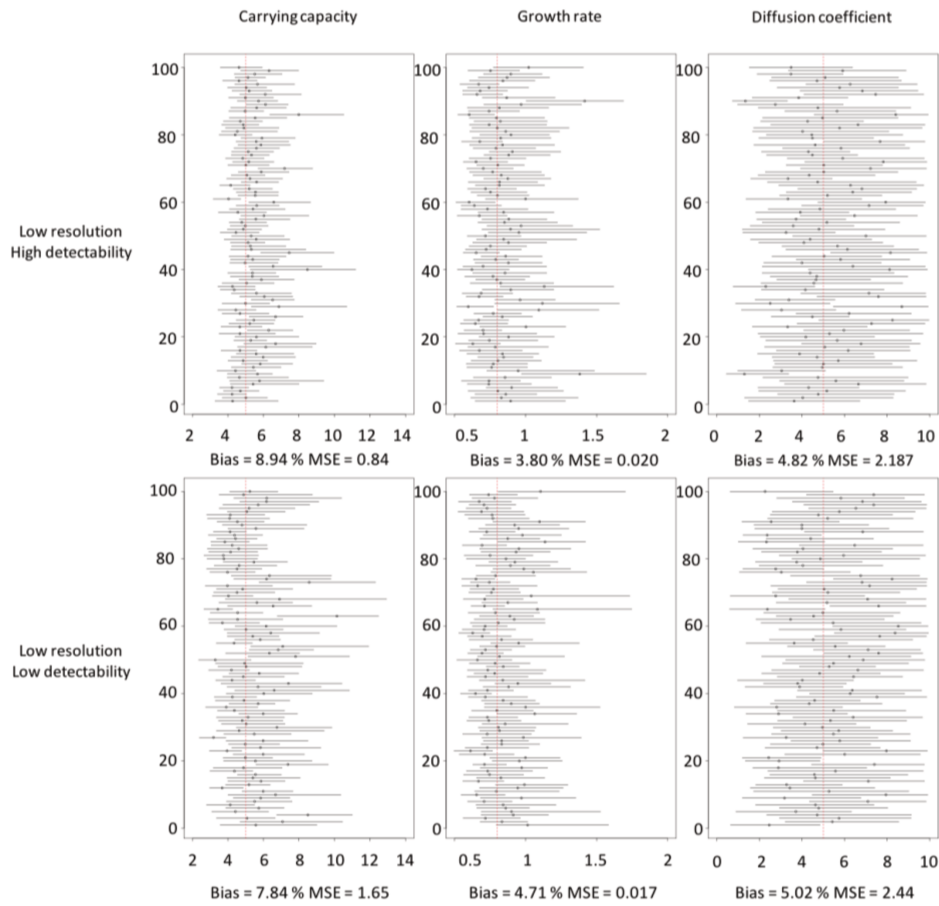
297 When looking at the model's ability to forecast abundance (Appendix 1), the true
298 abundance was always within the 95 % credible interval of the estimated abundance in all four
299 scenarios.



300

301 Figure 2: Performance of the model in the high resolution / high detectability scenario (top
302 panels) and in the high resolution / low detectability scenario (bottom panels). For each of the
303 100 simulated datasets (on the Y-axis), we displayed the median (circle) and the 95% credible
304 interval (horizontal solid line) of the parameter. The actual value of the parameter is given by

305 the vertical dashed red line. The estimated bias and MSE are provided in the legend of the X-
306 axis.



307

308 Figure 3: Performance of the model in the low resolution / high detectability scenario (top
309 panels) and in the low resolution / low detectability scenario (bottom panels). For each of the
310 100 simulated datasets (on the Y-axis), we displayed the median (circle) and the 95% credible
311 interval (horizontal solid line) of the parameter. The actual value of the parameter is given by
312 the vertical dashed red line. The estimated bias and MSE are provided in the legend of the X-
313 axis.

314

315 2- Wolf case study

316 According to our model, the detection probability increased when the sampling effort increased
317 and decreased when road density increased (Fig. 4 and Appendix 4). We found that the logistic
318 growth rate increased when the forest cover increased. The carrying capacity was estimated
319 around 1 individual per 100 km² site (9.91×10^{-3} CRI: 8.87×10^{-3} ; 1.10×10^{-2}). Last, the diffusion
320 parameter decreased as human density increased until its average of 177 inhabitants/km².
321 Beyond this threshold, when human density increased, the diffusion parameter increased and
322 reached its maximum plateau.

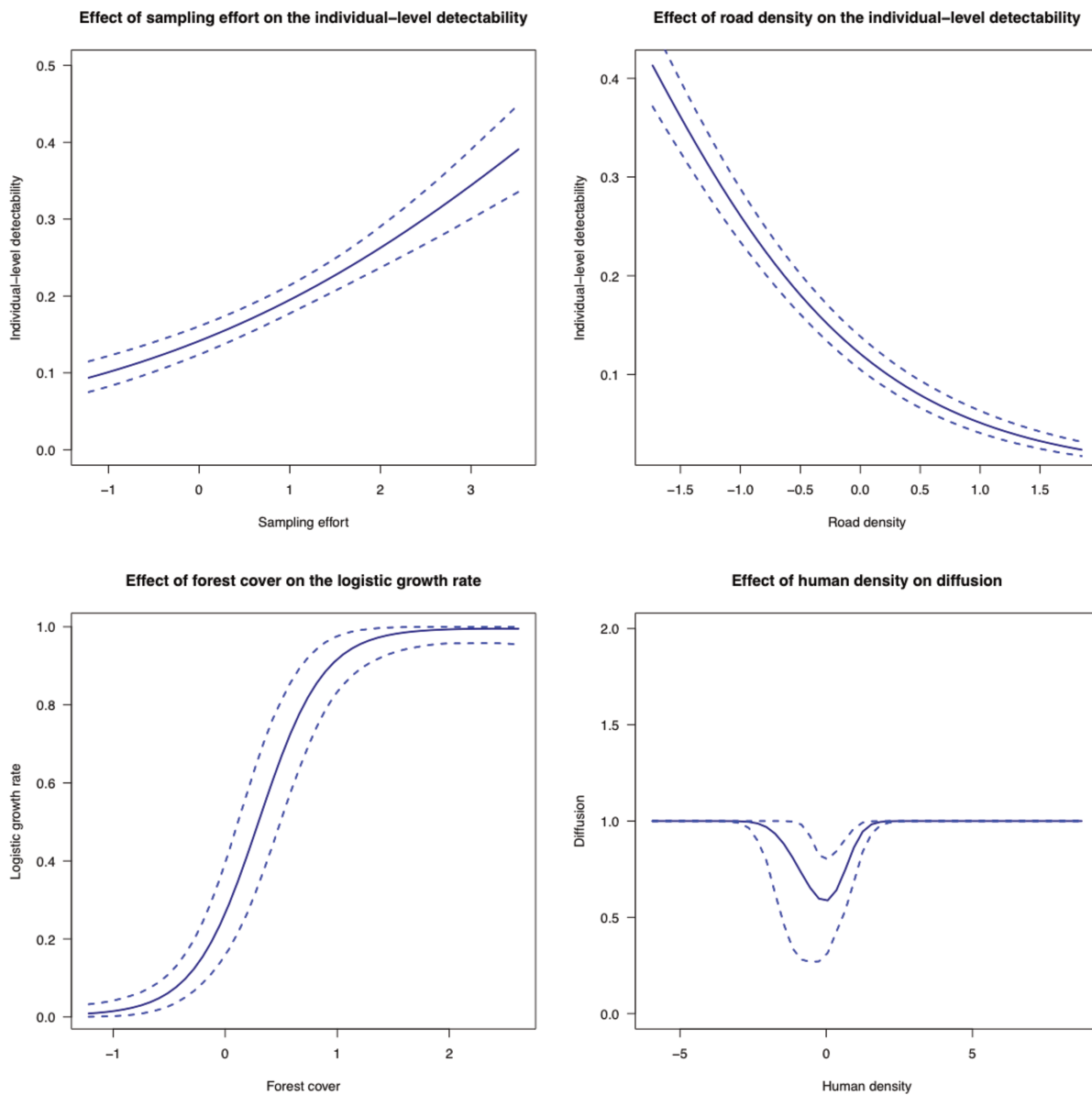


Figure 4: Estimated relationship between individual-level detectability and i) sampling effort (top left) or ii) road density (top right), between logistic growth rate and forest cover (bottom left) and between diffusion and human density (bottom right).

323 We estimated an abundance per site varying between 0 and 3 per 100 km² (Fig. 5, Appendix 2
 324 for the credible intervals and Appendix 3 for the resulting total abundance). Overall, the spatio-
 325 temporal trends in estimated abundance matched relatively well the trends in actual detections
 326 and non-detections (Fig. 5). In the northern part of the study area though, we estimated a non-
 327 null abundance at sites where no detections were made in the last four years of the study.

328 Turning to the forecasting exercise now, we predicted that a large part of sites with a
329 forecasted occupancy probability > 0.5 were indeed detected occupied in year 2016 (Fig. 6).
330 Amongst the 137 sites that were detected occupied in 2016, we found only 7 of them in the
331 South-Western part of the study area were forecasted with a low occupancy probability.
332 However, the model forecasted a higher number of sites with a high occupancy probability than
333 the number of detected occupied sites. Most of the sites with a high occupancy probability had
334 also the highest uncertainty with a standard deviation of 0.25.
335

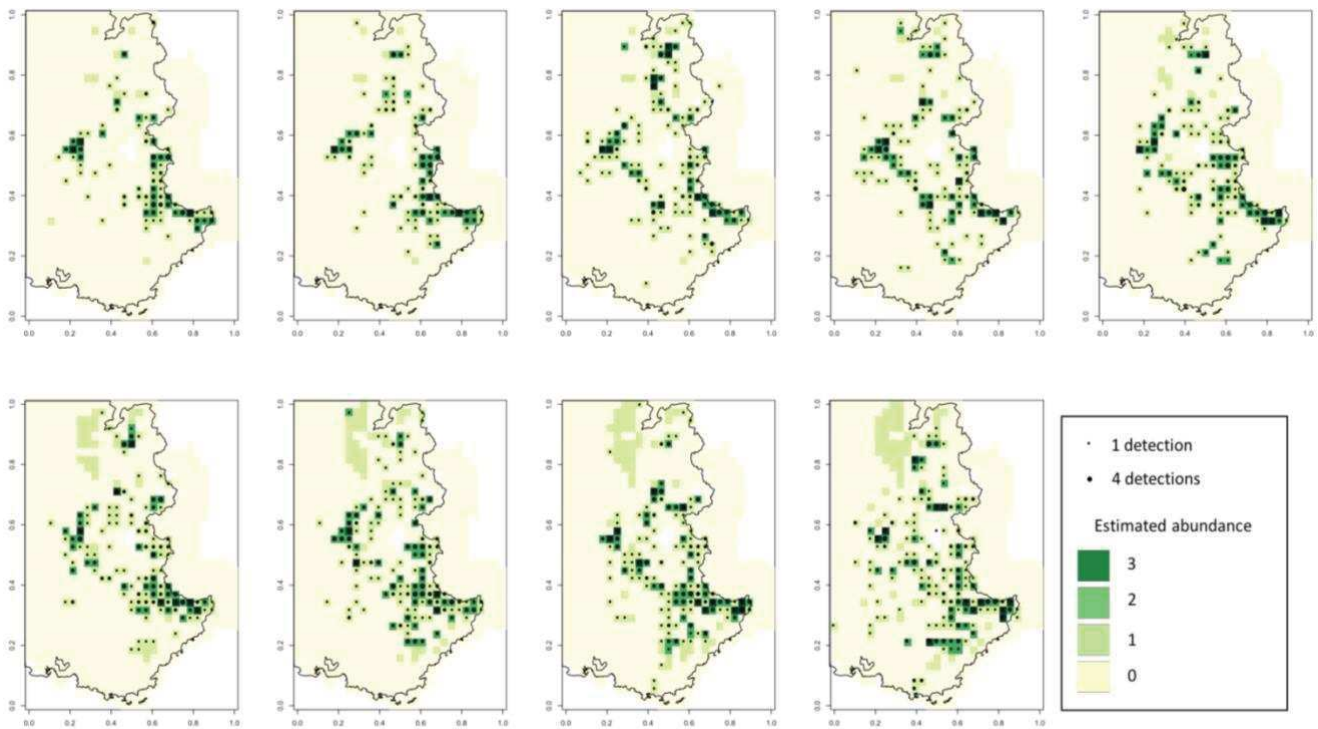
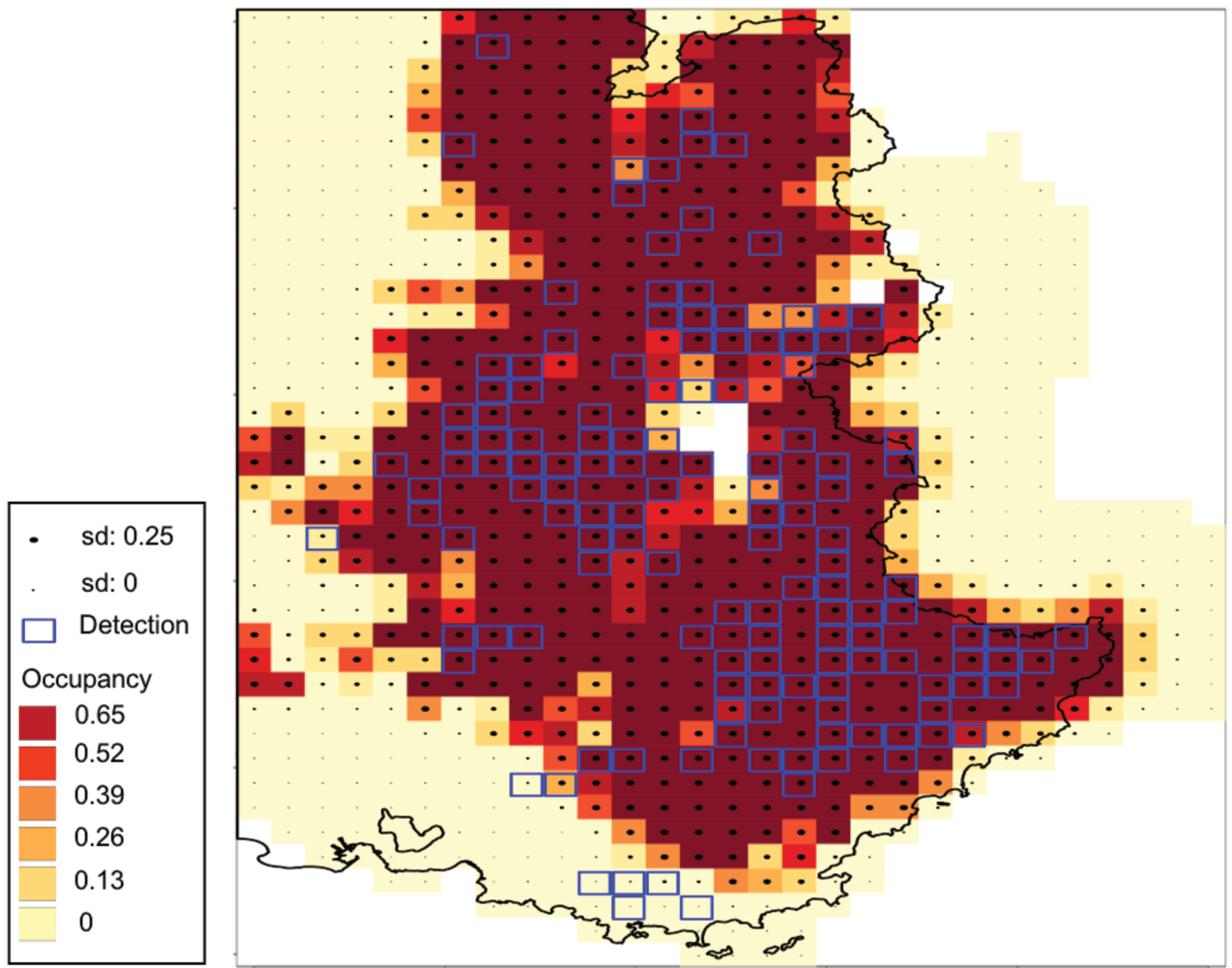


Figure 5: Maps of the estimated abundance of wolves per 100 km² site in South-East France between 2007 and 2015.

Black dots represent detections in a year.



337

338 Figure 6: Map of the forecasted probability of occupancy for the year 2016 obtained from our
 339 mechanistic-statistical model fitted to the 2007-2015 period. The blue squares represent sites
 340 where detections occurred in 2016 and the black dots capture the prediction uncertainty, with
 341 the size of a black dot proportional to the standard deviation of the forecasted occupancy in the
 342 corresponding site (varying between 0 and 0.25).

343

344 **Discussion**

345

346 We estimated the distribution of wolves using an SDM explicitly incorporating biological
 347 mechanisms and making optimal use of the information contained in species detections and

348 non-detections. Besides, we explored the possibility of forecasting the potential future
349 distribution of a large carnivore, which could be used to target management areas or focus on
350 potential conflictual areas (Marucco & Mcintire 2010; Eriksson & Dalerum 2018).

351

352 *Simulations*

353

354 In the simulation study, we showed that ecological parameters were sensitive to the way we
355 discretized space to solve the PDE. Our model performed well when the resolution was high,
356 with less biased (under 5%) and more precise parameter estimates than in the low-resolution
357 scenario. We note however that the low-resolution scenario was an unrealistic example used to
358 test the model in extreme conditions. Furthermore, our model was sensitive to the species-level
359 detectability but only when the resolution was high. We explain this pattern by the fact that
360 resolution had a stronger effect on the parameter estimates than the detectability and the effect
361 of low detectability was hidden by the effect of the low-resolution grid.

362

363 *Wolf study*

364

365 We found that the logistic growth rate increased when forest cover increased. Although wolves
366 can adapt to various ecosystems, this pattern also matches with the suitable habitats of the key
367 prey species for wolves (Darmon et al. 2012). We found that when human density increased,
368 until the average value of human density, the diffusion coefficient decreased, leading to less
369 individuals diffusing in a neighboring site with a higher human density. However, above this
370 average, the diffusion coefficient increased again. These contrasted patterns might be the result
371 of two underlying mechanisms. First, the fact the diffusion coefficient decreased is in agreement
372 with previous studies demonstrating the influence of human density on wolf presence (Corsi et
373 al. 1999; Marucco & Mcintire 2010). Second, beyond the average value of human density, the

374 diffusion coefficient might increase due to the increase of linear features, which have been
375 found to be selected over natural linear features for wolves' movements (Newton et al. 2017).

376 As expected, we found that when sampling effort increased, the individual-level
377 detectability increased, while it decreased when road density increased. We also expected that
378 road density would influence wolf detectability by facilitating observers survey and site
379 accessibility. Other studies have found that linear features also facilitate wolf travel and prey
380 encounter rate (Dickie et al. 2017; Newton et al. 2017). On the contrary, we found that the
381 increase in road density negatively affected the species detection. This result was found in
382 previous studies as well (Falcucci et al. 2013; Votsi et al. 2016; Louvrier et al. 2018)
383 corroborating the fact that wolves avoid roads and leave fewer marks in sites with highly
384 frequented roads or pathways (Whittington et al. 2005).

385 We estimated carrying capacity at 1 individual per 10x10 km site. Wolves pack size was
386 documented on average at 3.8 individuals per pack in France (Duchamp et al, 2012) varying
387 from 2 to a dozen individuals. Considering the average wolf territory size commonly reported
388 between 100 and 400 km² in western and central Europe (Ciucci et al. 2009; Mech & Boitani
389 2010; Duchamp et al. 2012), our estimate covers the standard range of wolf densities reported
390 elsewhere (Mech & Boitani for a review).

391 392 *Model Assumptions*

393
394 We built our model based on several assumptions that need to be discussed. We assumed that
395 the sampling effort was constant across surveys and that the individual-level detectability and
396 the local abundance remained unchanged. First, it is likely that the sampling effort varies
397 between surveys (months) due to human factors. However, we could only quantify the sampling
398 effort between years, and had no information at the month level. Second, it is also likely that
399 the individual-level detectability varies between months partly due to the varying sampling

400 effort between months, but also to environmental conditions, such a snow cover represented by
401 the month of survey (Louvrier et al. 2018). Third, the local abundance at a site is also likely to
402 change between surveys. The choice to consider the wintering data survey, during which the
403 social organization of packs is the most stable (Mech & Boitani 2010), may prevent a large part
404 of this sampling heterogeneity according the sampling protocol implemented in the Alps by the
405 wolf network (Duchamp et al, 2012). However, we cannot exclude that mortality or movements
406 occur inside or outside the sites. In this case, the estimates for local abundance can be
407 overestimated as the same individuals can be detected in two neighboring sites for instance.,
408 The distribution should in any case be interpreted cautiously and considered as area of use
409 (Mackenzie 2006).

410 Under the Royle-Nichols model, the species-level detectability is function of the
411 individual-level detectability, but the relationship between these two parameters is not always
412 linear and depends on the abundance of individuals at a site. If abundance is large (i.e., above
413 50 individuals), then individuals can be detected during all surveys, and no variability in the
414 species-level detectability will be observed, which leads to the inability to characterize the
415 distribution of abundances (Royle & Nichols 2003). Finally, our approach was based on a
416 logistic growth, but other forms of growth should be investigated, especially growth accounting
417 for an Allee effect, of particular relevance for wolves for which the probability of finding a
418 mate decreases in areas with low density ((Hurford et al. 2006).

419 Another assumption relies on the model construction considering the diffusion equally
420 for all individuals. Wolves have a strong social organization in packs and future works may
421 consider the social aggregation of individuals when modeling the dynamic of the wolf
422 distribution (see for instance Lewis et al. 1997 and Potts & Lewis 2014).

423 We need to highlight here the fact that our model was realistic because we fitted it on
424 data from the core distribution of wolves in France. However, if we had extended our model to

425 the whole country, we would expect less realistic estimates due to the fact that wolves not only
426 disperse at short distance but also at long distance, especially on colonization fronts (Mech and
427 Boitani 2010). In Louvrier et al. (2018), we found that the number of observed occupied sites
428 at long distance also influenced the probability for a site to be occupied. Our model was
429 deterministic but if we were to extend our model to the whole country, we would need to
430 account for stochasticity in events when the population is a low density (Hurford et al. 2006).

431

432

433 *Comparison with dynamic site-occupancy models*

434

435 In Louvrier et al. (2018), a dynamic site-occupancy model was fitted to the same dataset, at a
436 national level and between 1994-2016. We found in this previous study that when forest cover
437 increased, the probability for a site to be colonized the year after increased as well. This can be
438 related to the logistic growth rate parameter, because once a site is colonized, population will
439 start growing. We found the same effects of sampling effort and road density on the species-
440 level detectability, which can be explained by the fact that maximum abundance per site is low
441 enough to guarantee a linear correspondence between species- detectability and individual-level
442 detectability. In comparison with the map of occupancy estimated with a dynamic site
443 occupancy model (top right panel of Figure 7 in Louvrier et al. 2018), we found that the
444 mechanistic approach predicted more sites with an average occupancy probability of 0.6 than
445 the dynamic site-occupancy model. The latter approach estimated a smaller number of occupied
446 sites but with a higher occupancy probability, up to 1. This difference could be explained by
447 the fact that occupancy models are regression-type models, which means that the estimated
448 occupancy is linked to the data, while our mechanistic approach is based on a continuous model
449 over time model which allows the spreading of individuals over several sites without having to
450 be detected. Another explanation could be that we assumed a Poisson distribution for the
451 number of individuals per site in our mechanistic model. However, there could be an issue with

452 adjusting our model to the data, for example, due to over-dispersion in relation to the
453 dependence between individuals due to their social structure. A first way to overcome this
454 problem is to use a negative binomial distribution to relax the constraint of equal mean and
455 variance inherent to the Poisson distribution (White & Bennetts 1996). Another approach would
456 be to directly model the dependence between individuals by explaining the pack structure in
457 the mechanistic part of our model (Lewis et al. 1997).

458

459 *Forecasting capacities*

460 In the current context of fast-changing environments, predicting the future distribution or
461 abundance of species is an increasing challenge in the field of ecology, where ecologists are
462 calling for a more “predictive ecology” (Mouquet et al. 2015; Houlahan et al. 2017; Maris et
463 al. 2018). Ecological forecasting is the process of predicting the state of an ecological system
464 with fully specified uncertainties (Clark et al. 2001). Forecasts should therefore be probabilistic
465 (Gneiting & Katzfuss 2014) to provide reliable uncertainties. Not accounting for uncertainties
466 associated with predictions of future change in distributions can lead to misguided decisions by
467 policymakers or managers (Gauthier et al. 2016). The Bayesian method provides a natural
468 framework for making probabilistic forecasts (Hefley et al. 2017). We demonstrated using
469 simulations that our model had satisfying forecasting capabilities. When we applied our
470 approach to the wolf, we produced relatively correct forecasts for the presence of wolves. In
471 2016, 137 sites were detected as being occupied, out of which 7 sites were not forecasted as
472 occupied by our model. These sites were found at the edge of the distribution core in the South-
473 Western part of the study site. This part of the distribution was recently colonized by wolves
474 with first detections of wolves occurred there in 2014 and 2015 for the first time. Wolves are
475 highly flexible and can live in various areas from maize cultures to high mountains (Kaczensky
476 et al. 2012). This South-Western part are places where forest cover is lower and human density

477 is higher than in the Alpine range. In the future we might expect the effects of forest cover to
478 be weaker as wolves expand in different landscapes. Overall, we forecasted more sites with a
479 relatively high probability of being occupied than the number of sites detected occupied in
480 2016, which can be explained by the fact that our model accounts for imperfect wolf detection.

481

482 **Conclusion**

483 Mechanistic-statistical models are valuable tools to bring insight into the dynamic of species
484 distribution. However, ecologists are often faced with cryptic species with detectability less
485 than one. Here we developed a mechanistic-statistical model accounting for imperfect detection
486 for wolf management in France. The model is flexible and can be used in a variety of contexts
487 to assess the dynamic of species distribution by amending the observation process if required.
488 By forecasting the distribution of wolves in France, we illustrate that our approach may provide
489 a new tool in the context of the management of a species with possible conflictual interactions
490 with human activities. Our approach resonates with the adaptive management framework where
491 ecologists need to make decisions based on yearly estimates of population abundance and
492 distribution (Marescot et al. 2013).

