Approche intégrative de la stratégie de conservation du Râle des genêts

Yoan Fourcade

Thèse de Doctorat

Yoan FOURCADE

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le label de L'Université Nantes Angers Le Mans

École doctorale : *Végétal, Environnement, Nutrition, Agroalimentaire, Mer (VENAM)*

Discipline : *Biologie des Populations et Ecologie*
Spécialité : *Environnement*
Unité de recherche : *Groupe écologie et conservation des vertébrés (GECCO), Univ. Angers*

Soutenue le *2 septembre 2014*
N° d'ordre : *1433*

Approche intégrative de la stratégie de conservation du Râle des genêts

JURY

Rapporteurs :	**Éric PETIT,** Directeur de recherche, INRA **Romain JUILLARD**, Professeur, MNHN
Examinateurs :	**Pierre-André CROCHET**, Chargé de recherche, CNRS **Gabriele SORCI**, Directeur de recherche, CNRS
Directeur de Thèse :	**Bruno LE CAM**, Directeur de recherche, INRA
Co-directeur de Thèse :	**Jean SECONDI**, Maître de conférences, Université d'Angers

Remerciements

Les traditionnels remerciements vont bien évidemment en premier lieu à Jean Secondi qui m'a encadré durant ces 3 ans et demi. Là où bon nombre de thésards finissent éreintés par le stress, épuisés par la pression, et parfois avec des envies d'étrangler leur(s) chef(s), j'ai l'impression d'être particulièrement chanceux d'avoir pu mener et terminer cette thèse sereinement. Un encadrement de qualité, tant humainement que professionnellement, y est sans aucun doute pour beaucoup.

Merci également à David Richardson d'avoir participé à l'encadrement d'une partie du travail et de m'avoir accueilli à l'université de Norwich durant quelques mois.

Merci à Bruno Le Cam d'avoir accepté de prendre la responsabilité administrative de la thèse.

Merci aux membres du jury qui ont accepté de juger mon travail : Éric Petit, Romain Juillard, Pierre-André Crochet et Gabriele Sorci.

Merci à Christophe Lemaire qui, après m'avoir fait entrer dans le monde de la recherche en M2, a encore été là pour prodiguer quelques bons conseils.

Je remercie bien entendu l'ensemble des membres ou anciens membres de l'équipe GECCO que j'ai côtoyés durant ma thèse : Damien Picard, Stéphane Sourice, Olivier Pays, Sébastien Maugenest, Audrey Chaput-Bardy, Pierre-Cyril Renaud, Cécile Patrelle, Mireille Bellembou, Sophie Lardy, Aurélien Besnard, François-René Favreau, Mélissa Martin, Gwendolen Rodgers, Sandra Guérin, et tous les stagiaires qui sont passés par le labo.

Lors de mon séjour à Norwich, plusieurs personnes ont contribué à me former, me conseiller ou m'héberger, et je les en remercie : Dave et Laura Wright, Dan Locke et Jemma Mickleburgh, Catalina Gonzalez-Quevedo, Danni Gilroy, Karl Philipps et Lewis Spurgin.

Les échantillons de Râle des genêts ne se récoltant pas tout seuls, merci à ceux qui ont contribué à l'effort d'échantillonnage, que ce soit avec moi sur le terrain : Gilles Mourgaud, Emmanuel Séchet, Edouard Beslot en BVA, Serguei Fokin et Peter Sverev en Russie, ainsi que tous les autres, dans leur pays respectif, qui ont donné de leur personne : Oskars Keiss, Michal Budka, Attila Sandor, Sandor Boldogh, Peter Olsson, Paolo Vasca, Patrik Rhönnstad, Susanne Arbeiter, Jochen Bellebaum, Rhys Green, Joachim Sadlik.

Merci d'ailleurs au Râle des genêts d'être toujours présent en France à l'heure où je finis ma thèse, il s'en est fallu de peu…

Merci aux collègues allemands Jan Engler et Dennis Rödder de m'avoir guidé dans mon apprentissage des techniques de modélisation de distribution.

Merci aux très nombreuses personnes que je ne rencontrerai jamais mais qui, par leurs observations, ont contribué à alimenter les bases de données qui ont permis de réaliser ce travail de modélisation.

Puisqu'on en arrive aux remerciements plus personnels, je me contenterai de nommer les Docteurs Nicolas Daviaud et Clément Chevalier, qui savent qu'ils seront toujours cités en premiers dans ma liste d'amis. Pour ne pas faire de jaloux et d'oublis, les autres devront se contenter de remerciements collectifs mais sincères.

Merci à tous ceux qui auraient voulu être cités mais que j'ai oubliés ou négligés. Considérez-vous quand même comme remerciés, y'a pas de raison.

Et puis merci pour tout à Laura…

Cette thèse est financée par le Plan Loire Grandeur Nature, le Fond Européen de Développement Régional, Angers Loire Métropole, la région Pays de la Loire, le conseil général de Maine et Loire et la DREAL Pays de la Loire

Sommaire

INTRODUCTION

1. INTRODUCTION

1.1. CONTEXTE, OBJECTIFS ET STRUCTURE DE LA THESE

La biodiversité connait une crise sans précédent (Barnosky *et al.* 2011), principalement causée par la perte des habitats naturels, convertis pour l'activité humaine. La principale cause de conversion des habitats est leur utilisation pour l'activité agricole. Alors que les rendements agricoles augmentent, notamment pour répondre à la pression démographique (Phalan *et al.* 2014), les espèces inféodées aux milieux agricoles connaissent un déclin majeur dû à l'intensification des pratiques (Donald *et al.* 2001). Malgré l'urgence de la situation, la conservation des espèces à large distribution se heurte à la multiplicité des acteurs impliqués dans la collecte de données et dans la mise en place d'actions de protection. Dans le même temps, l'hétérogénéité des situations et de l'écologie de l'espèce au sein de son aire de distribution complique la pose d'un diagnostic fiable. C'est le cas des espèces d'oiseaux paléarctiques, dont l'aire de répartition recouvre des situations socio-économiques très variables qui impactent directement la répartition des menaces et des actions consacrées à la conservation de ces espèces.