493

494

495 **References**

496

497 Araújo MB, Guisan A. 2006. Five (or so) challenges for species distribution modelling.
498 *Journal of Biogeography* 33:1677–1688.

499 Chapron G et al. 2014. Recovery of large carnivores in Europe's modern human-dominated
500 landscapes. *Science* 346:1517–1519.

501 Ciucci P, Reggioni W, Maiorano L, Boitani L. 2009. Long-Distance Dispersal of a Rescued
502 Wolf From the Northern Apennines to the Western Alps. *Journal of Wildlife Management*
503 73:1300–1306.

504 Clark JS et al. 2001. Ecological Forecasts: An Emerging Imperative. *Science, New Series*
505 293:657–660.

506 Corsi F, Dupre E, Boitani L. 1999. A Large-Scale Model of Wolf Distribution in Italy for
507 Conservation Planning. *Conservation Biology* 13:150–159.

508 Darmon G, Calenge C, Loison A, Jullien J-M, Maillard D, Lopez J-F. 2012. Spatial
509 distribution and habitat selection in coexisting species of mountain ungulates. *Ecography*
510 35:44–53.

511 Dickie M, Serrouya R, McNay RS, Boutin S. 2017. Faster and farther: wolf movement on
512 linear features and implications for hunting behaviour. *Journal of Applied Ecology* 54:253–
513 263.

514 Duchamp C et al. 2012. A dual frame survey to assess time- and space-related changes of the
515 colonizing wolf population in France. *Hystrix* 23:14–28.

516 Elith J, Leathwick JR. 2009. Species Distribution Models : Ecological Explanation and
517 Prediction Across Space and Time. *Annual review of ecology, evolution, and systematics*
518 40:677–697.

519 Eriksson T, Dalerum F. 2018. Identifying potential areas for an expanding wolf population in
520 Sweden. *Biological Conservation* 220:170–181. Elsevier.

521 Falcucci A, Maiorano L, Tempio G, Boitani L, Ciucci P. 2013. Modeling the potential
522 distribution for a range-expanding species: Wolf recolonization of the Alpine range.
523 *Biological Conservation* 158:63–72.

524 Gauthier G, Péron G, Lebreton J-D, Grenier P, van Oudenhove L. 2016. Partitioning
525 prediction uncertainty in climate-dependent population models. *Proceedings of the Royal*
526 *Society B: Biological Sciences* 283:20162353.

527 Gelman A, Shirley K. 2011. Inference from simulations and monitoring convergence.
528 *Handbook of Markov Chain Monte Carlo* 6:163–174.

529 Gneiting T, Katzfuss M. 2014. Probabilistic forecasting. *Annual Review of Statistics and Its*
530 *Application* 1:125–151.

531 Guillera-Arroita G. 2017. Modelling of species distributions, range dynamics and
532 communities under imperfect detection: advances, challenges and opportunities. *Ecography*
533 40:281–295.

534 Guillera-arroita G, Lahoz-monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, Mccarthy MA,
535 Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose ? Matching data
536 and models to applications. *Global Ecology and Biogeography* 24:276–292.

537 Guisan A, Thuiller W. 2005. Predicting species distribution : offering more than simple
538 habitat models. *Ecology Letters* 8:993–1009.

539 Hefley TJ, Hooten MB. 2016. Hierarchical Species Distribution Models. *Current Landscape*
540 *Ecology Reports* 1:87–97.

541 Hefley TJ, Hooten MB, Russell RE, Walsh DP, Powell JA. 2017. When mechanism matters:
542 Bayesian forecasting using models of ecological diffusion. *Ecology Letters* 20:640–650.

543 Hobbs NT, Hooten MB. 2015. Bayesian models: a statistical primer for ecologists. Princeton
544 University Press.

545 Hooten MB, Garlick MJ, Powell JA. 2013. Computationally Efficient Statistical Differential
546 Equation Modeling Using Homogenization. *Journal of Agricultural, Biological, and*
547 *Environmental Statistics* 18:405–428.

548 Houlahan JE, McKinney ST, Anderson TM, McGill BJ. 2017. The priority of prediction in

549 ecological understanding. *Oikos* 126:1–7.

550 Hurford A, Hebblewhite M, Lewis MA. 2006. A spatially explicit model for an Allee effect :
551 Why wolves recolonize so slowly in Greater Yellowstone. *Theoretical population biology*
552 70:244–254.

553 Jeschke JM, Strayer DL. 2006. Determinants of vertebrate invasion success in Europe and
554 North America. *Global Change Biology* 12:1608–1619.

555 Whittington J, Cassady St. Clair C, Mercer G. 2005. Spatial Responses of Wolves to Roads
556 and Trails in Mountain Valleys. *Ecological Applications* 15:543–553.

557 Kaczensky P, Chapron G, von Arx M, Hubert D, Andren H, Linnell JDC. 2012. Status,
558 management and distribution of large carnivores–bear, lynx, wolf & wolverine–in Europe.
559 Report. European Commission, Brussels.

560 Kéry M. 2011. Towards the modelling of true species distributions. *Journal of Biogeography*
561 38:617–618.

562 Kéry M, Schaub M. 2011. Bayesian population analysis using WinBUGS: a hierarchical
563 perspective. Academic Press.

564 Koontz MJ, Oldfather MF, Melbourne BA, Hufbauer RA. 2017. Parsing propagule pressure:
565 Number, not size, of introductions drives colonization success in a novel environment.
566 bioRxiv 108324.

567 Koshkina V, Wang Y, Gordon A, Dorazio RM, White M, Stone L. 2017. Integrated species
568 distribution models : combining presence-background data and site-occupancy data with
569 imperfect detection. *Methods in Ecology and Evolution* 8:420–430.

570 Lahoz-Monfort JJ, Guillera-arroita G, Wintle BA. 2014. Imperfect detection impacts the
571 performance of species distribution models. *Global Ecology and Biogeography* 23:504–515.

572 Lewis MA, White KAJ, Murray JD. 1997. Analysis of a model for wolf territories. *Journal of*
573 *Mathematical Biology* 35:749–774.

574 Louvrier J, Duchamp C, Lauret V, Marboutin E, Cubaynes S, Choquet R, Miquel C, Gimenez
575 O. 2018. Mapping and explaining wolf recolonization in France using dynamic occupancy
576 models and opportunistic data. *Ecography* 41:647–660.

577 Lunn D, Miller S, Unit MRCB, Way R. 2010. OpenBUGS Differential Equation Solver
578 Advanced applications:0–29.

579 Mackenzie DI. 2006. Modeling the probability of resource use: The effect of, and dealing
580 with, detecting a species imperfectly. *Journal of Wildlife Management* 70:367–374.

581 Marescot L, Chapron G, Chadès I, Fackler PL, Duchamp C, Marboutin E, Gimenez O. 2013.
582 Complex decisions made simple: a primer on stochastic dynamic programming. *Methods in*
583 *Ecology and Evolution* 4:872–884.

584 Maris V, Huneman P, Coreau A, Kéfi S, Pradel R, Devictor V. 2018. Prediction in ecology:
585 promises, obstacles and clarifications. *Oikos* 127:171–183.

586 Marucco F, Mcintire EJB. 2010. Predicting spatio-temporal recolonization of large carnivore
587 populations and livestock depredation risk : wolves in the Italian Alps. *Journal of Applied*
588 *Ecology* 47:789–798.

589 Mech LD, Boitani L. 2010. Wolves: behavior, ecology, and conservation. University of
590 Chicago Press.

591 Mouquet N et al. 2015. Predictive ecology in a changing world. *Journal of Applied Ecology*
592 52:1293–1310.

593 Newton EJ, Patterson BR, Anderson ML, Rodgers AR, Vander Vennen LM, Fryxell JM.
594 2017. Compensatory selection for roads over natural linear features by wolves in northern
595 Ontario: Implications for caribou conservation. *PLOS ONE* 12:e0186525.

596 Petzold L. 1983. Automatic selection of methods for solving stiff and nonstiff systems of
597 ordinary differential equations. *SIAM journal on scientific and statistical computing* 4:136–
598 148.

599 Potts JR, Lewis MA. 2014. How do animal territories form and change? Lessons from 20
600 years of mechanistic modelling. *Proceedings of the Royal Society B: Biological Sciences*
601 281:20140231.

602 Ricciardi A. 2007. Are Modern Biological Invasions an Unprecedented Form of Global
603 Change? *Conservation Biology* 21:329–336.

604 Roques L, Bonnefon O. 2016. Modelling Population Dynamics in Realistic Landscapes with
605 Linear Elements: A Mechanistic-Statistical Reaction-Diffusion Approach. *PLOS ONE*
606 11:e0151217.

607 Royle JA, Nichols JD. 2003. Estimating abundance from repeated presence-absence data or
608 point counts. *Ecology* 84:777–790.

609 Sakai AK et al. 2001. The Population Biology of Invasive Species. *Annual Review of*
610 *Ecology and Systematics* 32:305–332.

611 Schmeller DS et al. 2009. Ventajas del monitoreo de biodiversidad basado en voluntarios en
612 Europa. *Conservation Biology* 23:307–316.

613 Soetaert K, Petzoldt T, Setzer RW. 2010. Solving Differential Equations in R : Package
614 deSolve. *Journal of Statistical Software* 33.

615 Soubeyrand S, Roques L. 2014. Parameter estimation for reaction-diffusion models of
616 biological invasions. *Population Ecology* 56:427–434.

617 Valière N, Fumagalli L, Gielly L, Miquel C, Lequette B, Poulle M-L, Weber J-M, Arlettaz R,
618 Taberlet P. 2003. Long-distance wolf recolonization of France and Switzerland inferred from
619 non-invasive genetic sampling over a period of 10 years. *Animal Conservation* 6:83–92.

620 Van Strien AJ, Van Swaay CAM, Termaat T. 2013. Opportunistic citizen science data of
621 animal species produce reliable estimates of distribution trends if analysed with occupancy
622 models. *Journal of Applied Ecology* 50:1450–1458.

623 Votsi N-EP, Zomeni MS, Pantis JD. 2016. Evaluating the Effectiveness of Natura 2000
624 Network for Wolf Conservation: A Case-Study in Greece. *Environmental Management*
625 57:257–270.

626 White GC, Bennetts RE. 1996. Analysis of Frequency Count Data Using the Negative
627 Binomial Distribution. *Ecology* 77:2549–2557.

628 Wikle C, Berliner LM, Cressie N. 1998. Hierarchical Bayesian space-time models.
629 *Environmental and Ecological Statistics* 5:117–154.

630 Wikle CK. 2003. Hierarchical bayesian models for predicting the spread of ecological
631 processes. *Ecology* 84:1382–1394.

632 Wikle CK, Hooten MB. 2010. A general science-based framework for dynamical spatio-
633 temporal models. *Test* 19:417–451.

- 634 Williams PJ, Hooten MB, Womble JN, Esslinger GG, Bower MR. 2018. Monitoring dynamic
635 spatio-temporal ecological processes optimally. *Ecology* 99:524–535.
- 636 Williams PJ, Hooten MB, Womble JN, Esslinger GG, Bower MR, Hefley TJ. 2017. An
637 integrated data model to estimate spatiotemporal occupancy, abundance, and colonization
638 dynamics. *Ecology* 98:328–336.
- 639 Morin X, Thuiller W. 2009. Comparing Niche- and Process-Based Models to Reduce
640 Prediction Uncertainty in Species Range Shifts under Climate Change. *Ecology* 90:1301–
641 1313.
- 642 Yackulic CB, Nichols JD, Reid J, Der R. 2015. To predict the niche, model colonization and
643 extinction. *Ecology* 96:16–23.
- 644
- 645

10. Discussion

10.a. Synthèse de mon travail et intérêt

L'évaluation de la répartition spatiale des grands carnivores est une information centrale pour évaluer leur état de conservation (IUCN 2012; Jedrzejewski et al. 2018), identifier les zones de conflit potentiel (Marucco & McIntire 2010) et comprendre les mécanismes sous-jacents à la dynamique de leur distribution pour aider à une meilleure gestion (Eriksson & Dalerum 2018). Dans le cadre de ma thèse, j'ai contribué à identifier les déterminants des habitats favorables ainsi qu'à établir la dynamique de distribution de grands carnivores en France, dans les pays alpins ainsi qu'en Norvège, et ce grâce aux modèles d'occupancy. Nous avons démontré qu'il était possible d'utiliser ces modèles avec des données en grande partie issues de suivis sans protocoles stricts, lorsque l'effort d'échantillonnage était hétérogène, avec des données incertaines contenant des faux-positifs, ou des données issues de différents suivis.

Dans le premier article, nous avons utilisé des modèles dynamiques d'occupancy dans lesquels nous avons intégré une mesure de l'effort d'échantillonnage mesurée à posteriori. Nous avons démontré que sans cette mesure de l'échantillonnage, la distribution du loup était sous-estimée. Mes résultats ont apporté de nouvelles connaissances sur les processus régissant la dynamique de recolonisation du loup en France tout en prenant en compte la détection imparfaite de l'espèce et un effort d'échantillonnage hétérogène dans le temps et l'espace. L'utilisation ici de données opportunistes est particulièrement intéressante pour optimiser les plans de gestions pour le suivi de la dynamique spatiale de ces espèces cryptiques et aux enjeux sociétaux importants.

Dans le deuxième article, j'ai démontré qu'il était possible d'utiliser les données incertaines traditionnellement écartées des analyses, en considérant un modèle d'occupancy dynamique prenant en compte les faux-positifs dans les données incertaines. A l'aide de simulations, j'ai illustré le fait qu'inclure les données incertaines pouvait améliorer les estimations de paramètres écologiques en produisant des estimations moins biaisées et plus précises qu'un modèle ajusté sur seulement les données certaines. De plus, en ajoutant les données incertaines issues du suivi du lynx à ce modèle, nous avons trouvé que les données incertaines pouvaient apporter de nouvelles informations sur la distribution du lynx dans les pays alpins, dans des endroits où il est difficile d'obtenir des données certaines.

Dans le troisième article, nous avons développé des modèles d'occupancy qui prennent en compte à la fois l'hétérogénéité dans la probabilité de détection et les potentiels faux positifs issus des données incertaines. À l'aide de simulations, nous avons montré que lorsqu'elle est ignorée, l'hétérogénéité spatiale dans la détection induit des biais dans l'estimation des paramètres écologiques. En appliquant cette approche aux données issues du suivi du loup, nous avons également constaté que ce modèle pouvait être utilisé pour capturer une partie de l'hétérogénéité attribuable à un effort d'échantillonnage variable dans l'espace, un effort qui peut s'avérer difficile voire impossible à mesurer sur le terrain dans un cadre de suivi à grande échelle.

Dans le quatrième article, nous avons fait la démonstration de l'utilisation d'un modèle de distribution d'espèce utilisant simultanément les données de « presence-background »

Discussion

issues de la chasse du lynx en Norvège et les données de détections issues du suivi des groupes familiaux du lynx. Nous avons appliqué ce nouveau modèle au lynx et avons mis en évidence des facteurs environnementaux qui influencent la répartition du lynx sur plusieurs années. Le modèle est formulé comme un processus ponctuel et offre ainsi la souplesse méthodologique nécessaire pour vérifier les résidus spatiaux et utiliser la technique LASSO pour sélectionner automatiquement les facteurs supposés avoir une influence sur les paramètres écologiques ainsi que sur les paramètres liés à l'observation de l'espèce.

Dans le cinquième article, nous avons estimé la répartition des loups en France à l'aide d'un modèle mécanistico-statistique intégrant explicitement les mécanismes biologiques de la diffusion écologique et faisant un usage optimal de l'information contenue dans les détections et les non-détections du suivi du loup en France. En outre, nous avons montré le potentiel de notre approche pour prédire la distribution future potentielle du loup à court terme, un élément qui pourrait contribuer à la gestion pour cibler des zones de gestion ou se concentrer sur des zones de conflit potentiel.

10.b. Interprétation et utilisation des résultats

MacKenzie et al. (2002) définissent l'occupancy comme la proportion de sites occupés par l'espèce étudiée. L'occupancy est par extension la probabilité que l'espèce étudiée occupe un site pendant une période donnée au cours de laquelle l'état du site – occupé ou pas – reste inchangé (Bailey et al. 2014). Le protocole d'échantillonnage classique utilisé pour établir l'occupancy d'une espèce se déroule en plusieurs étapes. Tout d'abord, on identifie l'aire d'étude sur laquelle l'occupancy sera établie, puis on sélectionne des sites qui seront échantillonnés, c'est-à-dire visités plusieurs fois au cours d'une période de temps choisie de telle sorte que le statut du site ne changera pas entre les visites (MacKenzie & Royle 2005). Au cours des visites, l'état observé est donc noté, à savoir si l'espèce y a été détectée ou non. Beaucoup d'études ont été réalisées afin d'établir un protocole d'étude adéquat aux modèles d'occupancy (MacKenzie & Royle 2005; MacKenzie et al. 2006; Bailey et al. 2007; Guillera-Arroita & Lahoz-Monfort 2012). Dans le cadre de ma thèse, les données opportunistes ont été récoltées sans protocole formel d'échantillonnage des sites spatialisés., une discussion est donc nécessaire sur l'interprétation des résultats obtenus sur modèles d'occupancy utilisés. En particulier, l'occupancy que nous avons obtenue correspond-elle à l'occupancy que MacKenzie et al. (2002) ont défini ? Outre les aspects méthodologiques, je discuterai dans un deuxième temps de la signification écologique de nos résultats. Enfin, dans un troisième temps, je discuterai les implications de nos résultats en termes de gestion des espèces et de leur suivi.

10.b.1. Interprétation de l'occupancy

L'un des critères utilisés pour évaluer l'état de conservation d'une espèce sur la liste rouge de l'UICN est l'aire occupée par une espèce (AOO pour « Area Of Occupancy ») définie comme la zone « occupée par un taxon, à l'exclusion des cas de vagabondage » (IUCN 2012). L'aire occupée représente donc la superficie de l'habitat adaptée à l'espèce réellement occupée par le taxon. L'UICN suggère plusieurs méthodes pour estimer l'AOO, et en pratique la méthode la plus courante consiste à « compter le nombre de sites occupés dans une grille uniforme couvrant l'ensemble de l'aire de répartition d'un taxon » (IUCN 2012). Quel lien peut-on faire

Discussion

avec l'AOO et l'occupancy telle que nous l'avons considérée dans cette thèse ? Avant de répondre à cette question, il convient de revenir sur l'interprétation de l'occupancy.

Une hypothèse importante sur laquelle reposent les modèles d'occupancy est que le statut d'un site visité ne change pas entre les visites (MacKenzie et al. 2002). Cela signifie que la taille du site et les occasions secondaires doivent être choisies de telle sorte à respecter cette hypothèse. MacKenzie & Nichols (2004) ont suggéré que cette hypothèse pouvait être assouplie à condition que les changements d'état des sites se produisent de manière aléatoire, c'est-à-dire que la probabilité d'occupancy dans un intervalle de temps ne dépende pas de l'état d'occupation d'un site dans l'intervalle précédent. Sous cette condition, les modèles d'occupancy peuvent toujours être appliqués, et l'occupancy s'interprète alors comme l'utilisation de l'espace. L'assouplissement de cette hypothèse modifie donc l'interprétation du paramètre d'occupancy. La proportion de l'espace utilisé par une espèce sera souvent plus grande que la proportion de la superficie où l'espèce est physiquement présente (MacKenzie & Royle 2005). Par conséquent, si l'occupancy à un moment donné est l'indicateur souhaité, il faut alors procéder à des relevés répétés le plus rapidement possible (peut-être au cours de la même visite) afin de réduire les risques que l'espèce se déplace entre les sites (MacKenzie & Royle 2005). Dans notre cas, les grands carnivores sont des espèces très mobiles vivant à faible densité dont les territoires ont des tailles variables bien que proches de la taille des sites que nous avons sélectionnés. Nous pouvons ici conclure que ce que j'ai modélisé n'était donc pas à proprement parler « l'occupancy » (selon MacKenzie et al. 2002) mais bien l'utilisation de l'espace.

Le problème de l'échelle des sites choisis a également été souligné par Breiner & Bergamini (2018) dans le cas de l'établissement de l'AOO pour l'UICN. En effet, le choix de la taille des sites peut apparaître problématique car les modèles d'occupancy sont dépendants de l'échelle choisie, autrement dit plus l'échelle à laquelle les distributions ou les habitats des taxons sont cartographiés est fine, plus la zone estimée occupée sera petite. Inversement, établir la distribution à échelle grossière estimera moins de zones inoccupées (Marboutin et al. 2011; Steenweg et al. 2018). Le choix de l'échelle à laquelle la distribution est estimée peut donc influencer les résultats des évaluations de la liste rouge de l'UICN. Dans l'article 4, nous avons cependant démontré qu'en ré-exprimant les modèles d'occupancy comme des processus ponctuels, l'on pouvait se dédouaner de cet effet de dépendance des estimations à la taille des sites. En effet, les paramètres d'un processus ponctuel ne dépendent pas de l'échelle spatiale puisque les sorties de ces modèles sont exprimées en intensité par surface par opposition à une intensité par site (Renner & Warton 2013). Par conséquent, utiliser le cadre de modélisation des processus ponctuels présente un avantage pour prédire la présence d'individus pour toute taille de site d'échantillonnage (Dorazio 2014; Koshkina et al. 2017).

Dans mon dernier article, j'ai développé un modèle mécanistico-statistique faisant le lien entre les mécanismes de diffusion de la population de loups en France et la probabilité que l'espèce soit détectée à un site. Afin de faire des prédictions, j'ai défini la probabilité qu'un site soit occupé comme la probabilité qu'il y ait au moins un individu sur le site. Or nous avons constaté que le modèle mécanistico-statistique prédit plus de sites avec une probabilité moyenne d'occupancy d'environ 0.6 que le modèle de site-occupancy dynamique qui, lui, estime des sites avec une probabilité d'occupancy plus marquée, jusqu'à 1, mais pour un nombre inférieur de sites. Cette différence pourrait s'expliquer par le fait que les modèles

Discussion

d'occupancy sont des modèles de régression, ce qui signifie que l'occupancy estimée est liée aux données, alors que l'approche mécaniste est basée sur un modèle continu dans le temps qui permet de répartir les individus sur plusieurs sites sans avoir à être détectés. Une autre explication pourrait résider dans le fait que nous avons fait l'hypothèse d'une distribution de Poisson pour les effectifs par site dans notre modèle mécanistico-statistique. Or il pourrait exister un problème d'ajustement de notre modèle aux données dû par exemple au phénomène de surdispersion lié à la dépendance entre individus du fait de la structure sociale du loup. Une première piste pour résoudre ce problème consiste à utiliser une distribution négative binomiale pour relâcher la contrainte d'égalité entre espérance et variance des effectifs, propre à la distribution de Poisson (White & Bennetts 1996). Nos essais d'implémentation sont pour l'instant restés vains, en raison de problèmes sévères de convergence. Une autre approche consisterait à modéliser directement la dépendance entre individus en explicitant dans la partie mécaniste de notre modèle la structure en meute (Lewis et al. 1997).

L'avantage d'avoir travaillé avec une approche bayésienne est que nous avons pu estimer directement les états des sites (Royle & Kéry 2007), comme les autres paramètres des modèles (écologiques ou de détection). Dans l'étude de la dynamique de colonisation du loup en France et celle du lynx dans les pays alpins, nous étions effectivement intéressés par produire des cartes de l'occupancy réalisée. Il convient de distinguer la variable aléatoire occupancy (latente) de la probabilité d'occupancy qui permet d'obtenir l'occupancy potentielle et les relations avec les variables environnementales, autrement dit l'habitat favorable. Si la probabilité d'occupancy s'obtient directement avec l'approche fréquentiste, il est difficile d'estimer la variable latente occupancy à moins d'utiliser des algorithmes spécifiques aux modèles de Markov cachés, comme l'algorithme de Viterbi que nous avons utilisé dans l'article 3. Il est aussi possible d'obtenir la probabilité d'occupancy sachant que l'espèce n'a jamais été détectée sur le site (la probabilité d'occupancy conditionnelle) via une application du théorème de Bayes (MacKenzie et al. 2002).

10.b.2. Interprétation des variables écologiques

Dans le cadre de ma thèse, j'ai identifié des variables environnementales avec une influence significative sur les paramètres écologiques. Je renvoie aux articles 1, 2 et 4 pour une discussion détaillée de ces effets. J'aimerais ici discuter plutôt du fait que les variables environnementales peuvent aussi ne pas représenter entièrement tous les habitats favorables pour l'espèce étudiée. Tout d'abord, Mackenzie & Royle (2005) ont souligné le problème que lors de protocoles classiques d'occupancy, les sites à partir desquels les données sont recueillies peuvent ne représenter qu'une fraction de l'ensemble de la distribution potentielle. Il est donc nécessaire que la sélection des sites permette de généraliser les résultats de l'analyse des données à l'ensemble de l'aire d'étude. Dans le cadre de ma thèse, les sites ont été sélectionnés à l'échelle des pays, comme en Norvège, voire sur plusieurs pays comme pour les pays alpins. En France, pour modéliser la distribution dynamique du loup, j'ai choisi une aire d'étude suffisamment grande pour couvrir suffisamment d'habitats différents. L'utilisation des données participatives représente de ce point de vue un avantage car elles permettent de couvrir une large couverture spatiale et donc de recouvrir différents habitats (Dickinson et al. 2012), moyennant le fait de pouvoir contrôler l'effort d'échantillonnage.

Discussion

Toutefois, dans les trois cas d'étude, bien que nous ayons choisi une grille d'étude couvrant une large partie des pays concernés, il est possible que les effets des variables environnementales que nous avons obtenus soient le résultat de l'histoire de la recolonisation de ces espèces (Chapron et al. 2014). Par exemple, les loups en France sont réapparus dans les Alpes en partant de l'Italie via le corridor naturel que constituent les Apennins (Fabbri et al. 2007). Par conséquent, l'effet de l'altitude moyenne sur la probabilité de colonisation d'un site par le loup en France peut être lié à l'histoire de ce processus naturel de recolonisation. Nous pensons qu'à l'avenir, l'effet de l'altitude sera sûrement atténué lorsque les loups commenceront à coloniser les plaines. L'interprétation des variables environnementales est donc à faire en regard de l'histoire de la colonisation, et le modèle ne devrait donc pas être utilisé pour réaliser des prédictions à long terme.

Malgré nos efforts pour capturer la variation spatiale dans les probabilités d'occupancy et d'extinction/colonisation, pour les trois jeux de données, nous n'avons pas eu accès à la présence des proies, qui représente toutefois un facteur important sur la présence des carnivores tels que le loup et le lynx (Breitenmoser 1998; Mech & Boitani 2010; Mattisson et al. 2011; Duchamp et al. 2012; Molinari-Jobin et al. 2012). Pour faire face à ce défaut, nous avons donc utilisé la densité de forêt sur les sites comme indicateur de la présence de proies. Pour modéliser la présence du lynx en Norvège, nous avons une information de la présence de proies, qui se présentait sous forme binaire, absence ou présence de proies sauvages à un site, mais nous n'avons pas d'information sur la variation de la densité d'effectifs des proies. Nous avons donc utilisé la productivité primaire comme proxy de la variation de densité des proies. Dans tous les modèles que nous avons établis, nous avons trouvé que la présence du loup et du lynx augmentait lorsque le pourcentage de couverture forestière augmentait, ce qui va dans le sens d'un lien entre la présence des carnivores et leurs proies. Cependant, pour de futures recherches, je pense qu'il serait intéressant de travailler vers des mesures de présence de proies à l'échelle des sites afin de pouvoir affiner les estimations de présence du loup et du lynx.

10.b.3. Utilisation des résultats en termes de suivi

L'un des points communs à tous mes travaux de thèse est le fait que nous avons mis en évidence l'importance de prendre en compte la manière dont le loup et le lynx sont suivis dans le but d'établir leur répartition spatiale. Comme nous l'avons vu, le suivi de populations de grands carnivores est crucial pour éclairer les mesures de conservation (IUCN 2012), comme c'est le cas en général en gestion (Nichols & Williams 2006). L'état d'une population et les facteurs qui influencent cet état sont régulièrement évalués à l'aide d'indicateurs de l'occupancy (IUCN 2012). Ces indicateurs varient en termes de qualité et de quantité de données nécessaires pour les mesurer, et la stratégie de suivi appropriée pour une population est souvent difficile à mettre en place en termes de coût et de logistique (Pollock et al. 2002). La mise en place d'un suivi adapté peut parfois nécessiter de choisir entre l'indicateur occupancy vs. les effectifs de la population, tout en prenant en compte les coûts économiques et logistiques associés à ces suivis (Ward et al. 2017). L'occupancy et les effectifs sont influencés par des facteurs environnementaux différents, et peuvent donc être des mesures complémentaires dans l'évaluation de l'état de conservation des populations et des tendances associées (Gaston et al. 2000; Steenweg et al. 2018). Le suivi nécessaire pour mesurer les effectifs d'une population peut parfois exiger des ressources plus importantes telles que des

Discussion

piégeages photos et des analyses génétiques coûteuses (Gaston et al. 2000; Kéry et al. 2009). En revanche, mesurer l'occupancy est basé sur un suivi faisable à large échelle à un coût moindre, et ce suivi peut être utilisé sur des espèces cryptiques comme les grands carnivores (MacKenzie et al. 2002).

Outre les résultats sur la biologie de l'espèce et les avancés méthodologiques, j'aimerais m'attarder sur les leçons que l'on peut tirer par rapport au suivi du loup en France et du lynx dans les pays alpins et en Norvège. Dans le cas du suivi du loup en France, il est apparu que la probabilité de détecter l'espèce était hétérogène dans l'espace et dans le temps. Grâce aux informations sur les observateurs du réseau, nous avons réussi à établir une mesure de l'effort d'échantillonnage a posteriori, et nous avons trouvé que plus cet effort était grand, plus il y avait de chance pour que le loup soit détecté. Cependant, nous recommandons d'enregistrer l'activité des observateurs au sein du réseau lorsqu'ils vont prospecter sur le terrain, surtout sur les lieux où ils ne sont pas allés prospecter (Beirne & Lambin 2013). De plus, nous avons trouvé que l'hétérogénéité de la probabilité de détection était due à une forte hétérogénéité spatiale de l'effort d'échantillonnage dans l'article 3. Bien que nous ayons démontré qu'il était possible de mettre en place un modèle d'occupancy qui prenne en compte l'hétérogénéité dans la probabilité de détection, nous recommandons d'essayer de répartir des observateurs de manière plus homogène sur les zones où la probabilité de détecter l'espèce est plus faible qu'au cœur des Alpes afin de laisser peu de sites non échantillonnés sur le front de colonisation. En particulier, nous avons vu que le loup peut disperser à courte et longue distance (Mech & Boitani 2010, voir encadré No 1). Il est donc important de prendre ce facteur en compte dans le cadre de son suivi, c'est ce que l'ONCFS a mis en place depuis quelques années (Duchamp et al. 2012), où des observateurs sont formés dans des zones que le loup n'a pas encore recolonisées afin d'anticiper son retour éventuel et sa détection.

Dans le cas du suivi du lynx dans les pays alpins, nous avons confirmé le fait que la variation dans la probabilité de détecter le lynx était due à une variation de la détection en fonction du pays dans lequel le suivi se déroulait. Bien que tous les pays se soient entendus sur un suivi et une interprétation des données communes pour évaluer l'état et la répartition du lynx dans leurs pays respectifs, Molinari-Jobin et al. (2003) avaient déjà souligné le problème de l'homogénéisation des suivis lorsque plusieurs pays sont concernés. Dans cette étude, des conseils sur le renforcement du suivi du lynx tels que développer la communication au sein du réseau, informer les acteurs locaux sur la biologie de l'espèce, ont été développés pour chaque pays afin d'éviter une variabilité dans le suivi qui mène donc à une variabilité dans l'interprétation des données mais aussi dans la détection même de l'espèce. Le suivi sur plusieurs pays possède malgré tout certains avantages que je souhaite ici souligner. Tout d'abord, le suivi sur plusieurs pays a plus de sens au niveau de la biologie de l'espèce, le lynx étant présent dans plusieurs pays de l'arc alpin (Molinari-Jobin et al. 2003). Ensuite, la coopération internationale est essentielle pour la conservation des populations animales communes et encore plus pour résoudre les problèmes mutuels (Linnell & Boitani 2012). En effet, partager les données permet d'évaluer les mouvements de populations entre les pays, ainsi que de produire des estimations d'effectifs plus précis et moins biaisés que si les données étaient analysées séparément par pays (Bischof et al. 2016; Gervasi et al. 2016). Enfin, nous avons constaté que l'intégration des données incertaines produisait des sites estimés comme occupés qui ne le sont pas si ces données n'étaient pas utilisées. Estimer ces nouveaux sites occupés peut représenter une nouvelle information cruciale dans le cadre de la conservation

Discussion

d'une espèce comme le lynx qui rencontre des difficultés à coloniser de nouveaux habitats et qui vit à faibles effectifs. Ces nouveaux endroits peuvent représenter de nouvelles zones colonisées mais peuvent aussi représenter de potentiels conflits avec les activités humaines, identifier ces zones peut aussi permettre l'anticipation de l'émergence de nouvelles zones de conflits. Pour de futures études, il pourrait être intéressant de valider la présence de l'espèce, en réalisant de nouvelles visites ou en plaçant des pièges caméras sur des sites où uniquement des données incertaines ont été trouvées, afin de pouvoir obtenir des données certaines. De plus, ces zones estimées occupées à l'aide des données incertaines peuvent servir comme des zones d'alerte où l'espèce pourra être présente dans le futur de manière certaine.

Dans le cas du suivi du lynx en Norvège, plusieurs suivis sont réalisés, or nous avons démontré qu'avec deux suivis différents combinés dans un cadre commun, il était possible d'établir les tendances dans la répartition spatiale du lynx à l'échelle nationale et d'établir les facteurs sous-jacents gouvernant ce processus de manière plus précise que lorsque les deux sources de données étaient analysées séparément. De plus, le suivi du lynx en Norvège est similaire à celui réalisé en Suède, et il serait intéressant pour de futures recherches d'intégrer ensemble les suivis de ces deux pays également car la population de lynx est répartie spatialement sur toute la Scandinavie (Andrén et al. 2002; Gervasi et al. 2016).

Enfin, l'implication d'un large public d'observateurs contribuant à la récolte des données permet d'élargir la couverture de prospection suffisamment grande pour couvrir les zones d'études des grands carnivores qui vivent à de faibles densités sur de larges territoires de plusieurs centaines de km². De plus, comme nous l'avons vu dans le premier article, les grands carnivores ont une probabilité de détection faible et la probabilité de détecter l'espèce augmente avec le nombre d'observateurs qui viennent prospecter chaque site chaque année. Il est donc important de maintenir un certain nombre d'observateurs prospectant afin d'avoir une probabilité de détecter l'espèce suffisamment élevée – et dans tous les cas non nulle – pour pouvoir la détecter au moins dans toutes zones potentielles. Ici, nous avons vu l'importance que prenait la participation de personnes aux profils socio-professionnels variés dans le suivi des grands carnivores : détection des signes indirects de présence du loup en France, du lynx dans le massif alpin ainsi que des groupes familiaux en Norvège. La combinaison des sciences citoyennes avec les données obtenues selon des protocoles scientifiques offre donc un cadre utile pour évaluer les changements de distribution des grands carnivores à grande échelle spatiale et temporelle. L'identification des facteurs favorisant la détection de l'espèce est un enjeu majeur pour permettre l'amélioration des réseaux de suivi (Molinari-Jobin et al. 2018) tant dans la diversité de la prospection que dans son efficacité en termes de rapport coûts/bénéfices.

10.c. Vers une gestion adaptative des grands carnivores, sur la base de l'occupancy

Une des perspectives qui se dessine de mon travail visera à étudier la pertinence des approches de modélisation de la distribution d'espèces que j'ai développées pour une gestion appliquée des grands carnivores. Une des pistes que nous proposons consiste à utiliser l'occupancy comme un indicateur de l'état des populations en complément, voire en substitution, de l'indicateur basé sur les effectifs, coûteux en termes logistiques et financiers pour être maintenu à grande échelle. S'il s'agit sûrement de la seule option possible pour les populations

Discussion

de lynx en Scandinavie puisqu'il n'existe pas à l'heure actuelle de suivis permettant l'estimation des effectifs, la question se pose actuellement en France pour la gestion du loup. Les effectifs de loups en France sont estimés par des méthodes de capture-recapture qui nécessitent des données à l'échelle de l'individu, obtenues sur la base de suivi biomoléculaire non-invasif principalement (Miquel et al. 2006; Cubaynes et al. 2010). La complexité logistique et financière liée à ce monitoring à grande échelle nécessitant plusieurs milliers d'observateurs, qui limite les volumes d'analyse possibles a pour conséquence que ces effectifs estimés par capture-recapture sont obtenus avec un retard de quelques années. Ce processus rend donc difficile son implémentation directe dans le processus de décision de gestion réactive des populations. De plus, les méthodes d'estimation des effectifs n'intègrent (pour l'instant) pas l'information sur la localisation géographique des données récoltées, une contrainte qui réduit les possibilités d'une gestion spatialisée des populations. Pour résoudre ces deux problèmes, nous proposons d'explorer le recours aux modèles de distribution d'espèce qui intègrent naturellement la dimension spatiale des données de suivi de population du Réseau loup-lynx qui sont mises à jour annuellement afin de mettre en place une gestion adaptative basée sur les estimations spatiales de la distribution de la population plutôt que sur les effectifs.

Un autre avantage à l'utilisation de l'occupancy comme indicateur de l'état des populations réside dans la possibilité de mettre en place une gestion adaptative des populations. La gestion adaptative consiste à redéfinir régulièrement les moyens de gestion d'une espèce, ou de ses prélèvements, en fonction des estimations de l'état de la population, des nouvelles connaissances acquises sur son fonctionnement, et des objectifs fixés (tels que la diminution des impacts sur les élevages ou une croissance de population limitée) (Runge 2011). Il s'agit donc d'une démarche d'allers-retours qui consiste à prendre une décision à partir des connaissances disponibles, tout en prenant en compte l'incertitude associée (Williams 2009). L'adoption de cette démarche est aujourd'hui quasi unanime en biologie de la conservation pour des espèces menacées pour lesquelles il est nécessaire d'agir à court terme, même lorsque les menaces sont mal connues (Nichols & Williams 2006). La gestion adaptative est aussi applicable pour les espèces où des prélèvements sont effectués, comme les oiseaux d'eau (Madsen et al. 2017). Cette approche a récemment été recommandée dans le cadre d'une expertise collective sur la viabilité et le devenir de la population de loups en France à l'horizon 2030 commanditée par le ministère en charge de l'environnement (Duchamp et al. 2017).

Outre les mesures de prévention et de protection des troupeaux déployées sur une large partie de l'aire de présence régulière du loup, un des moyens de gestion utilisé par l'État français pour gérer les dégâts faits par le loup sur les troupeaux domestiques est le recours aux tirs de prélèvement dérogatoires. L'un des enjeux est de notamment stopper les foyers d'attaques et de décider comment effectuer ces prélèvements sans pour autant affecter la viabilité de la population et le statut de conservation favorable tel que défini par les textes légaux européens (Habitat directive 1992). Dans le contexte d'une population en colonisation encore actuelle, la gestion adaptative, basée sur la connaissance de l'espèce et l'évaluation de l'impact des prélèvements, permettrait de mettre en place des moyens de gestion afin de diminuer les impacts sur les activités ajustés au mieux à la biologie de l'espèce ainsi qu'aux objectifs de gestion.

Discussion

Le cœur de l'approche de gestion adaptative nécessite le développement d'un modèle qui capture le fonctionnement des populations en réponse aux perturbations, ici les prélèvements (Marescot et al. 2013). Nous proposons ici d'utiliser l'occupancy comme alternative aux effectifs pour assurer une aide à la décision mieux ciblées aux enjeux de territoire. La démarche consisterait alors à utiliser l'occupancy prédite selon un modèle pour cibler de futures zones potentiellement conflictuelles afin de prévenir les dégâts sur le bétail. L'approche mise à jour sur des pas de temps encore à définir permettrait d'adapter la gestion chaque année via la comparaison de la répartition prédite et la répartition estimée sur les données récoltées. Quel modèle utiliser ? Le choix du modèle dynamique d'occupancy développé dans l'article 1 semble difficile. En effet, il ne peut en l'état être utilisé qu'à des fins explicatives car la probabilité qu'un site soit colonisé dépend entre autres du nombre de sites voisins occupés à courte et longue distance, une information qui n'existe par définition pas encore quand on prédit à l'année suivante. Une piste prometteuse consistera alors à explorer l'utilisation du modèle mécanistico-statistique développé dans l'article 5 puisque l'approche modélise explicitement la diffusion écologique du loup et semble montrer de bonnes capacités prédictives à court terme. Cette piste fait partie d'une des propositions à explorer dans le cadre du nouveau plan 'action loup' gouvernemental et un nouvel axe du travail de thèse d'Oksana Grente qui a commencé le 1^{er} octobre 2018.

Annexes

Annexe 1 : Article vulgarisé sur l'article 1, diffusé sur le site du CNRS

En direct des laboratoires

14 novembre 2017

« Loup y es-tu » ? De l'utilisation d'un réseau d'observateurs pour modéliser la dynamique de colonisation du loup en France

Les grands prédateurs recolonisent leur aire de répartition originelle dans une grande partie de l'Europe, retour qui s'accompagne d'interactions croissantes avec les activités humaines et notamment l'élevage. En France, le loup gris (*Canis lupus*) est de retour depuis le début des années 1990 en provenance d'Italie. Comprendre la dynamique de colonisation du loup peut permettre de mesurer plus finement son statut de conservation et de mieux cibler les zones potentiellement conflictuelles en lien avec les attaques sur les troupeaux. Toutefois, modéliser la distribution du loup (et des grands carnivores en général) présente une difficulté méthodologique majeure : du fait de sa mobilité sur de vastes domaines vitaux, de sa discrétion et de sa faible densité, il est possible de ne pas détecter l'espèce dans certaines zones et d'en conclure, à tort, que le loup en est absent – on parle de faux négatifs.



Un loup pris au piège photo placé par les correspondants du réseau dans les Alpes de Hautes Provence / Crédit : ONCFS SD04

Annexes

Pour s'attaquer à ce problème de détection imparfaite, nous avons fait appel au réseau national d'observateurs formés à la reconnaissance et au rapportage des signes de présence du loup, réseau animé à l'échelle de la France par l'Office National de la Chasse et de la Faune Sauvage. Grâce à une caractérisation du rayon de prospection des observateurs en fonction de la catégorie socio-professionnelle et du lieu d'habitation ou de travail, nous avons reconstruit une mesure *a posteriori* de la pression d'échantillonnage hétérogène dans le temps et l'espace. Des modèles statistiques dit « d'occupation dynamique » nous ont permis d'analyser l'ensemble des indices de présence de l'espèce récoltés par le Réseau, et ainsi de cartographier la dynamique du retour du loup entre 1994 et 2016 en tenant compte de cette détection imparfaite.



Un correspondant en action : relevé d'une trace dans la neige / Crédit : O. Salvador

Annexes

Nos résultats montrent que, au-delà de la distribution du loup détectée par le Réseau, il existe d'autres endroits pour lesquels une probabilité que l'espèce soit présente est non négligeable compte tenu de la pression d'observation et des caractéristiques environnementales. Nous montrons également que la probabilité qu'un site soit colonisé dépend du taux d'occupation des sites voisins (<20km) mais aussi des mécanismes de colonisation à longue distance (<150km) que le loup adopte pour éviter la concurrence avec les meutes voisines. L'altitude, le taux de couverture forestière et de terres agricoles viennent ensuite compléter le pouvoir du modèle pour expliquer la probabilité de présence de l'espèce.

Depuis les Alpes du Sud où l'espèce était détectée en 1994, celle-ci s'est étendue dans un premier temps vers le Nord des Alpes puis les Pyrénées au début des années 2000 via le Massif central. Depuis le début des années 2010, le loup a continué sa progression jusque dans le Nord-Est de la France. Le taux d'expansion géographique de l'espèce en France était à son maximum dans les premières années de colonisation notamment lié à la disponibilité des espaces vacants à cette époque. Une première phase de ralentissement de la colonisation s'est opérée jusqu'au début des années 2000, pour se stabiliser depuis à un taux d'environ 7 % de couverture supplémentaire par an, avec une variabilité d'une année sur l'autre et selon les régions.

Les enjeux de la conservation de l'espèce et de la gestion des conflits liés aux attaques pourraient trouver dans cette production un nouvel outil d'aide à la décision, pour autant que celui-ci s'inscrive dans une démarche de gestion adaptative avec une réévaluation des prédictions des modèles suite aux actions de gestion réalisées sur le terrain.

Ces travaux ont été mobilisés dans l'[expertise récente commanditée par le ministère sur le futur du loup](#) à laquelle nous avons participé.

Annexes Article 1

Table A1: Size of prospection areas as a function of socio-professional category of observers. Observers were classified according to 8 entities to capture the diversity of their professional and personal field activities. People working for the departmental authorities (Category 1) display a field effort all over that departmental area. Observers belonging to the category 2 are state employees affected to the protected area they are working in. Details were not given for Regional Natural Park agents and Natural reserve agents. Their prospection area corresponds to the mean area of the protected area they are affiliated to. ONCFS agents (category 8) are attributed half a French Department as field areas when assigned for species monitoring. ONF agents (category 9) are attributed 1/10 of a French Department. Farmers (category 4) and hunters (category 5) usually focus on the restricted area (“municipality”) where they farm, breed sheep or hunt. Scientists (category 3), members of a naturalist association (category 6) and volunteers (category 7) were given $\frac{1}{4}$ of their affiliated department as their main activity might not be focused on species monitoring.

Socio-professional category	Prospection area	Number of observers
Administrative (e.g., department authorities) (category 1)	Area of the affiliated French department	275
National Park agent (category 2)	Area of the affiliated National Park	323
Regional Natural Park agent (category 2)	450 km ²	87
Natural reserve agent (category 2)	10km ²	26
Scientist (category 3)	1/4 of the affiliated department's surface	10
Agricultural profession (category 4)	Area of the affiliated municipality	91
Hunter (category 5)	Area of the affiliated municipality	423
Member of a naturalist association (category 6)	1/4 of the affiliated department's surface	267
Volunteer (category 7)	1/4 of the affiliated department's surface	474
ONCFS agent (category 8)	1/2 of the affiliated department's surface	746
French Forest Agency (ONF) agent (category 9)	1/10 of the affiliated department's surface	354

Annexes

Table A2: French departments where observers were present in 2016 along with their area and the number of observers affiliated to each department depending on their socio-professional category. Categories 2, 4 and 5 are not shown because their prospection areas do not depend on the size of the affiliated department. See also Figure A1.

Department.code	Surface (km²)	Category.1	Category.3	Category.6	Category.7	Category.8	Category.9	Total
1	5762	4	0	6	10	20	10	76
3	7340	0	0	0	0	2	0	2
4	6925	23	0	8	42	39	27	174
5	5549	11	0	18	29	31	19	210
6	4299	14	1	0	35	41	2	208
7	5529	6	0	7	3	13	4	58
8	5229	4	0	1	3	6	5	38
9	4890	0	0	6	2	16	7	36
10	6004	0	1	1	0	3	0	6
11	6139	1	0	3	2	10	3	44
12	8735	4	0	0	3	19	4	50
13	5087	2	0	2	3	11	2	27
14	5548	0	0	0	0	1	0	1
15	5726	6	0	4	3	9	2	35
19	5857	0	0	0	0	1	0	1
21	8763	2	0	1	1	13	0	18
23	5565	0	0	0	0	2	0	2
24	9060	0	0	0	0	3	0	3
25	5234	7	0	13	5	23	13	90
26	6530	14	0	6	22	24	26	145
27	6040	0	0	0	0	1	0	1
29	6733	0	0	0	0	1	0	1
30	5853	4	0	3	0	12	1	47
31	6309	1	1	0	6	19	1	32
32	6257	0	0	0	1	10	0	11
33	10725	0	0	0	0	2	0	2
34	6101	7	2	4	2	21	1	43
36	6791	0	0	0	1	0	0	1
38	7431	6	1	23	37	41	32	211
39	4999	6	0	19	12	14	15	90
40	9243	0	0	0	0	6	0	6
42	4781	12	0	1	3	4	2	37
43	4977	3	0	0	0	9	1	43
44	6815	0	0	0	1	0	0	1

Department.code	Surface (km²)	Category.1	Category.3	Category.6	Category.7	Category.8	Category.9	Total
46	5217	0	0	0	0	7	0	7
47	5361	0	0	0	0	1	0	1
48	5167	18	0	7	10	17	8	111
51	8162	0	0	0	0	2	0	2
52	6211	1	0	0	0	10	0	11
54	5246	4	0	12	3	12	4	47
55	6211	6	0	4	2	14	5	47
56	6823	0	0	0	0	1	0	1
57	6216	3	0	6	8	17	1	38
58	6817	1	0	0	0	8	0	9
59	5743	0	0	0	1	0	0	1
63	7970	1	1	4	2	9	1	18
64	7645	0	0	0	2	13	0	16
65	4464	0	0	0	0	4	0	6
66	4116	7	0	12	24	15	15	102
67	4755	6	0	21	17	29	9	95
68	3525	29	0	24	59	18	10	163
69	3249	7	0	2	4	5	0	19
70	5360	5	0	9	3	8	7	48
71	8575	0	0	0	2	8	0	10
73	6028	17	1	6	32	24	54	225
74	4388	7	0	20	40	21	19	133
75	105	1	0	0	0	0	0	1
78	3284	0	0	0	1	0	0	1
80	6170	0	0	1	0	0	0	1
81	5758	2	0	0	1	9	5	34
82	3718	0	0	0	0	3	0	3
83	5973	7	2	2	6	21	4	64
84	3567	2	0	1	3	13	5	32
85	6720	0	0	0	0	1	0	1
87	5520	0	0	0	2	0	0	2
88	5874	9	0	10	21	15	28	107
89	7427	0	0	0	0	6	0	6
90	609	5	0	0	3	8	2	23
92	176	0	0	0	2	0	0	2



Figure A1: Map of French departments with the identity code used in Table A2.