A large échelle, la distribution d'une espèce peut présenter des discontinuités majeures ou au contraire constituer une seule entité connectée, en fonction de l'histoire phylogéographique de l'espèce ou de la répartition de barrières et corridors à la dispersion. De cette distribution et des mouvements des individus (couloirs de migration, modes et distances de dispersion) va dépendre la structure génétique globale des populations. Celle-ci peut être extrêmement variable y compris dans un même groupe d'espèces d'oiseaux paléarctiques (Zink *et al.* 2008). De plus, les exigences écologiques d'une espèce sont susceptibles de varier au gré d'adaptations aux conditions écologiques locales. Outre ces variations liées à l'écologie de l'espèce, l'intensité des menaces dépend souvent directement des politiques et de l'activité économique locales, rendant difficile l'évaluation des besoins de conservation et la mise en place de stratégies au niveau global. En amont de cela, une limite commune est que la quantité et la qualité des données disponibles varient également.

Dès lors, l'évaluation des besoins de conservation et la définition des actions à large échelle vont nécessiter de rassembler un large ensemble de connaissances sur l'écologie et la dynamique globale des populations. Pour des espèces dont la connaissance reste inégale au sein de leur aire de distribution, une approche intégrée de l'étude du fonctionnement des populations peut d'avérer nécessaire pour reconstituer l'ensemble des processus affectant la situation de conservation de l'espèce. D'autre part, la mise en place de programmes de conservation à l'échelle

locale ne peut s'affranchir de la compréhension du fonctionnement global des populations. Dans cette optique, cette thèse se propose de développer une approche globale et multi-approche de la conservation d'une espèce d'oiseau à large distribution continentale (Figure 1). Elle prend pour modèle une espèce migratrice paléarctique, le Râle des genêts (*Crex crex*), afin d'améliorer la connaissance des dynamiques de populations et de fournir des éléments scientifiques à la conservation locale et globale de l'espèce en période de reproduction.

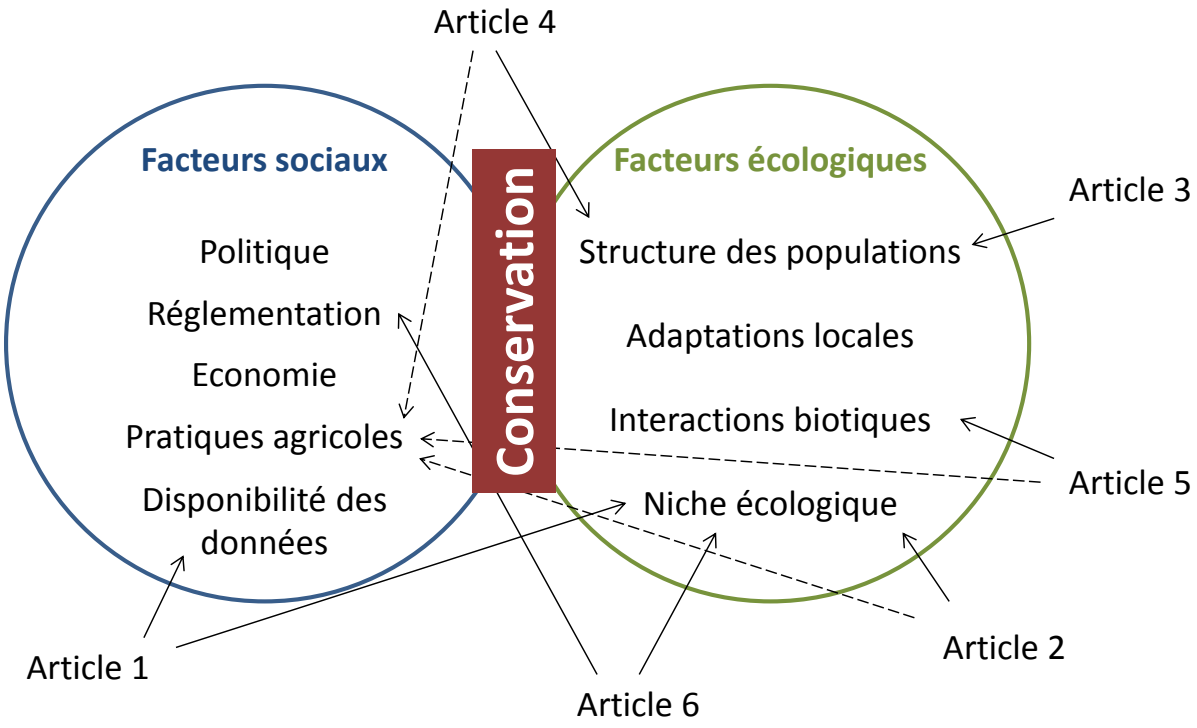


Figure 1: Facteurs sociaux et écologiques intervenant dans la problématique de conservation d'une espèce à large distribution continentale, et facteurs abordés par les différents chapitres de la thèse.

Le Râle des genêts est un oiseau prairial présent dans une grande partie du Paléarctique pendant sa période de reproduction. L'espèce a connu un déclin massif de ses effectifs en Europe de l'ouest lié à l'intensification agricole, et tout particulièrement la mécanisation des fauches ainsi que leur date trop précoce (Green *et al.* 1997a). A l'inverse, l'Europe de l'est et l'Asie, où se concentrent la grande majorité des effectifs de Râles des genêts, ont vu leurs populations se maintenir, voire augmenter depuis la déprise agricole qui a suivi la chute de l'URSS (Keišs 2005). Actuellement, la distribution du Râle des genêts est caractérisée par des différences démographiques considérables entre d'un côté le cœur et l'est de son aire de répartition où les

populations sont abondantes, et d'un autre côté l'ouest de sa distribution où les populations sont à faible effectifs, fragmentées et encore souvent en déclin.