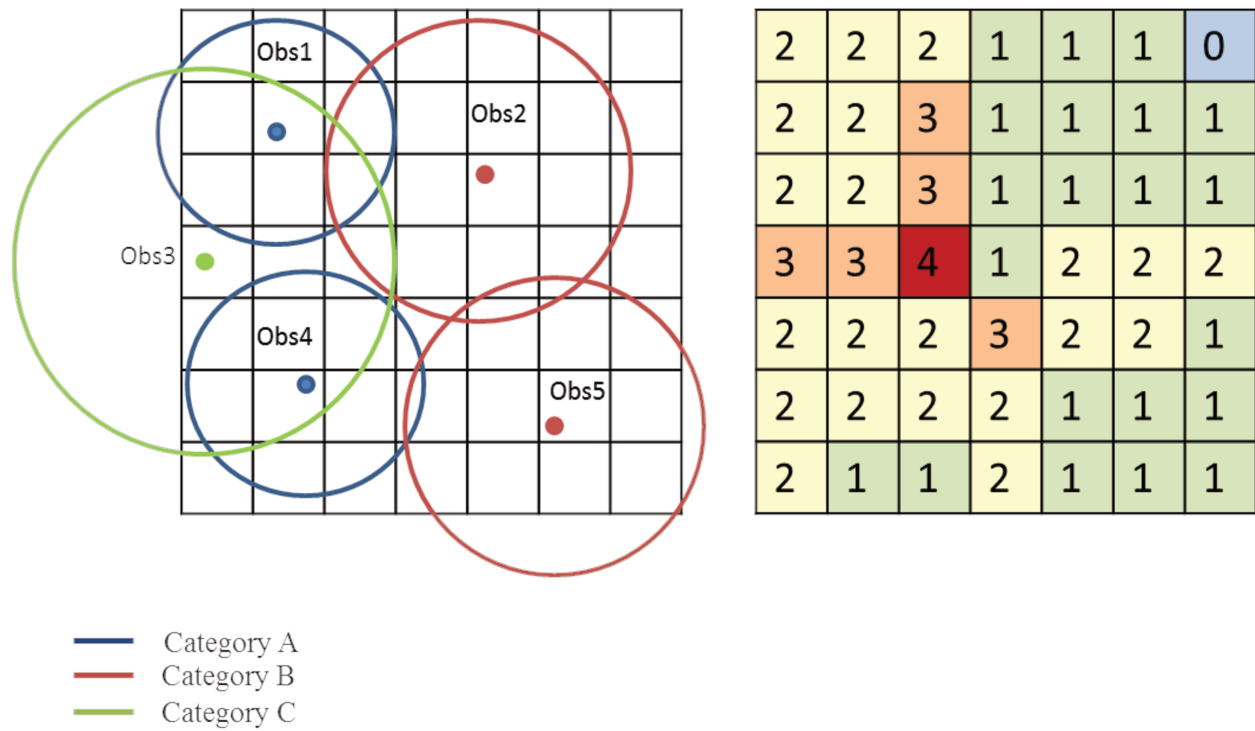


Figure A2: Schematic representation of how the sampling effort was calculated. Left: Observers were plotted according to their address. A circular buffer was affiliated to each observer with a surface equal to the prospection area following Tables A1 and A2. Right: resulting sampling effort calculated as the sum of observers sampling in each cell.

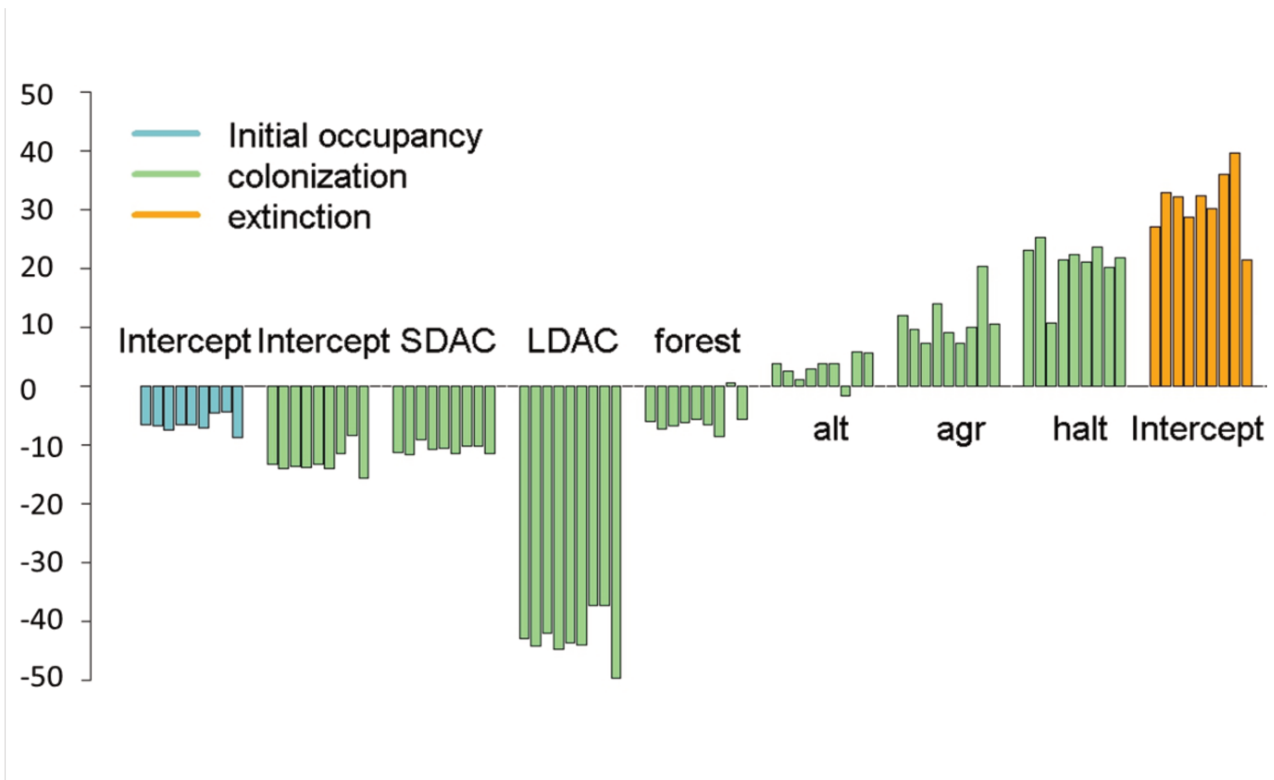


Figure A3: Analysis of sensitivity to the sampling effort definition. Each bar within a barplot represents the percentage of change (in %) in parameter estimates when compared to our original definition of the sampling effort after modification of the prospection area for one category of observers, in the following order: category 6: 1/10 of the affiliated department's surface; category 6: 1/2 of the affiliated department's surface; category 6: 100% of the affiliated department's surface; category 7: 1/10 of the affiliated department's surface; category 7: 1/2 of the affiliated department's surface; category 7: 100% of the affiliated department's surface; category 8: 1/10 of the affiliated department's surface; category 8: 1/4 of the affiliated department's surface; category 8: 100% of the affiliated department's surface. The model best supported by the data is used throughout these analyses.

Table C1: Top 10 models incorporating habitat covariates for the wolf detection/non-detection data. In the model structure, a 1/0 indicates the presence/absence of the corresponding covariate in the colonization, extinction or detection probability. Note that the intercept is always included in the model and therefore not represented in this notation.

Model structure												Posterior model probability
colonization						extinction			detection			
forest	agr	rock	halt	alt	Dbar r	SDAC	LDAC	Trend-year	SEff	Rden s	survey	
1	1	0	1	1	0	1	1	0	1	1	1	0.656
1	1	0	1	1	0	1	1	1	1	1	1	0.182
1	1	1	1	1	0	1	1	0	1	1	1	0.066
1	1	0	1	1	1	1	1	0	1	1	1	0.061
1	1	1	1	1	0	1	1	1	1	1	1	0.016
1	1	0	1	1	1	1	1	1	1	1	1	0.014
1	1	1	1	1	1	1	1	0	1	1	1	0.005
0	0	0	0	0	0	0	0	0	0	0	0	0.000
1	0	0	0	0	0	0	0	0	0	0	0	0.000
0	1	0	0	0	0	0	0	0	0	0	0	0.000

Table C2: Parameters estimates from the best dynamic site-occupancy model for wolf in France between 1994 and 2016. The median a posteriori is given with the associated standard deviation (SD). Occasions 2, 3 and 4 correspond to January, February and March. Estimates are given on a logit scale except for alpha.psi which is given on its natural scale, i.e. [0, 1].

Parameters	Median	SD	Interpretation
alpha.phi	-1.13	7.50×10^{-2}	Extinction intercept
alpha.psi	2.55×10^{-3}	1.49×10^{-3}	Initial occupancy
alpha.p	-1.98	8.04×10^{-2}	Detection intercept
beta.SEffp	0.37	2.56×10^{-2}	Effect of sampling effort on detection
beta.Rdensp	-0.33	4.00×10^{-2}	Effect of accessibility (road density) on detection
beta.occp2	0.47	6.42×10^{-2}	Effect of occasion 2 on detection
beta.occp3	0.40	6.36×10^{-2}	Effect of occasion 3 on detection
beta.occp4	0.38	6.42×10^{-2}	Effect on occasion 4 on detection
alpha.gamma	-5.65	8.79×10^{-2}	Colonization intercept
beta.agamma	0.49	8.93×10^{-2}	Effect of farmland cover on colonization
beta.SDACgamma	0.68	3.33×10^{-2}	Effect of short distance occupied neighboring cells on colonization
beta.LDACgamma	0.49	5.25×10^{-2}	Effect of long-distance occupied neighboring cells on colonization
beta.fgamma	0.64	6.34×10^{-2}	Effect of forest cover on colonization
beta.altgamma	0.59	6.06×10^{-2}	Effect of mean altitude on colonization
beta.haltgamma	-0.15	4.44×10^{-2}	Effect of high altitude proportion on colonization

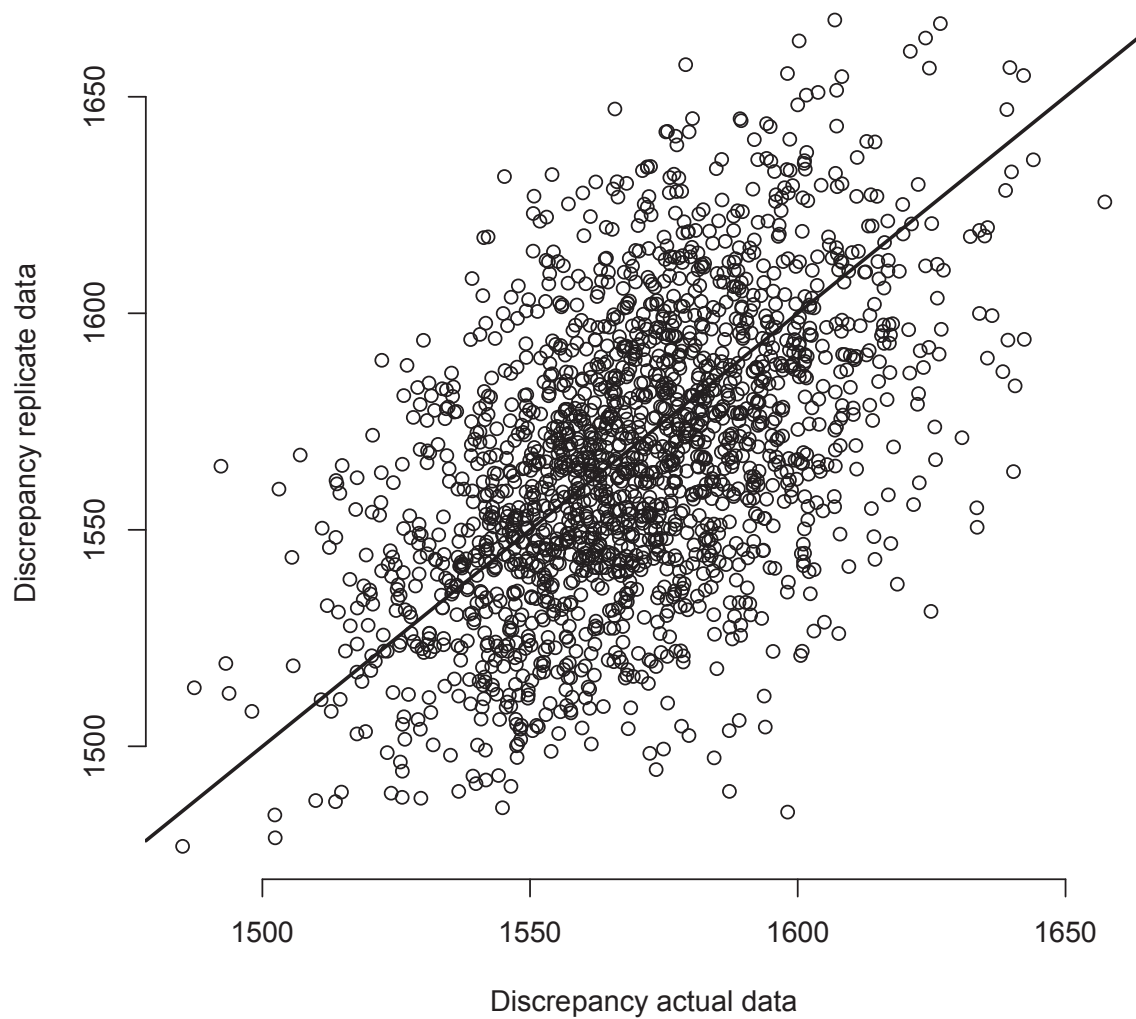


Figure B1: Results from posterior predictive checks for the dynamic occupancy model best supported by the wolf data. We show a scatterplot of the predicted chi-square discrepancy between simulated and expected data (on the Y axis) versus the observed chi-square discrepancy between expected and observed data (on the X axis) across MCMC samples. The Bayesian predictive p-value is 0.46 and represents the proportion of samples above the diagonal. Overall, the fit of the model seems satisfactory.

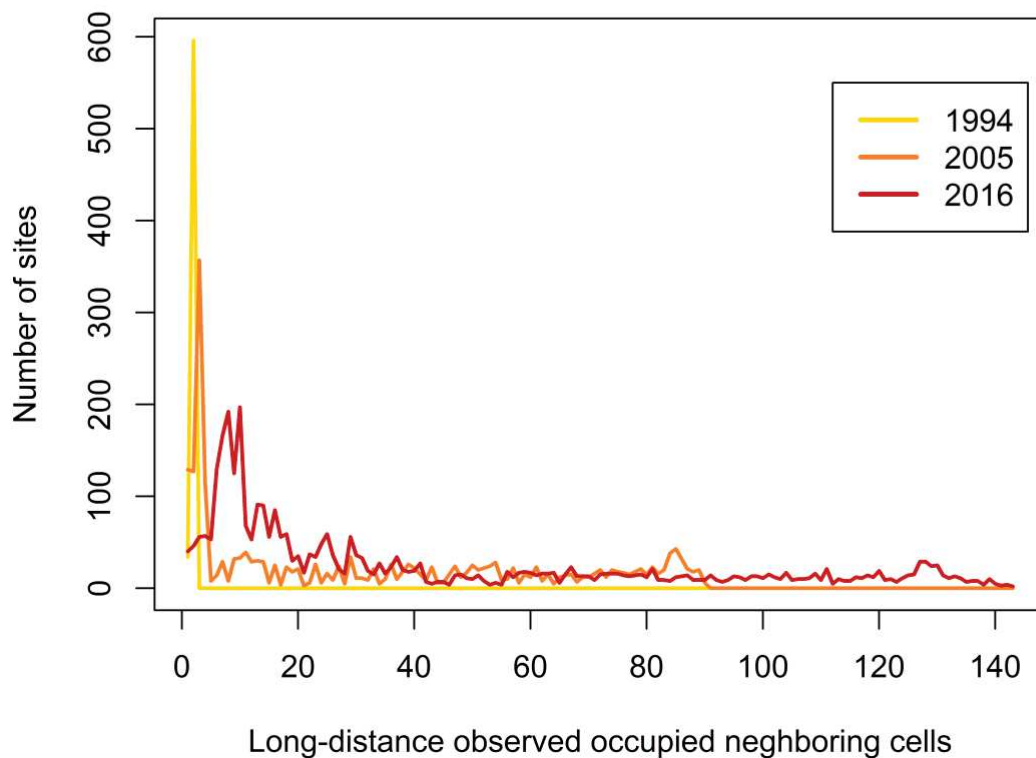
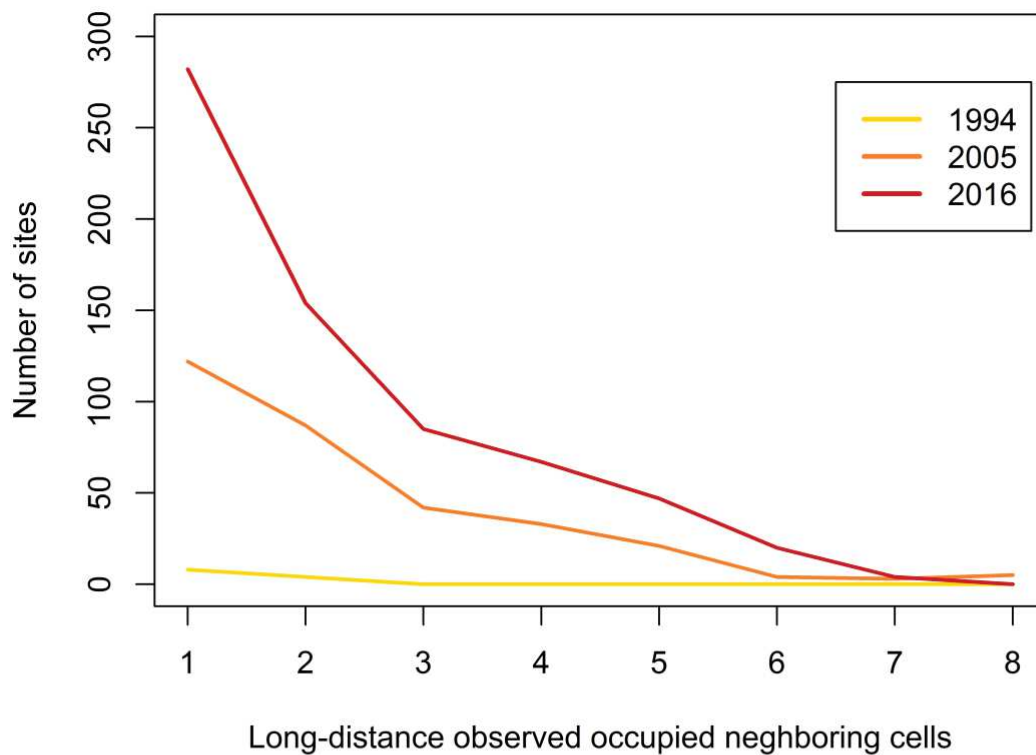


Figure D1: Number of sites having more than 0 observed occupied neighboring cells at short (contiguous cells) and long distance (between 10 km and 150 km) for 1994, 2005, and 2016.

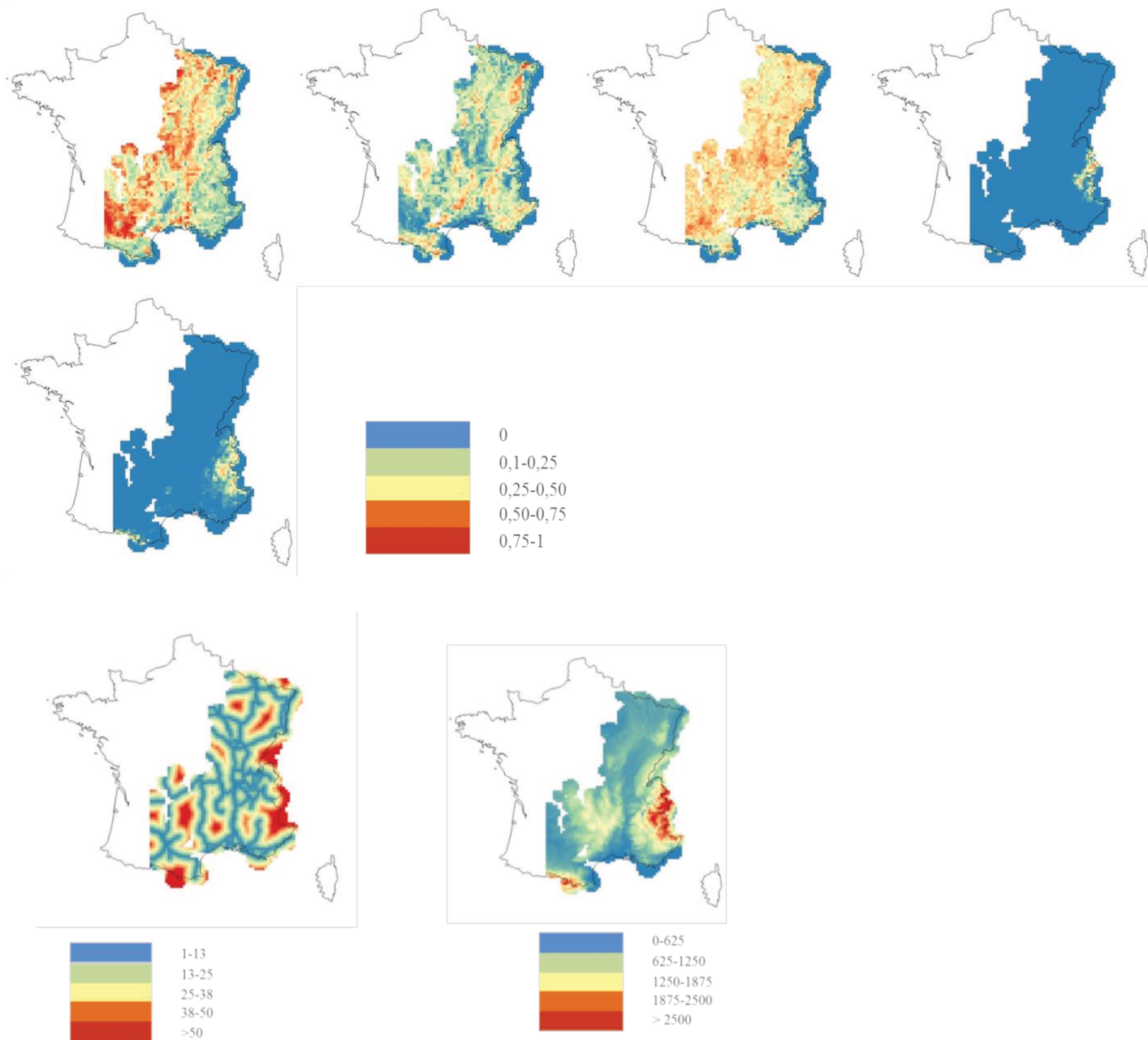


Figure D2: Maps of covariates tested in the study. First row from left to right: proportion of farmland cover (from 0 to 1); proportion of forest cover (from 0 to 1); road density (from 0 to 1); high altitude density (from 0 to 1). Second row: proportion of rock cover (from 0 to 1). Third row: left: distance from the cell center to the closest barrier (highway or river); right: mean altitude (in meters).

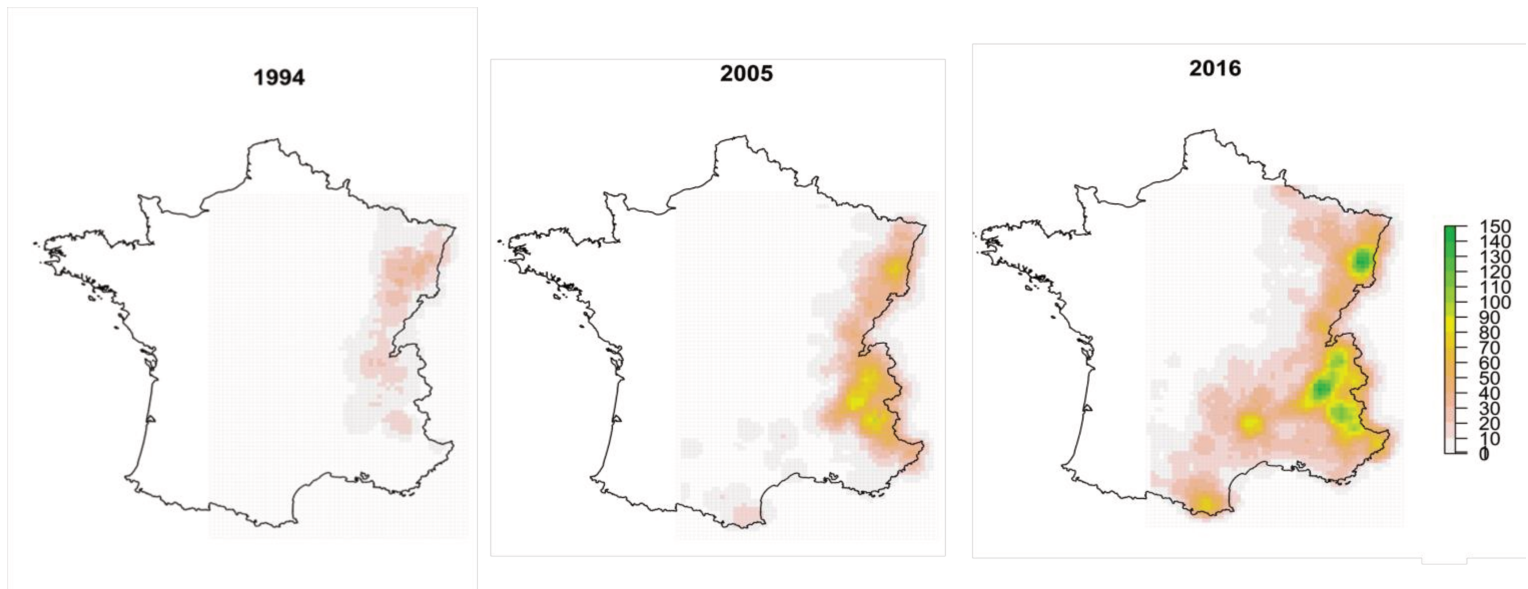
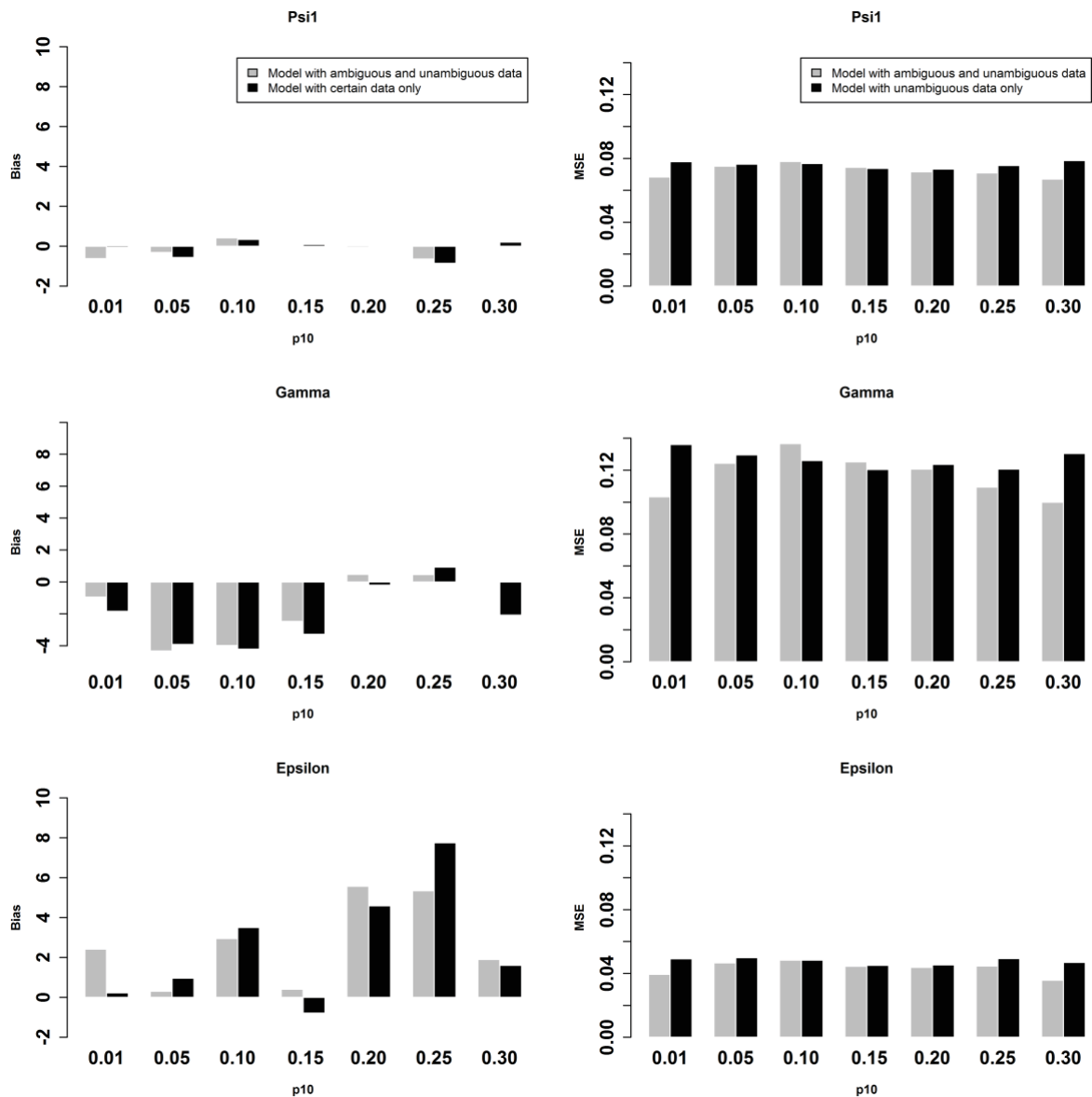


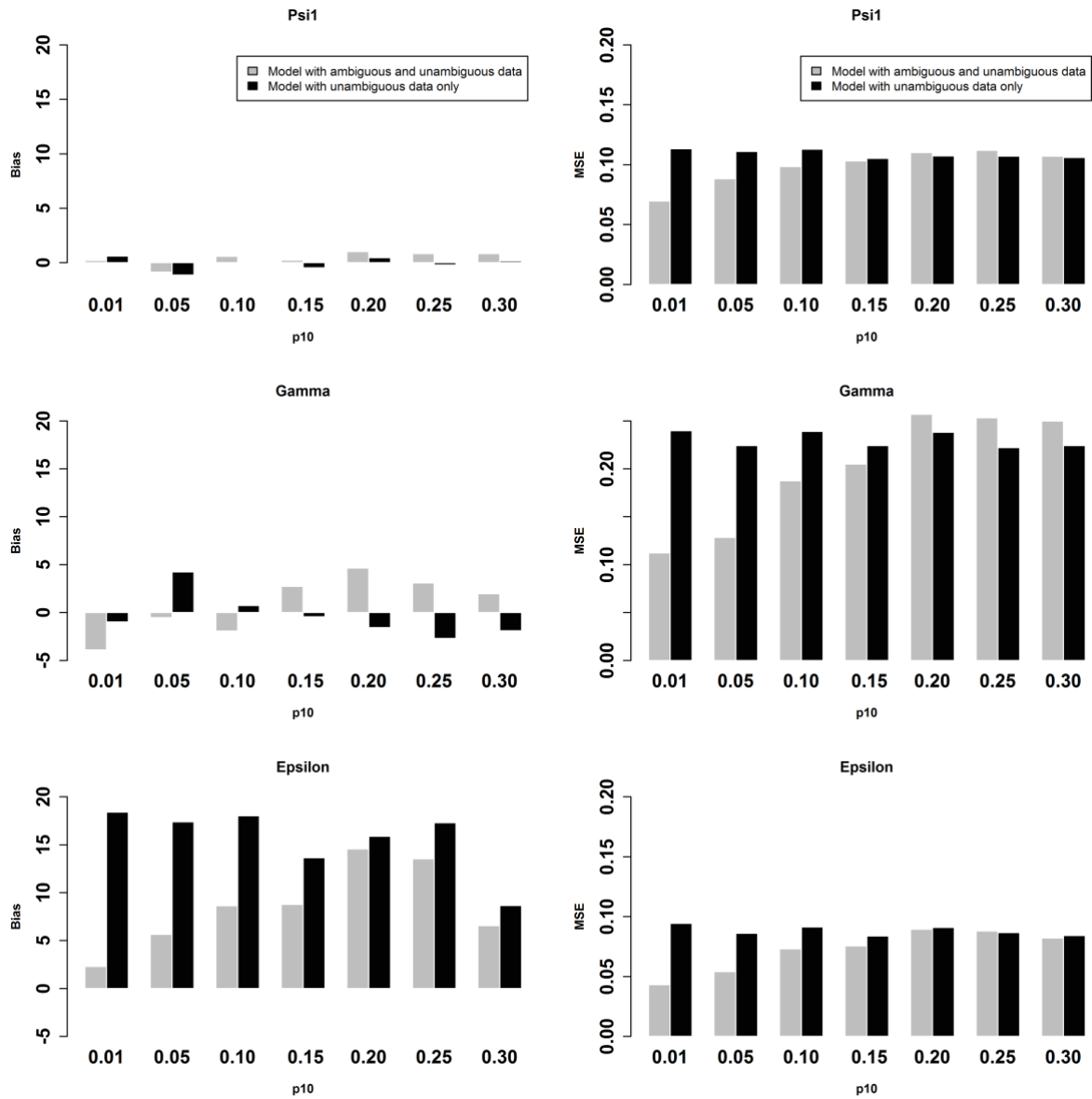
Figure D3: Maps of sampling effort for 1994, 2005, and 2016. Sampling effort was defined as the number of prospecting observers per site per year. In 1994, 1036 sites were prospected by at least one observer and 3547 were prospected in 2016. In 1994 only sites in North-Eastern part of France were prospected and in 2016 all Eastern France was prospected with some parts in South-West. Size of the prospection area depended on the socio- professional category of observers: observers from an administrative field (policemen for instance) were assigned a theoretical prospection area the size of the French department in which they were affiliated, observers from national parks, regional natural parks and natural reserves were assigned a prospection area the size of those areas. Observers from the farming profession and hunters were assigned a prospection area the size of the county they work in. Scientists, members of naturalist associations and private observers were assigned an area a quarter the surface of the French department where they found signs. Observers working in the French Game and Wildlife Agency (ONCFS) were assigned an area half the surface of the French department they work in. Observers from the French National Office for Forest (ONF) were assigned an area 1/10 the surface of the department they work in.

Annexes Article 2

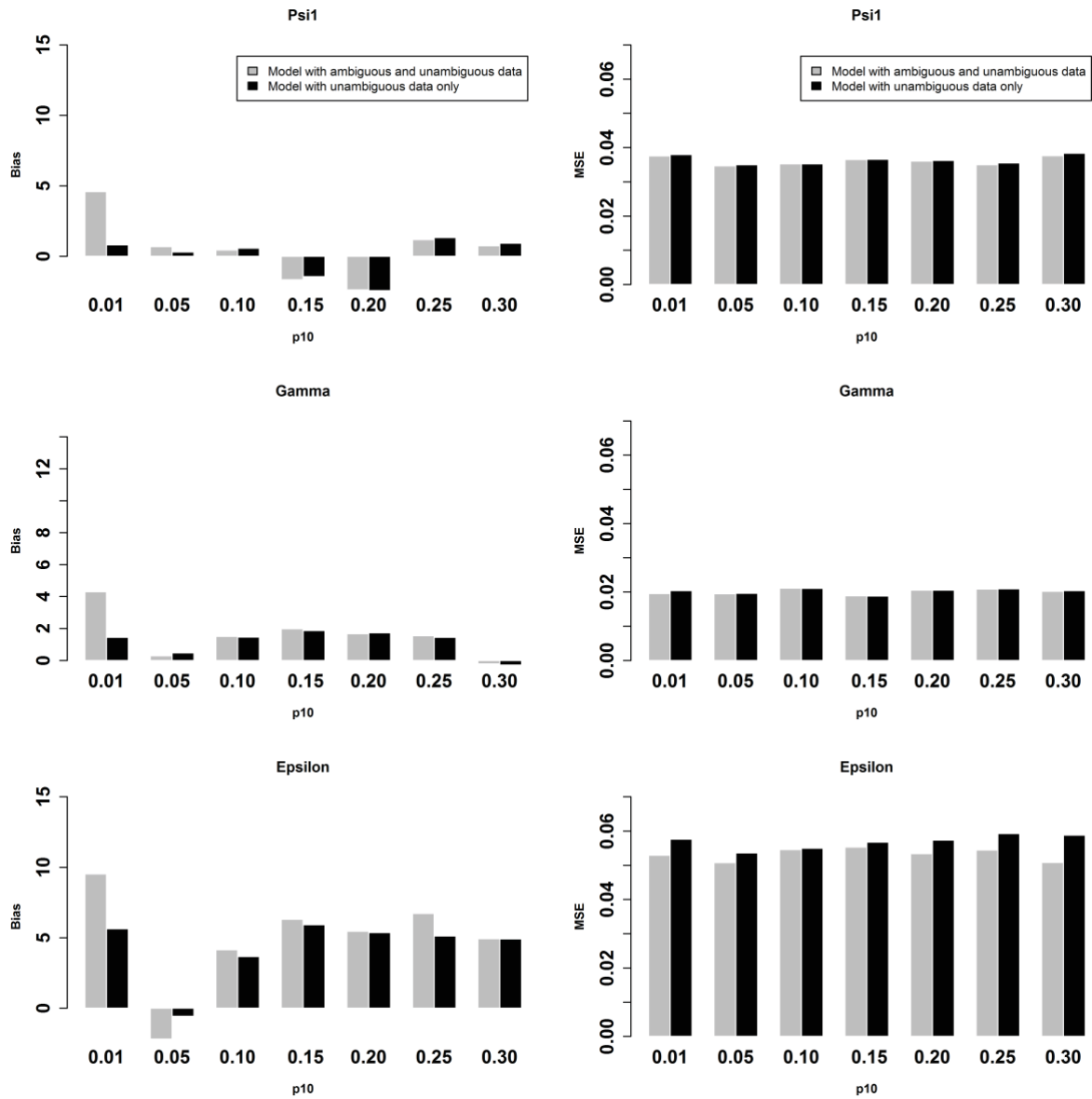
Appendix S3: Relative bias (in %) and Mean Square Error (MSE) from a dynamic occupancy model accounting for ambiguity and fitted with unambiguous and ambiguous data (MUA; grey bars) and a dynamic occupancy model dealing with unambiguous data only (MU; black bars) for the three ecological parameters, initial occupancy probability ψ_1 , colonization probability γ and extinction probability ϵ based on simulations in the HH, HL, LH and LL scenarios.



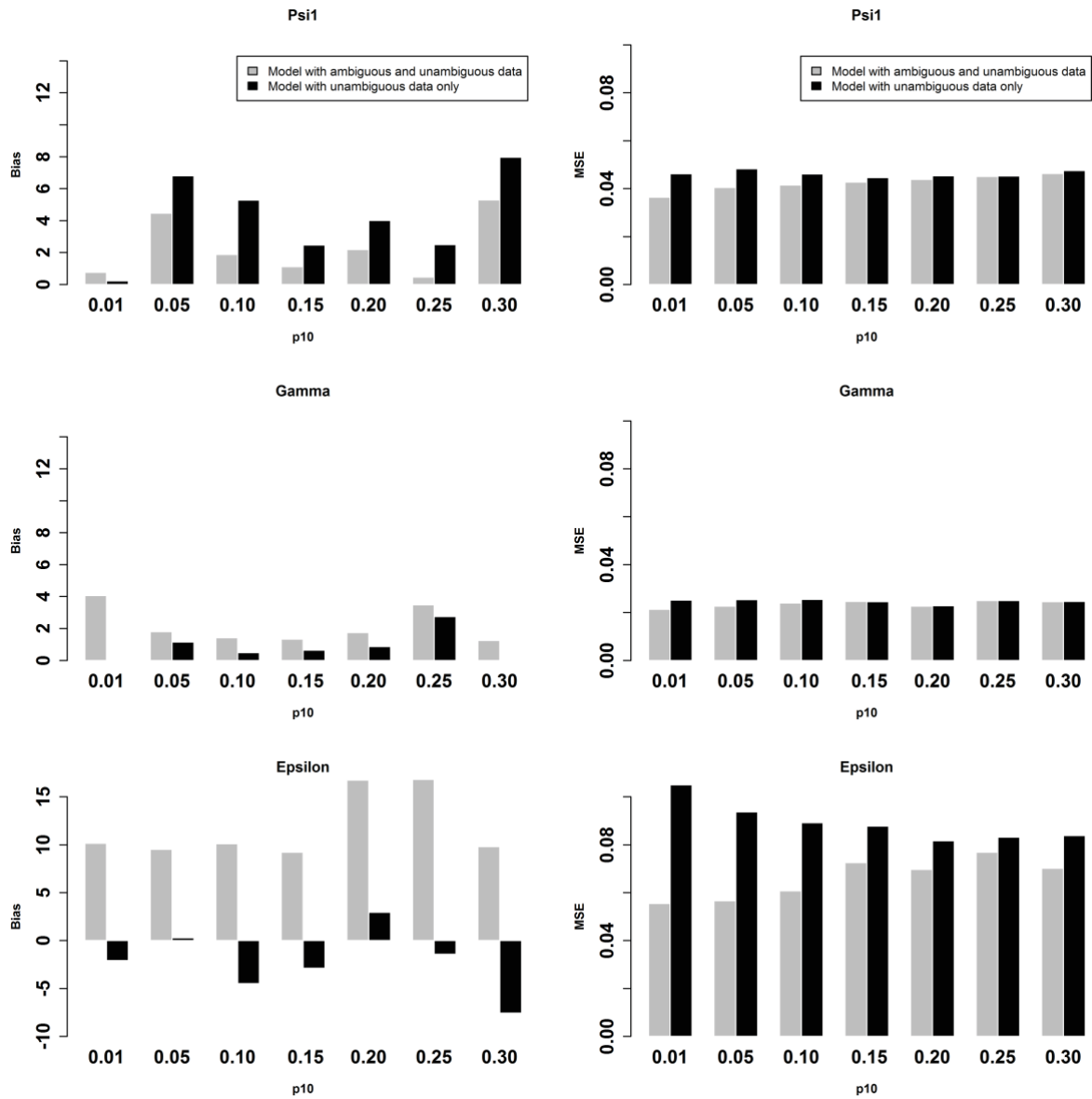
Relative bias (in %) and Mean Square Error (MSE) from a dynamic occupancy model accounting for ambiguity and fitted with unambiguous and ambiguous data (MUA; grey bars) and a dynamic occupancy model dealing with unambiguous data only (MU; black bars) for the three ecological parameters, initial occupancy probability ψ_1 , colonization probability γ and extinction probability ϵ based on simulations in the “high” occupancy “high” b scenario (HH).



Relative bias (in %) and Mean Square Error (MSE) from a dynamic occupancy model accounting for ambiguity and fitted with unambiguous and ambiguous data (MUA; grey bars) and a dynamic occupancy model dealing with unambiguous data only (MU; black bars) for the three ecological parameters, initial occupancy probability ψ_1 , colonization probability γ and extinction probability ϵ based on simulations in the “high” occupancy “low” b scenario (HL).



Relative bias (in %) and Mean Square Error (MSE) from a dynamic occupancy model accounting for ambiguity and fitted with unambiguous and ambiguous data (MUA; grey bars) and a dynamic occupancy model dealing with unambiguous data only (MU; black bars) for the three ecological parameters, initial occupancy probability ψ_1 , colonization probability γ and extinction probability ϵ based on simulations in the “low” occupancy “high” b scenario (LH).



Relative bias (in %) and Mean Square Error (MSE) from a dynamic occupancy model accounting for ambiguity and fitted with unambiguous and ambiguous data (MUA; grey bars) and a dynamic occupancy model dealing with unambiguous data only (MU; black bars) for the three ecological parameters, initial occupancy probability ψ_1 , colonization probability γ and extinction probability ϵ based on simulations in the “low” occupancy “low b” scenario (LL).

Annexes

Appendix S4: Estimates of parameters in the dynamic occupancy model accounting for unambiguous data (MU) on the left and in the dynamic occupancy model accounting for unambiguous and ambiguous data (MUA) on the right, with posterior means and standard deviations and the lower and upper bound of the 95% credible interval.

Model with unambiguous data only (MU)					Model with unambiguous and ambiguous data (MUA)				
Parameter	mean	sd	2.5%	97.5%	Parameter	mean	sd	2.5%	97.5%
Initial occupancy probability					Initial occupancy probability				
Intercept	-5.31	0.67	-6.77	-4.12	Intercept	-5.29	0.63	-6.64	-4.11
Effect of forest	0.76	0.27	0.26	1.34	Effect of forest	0.71	0.25	0.22	1.22
Effect of distance to nearest released site	-3.11	0.58	-4.32	-2.06	Effect of distance to nearest released site	-2.99	0.56	-4.15	-1.93
Extinction probability					Extinction probability				
Intercept	-0.53	0.32	-1.16	0.10	Intercept	0.23	0.21	-0.17	0.66
Effect of forest	0.29	0.14	0.01	0.5	Effect of forest	0.21	0.13	-0.04	0.48
Effect of number of observed occupied contiguous neighbors.	8.40	1.37	5.93	11.41	Effect of number of observed occupied contiguous neighbors.	7.87	1.21	5.77	10.56
Colonization probability					Colonization probability				
Intercept	-4.57	0.20	-4.91	-4.26	Intercept	-4.63	0.16	-4.97	-4.34
effect of elevation	-0.20	0.08	-0.37	-0.04	effect of elevation	-0.24	0.08	-0.40	-0.09
effect of human density	-3.78	1.45	-6.58	0.96	effect of human density	-3.07	1.39	-5.79	-0.42

Annexes

effect of distance to nearest released site	-0.88	0.10	-1.09	-0.69	effect of distance to nearest released site	-0.86	0.10	-1.08	-0.67
Effect of number of observed occupied contiguous neighbors.	6.98	0.59	5.83	8.15	Effect of number of observed occupied contiguous neighbors.	7.32	0.59	6.19	8.51

Annexes Article 3

Appendix A. Mathematical formulation and instructions for implementing an occupancy model incorporating misidentification and detection heterogeneity with altitude as a covariate on the initial occupancy probability in program E-SURGE.

Occupancy models dealing with misidentification and heterogeneity can be implemented in Program E-SURGE (Choquet et al. 2009) which is freely downloadable at <https://www.cefe.cnrs.fr/fr/actus/livres/34-french/recherche/bc/bbp/264-logiciels>. Below we provide the matrix patterns of state and event transitions for the occupancy model accounting for misidentification and heterogeneity developed in this paper. Statistical and software details can be found in E-SURGE manual (Choquet and Nogue 2011).

We considered 2 classes of sites, say A and B. We considered also 4 states: unoccupied in site A (U_A), unoccupied in site B (U_B), occupied in site A (O_A), occupied in site B (O_B) plus the state dead that is required by E-SURGE. Finally, we considered 3 events: not detected (0), detected unambiguous (1) and detected ambiguous (2).

We defined the class affiliation vector Π , the initial state matrix Φ , the transition matrix A , the detection matrix P and the classification matrix B . The parameter π (respectively $1 - \pi$) denotes the proportion of sites in class A (resp. class B), while ψ_{1A} (resp. ψ_{1B}) denotes the initial occupancy probability for sites in class A (resp. class B). The parameter γ_A (resp. γ_B) is the probability that an unoccupied site of class A (resp. B) at time k becomes occupied at time $k+1$ and ε_A (resp. ε_B) the probability that an occupied site of class A (resp. B) at time k becomes unoccupied at time $k+1$. The parameter p_{A10} (resp. p_{B10}) is the probability of incorrectly detecting the species at an unoccupied site A (resp. B) and p_{A11} is the probability of detecting the species at an occupied site A (resp. B). Finally, the parameter b is the probability to classify a true detection as unambiguous.

Based, on those parameters, we have:

Initial States (in two steps):

Class affiliation

At the initial state, sites can be of class A (resp. B) with the probability π (resp. $1 - \pi$)

$$\Pi = \begin{matrix} \text{class A} & \text{class B} \\ \begin{bmatrix} \pi & 1 - \pi \end{bmatrix} \end{matrix}$$

Initial state

At the second step, sites of class A. (resp. B) can be occupied with the probability ψ_{1A} (resp.

ψ_{1B})

$$\Phi = \begin{matrix} \text{class A} & \text{class B} \\ \begin{bmatrix} 1 - \psi_{1A} & 0 \\ 0 & 1 - \psi_{1B} \end{bmatrix} \end{matrix} \begin{matrix} UA & UB & OA & OB \\ \begin{bmatrix} \psi_{1A} & 0 \\ 0 & \psi_{1B} \end{bmatrix} \end{matrix}$$

State Transitions

The transition step in the occupancy framework describes the change of state across primary occasions or years. If a site of class A (resp. class B) is unoccupied at year k , it will become occupied at year $k+1$ with the colonization probability γ_A (resp. γ_B). On the opposite, if a site of class A (resp. class B) is occupied at year k , it will become unoccupied at year $k+1$ with the extinction probability ε_A (resp. ε_B) (Gimenez et al. 2014).

$$A = \begin{matrix} UA(k) & UB(k) & OA(k) & OB(k) \\ \begin{bmatrix} 1 - \gamma_A & 0 \\ 0 & 1 - \gamma_B \\ \varepsilon_A & 0 \\ 0 & \varepsilon_B \end{bmatrix} \end{matrix} \begin{matrix} UA(k+1) & UB(k+1) & OA(k+1) & OB(k+1) \\ \begin{bmatrix} \gamma_A & 0 \\ 0 & \gamma_B \\ 0 & 1 - \varepsilon_A \\ 0 & 1 - \varepsilon_B \end{bmatrix} \end{matrix}$$

Observation Events (in two steps):

Detection

The first step deals with detection. An observation made on an unoccupied site A (resp. B) is a false positive made with the probability p_{A10} (resp. p_{B10}). An observation made on an occupied site A (resp. B) is a true positive made with the probability p_{A11} (resp. p_{B11}).

$$P = \begin{matrix} & \textit{no detection} & \textit{true positive} & \textit{false positive} \\ \begin{matrix} UA \\ UA \\ OA \\ OB \end{matrix} & \begin{bmatrix} 1 - p_{A10} \\ 1 - p_{B10} \\ 1 - p_{A11} \\ 1 - p_{B11} \end{bmatrix} & \begin{bmatrix} 0 \\ 0 \\ p_{A11} \\ p_{B11} \end{bmatrix} & \begin{bmatrix} p_{A10} \\ p_{B10} \\ 0 \\ 0 \end{bmatrix} \end{matrix}$$

Classification

The second step deals with the classification of detections. If a true positive detection is made, it will be classified as unambiguous with probability b .

$$B = \begin{matrix} \textit{no detection} & 0 & 1 & 2 \\ \textit{true positive} & \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 - b \\ 1 \end{bmatrix} \\ \textit{false positive} & & & \end{matrix}$$

FITTING IN E-SURGE

The first step is to load the data:

Start » New session (preferably in the directory where your dataset is)

Data » Load data (Headed); select the dataset, here “data2013_wolf_12.txt” click on ‘select quantitative covariates’ if willing to include covariates. Here we selected ‘POscale’ and ‘altiscale’. Click on ‘validate’.

In the 'DATA' section in the main window, click the 'Modify' button and use 5 states and 1 age class.

The second step is to specify that we are in the occupancy framework:

To specify that we are in the occupancy framework, click the ‘Models’ button and specify ‘Markovian states only’ and select ‘Occupancy’.

The third step is to specify if we want to compute the confidence intervals of the models’ estimates:

In the 'Advanced Numerical' section in the main window, tick the 'Compute C-I (Hessian)' box to get confidence intervals

Then, the model specification procedure is decomposed into i) implementing the basic structural form of the matrices using the GEPAT interface, ii) setting linear model of each

parameter using the GEMACO interface and iii) fixing initial parameters using IVFV interface.

In the 'COMPUTE A MODEL' section in the main window, click on the 'Gepat' yellow button and use the following Matrix Patterns:

In the GEPAT module in E-SURGE, '*' entries denote the complement of the sum of positive row entries, and '-' entries denote zeroes. For the initial states matrices, the transition and event matrices introduced above, we have:

In the Gepat interface click on the window in front of 'Number of steps' and change the number to 2 for Initial state and Event.

Initial State (in two steps):

- Class affiliation vector

p *

- Initial state matrix

*	-	p	-
-	*	-	p

Transition

- Transition matrix

*	-	y	-	-
-	*	-	y	-
y	-	*	-	-
-	y	-	*	-
-	-	-	-	*

Event (in two steps):

- Detection matrix

*	-	b
*	-	b
*	b	-
*	b	-
*	-	-

- Classification matrix

*	-	-
-	b	*
-	-	*

Click Exit to go to the next step.

Annexes

In the GEMACO interface (Choquet 2008), predefined shortcuts are used to specify which parameters are to be time-constant, time-specific or state-specific (e.g. 'i' denotes constancy, 't' means time effect, 'from' means that parameters are not equal in each matrix row). In the model considered here, we used 'i' for initial parameters and 'from' for transitions which depend on the individual state.