Des mesures de conservations ont été prises dans la plupart des pays concernés par le déclin de l'espèce (Schäffer & Weisser 1996). Bien que celle-ci se soient avérées souvent insuffisantes pour enrayer ces tendances démographiques, l'exemple des opérations menées au Royaume-Uni montre que l'essentiel des actions envisageables sont déjà identifiées et pourraient remplir leur rôle si elles étaient plus efficacement implémentées (Stowe & Green 1997a). Dès lors, le présent travail n'aura pas pour but de trouver des solutions innovantes à la problématique de conservation locale du Râle des genêts, ni même d'évaluer directement les mesures mises en place. Par contre, l'espèce présente des caractéristiques intéressantes pour étudier, via des approches variées à différentes échelles, comment des pressions anthropiques variables à l'échelle continentale structurent les relations entre populations et menacent les plus petites populations. Dans cette optique, la thèse s'articule autour de quatre thématiques principales, de la plus large à la plus fine échelle, qui permettent de décomposer les mécanismes à l'œuvre et de tirer des conclusions utiles à la conservation tant globale que locale de l'espèce.

Dans un premier temps, une approche de modélisation de distribution sera menée afin de renforcer les connaissances, jusque-là incertaines, relatives à l'aire de reproduction du Râle des genêts (**Chapitre 1**). La robustesse de la méthode statistique employée étant influencée par l'utilisation d'un jeu de donnée spatialement biaisé, une première étape consistera à identifier les meilleures techniques permettant de contrer ce biais dans la procédure de modélisation (*Article 1*). Ces méthodes seront alors utilisées pour la modélisation de la distribution du Râle de genêts, afin de confronter une approche de modélisation à une distribution estimée à dire d'experts. Cette analyse permettra également d'identifier le cœur de son aire de distribution ainsi que des éventuelles discontinuités dans sa répartition (*Article 2*).

Une fois la distribution globale de l'espèce caractérisée, le **Chapitre 2** s'attachera à identifier la structuration spatiale des populations à l'échelle européenne. Ainsi, seront étudiées à la fois les variations géographiques du chant des mâles (*Article 3*) ainsi que les variations génétiques entre populations (*Article 4*). Les analyses génétiques permettront ainsi de décrire tout d'abord la structure génétique proprement dite de l'espèce, c'est-à-dire le degré d'isolement et les relations entre les différentes populations européennes, et notamment entre les populations périphériques et le cœur de la distribution de du Râle des genêts. Cette approche permettra également d'évaluer les traces génétiques laissées par les variations de tailles de populations sous

l'action anthropique, notamment en termes de diversité génétique, paramètre crucial pour les petites populations déclinantes.

Dans un troisième temps, une analyse des pathogènes, ici les malarias aviaires, sera menée afin de déterminer comment les variations génétiques et écologiques observées à l'échelle continentale influencent la prévalence parasitaire dans les populations de Râles des genêts (**Chapitre 3**, *Article 5*). En effet, on pose l'hypothèse que des réductions de population telles que celle observée en France étaient susceptibles, via une diminution de la diversité génétique, d'augmenter la susceptibilité des individus aux infections parasitaires. En fonction des résultats génétiques obtenus au chapitre 2 et de divers paramètres écologiques, les variations de prévalence et les facteurs qui les déterminent seront caractérisés.

Enfin, le dernier chapitre présentera une évaluation de la capacité du Râle des genêts à œuvrer comme une espèce parapluie pour la conservation de l'écosystème prairial (**Chapitre 4**, *Article 6*). En prenant l'exemple de la communauté d'oiseaux prairiaux présents dans les grands ensembles paysagers continus du Maine-et-Loire et de Loire Atlantique, le recouvrement des niches de 5 espèces d'intérêt (Râle des genêts et quatre passereaux) dans l'espace environnemental et géographique sera évalué afin de déterminer si la conservation de l'une ou l'autre de ces espèces permettrait de protéger efficacement les autres. Ce travail se positionne dans le cadre des questions sur les stratégies de conservation espèce centrée *vs.* communauté centrée ou services centrés.

1.2. CONSERVATION A L'ECHELLE DES POPULATIONS

L'un des pré-requis essentiels de la biologie de la conservation est d'identifier les entités qui devront faire l'objet d'une évaluation et, éventuellement, d'une mesure de conservation si nécessaire. Traditionnellement, le niveau supérieur de gestion est celui de l'espèce, qui sert de base à la plupart des efforts de conservation, et sur lequel se concentre l'attention du grand public (Verissimo *et al.* 2011). Toutefois, il peut exister des discontinuités spatiales, écologiques ou génétiques au sein d'une espèce qui vont conditionner des réponses différentes aux menaces et actions selon les populations. L'essentiel des thèmes développés au cours de cette thèse s'articulera donc autour de la notion de population, et notamment de l'analyse des processus biogéographiques influençant leur distribution et les caractéristiques génétiques.

1.2.1. Déterminer la distribution occupée par l'espèce

Parmi les données les plus élémentaires relatives à l'écologie d'une espèce, la connaissance de son aire de répartition et de ses déterminants est essentielle dans l'évaluation des menaces qui pèsent sur cette espèce et pour envisager des actions de conservations. Pourtant, des données d'occurrence seules s'avèrent insuffisantes pour refléter à elles-seules la distribution d'une espèce, d'autant plus lorsqu'elles ont été collectées de manière opportunistes ou biaisées. Des méthodes de **modélisation de distribution** ont donc été mises en place afin de disposer d'outils statistiques permettant de déterminer de façon la plus objective possible la distribution occupée ou potentielle d'une espèce d'intérêt. Toutefois, ces outils ne corrigent pas ni même ne tiennent compte du biais d'échantillonnage initial. Cette limite pourtant identifiée n'a été jusqu'à très récemment pas considérée dans l'évaluation des modèles.