In the 'COMPUTE A MODEL' section in the main window, click on the 'Gemaco' green button and use the following syntax in the 'Model definition' dialog box (in the right bottom corner):

GEMACO:

IS: step 1: i

Because we were interested in incorporating heterogeneity in the detection process only, we considered $\psi_{1A} = \psi_{1B}$ here with an effect of altitude as a covariate.

IS: step 2: i + altiscale

Because we were interested in static occupancy models here we considered $\varepsilon_A = \varepsilon_B = \gamma_A = \gamma_B$ and set their values to zero in the IVFV interface.

T: i

We wanted to get all detection probabilities different:

E: step : f

E: step 2: i

To check that the model is correctly implemented:

Gemaco » Call Gemaco (all phrases) or Ctr+G, then click Exit

In the 'COMPUTE A MODEL' section in the main window, click on the 'IVFV' pink button and specify parameters value if you want some to be fixed:

In the IVFV interface, for the Transition as we have one parameter, one needs to set it at 0:

$\beta_{\#1} = 0$ to remain in a static occupancy model.

In the 'COMPUTE A MODEL' section in the main window, click on the 'RUN' red button to fit the model to the simulated dataset.

When the dialog box pops up, modify the model name if needed, then click OK

In the 'Output' section of the main window, click on 'Selected Model Results (.out)' to get the results. More precisely, check out the 'Beta' section in the output file. The

eight lines below are organized as follows: π_i , ψ_i , effect of altitude on ψ_1 , p_{A10} , p_{B10} , p_{A11} , p_{B11} , and b . To check the likelihood estimate, check the 'deviance' section in the output file.

Note: Miller et al. (2011) developed a multiple detection method model to account for multiple detection methods employed on unique sampling occasions. The general HMM framework we introduce can be used with a multi-method model. The decomposition of the observation matrix is actually the same, only the implementation in program E-SURGE would differ. For instance, for the surveys during which only unambiguous data happen for sites of class A, we would have to fix p_{A10} and $p_{A11}*(1-b)$ to zero during those surveys. For surveys during which only ambiguous data happen then we would have to fix $b*p_{A11}$ to zero. Same for occasions for sites B.

REFERENCES:

- Choquet R (2008) Automatic generation of multistate capture recapture models. The Canadian Journal of Statistics 36: 43-57.
- Choquet R, Rouan L, Pradel R (2009) Program E-Surge: A Software Application for Fitting Multievent Models. In: Thomson DL, Cooch EG, Conroy MJ (eds) Modeling Demographic Processes In Marked Populations. Springer, New York, pp 845–865
- Choquet R, Nogue, E (2011) E-SURGE user's manual. CEFÉ, Montpellier, France.
- Gimenez, O., Blanc, L., Besnard, A., Pradel, R., Doherty, P.F.J., Marboutin, E. & Choquet, R. (2014) Fitting occupancy models with E-SURGE : hidden Markov modelling of presence – absence data. *Methods in Ecology and Evolution*, 5, 592–597.
- Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Weir, L.A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, 92, 1422–1428.

Appendix B. Mathematical formulation and instructions for implementing an occupancy model incorporating misidentification only with altitude as a covariate on the initial occupancy probability and sampling effort as a covariate on detection probabilities in program E-SURGE.

Occupancy models dealing with misidentification and heterogeneity can be implemented in Program E-SURGE (Choquet et al. 2009) which is freely downloadable at <https://www.cefe.cnrs.fr/fr/actus/livres/34-french/recherche/bc/bbp/264-logiciels>. Below we provide the matrix patterns of state and event transitions for the occupancy model accounting for misidentification only developed in this paper. Statistical and software details can be found in E-SURGE manual (Choquet and Nogue 2011).

We considered 2 states: unoccupied and occupied plus the state dead in E-SURGE. We also considered 3 events: not detected (0), detected unambiguous (1) and detected ambiguous (2).

We defined the initial state vector Ψ , the transition matrix A , the detection matrix P and the classification matrix B . The parameter ψ_1 denotes the initial occupancy probability. The parameter γ is the probability that an unoccupied site at time k becomes occupied at time $k+1$ and ε the probability that an occupied site at time k becomes unoccupied at time $k+1$. The parameter p_{10} is the probability of incorrectly detecting the species at an unoccupied site and p_{11} is the probability correctly detecting the species at an occupied site. Finally, the parameter b is the probability to classify a true detection as unambiguous.

Based, on those parameters, we have:

Initial States:

Initial state

At the initial step sites can be occupied with the probability ψ_1

$$\Psi = \begin{matrix} \textit{unoccupied} & \textit{occupied} \\ [1 - \psi_1 & \psi_1] \end{matrix}$$

State Transitions

The transition step in occupancy framework describes the change of state across primary occasions or years. If a site is unoccupied at year k , it will become occupied at year $k+1$ with the colonization probability γ . On the opposite if a site is occupied at year k , it will become unoccupied at year $k+1$ with the extinction probability ε (Gimenez et al. 2014).

$$A = \begin{matrix} & \begin{matrix} \text{unoccupied at } k+1 & \text{occupied at } k+1 \end{matrix} \\ \begin{matrix} \text{unoccupied at } k \\ \text{occupied at } k \end{matrix} & \begin{bmatrix} 1-\gamma & \gamma \\ \varepsilon & 1-\varepsilon \end{bmatrix} \end{matrix}$$

Observation Events (in two steps):

Detection

The first step deals with detection. An observation made on an unoccupied site is a false positive made with the probability p_{10} . An observation made on an occupied site is a true positive made with the probability p_{11} .

$$P = \begin{matrix} & \begin{matrix} \text{no detection} & \text{true positive detection} & \text{false positive detection} \end{matrix} \\ \begin{matrix} \text{unoccupied} \\ \text{occupied} \end{matrix} & \begin{bmatrix} 1-p_{10} & 0 & p_{10} \\ 1-p_{11} & p_{11} & 0 \end{bmatrix} \end{matrix}$$

Classification

The second step deals with the classification of detections. If a true positive detection is made, it will be classified as unambiguous with probability b .

$$B = \begin{matrix} & \begin{matrix} \text{no detection} \\ \text{true positive} \\ \text{false positive} \end{matrix} & \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 0 & b & 1-b \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

FITTING IN E-SURGE

The first step is to load the data:

Start » New session (preferably in the directory where your dataset is)

Data » Load data (Headed); select the dataset, here “data2013_wolf_12.txt” click on ‘select quantitative covariates’ if willing to include covariates. Here we selected ‘POscale’ and ‘altiscale’. Click on ‘validate’.

In the 'DATA' section in the main window, click the 'Modify' button and use 5 states and 1 age class.

Annexes

The second step is to specify that we are in the occupancy framework:

To specify that we are in the occupancy framework, click the 'Models' button and specify 'Markovian states only' and select 'Occupancy'.

The third step is to specify if we want to compute the confidence intervals of the models' estimates:

In the 'Advanced Numerical' section in the main window, tick the 'Compute C-I (Hessian)' box to get confidence intervals

Then, the model specification procedure is decomposed into i) implementing the basic structural form of the matrices using the GEPAT interface, ii) setting linear model of each parameter using the GEMACO interface and iii) fixing initial parameters using IVFV interface .

In the 'COMPUTE A MODEL' section in the main window, click on the 'Gepat' yellow button and use the following Matrix Patterns:

In the GEPAT module in E-SURGE, '*' entries denote the complement of the sum of positive row entries, and '-' entries denote zeroes. For the initial states matrices, the transition and event matrices introduced above, we have:

Initial State:

- Initial state vector
* p

Transition

- Transition matrix
* y -
y * -
- - -

Event(in two steps):

- Detection matrix
* - b
* b -
* - -

Annexes

- Classification matrix

*	-	-
-	b	*
-	-	*

Click Exit to go to the next step.

In the GEMACO interface (Choquet 2008), predefined shortcuts are used to specify which parameters are to be time-constant, time-specific or state-specific (e.g. ‘i’ denotes constancy, ‘t’ means time effect, ‘from’ means that parameters are not equal in each matrix row. In the model considered here, we used ‘i’ for initial parameters and ‘from’ for transitions which depend on the individual state.

In the 'COMPUTE A MODEL' section in the main window, click on the 'Gemaco' green button and use the following syntax in the 'Model definition' dialog box (in the right bottom corner):

GEMACO:

We wanted to test for the effect of altitude on the initial occupancy probability:

IS: i + altiscale

Because we were interested in static occupancy models here we considered $\varepsilon = \gamma$ and set their values to zero in the IVFV interface.

T: i

We wanted to get all detection probabilities different with sampling effort as a covariate on both detection probabilities:

E: step : f.[i+POscale]

E: step 2: i

To check that the model is correctly implemented:

Gemaco » Call Gemaco (all phrases) or Ctr+G, then click Exit

In the 'COMPUTE A MODEL' section in the main window, click on the 'IVFV' pink button and specify parameters value if you want some to be fixed:

In the IVFV interface, for the Transition as we have one parameter, one needs to set it at 0:

$\beta_{\#1} = 0$ to remain in a static occupancy model.

In the 'COMPUTE A MODEL' section in the main window, click on the 'RUN' red button to fit the model to the simulated dataset.

When the dialog box pops up, modify the model name if needed, then click OK

In the 'Output' section of the main window, click on 'Selected Model Results (.out)' to get the results. More precisely, check out the 'Beta' section in the output file. The eight lines below are organized as follows: ψ , effect of altitude on ψ_1 , p_{10} , p_{11} , b , the effect of sampling effort on p_{10} and the effect of sampling effort on p_{11} . To check the likelihood estimate, check the 'deviance' section in the output file.

REFERENCES:

- Choquet R (2008) Automatic generation of multistate capture recapture models. The Canadian Journal of Statistics 36: 43-57.
- Choquet R, Rouan L, Pradel R (2009) Program E-Surge: A Software Application for Fitting Multievent Models. In: Thomson DL, Cooch EG, Conroy MJ (eds) Modeling Demographic Processes In Marked Populations. Springer, New York, pp 845–865
- Choquet R, Nogue, E (2011) E-SURGE user's manual. CEFE, Montpellier, France.
- Gimenez, O., Blanc, L., Besnard, A., Pradel, R., Doherty, P.F.J., Marboutin, E. & Choquet, R. (2014) Fitting occupancy models with E-SURGE : hidden Markov modelling of presence – absence data. Methods in Ecology and Evolution, 5, 592–597.

Annexes

Appendix E: Results of the simulation study to assess the performance of the misidentification occupancy model accounting for heterogeneity (MMH) vs. the model without heterogeneity (MMO). The first column corresponds to the heterogeneity coefficient calculated with π the proportion of sites of class A and p_{A11} the probability of making true positive on sites of class A and p_{B11} the probability of making true positive on sites of class B. The differences of AIC (ΔAIC) between the two models, obtained from 200 simulations for each scenario, are provided. Estimation accuracy (RMSE) and relative bias for the occupancy probability are both provided as measures of model performance.

Heterogeneity coefficient	π	p_{A11}	p_{B11}	ΔAIC	Sd (ΔAIC)	RMSE(ψ_1) MMH	RMSE(ψ_1) MMO	Relative bias(ψ_1) MMH	Relative bias(ψ_1) MMO
0.24	0.2	0.1	0.7	81.49	19.54	0.04	0.08	0.12	-8.41
0.49	0.5	0.1	0.8	252.99	30.79	0.03	0.09	1.14	-10.91
0.53	0.8	0.1	0.9	305.96	37.96	0.05	0.06	-1.96	-3.12

Annexes

Appendix F: Results of the simulation study to assess the performance of the misidentification occupancy model accounting for heterogeneity (MMH) vs. the model without heterogeneity (MMO) with 3 occasions. The first column corresponds to the heterogeneity coefficient calculated with π the proportion of sites of class A and p_{A10} the probability of making true positive on sites of class A and p_{B10} the probability of making true positive on sites of class B. The differences of AIC (ΔAIC) between the two models, obtained from 200 simulations for each scenario, are provided. Estimation accuracy (RMSE) and relative bias for the occupancy probability are both provided as measures of model performance.

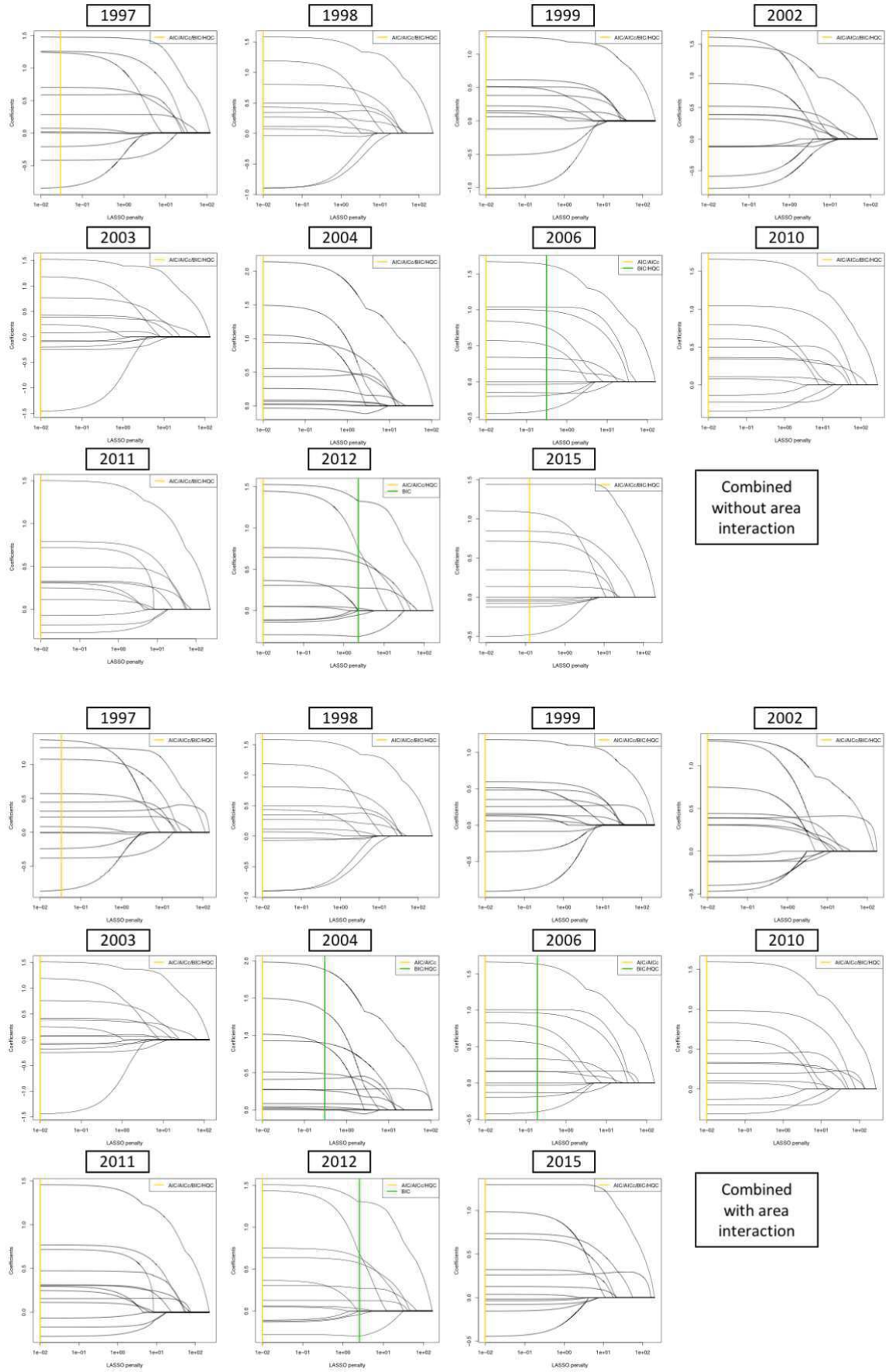
Heterogeneity coefficient	π	p_{A10}	p_{B10}	ΔAIC	Sd (ΔAIC)	RMSE(ψ_1) MMH	RMSE(ψ_1) MMO	Relative bias(ψ_1) MMH	Relative bias(ψ_1) MMO
0.24	0.2	0.1	0.7	-4.65	1.84	0.12	0.07	9.96	7.28
0.49	0.5	0.1	0.8	-3.33	2.95	0.14	0.09	15.61	10.24
0.53	0.8	0.1	0.9	-2.98	2.49	0.10	0.08	7.04	4.79

Annexes Article 4

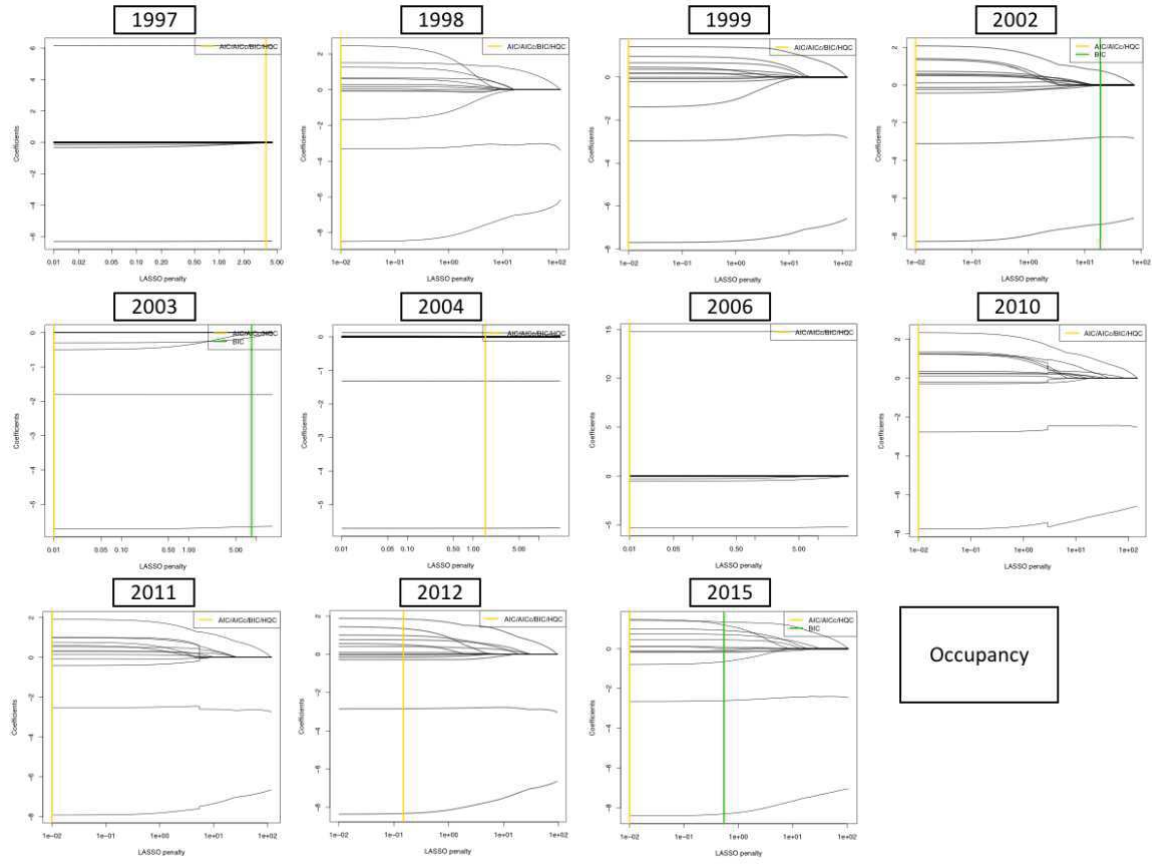
Appendix 1: Comparison of Pearson residuals for the PPM only model not accounting for spatial dependence and the PPM only accounting for spatial dependence.

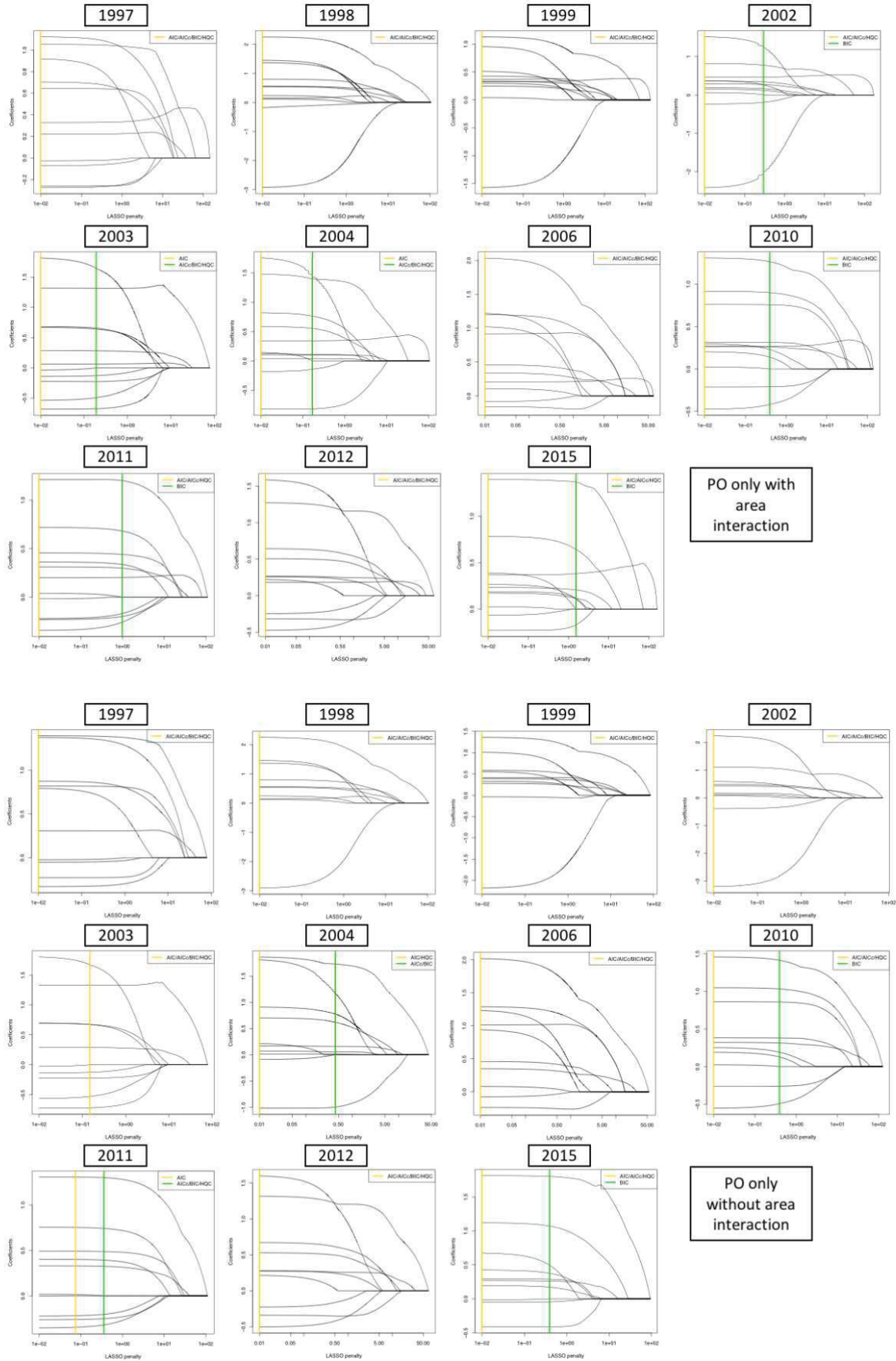
Year	Without area interaction				With area interaction				
	Sum of Pearson residuals in entire window	Area of entire window	Quadrature area	Range of smoothed field	Area interaction radius	Sum of Pearson residuals in clipped window	Area of clipped window	Quadrature area	Range of smoothed field
1997	-329.9	350000	350000	-0.004285 : 0.004462	7	-359.5	349100	349000	-0.004487 : 0.00429
1998	-143.3	350000	350000	-0.004939 : 0.004263	1	-58.31	349100	349100	-0.005336 : 0.008158
1999	-40.34	350000	350000	-0.004384 : 0.003941	5	-23.47	349100	349100	-0.004407 : 0.004018
2002	-61.9	350000	350000	-0.006553 : 0.006637	6	-5.188	349100	349100	- 0.006321 : 0.0068
2003	247.5	350000	350000	-0.004952 : 0.006742	3	-153	349100	349000	-0.005359 : 0.005097
2004	-232.5	350000	350000	-0.004583 : 0.002849	7	-354.3	349100	349100	-0.004168 : 0.002875
2006	-250	350000	350000	-0.00501 : 0.005936	5	-390.3	349100	349000	-0.005833 : 0.006349
2010	-129.4	350000	350000	-0.005193 : 0.006978	5	-315.4	349100	348900	-0.005315 : 0.006617
2011	-130.5	350000	350000	-0.00508 : 0.007895	6	-465.2	349100	349100	-0.005674 : 0.004231
2012	-179.1	350000	350000	-0.0037 : 0.005036	5	-220.9	349100	349000	-0.003959 : 0.00574
2015	380.3	350000	350000	-0.003725 : 0.008094	10	-120.1	349100	349000	-0.003961 : 0.005246

Appendix 2: Variable selection using the LASSO technique (Renner et al. 2015) for all the years. The Poisson PPM likelihood is augmented with a penalty function. On the plot, each curve represents a variable in the model. The x axis is a function of the regularization penalty parameter. The y axis gives the value of the variables. The vertical line shows, depending on the selection criteria, the optimal value of the penalty parameter and the values of the selected variables.