1.2.1.1. Modélisation de distribution : principe

Les techniques de modélisation de distribution se fondent sur le concept de **niche écologique**. En effet, l'idée centrale de ces méthodes est de déterminer la niche d'une espèce, soit l'ensemble des facteurs écologiques qui permettent à cette espèce de persister, afin d'extrapoler son aire de répartition spatiale (Guisan & Zimmermann 2000). Bien que les concepts soient nombreux, Hutchinson (1957) a profondément marqué la notion de niche écologique et sa définition constitue le cadre théorique sur lequel s'appuie la relation niche-distribution. Il définit la niche fondamentale d'une espèce comme un hyper-volume à *n* dimensions, dont chaque point

représente une combinaison de paramètres environnementaux au sein de laquelle l'espèce pourrait exister indéfiniment. Cette **niche fondamentale**, uniquement déterminée par des variables abiotiques, a reçu le nom de niche Grinnellienne (James *et al.* 1984), puisqu'elle se rapproche de l'un des premiers concepts de niche formulé par Grinnell (1917). Cependant, il apparait rapidement que la niche d'une espèce est également influencée par la présence d'autres espèces (compétiteurs, prédateurs, etc.) : en effet, lorsque la niche fondamentale de deux espèces se superpose, l'une des deux espèces verra sa niche réduite par exclusion compétitive. En conséquence, la **niche réalisée** est nécessairement plus petite que la niche fondamentale. L'ensemble des facteurs biotiques qui déterminent la niche réalisée, sont à rapprocher du concept de niche Eltonienne (Elton 1927), qui décrit la niche écologique comme l'influence qu'une espèce a sur son environnement.

Les interactions entre niche fondamentale et réalisée, et entre niche et distribution, ont été formalisées sous la forme du **diagramme BAM** (Biotic, Abiotic, Movement) (Soberón & Peterson 2005; Soberón 2007; Soberón & Nakamura 2009). Cette représentation heuristique (Figure 2) décrit les facteurs déterminant la distribution d'une espèce de la façon suivante. Au sein d'un espace géographique disponible **G**, **A** représente la région où l'on retrouve les conditions environnementales nécessaires à la persistance de l'espèce, soit la niche fondamentale. Le sous-ensemble **B** représente la région où les conditions biotiques (compétiteurs, prédateurs, parasites) sont favorables à l'espèce. Un nouveau facteur est alors ajouté, **M**, qui représente l'espace géographique colonisable par l'espèce. Dès lors, la niche réalisée, réellement occupée, correspond à l'intersection $\boldsymbol{G_0} = A \cap B \cap M$. Dans ce cadre, l'ensemble $\boldsymbol{G_1} = A \cap B$ représente une zone potentiellement colonisable si les conditions venaient à permettre à l'espèce de l'atteindre (par exemple un transport par l'homme). Il est à noter que des individus de l'espèce considérée peuvent être retrouvés potentiellement dans l'espace **M** tout entier dans un système source-puits où des individus pourraient disperser vers des régions situées hors de leur niche, mais sans possibilité d'y persister (Pulliam 2000).

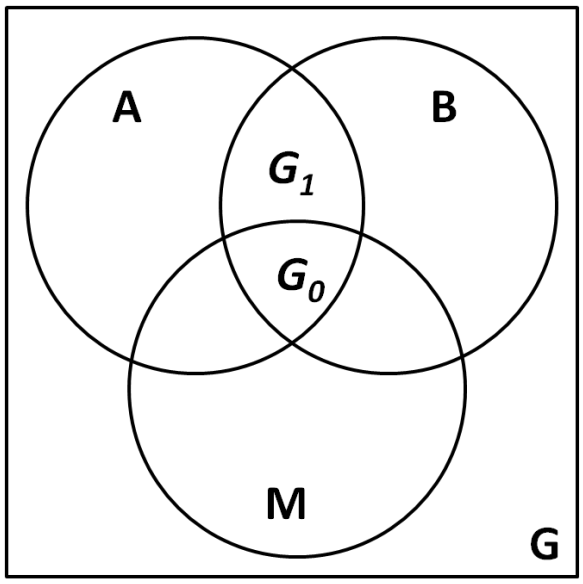


Figure 2: Diagramme de Venn décrivant les facteurs déterminant la distribution d'une espèce. G : espace géographique disponible, M : espace atteignable par l'espèce, A : espace contenant les conditions environnementales requises, B : espace contenant les conditions biotiques favorables. D'après Soberón (2007)

Dans ce contexte, la modélisation de distribution a pour but d'identifier la réponse d'une espèce à un ensemble de variables sélectionnées afin de projeter spatialement des probabilités de présence relatives en chaque point de la zone considérée. Concrètement, et bien que la nature réelle de la niche modélisée demeure parfois incertaine et dépendant des données utilisées (Saupe *et al.* 2012), la modélisation tendra donc à représenter uniquement la niche fondamentale (**A**), les interactions biotiques (**B**) étant quasiment impossible à implémenter dans ce type d'approche (Godsoe & Harmon 2012). Il a été suggéré qu'à une échelle large et relativement grossière, telle que la distribution globale d'une espèce, ce type d'interaction a très peu d'importance comparativement aux facteurs environnementaux tels que le climat (*Eltonian noise hypothesis*) (Soberón & Nakamura 2009). Toutefois, des exemples tendent à montrer que les interactions entre espèces peuvent largement façonner la façon dont des espèces proches se distribuent les unes par rapport aux autres, notamment dans les zones de contacts entre espèces parapatriques (Engler *et al.* 2013).

Quelle que soit l'algorithme utilisé, deux types de données sont nécessaires :

(i) Des **données de présence**, c'est-à-dire des coordonnées spatiales auxquelles l'occurrence de l'espèce a été enregistrée. Idéalement, on peut également utiliser des données d'absences, mais celles-ci étant extrêmement rares à l'échelle d'une aire de distribution, la

modélisation en présence-absence ne sera pas considérée ici. A l'échelle globale, la mise en place de bases de données en ligne telles que le projet GBIF (*Global Biodiversity Information Facility*, www.gbif.org) a grandement amélioré la disponibilité de ces données de présence, ce qui a largement contribué à l'essor des méthodes de modélisation de distribution. Cependant, la qualité de ces jeux de données est très variable et est susceptible d'être largement biaisée spatialement (Beck *et al.* 2013), ce qui peut contribuer à diminuer la performance du modèle (Leitão *et al.* 2011; Kramer-Schadt *et al.* 2013).

(ii) Des **données environnementales**, choisies car elles contribuent à la détermination de la niche écologique de l'espèce modélisée et donc à sa distribution. Elles se présentent la plupart du temps sous la forme de couches utilisables au sein d'un système d'information géographique (SIG) (Kozak *et al.* 2008), dont les caractéristiques de résolution et d'étendue se retrouveront dans la distribution modélisée. Encore une fois, la mise à disposition en ligne de données, climatiques (par exemple le projet WorldClim, www.worldclim.org) (Hijmans *et al.* 2005), topographiques (par exemple le projet *Shuttle Radar Topography Mission* SRTM) ou satellites (données SPOT ou MODIS par exemple), s'est développée parallèlement à la croissance des approches de modélisation de distribution. Comme pour les jeux de données d'occurrence, la facilité avec laquelle les données environnementales peuvent être téléchargées est source de biais si elles ne sont pas évaluées avec attention. En effet, les couches environnementales utilisées doivent bien évidemment être pertinentes pour l'écologie de l'espèce considérée (Pérez-Rodríguez *et al.* 2013) et présenter le moins de corrélations possibles entre elles (Braunisch *et al.* 2013), sous peine de produire un modèle peu fiable. Les données climatiques, par exemple, présentent couramment des corrélations fortes obligeant à une présélection des prédicteurs les moins corrélés.

De très nombreuses méthodes statistiques permettent de modéliser la distribution d'une espèce. Les premières approches se basaient sur des méthodes de régression classiques (Manly *et al.* 1993) ou d'enveloppes autour de points d'occurrences (Busby 1991; Carpenter *et al.* 1993). Cependant, le développement considérable des techniques de modélisation de distribution est allé de pair avec l'apparition d'approches algorithmiques permettant de modéliser des réponses écologiques plus complexes, telles que les méthodes de réseau de neurones artificiels (Ripley 1996), d'arbres de classification (Breiman *et al.* 1984; Breiman 2001a) ou d'entropie maximum (MAXENT) (Phillips *et al.* 2004). Devant le vaste éventail de méthodes disponibles et la difficulté d'en identifier une meilleure que les autres, un compromis peut être trouvé en réalisant un modèle consensus issu de la combinaison des résultats de différentes méthodes, pondéré par une évaluation de la performance de chacun des modèles individuels (Thuiller *et al.* 2009). Toutefois,

ces approches nécessitent de maitriser un ensemble de méthodes très variées aux résultats parfois assez largement variables, rendant relativement incertain la fiabilité du modèle consensus. Les études qui se sont intéressées à l'évaluation comparée de différents algorithmes concluent que **MAXENT** offre les meilleurs résultats dans la plupart des cas (Elith *et al.* 2006; Tsoar *et al.* 2007). C'est pourquoi cette unique méthode sera retenue dans les approches de modélisation de distribution développées au cours cette thèse.

Plusieurs méthodes de modélisation en présence seule reprennent les principes de la modélisation en présence-absence, en substituant aux données d'absence des points de pseudo-absence ou de *background*, tirés aléatoirement au sein de la zone d'étude et représentant un échantillon des conditions environnementales disponibles. Dans le cas de MAXENT, l'algorithme compare les densités de probabilité dans l'espace environnemental des points de présence et du *background* (Figure 3), pour déterminer un indice d'adéquation de l'habitat avec la niche de l'espèce en chaque point de l'environnement et la projeter spatialement (Elith *et al.* 2011; Merow *et al.* 2013). En l'absence d'information sur l'occurrence de l'espèce dans la zone considérée, les résultats de la procédure de modélisation doivent être interprétés en tant qu'indice d'habitat favorable et non pas de probabilité de présence en tant que tel. En plus de cela, MAXENT, comme d'autres méthodes de modélisation de distribution, estime l'importance de chaque prédicteur dans la distribution et produit une courbe de réponse de l'espèce pour chaque variable environnementale. Ces différentes sorties peuvent ainsi être utilisées afin de prédire la distribution de l'espèce modélisée et les variables qui la déterminent.

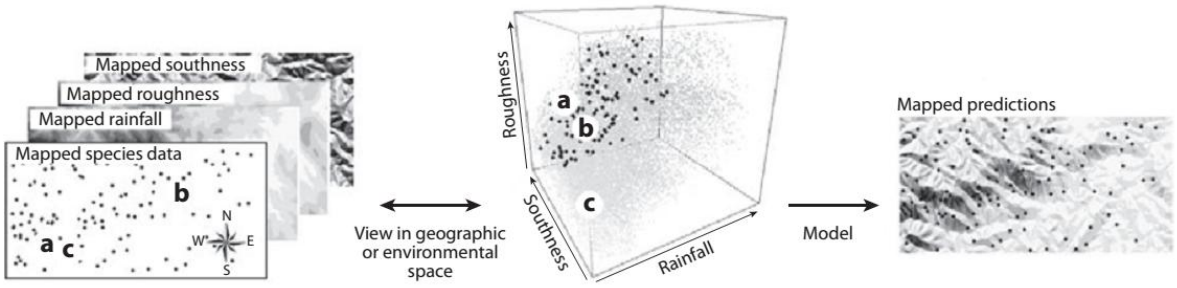


Figure 3: Représentation simplifiée de la procédure de modélisation MAXENT: à partir d'un ensemble de variables environnementales définies par l'utilisateur et d'occurrences de l'espèce à modéliser, l'algorithme compare la distribution dans l'espace environnemental du jeu de données de présence et du *background*. La détermination de la niche écologique permet alors d'estimer la probabilité de présence relative de l'espèce en chaque point de l'espace géographique. D'après Elith & Leathwick (2009).