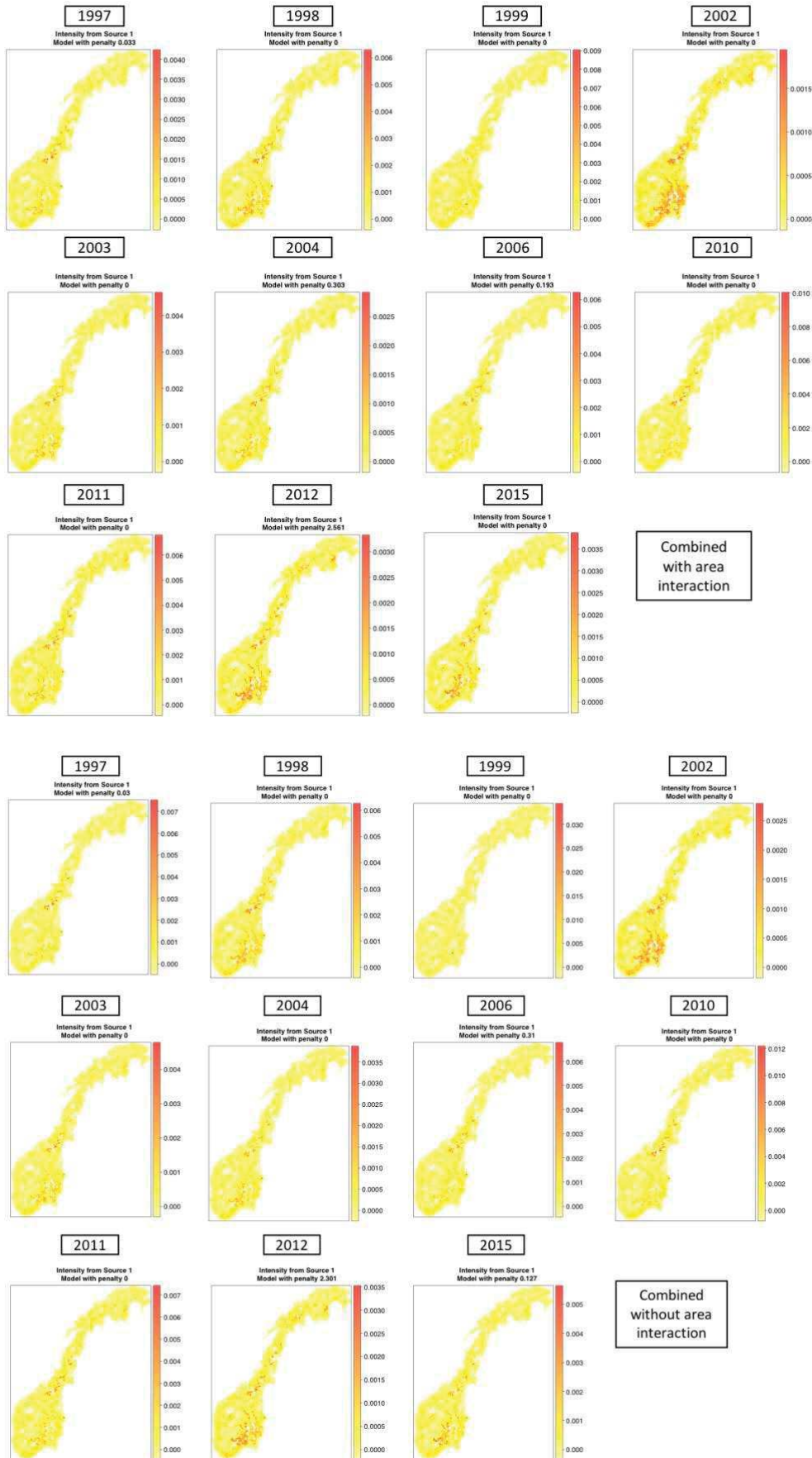


Annexes

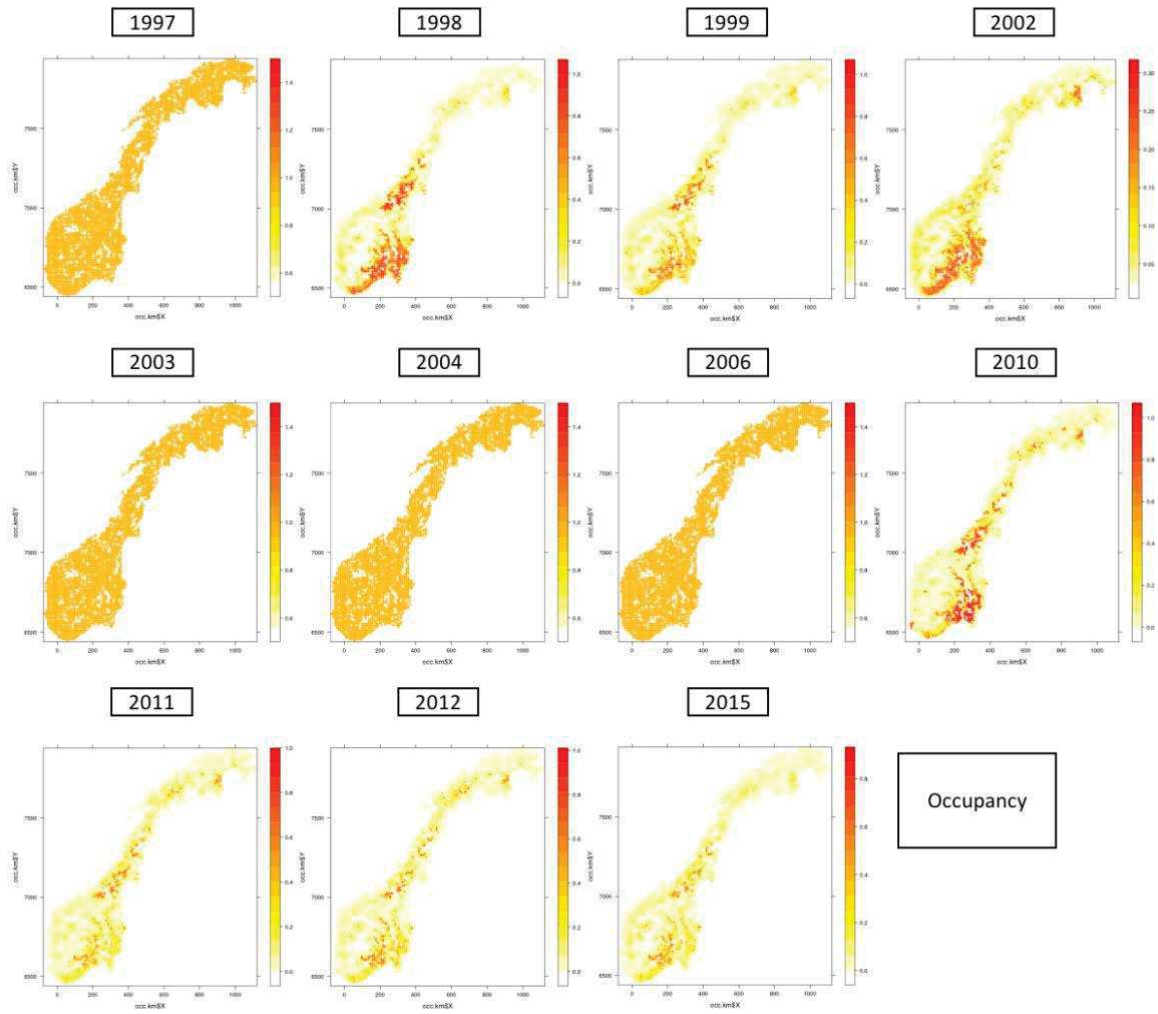


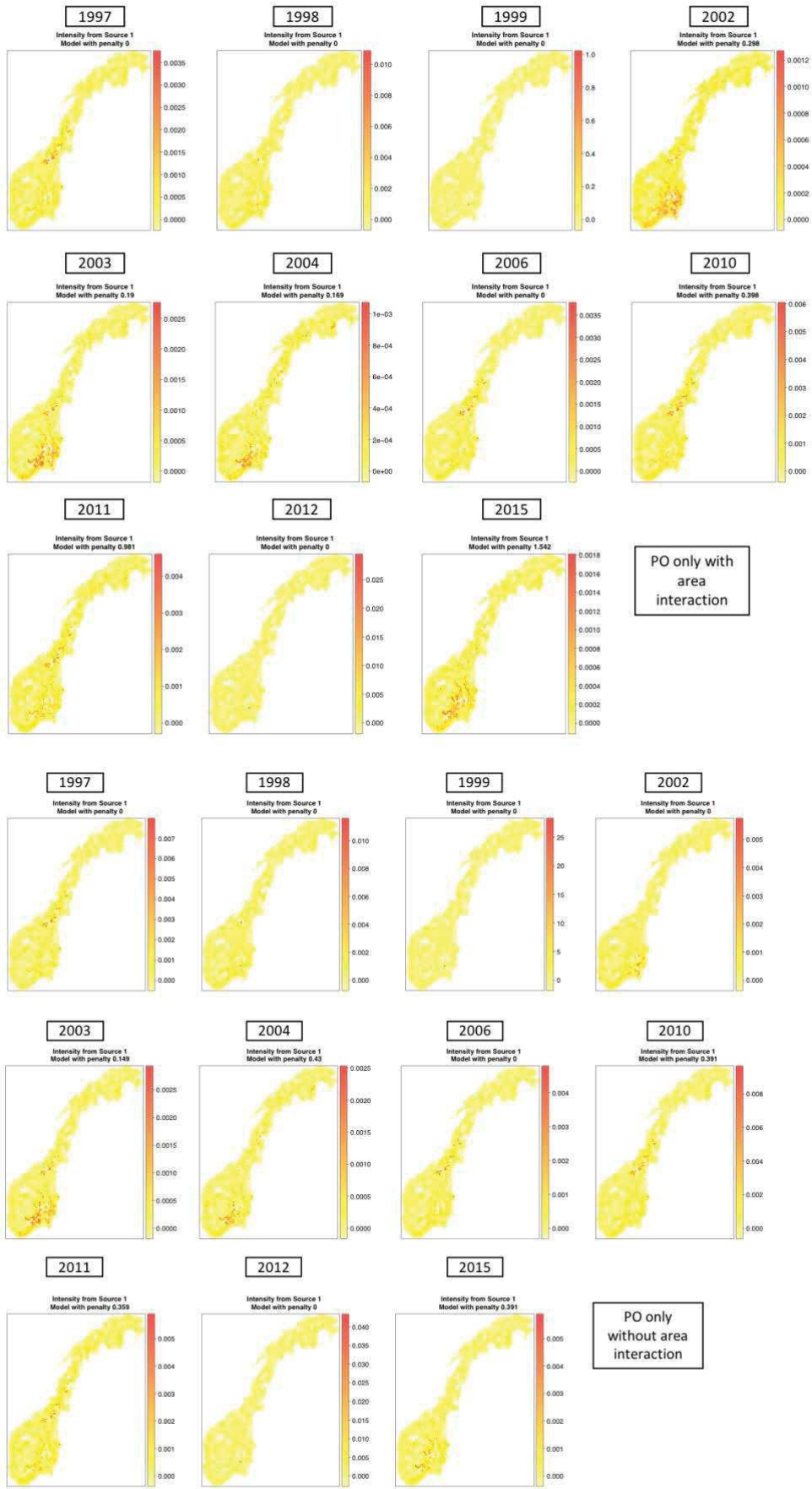


Appendix 3: Intensity maps from the combined model accounting and not accounting for spatial dependence for all the years, probability of occupancy maps for all the years and intensity maps from the PPM model accounting and not accounting for spatial dependence for all the years



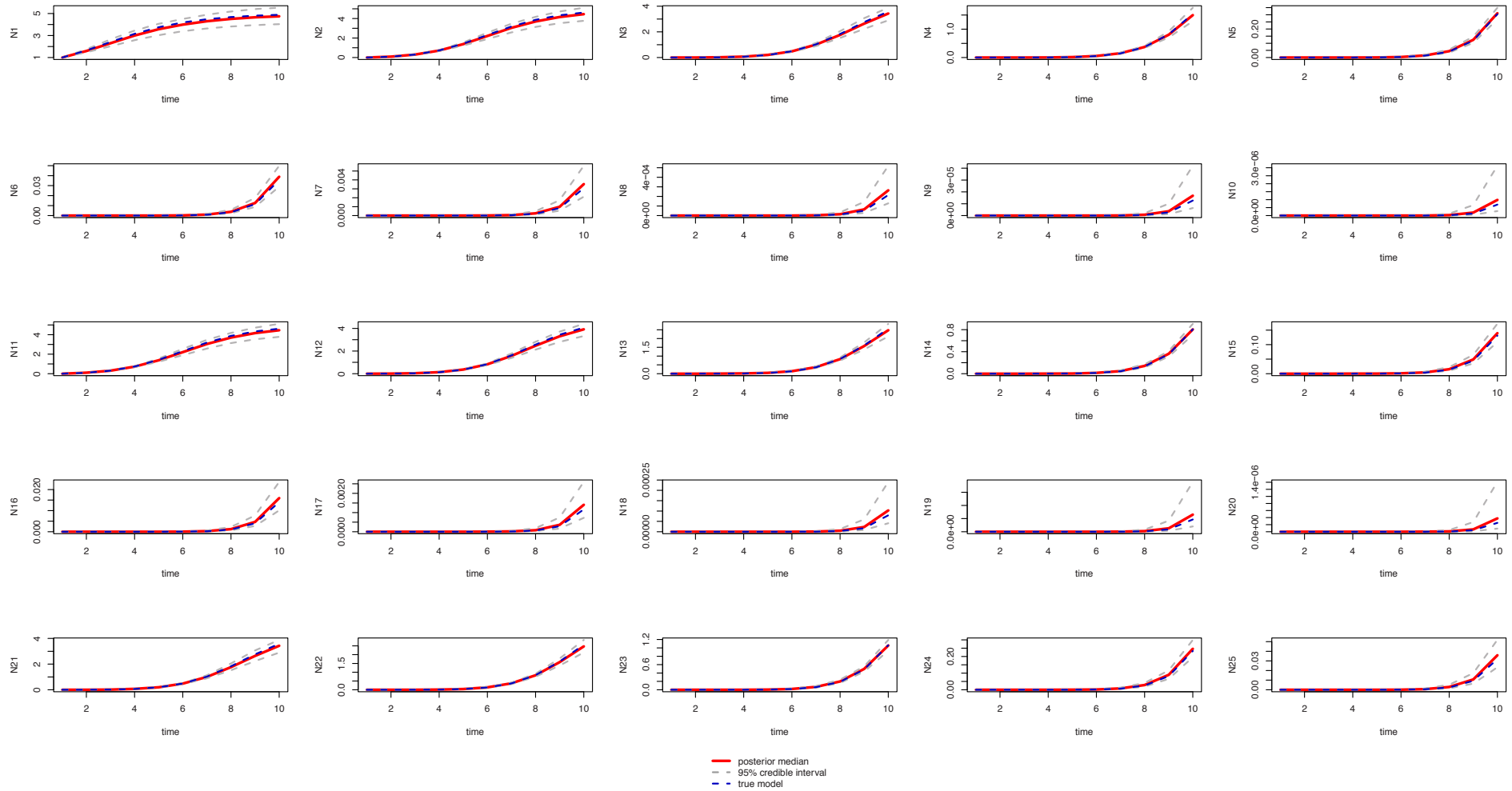
Annexes





Annexes Article 5

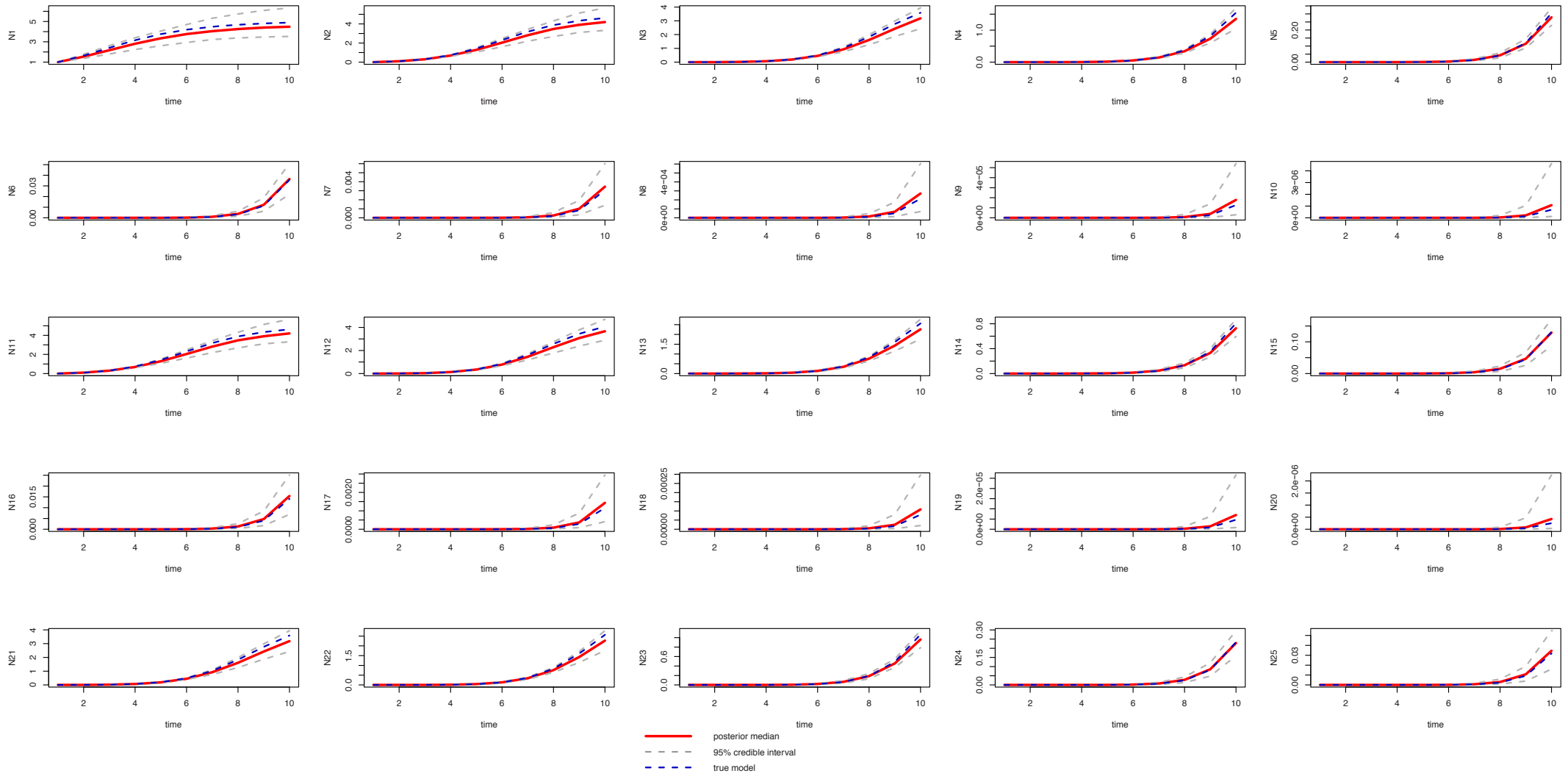
Appendix 1: Estimated abundance evolution for 10 years from the posterior median (red solid line) and the 95 % credible intervals (grey dashed line) in comparison with the true abundance (blue dashed line) for the first 25 sites in the two “high resolution” scenarios and the 16 sites in the two “low resolution” scenarios.



HR-HD

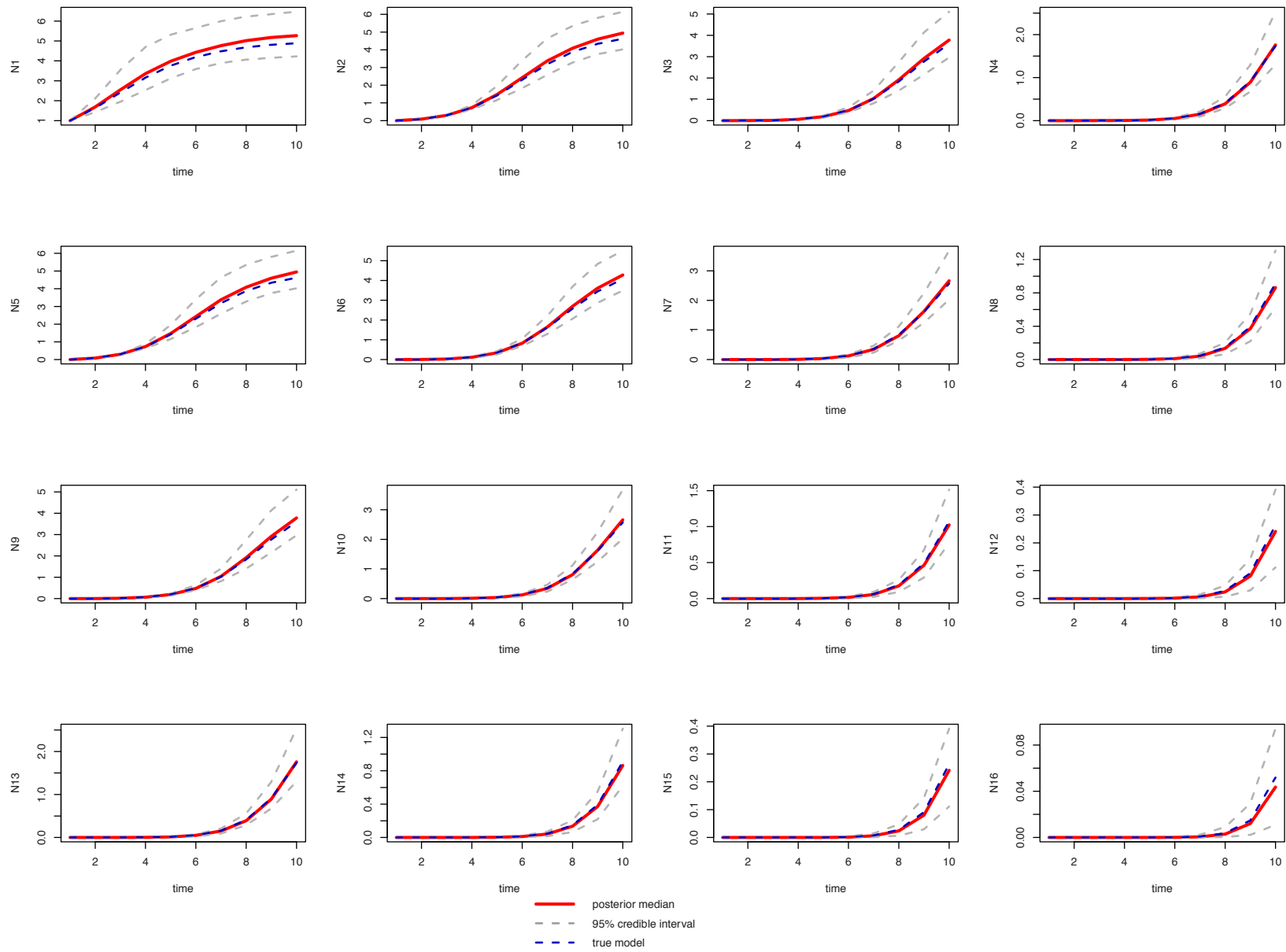
Annexes

HR-LD



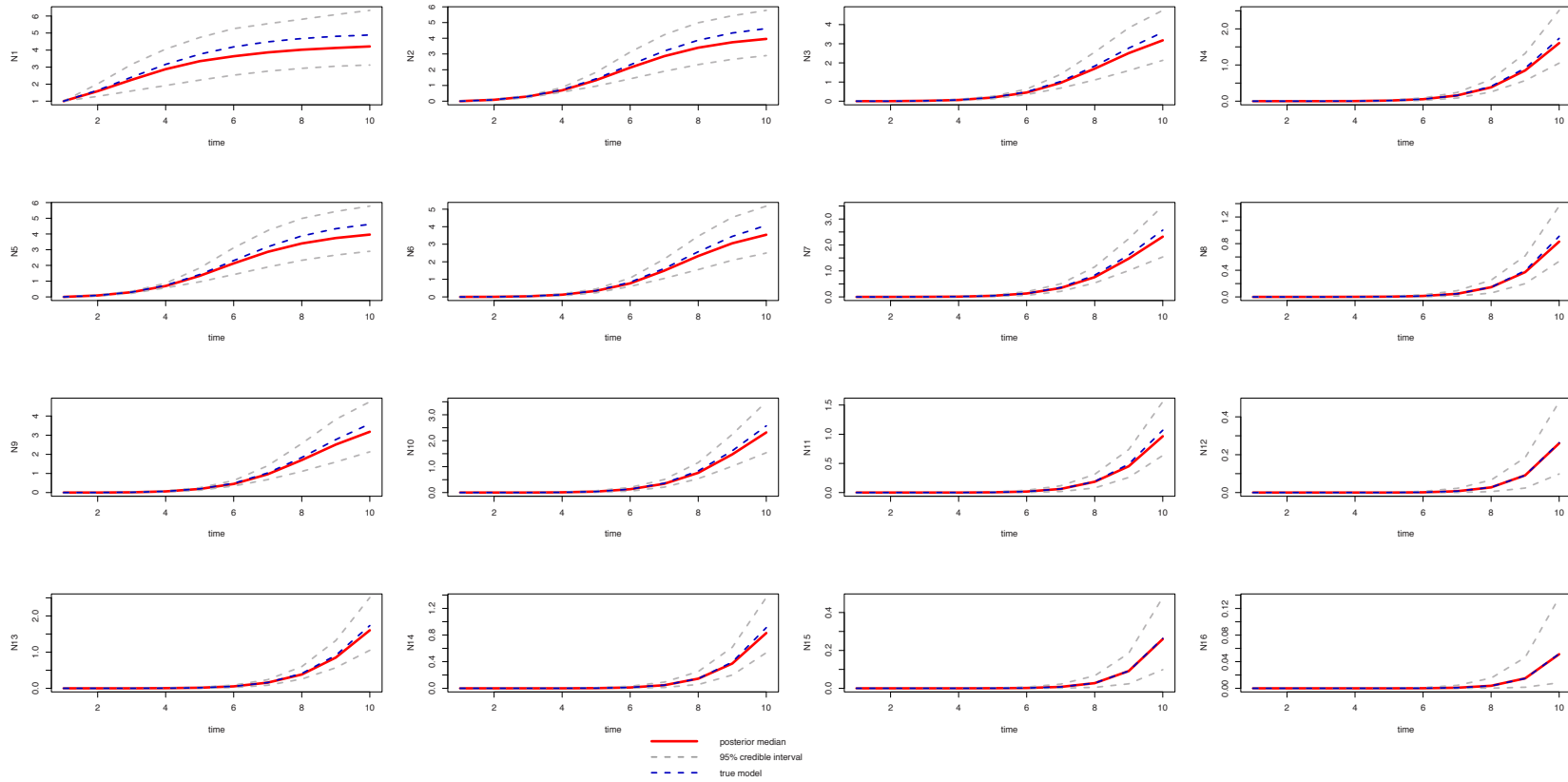
Annexes

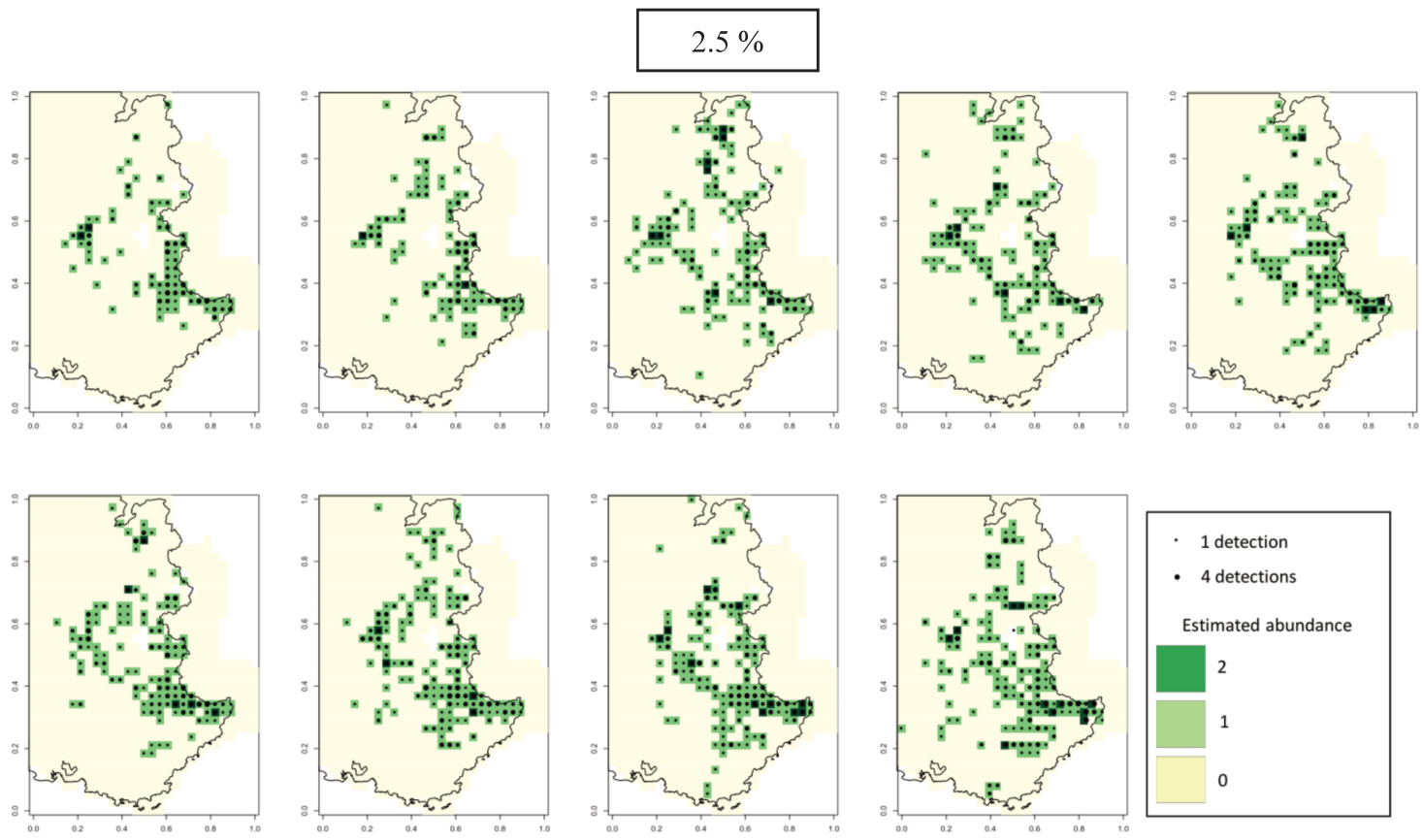
LR-HD



Annexes

LR-LD





Appendix 2: Maps of the quantiles of the estimated abundance of wolves per site in South-East France between 2007 and 2015. Black dots represent detections during a year.

Appendix 3: Median and 95% credible interval of the yearly wolf abundance estimates in France between 2007 and 2015 in South-Eastern France.

	2.50%	50%	97.50%
Year	Estimated abundance		
2007	81.0	121	313.0
2008	95.0	137	330.0
2009	134.0	180	441.0
2010	126.0	177	488.8
2011	124.0	179	566.8
2012	127.0	198	630.8
2013	140.0	228	710.0
2014	146.2	245	780.0
2015	159.0	265	842.0

Appendix 4: Median and 95% credibility intervals for the parameters and the effects of ecological variables on wolf distribution dynamics between 2007 and 2015 in South-Eastern France. A * sign indicates that the credible interval of the corresponding parameter does not contain zero.

	2.50%	50%	97.50%
Species-level detectability q			
Intercept*	-2.19	-2.03	-1.86
Linear effect of sampling effort*	0.30	0.39	0.47
Quadratic effect of sampling effort*	-1.05	-0.94	-0.84
Logistic growth rate R			
Intercept*	-1.67	-1.04	-0.42
Linear effect of forest cover*	2.60	3.43	4.49
Quadratic effect of forest cover	-0.90	0.09	1.03
Carrying capacity K			
Intercept*	8.88×10^{-2}	9.90×10^{-2}	1.10×10^{-2}
Diffusion parameter D			
Intercept	-0.11	1.16	2.89
Linear effect of human density	-1.34	0.46	1.89
Quadratic effect of human density*	0.18	1.17	2.76

Annexes

Les scripts utilisés pour les articles 2 et 3 peuvent se trouver en suivant ce lien :
<https://github.com/JulieLouvrier/appendices-PhD>

Références

- Andrén H, Linnell JDC, Liberg O, Ahlqvist P, Andersen R, Danell A, Franzén R, Kvam T, Odden J, Segerström P. 2002. Estimating total lynx *Lynx lynx* population size from censuses of family groups. *Wildlife Biology* 8:299–306.
- Bailey LL, Hines JE, Nichols JD, MacKenzie DI. 2007. Sampling Design Trade-Offs in Occupancy Studies with Imperfect Detection: Examples and Software. *Ecological Applications* 17:281–290.
- Bailey LL, Mackenzie DI, Nichols JD. 2014. Advances and applications of occupancy models. *Methods in Ecology and Evolution* 5:1269–1279.
- Basille M, Herfindal I, Santin-Janin H, Linnell JDC, Odden J, Andersen R, Arild Høgda K, Gaillard J-M. 2009. What shapes Eurasian lynx distribution in human dominated landscapes: selecting prey or avoiding people? *Ecography* 32:683–691.
- Basille M, Van Moorter B, Herfindal I, Martin J, Linnell JDC, Odden J, Andersen R, Gaillard J-M. 2013. Selecting Habitat to Survive: The Impact of Road Density on Survival in a Large Carnivore. *PLoS ONE* 8:e65493.
- Beirne C, Lambin X. 2013. Understanding the Determinants of Volunteer Retention Through Capture-Recapture Analysis: Answering Social Science Questions Using a Wildlife Ecology Toolkit. *Conservation Letters* 6:391–401.
- Bischof R, Brøseth H, Gimenez O. 2016. Wildlife in a Politically Divided World: Insularism Inflates Estimates of Brown Bear Abundance. *Conservation Letters* 9:122–130.
- Bischof R, Nilsen EB, Brøseth H, Männil P, Ozoliņš J, Linnell JDC. 2012. Implementation uncertainty when using recreational hunting to manage carnivores: Carnivore hunting quotas. *Journal of Applied Ecology* 49:824–832.
- Boitani L. 1995. Ecological and cultural diversities in the evolution of wolf-human relationship. *Ecology and conservation of wolves in a changing world*:3–11.
- Bouyer Y, Gervasi V, Poncin P, Beudels-Jamar RC, Odden J, Linnell JDC. 2015. Tolerance to anthropogenic disturbance by a large carnivore: the case of Eurasian lynx in south-eastern Norway: Lynx tolerance to human disturbance in Norway. *Animal Conservation* 18:271–278.
- Breiner FT, Bergamini A. 2018. Improving the estimation of area of occupancy for IUCN Red List assessments by using a circular buffer approach. *Biodiversity and Conservation* 27:2443–2448.
- Breitenmoser U. 1998. Large predators in the alps: the fall and rise of man's competitors. *Biological Conservation* 83:279–289.
- Breitenmoser U, al. 2000. Action plan for the conservation of the Eurasian lynx in Europe (*Lynx lynx*). Council of Europe.
- Chapron G et al. 2014. Recovery of large carnivores in Europe's modern human-dominated landscapes. *Science* 346:1517–1519.
- Cohn JP. 2008. Citizen Science: Can Volunteers Do Real Research? *BioScience* 58:192–197.
- Cubaynes S et al. 2010. Importance of Accounting for Detection Heterogeneity When

Références

- Estimating Abundance: the Case of French Wolves. *Conservation Biology* 24:621–626.
- Dickinson JL, Shirk J, Bonter D, Bonney R, Crain RL, Martin J, Phillips T, Purcell K. 2012. The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment* 10:291–297.
- Directive H. 1992. Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Union* 206:7–50.
- Dorazio RM. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography* 23:1472–1484.
- Duchamp C et al. 2012. A dual frame survey to assess time- and space-related changes of the colonizing wolf population in France. *Hystrix* 23:14–28.
- Duchamp C, Chapron G, Gimenez O, Robert A, Sarrazin F, Beudels-Jamar R, Le Maho Y. 2017. Expertise collective scientifique sur la viabilité et le devenir de la population de loups en France à long terme sous la coordination ONCFS-MNHN de: Guinot-Ghestem M. Haffner P., Marboutin É., Rousset G., Savouret-Soubelet A., Sibley JP, Trudelle L.(par or. alph.).
- Elith J et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- Elith J, Leathwick JR. 2009. Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual review of ecology, evolution, and systematics* 40:677–697.
- Eriksson T, Dalerum F. 2018. Identifying potential areas for an expanding wolf population in Sweden. *Biological Conservation* 220:170–181.
- Fabbri E et al. 2007. From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. *Molecular Ecology* 16:1661–1671.
- Ford AT. 2017. Opportunities and challenges for the study and conservation of large carnivores. *Food Webs* 12:107–109.
- Gaston KJ, Blackburn TM, Greenwood JJ, Gregory RD, Quinn RM, Lawton JH. 2000. Abundance–occupancy relationships. *Journal of Applied Ecology* 37:39–59.
- Gervasi V, Brøseth H, Gimenez O, Nilsen EB, Odden J, Flagstad Ø, Linnell JDC. 2016. Sharing data improves monitoring of trans-boundary populations: the case of wolverines in central Scandinavia. *Wildlife Biology* 22:95–106.
- Gervasi V, Nilsen EB, Odden J, Bouyer Y, Linnell JDC. 2014. The spatio-temporal distribution of wild and domestic ungulates modulates lynx kill rates in a multi-use landscape: Lynx prey selection in a multi-use landscape. *Journal of Zoology* 292:175–183.
- Gimenez O, Blanc L, Besnard A, Pradel R, Doherty PF, Marboutin E, Choquet R. 2014. Fitting occupancy models with E-SURGE: Hidden Markov modelling of presence-absence data. *Methods in Ecology and Evolution* 5:592–597.
- Gimenez O, Lebreton J, Gaillard J, Choquet R, Pradel R. 2012. Estimating demographic parameters using hidden process dynamic models. *Theoretical Population Biology* 82:307–316.
- Godet L. 2010. La « nature ordinaire » dans le monde occidental. *Espace géographique* 39:295–308.

Références

- Grinnell J. 1917. The Niche-Relationships of the California Thrasher. *The Auk* 34:427–433.
- Guillera-Arroita G, Lahoz-Monfort JJ. 2012. Designing studies to detect differences in species occupancy: power analysis under imperfect detection: Power analysis for occupancy studies. *Methods in Ecology and Evolution* 3:860–869.
- Guillera-arroita G, Lahoz-monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, Mccarthy MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose ? Matching data and models to applications. *Global Ecology and Biogeography* 24:276–292.
- Guisan A, Thuiller W, Zimmermann NE. 2017. *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.
- Hefley TJ, Hooten MB. 2016. Hierarchical Species Distribution Models. *Current Landscape Ecology Reports* 1:87–97.
- Herfindal I, Linnell JDC, Odden J, Nilsen EB, Andersen R. 2005. Prey density, environmental productivity and home-range size in the Eurasian lynx (*Lynx lynx*). *Journal of Zoology* 265:63–71.
- Hutchinson G. 1957. The multivariate niche. Pages 415–421. *Cold Spring Harbor Symposia on Quantitative Biology*.
- IUCN. 2012. *IUCN Red List Categories and Criteria: version 3.1, 2nd edn*. Gland, Switzerland and Cambridge, UK.
- Jedrzejewski W et al. 2018. Estimating large carnivore populations at global scale based on spatial predictions of density and distribution – Application to the jaguar (*Panthera onca*). *PLoS ONE* 13:e0194719.
- Jeschke JM, Strayer DL. 2006. Determinants of vertebrate invasion success in Europe and North America. *Global Change Biology* 12:1608–1619.
- Kaczensky P. 1999. Large Carnivore Depredation on Livestock in Europe. *Ursus* 11:59–71.
- Kearney M, Porter W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* 12:334–350.
- Kéry M. 2011. Towards the modelling of true species distributions. *Journal of Biogeography* 38:617–618.
- Kéry M, Dorazio RM, Soldaat L, Van Strien A, Zuiderwijk A, Royle JA. 2009. Trend estimation in populations with imperfect detection. *Journal of Applied Ecology* 46:1163–1172.
- Kéry M, Schaub M. 2011. *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press.
- Koshkina V, Wang Y, Gordon A, Dorazio RM, White M, Stone L. 2017. Integrated species distribution models : combining presence-background data and site-occupancy data with imperfect detection. *Methods in Ecology and Evolution* 8:420–430.
- Kosmala M, Wiggins A, Swanson A, Simmons B. 2016. Assessing data quality in citizen science. *Frontiers in Ecology and the Environment* 14:551–560.
- Kuijper DPJ, Sahlén E, Elmhagen B, Chamaillé-Jammes S, Sand H, Lone K, Cromsigt JPM. 2016. Paws without claws? Ecological effects of large carnivores in anthropogenic landscapes. *Proceedings of the Royal Society B: Biological Sciences* 283:20161625.

Références

- Lahoz-Monfort JJ, Guillera-arroita G, Wintle BA. 2014. Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography* 23:504–515.
- Layman CA et al. 2015. A primer on the history of food web ecology: Fundamental contributions of fourteen researchers. *Food Webs* 4:14–24.
- Leopold A, Sowls LK, Spencer DL. 1947. A Survey of Over-Populated Deer Ranges in the United States. *The Journal of Wildlife Management* 11:162.
- Lescureux N, Linnell JDC. 2014. Warring brothers: The complex interactions between wolves (*Canis lupus*) and dogs (*Canis familiaris*) in a conservation context. *Biological Conservation* 171:232–245.
- Lewis MA, White KAJ, Murray JD. 1997. Analysis of a model for wolf territories. *Journal of Mathematical Biology* 35:749–774.
- Linden DW, Fuller AK, Royle JA, Hare MP. 2017. Examining the occupancy-density relationship for a low-density carnivore. *Journal of Applied Ecology* 54:2043–2052.
- Linnell JDC, Boitani L. 2012. Building biological realism into wolf management policy: The development of the population approach in Europe. *Hystrix* 23:80–91.
- Linnell JDC, Breitenmoser U, Breitenmoser-Wrsten C, Odden J, von Arx M. 2009. Recovery of Eurasian Lynx in Europe: What Part has Reintroduction Played? Pages 72–91 in M. W. Hayward and M. J. Somers, editors. *Reintroduction of Top-Order Predators*. Wiley-Blackwell, Oxford, UK.
- Linnell JDC, Broseth H, Odden J, Nilsen EB. 2010. Sustainably Harvesting a Large Carnivore? Development of Eurasian Lynx Populations in Norway During 160 Years of Shifting Policy. *Environmental Management* 45:1142–1154.
- Linnell JDC, Cretois B. 2018. The revival of wolves and other large predators and its impact on farmers and their livelihood in rural regions of Europe. Research for AGRI committee. European Parliament, Policy Department for Structural and Cohesion Policies, Brussels.
- Linnell JDC, Fiske P, Odden J, Brøseth H, Herfindal I, Andersen R. 2007. An evaluation of structured snow-track surveys to monitor Eurasian *lynx Lynx lynx* populations. *Wildlife Biology* 13:456–466.
- MacKenzie DI. 2006. Modeling the probability of resource use: The effect of, and dealing with, detecting a species imperfectly. *Journal of Wildlife Management* 70:367–374.
- MacKenzie DI, Nichols JD. 2004. Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation* 27:461–467.
- MacKenzie DI, Nichols JD, Hines JE, Knutson MG, Franklin AB. 2003. Estimating occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle JA, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. 2006. *Occupancy Estimation and Modeling*. Academic Press.
- MacKenzie DI, Royle JA. 2005. Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology* 42:1105–1114.

Références

- Madsen J, Williams JH, Johnson FA, Tombre IM, Dereliev S, Kuijken E. 2017. Implementation of the first adaptive management plan for a European migratory waterbird population: The case of the Svalbard pink-footed goose *Anser brachyrhynchus*. *Ambio* 46:275–289.
- Marboutin E, Pruszek M, Calenge C, Duchamp C. 2011. On the effects of grid size and shape when mapping the distribution range of a recolonising wolf (*Canis lupus*) population. *European Journal of Wildlife Research* 57:457–465.
- Marescot L, Chapron G, Chadès I, Fackler PL, Duchamp C, Marboutin E, Gimenez O. 2013. Complex decisions made simple: a primer on stochastic dynamic programming. *Methods in Ecology and Evolution* 4:872–884.
- Marucco F, Mcintire EJB. 2010. Predicting spatio-temporal recolonization of large carnivore populations and livestock depredation risk : wolves in the Italian Alps. *Journal of Applied Ecology* 47:789–798.
- Mattisson J, Odden J, Nilsen EB, Linnell JDC, Persson J, Andrén H. 2011. Factors affecting Eurasian lynx kill rates on semi-domestic reindeer in northern Scandinavia: Can ecological research contribute to the development of a fair compensation system? *Biological Conservation* 144:3009–3017.
- Mckinley DC et al. 2017. Citizen science can improve conservation science , natural resource management , and environmental protection. *Biological Conservation* 208:15–28.
- Mech LD, Boitani L. 2010. *Wolves: behavior, ecology, and conservation*. University of Chicago Press.
- MEDEEM M. 2008. National Wolf Action Plan in the French context of substantial and traditional livestock farming. *Ministries of Ecology and Agriculture (Ed.):1–88*.
- Miller DA, Nichols JD, McClintock BT, Grant EHC, Bailey LL, Weir LA. 2011. Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology* 92:1422–1428.
- Milner JM, Nilsen EB, Andreassen HP. 2007. Demographic Side Effects of Selective Hunting in Ungulates and Carnivores. *Conservation Biology* 21:36–47.
- Miquel C, Bellemain E, Poillot C, BessièRe J, Durand A, Taberlet P. 2006. Quality indexes to assess the reliability of genotypes in studies using noninvasive sampling and multiple-tube approach. *Molecular Ecology Notes* 6:985–988.
- Molinari-Jobin A et al. 2003. the pan-Alpine conservation strategy for the lynx. Council of Europe.
- Molinari-Jobin A et al. 2010. Recovery of the Alpine lynx *Lynx lynx* metapopulation. *Oryx* 44:267–275. Cambridge University Press.
- Molinari-Jobin A et al. 2012. Monitoring in the presence of species misidentification: The case of the Eurasian lynx in the Alps. *Animal Conservation* 15:266–273.
- Molinari-Jobin A et al. 2018. Mapping range dynamics from opportunistic data: Spatiotemporal modelling of the lynx distribution in the Alps over 21 years. *Animal Conservation* 21:168–180.
- Molinari-Jobin A, Zimmermann F, Angst C, Breitenmoser-Würsten C, Capt S, Breitenmoser U. 2006. Status and distribution of the lynx in the Swiss Alps 200-2004. *Acta Biologica Slovenica*

Références

49:3–11.

Nichols JD, Williams BK. 2006. Monitoring for conservation. *Trends in Ecology and Evolution* 21:668–673.

Nilsen EB, Linnell JDC, Odden J, Andersen R. 2009. Climate, season, and social status modulate the functional response of an efficient stalking predator: the Eurasian lynx. *Journal of Animal Ecology* 78:741–751.

Odden J, Linnell JDC, Moa PF, Herfindal I, Kvam T, Andersen R. 2002. Lynx Depredation on Domestic Sheep in Norway. *The Journal of Wildlife Management* 66:98–105.

Packer C et al. 2013. Conserving large carnivores: Dollars and fence. *Ecology Letters* 16:635–641.

Pedersen VA, Linnell JDC, Andersen R, Andrén H, Lindén M, Segerström P. 1999. Winter lynx *Lynx lynx* predation on semi-domestic reindeer *Rangifer tarandus* in northern Sweden. *Wildlife Biology* 5:203–211.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.

Pollock KH, Nichols JD, Simons TR, Farnsworth GL, Bailey LL, Sauer JR. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13:105–119.

Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, Popovic G, Warton DI. 2015. Point process models for presence-only analysis. *Methods in Ecology and Evolution* 6:366–379.

Renner IW, Warton DI. 2013. Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology. *Biometrics* 69:274–281.

Ripple WJ, Beschta RL. 2005. Linking Wolves and Plants: Aldo Leopold on Trophic Cascades. *BioScience* 55:613.

Ripple WJ, Beschta RL, Fortin JK, Robbins CT. 2014. Trophic cascades from wolves to grizzly bears in Yellowstone. *Journal of Animal Ecology* 83:223–233.

Rota CT, Fletcher Jr RJ, Dorazio RM, Betts MG. 2009. Occupancy estimation and the closure assumption. *Journal of Applied Ecology* 46:1173–1181.

Royle JA, Kéry M. 2007. A Bayesian state-space formulation of dynamic occupancy models. *Ecology* 88:1813–1823.

Runge MC. 2011. An Introduction to Adaptive Management for Threatened and Endangered Species. *Journal of Fish and Wildlife Management* 2:220–233.

Schnidrig R, Nienhuis C, Imhof R, Bürki R, Breitenmoser U. 2016. Lynx in the Alps: recommendations for an internationally coordinated management. RowAlps Report Objective 3. KORA Report.

Silvertown J. 2009. A new dawn for citizen science. *Trends in Ecology and Evolution* 24:467–471.

Skogen K, Haaland H. 2001. En ulvehistorie fra Østfold: Samarbeid og konflikter mellom forvaltning, forskning og lokalbefolkning.

Références

- Steenweg R, Hebblewhite M, Whittington J, Lukacs P, McKelvey K. 2018. Sampling scales define occupancy and underlying occupancy – abundance relationships in animals. *Ecology* 99:172–183.
- Sunde P, Kvam T, Moa P, Negård A, Overskaug K. 2000. Space use by Eurasian lynxes *Lynx lynx* in central Norway. *Acta Theriologica* 45:507–524.
- Treves A. 2009. Hunting for large carnivore conservation: Hunting large carnivores. *Journal of Applied Ecology* 46:1350–1356.
- Trouwborst A. 2010. Managing the Carnivore Comeback: International and EU Species Protection Law and the Return of Lynx, Wolf and Bear to Western Europe. *Journal of Environmental Law* 22:347–372.
- Valière N, Fumagalli L, Gielly L, Miquel C, Lequette B, Poulle M-L, Weber J-M, Arlettaz R, Taberlet P. 2003. Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation* 6:83–92.
- Van Strien AJ, Van Swaay CAM, Termaat T. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology* 50:1450–1458.
- Wabakken P, Sand H, Liberg O, Bjärvall A. 2001. The recovery, distribution, and population dynamics of wolves on the Scandinavian peninsula, 1978-1998. *Canadian Journal of Zoology* 79:710–725.
- Ward RJ, Griffiths RA, Wilkinson JW, Cornish N. 2017. Optimising monitoring efforts for secretive snakes: a comparison of occupancy and N-mixture models for assessment of population status. *Scientific Reports* 7:18074.
- White GC, Bennetts RE. 1996. Analysis of Frequency Count Data Using the Negative Binomial Distribution. *Ecology* 77:2549–2557.
- Williams BK. 2009. Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling* 220:830–840.
- Yackulic CB, Nichols JD, Reid J, Der R. 2015. To predict the niche, model colonization and extinction. *Ecology* 96:16–23.
- Zimmermann F, Molinari-Jobin A, Ryser A, Breitenmoser-Würsten C, Breitenmoser U. 2011. Status and distribution of the lynx in the Swiss Alps 2005-2009. *Acta biologica slovenica* 54:73–80.
- Zurell D, Jeltsch F, Dormann CF, Schröder B. 2009. Static species distribution models in dynamically changing systems: How good can predictions really be? *Ecography* 32:733–744.

Table des figures

Figure 1: Cartes des détections certaines (en bleu) et incertaines (en rouge) du loup en France entre 1994 et 2016. La grille correspond aux sites de 10 x 10km considérés comme unités d'échantillonnage.	27
Figure 2: Cartes des détections certaines (bleu) et incertaines (rouge) du lynx dans les pays alpins entre 1995 et 2013. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage.	33
Figure 3 : Cartes des détections de groupes familiaux du lynx en Norvège entre 1997 et 2016. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage.	39
Figure 4 : Cartes des lynx chassés en Norvège entre 1997 et 2016. La grille correspond aux sites de 10 x 10km qui sont les unités d'échantillonnage. Ici, les détections n'ont pas été ramenées à l'échelle d'un site de 10x10km mais les variables environnementales ont été calculées à l'échelle d'un site de 10x10 km	40