1.2.1.2. Intérêt en biologie de la conservation

Sous l'action des changements globaux, les espèces tendent à modifier leur distribution. Ainsi, la pression anthropique dans l'occupation du sol entraine la fragmentation des habitats (Fischer & Lindenmayer 2007), tandis que des déplacements de distribution vers des altitudes ou des latitudes plus élevées sont observés sous l'action du réchauffement climatique (Walther *et al.* 2002). De plus, le déclin général de la biodiversité est largement associé à des réductions d'aire de répartition. En conséquence, la fraction restante de l'aire de distribution d'une espèce ou sa réduction dans un intervalle de temps donné font partie des critères d'évaluation du statut de conservation de l'espèce par l'IUCN (*International Union for Conservation of Nature*) (Mace *et al.* 2008). Dès lors, la question de la distribution des espèces apparait comme un élément d'intérêt majeur en **biologie de la conservation**.

Les distributions d'espèces ont longtemps été bâties de manières relativement simplistes en traçant un polygone telle qu'une enveloppe convexe autour des points de présence connus de l'espèce (Burgman & Fox 2003). Cette approche domine encore largement les données de distribution spatiales utilisées par l'IUCN, bien que l'utilisation de méthodes statistiques soit maintenant encouragée (IUCN Standards and Petitions Subcommittee 2010). D'autre part, des approches basées sur des opinions d'expert ont également été utilisées pour bâtir des prédictions spatiales de présence d'une espèce, avec plus ou moins d'efficacité (O'Leary *et al.* 2009; Murray *et al.* 2009). Toutefois, les dires d'expert souffrent d'un biais inhérent à la personnalité des experts contactés, aux interactions sociales intervenant au sein d'un panel ou à la méthode d'extraction de la connaissance experte (Burgman *et al.* 2011; Martin *et al.* 2012). Ainsi, les méthodes de modélisation de distribution apparaissent ainsi comme un outil permettant de guider des actions de conservations à partir d'outils plus robustes scientifiquement (Guisan *et al.* 2013).

Ces approches sont susceptibles d'intervenir à différents niveaux des processus de décisions aboutissant à la mise en place de mesures de conservation. Les domaines d'application de la modélisation de distribution sont nombreux en biologie de la conservation. En particulier, cinq thématiques principales bénéficient de l'utilisation de ces outils statistiques :

(i) Tout d'abord, dans certains cas, l'amélioration des connaissances sur la distribution d'une espèce présente un intérêt en tant que tel. La modélisation permet alors **d'identifier les habitats** les plus critiques pour la persistance de l'espèce ou de mettre en évidence des contractions/expansions dans la répartition de cette espèce, et ainsi d'alerter d'un déclin ou de guider des actions de gestion de l'occupation du sol (Guisan *et al.* 2013).

(ii) En combinant la modélisation de plusieurs espèces, il est également possible d'identifier des zones d'intérêt présentant une forte richesse spécifique, et ainsi de définir des *hotspots* de diversité pouvant aboutir à la mise en place de **réserves naturelles** (Moilanen 2005; Esselman & Allan 2011).

(iii) La modélisation de distribution s'avère également particulièrement intéressante afin de prédire les capacités d'expansion d'une **espèce invasive** (Ficetola *et al.* 2007; Jimenez-Valverde *et al.* 2011; Hill *et al.* 2012).

(iv) Le même type de procédure peut être envisagé pour prédire les capacités de recolonisations d'une espèce introduite lors d'une action de **translocation** (Chauvenet *et al.* 2013).

(v) D'autre part, en utilisant des données climatiques prédites pour les années futures sous différents scénarios, il est possible d'estimer les changements de distribution à venir sous l'action du **réchauffement climatique** (Araújo *et al.* 2004; Forester *et al.* 2013).

Ces trois dernières applications reposent sur l'idée essentielle que la niche écologique d'une espèce modélisée dans un espace spatio-temporel donné, à l'aide de données disponibles dans l'habitat présent, peut être **projetée** dans un nouvel espace géographique ou dans le futur. Souvent, les conditions environnementales dans lesquelles le modèle est projeté diffèrent de celles utilisées pour construire le modèle de distribution (par exemple, lorsque des combinaisons de valeurs environnementales dans le climat futur ou dans la zone d'invasion n'existaient pas dans la niche originale de l'espèce). Dans ce cas, les prédictions peuvent souffrir d'une certaine incertitude en fonction de la capacité de l'algorithme utilisé à extrapoler à de nouvelles données, ainsi que de la possibilité pour une espèce de changer sensiblement de niche au cours du temps (Elith *et al.* 2010).

1.2.2. Définir des unités de conservation au sein de cette distribution

1.2.2.1. Populations d'intérêt pour la conservation et unités évolutives significatives

Au sein d'une espèce donnée, les enjeux et actions de conservation se déroulent essentiellement au niveau de la population, mais la définition et le choix des populations d'intérêt demeurent ambigus. Dans le but de définir objectivement des entités de conservation, le concept d'unité

évolutive significative (***Evolutionarily Significant Unit***, ESU) a été développé (Ryder 1986). Sachant que l'échelle de l'espèce ne permet pas de refléter efficacement la diversité génétique naturelle, ce concept doit permettre d'identifier des groupes, à l'échelle infra-spécifique, qui serviront de base à la priorisation des actions de conservations. La définition ces groupes présentant des caractéristiques homogènes permet notamment de cibler des populations d'intérêt pour la conservation ou par exemple d'éviter des erreurs de translocation entre groupes génétiques distincts.

Selon sa définition originale, une unité évolutive significative représente un sous-groupe d'individus au sein d'une espèce dont les caractéristiques génétiques sont importantes pour le présent et le futur de l'espèce (Ryder 1986). Bien que cette définition pose les bases du concept, elle offre peu de directives pour identifier ces unités. Afin d'implémenter ce concept dans la législation américaine, Waples (1991) propose une définition plus précise. Ainsi, une unité évolutive significative serait une population ou un groupe de populations qui présente un **isolement reproducteur** substantiel par rapport aux autres populations de la même espèce, et qui représente une part importante de l'héritage évolutif de l'espèce. Cette définition se base comme la première sur l'idée qu'une ESU est un groupe relativement isolé, produit d'une histoire évolutive commune et réservoir génétique pour de futures adaptations (Waples 1995). Encore une fois, bien que cette définition soit conceptuellement intéressante et facilement adaptable à de nombreux cas concrets, sa subjectivité (isolement « substantiel », part « importante » de l'héritage évolutif) est un obstacle à son utilisation systématique. Afin d'améliorer l'objectivité du concept, des auteurs ont alors proposé des critères génétiques, et notamment phylogéographiques, pour identifier les unités de conservations. Ainsi, Dizon *et al.* (1992) définissent différentes catégories d'ESU sur la base de l'**isolement géographique** et des **divergences génétiques**, préférentiellement adaptatives, entre entités. De la même façon, l'approche de Moritz (1994) est de considérer comme entités évolutives distinctes des **groupes réciproquement monophylétiques** au niveau mitochondrial, et qui présentent une divergence significative des fréquences alléliques à des locus nucléaires. Cette définition permet de s'affranchir de données phénotypiques, offre un critère qualificatif pour la définition des ESU et parait ainsi plus facilement implémentable dans des mesures de gestion. Pourtant, les critères précédemment décrits présentent également un certain nombre d'inconvénients. En effet, le critère de monophylie mitochondrial est tellement strict qu'il aboutit dans de nombreux cas à identifier des espèces entières comme ESU, rendant peu utile cette approche (Waples 1995). De plus, des adaptations locales importantes peuvent avoir émergé lors de séparations de populations récentes, mais sans nécessairement résulter en une ségrégation stricte des lignées mitochondriales.

Devant la difficulté de trouver une définition simple d'une unité évolutive pertinente pour la conservation, des concepts plus intégratifs ont été développés au début des années 2000, prenant en compte l'ensemble des aspects précédemment décrits. Ainsi, Crandall *et al.* (2000) proposent de définir plusieurs niveaux de différentiations entre populations, basés sur des hypothèses testables, tant au plan écologique que génétique. Ce concept se base sur le principe d'**interchangeabilité des populations** (Templeton 1999). Deux populations sont dites écologiquement interchangeables si les individus peuvent librement passer de l'une à l'autre et occuper la même niche écologique et subir les mêmes contraintes sélectives. De même, l'interchangeabilité génétique se définit par un flux de gènes important entre populations, c'est-à-dire par exemple plus d'un migrant efficace par génération, pas d'allèles propres à chacune des populations, pas de divergence liée à une barrière géographiques. En testant l'hypothèse nulle de l'interchangeabilité au niveau écologique, génétique, et à différentes échelles temporelles, cette approche permet de classifier des unités évolutives du plus divergent (espèces distinctes) au plus uniforme (une seule population). Au vu de la diversité des définitions et concepts, Fraser & Bernatchez (2001) introduisent la notion de **conservation évolutive adaptative**, une approche plus flexible permettant d'utiliser, en fonction du taxon et du contexte, l'un ou l'autre des critères précédemment décrits pour définir des unités de conservations évolutives. Parallèlement, s'est développé le concept d'unités de gestions (*Management Units, MU*) qui cette fois-ci ne prend pas en compte l'histoire évolutive des populations mais regroupe simplement un ensemble d'individus présentant une différenciation génétique significative par rapport aux autres unités de gestion, quelle que soit la phylogénie des allèles impliquées (Moritz 1994). Concrètement, il s'agira en fait de populations suffisamment isolées pour être indépendantes démographiquement, c'est-à-dire que la dynamique de ces populations est uniquement déterminée par leur taux de survie et de natalité, sans influence de l'immigration (Palsbøll *et al.* 2007).

1.2.2.2. Identifier des populations et quantifier leur isolement

L'identification d'unités de conservations, et par extension de populations, repose essentiellement sur la caractérisation d'entités présentant un isolement relatif les unes par rapport aux autres. Une population peut ainsi être définie comme un groupe d'individus d'une même espèce maintenant une certaine cohésion écologique (co-occurrence spatio-temporelle d'individus en interactions) ou évolutive (reproduction panmictique entre individus) (Waples & Gaggiotti 2006). L'analyse des similarités et discontinuités génétiques entre des échantillons répartis spatialement représente un puissant outil pour identifier ces groupes cohérents. Deux approches différentes peuvent

permettre d'évaluer l'isolement entre populations en fonction du type de marqueurs génétiques utilisés : soit des marqueurs sélectivement neutres, soit des marqueurs sous sélection révélateurs d'adaptations à des conditions locales particulières.

La différenciation entre populations, et ainsi la définition d'unités de conservation, est la plupart du temps estimée via les variations génétiques à des **locus neutres**. En d'autres termes, on s'intéresse dans ce cas aux locus dont les variations de fréquences alléliques entre populations sont seulement dépendantes de la balance entre migration, mutation et dérive génétique. La mutation, survenant à faible fréquence et éventuellement négligeable à échelle temporelle courte, contribue à l'apparition de nouveaux allèles dans le pool génétique. La dérive génétique, quant à elle, entraine la fixation ou la disparition d'allèles à une vitesse fonction de la taille efficace de la population, du fait de l'échantillonnage aléatoire des gamètes lors de la reproduction. D'autre part, le flux de gènes causé par la migration d'individus entre populations contribue à l'homogénéisation génétique entre groupes (Hedrick 2011). Dès lors, les variations génétiques spatiales à des locus affectés seulement par ces trois processus permettent d'estimer le degré de différenciation/similarité entre groupes et ainsi de définir des entités possédant potentiellement des caractéristiques propres à considérer distinctement du point de vue de leur conservation. L'utilisation de marqueurs nucléaires très variables autorise la description fine des processus de différenciation génétique (Hedrick 2001). Parmi les marqueurs les plus utilisés figurent les microsatellites, séquences répétée di- ou tri-nucléotides (Balloux & Lugon-Moulin 2002; Selkoe & Toonen 2006), et les SNP (*Single-Nucleotide Polymorphism*), c'est-à-dire des variations nucléotidiques au sein du génome (Helyar *et al.* 2011). En plus des marqueurs nucléaires, des marqueurs mitochondriaux sont couramment utilisés, tout particulièrement dans des applications phylogéographiques où l'absence de recombinaison permet de retracer plus facilement l'historique de la distribution de lignées génétiques (Zink & Barrowclough 2008).

La structure génétique des populations, c'est-à-dire l'existence de différences génétiques significatives entre groupes, peut se tester par différentes **méthodes analytiques**. La plus courante et la plus ancienne est le calcul du F_{ST}, un indice de différenciation génétique entre populations. Cet indice est défini dans sa forme originale par la probabilité pour deux allèles d'un génotype d'être identiques par descendance, c'est-à-dire des copies d'un même allèle issues de la même population ancestrale (Wright 1951). Différentes définitions et estimations du F_{ST} ont émergées depuis. Parmi les estimateurs de F_{ST} les plus employées, on peut citer le θ de Weir et Cockerham (1984) ou le G_{ST} de Nei (1973). Ce dernier représente le déficit en hétérozygote dans un modèle en îles dû à la structuration génétique : $G_{ST} = (H_T - H_S)/H_T$, où H_T est

l'hétérozygotie attendue totale et H_S l'hétérozygotie attendue intra-population. Le F_{ST} peut également être relié au temps de coalescence entre allèles (Slatkin 1995) ou encore au flux de gènes dans un modèle en île à l'équilibre (Dobzhansky & Wright 1941) par la relation $F_{ST} = 1/(1 + 4N(m + \mu))$ où N est la taille efficace, m le taux de migration et μ le taux de mutation, souvent considéré négligeable et fixé à 0. D'autres estimateurs de différenciation ont vu le jour pour combler les lacunes du F_{ST} (valeur maximum dépendante de l'hétérozygotie des marqueurs employés), tels que le G'_{ST} (Hedrick 2005) ou le D (Jost 2008).

Au-delà de simples indices de différenciation ou de méthodes par ACP, des méthodes de description plus fines de la **structure génétique** ont été développés, notamment des algorithmes bayésiens de regroupement par groupes génétiques (*clustering*) (François & Durand 2010). La plus populaire de ces méthodes, implémentée dans le programme STRUCTURE, estime pour un nombre de groupes génétiques K donné, les fréquence alléliques dans chacun des groupes en partant de l'hypothèse que chaque groupe est à l'équilibre de liaison et de Hardy-Weinberg (Pritchard *et al.* 2000). Dès lors, le programme affecte à chaque individu génotypé la probabilité d'appartenance à chacun des groupes. Cette approche permet, en faisant varier K, d'estimer le nombre de groupes génétiques existants dans un échantillon d'individus réparti spatialement (Evanno *et al.* 2005) et d'évaluer l'intensité de la structure génétique, voire d'identifier des migrants. La robustesse de la méthode a été considérablement améliorée par l'implémentation de la localisation géographique des individus comme *a priori* lors de l'analyse, permettant de détecter des structure génétiques très faibles (Hubisz *et al.* 2009). D'autres méthodes bayésiennes spatialement explicites de regroupement génétique visent à identifier la structuration génétique spatiale de populations, telles que GENELAND (Guillot *et al.* 2005), BAPS (Corander *et al.* 2007) ou encore TESS (Chen *et al.* 2007).

Au-delà de l'identification de groupes génétiques cohérents, il est possible de caractériser plus finement les processus génétiques spatiaux. Des méthodes basé sur des reconstructions phylogénétiques par coalescence (Kingman 1982) permettent ainsi d'estimer un grand nombre de paramètres de génétique des populations. Il est notamment possible d'estimer les taux de migration entre groupes et les tailles efficaces de populations, par exemple à l'aide des algorithmes implémentés dans le programme MIGRATE (Beerli 2006). De même, on peut identifier des réductions/expansions récentes de populations en comparant l'hétérozygotie observée par rapport à une distribution attendue simulée par coalescence (méthode implémentée dans le logiciel BOTTLENECK (Piry *et al.* 1999)).

Toutefois, la reconstruction d'histoires démographiques complexes et l'estimation simultanée d'un grand nombre de paramètres s'avèrent souvent impossible à résoudre mathématiquement. Dans ce contexte se sont développées les méthodes dites d'*Approximate Bayesian Computation* (**ABC**) qui permettent de tester la probabilité de différents scénarios face à des données empiriques et d'estimer les paramètres associés (Beaumont *et al.* 2002). L'approche ABC nécessite en premier lieu de simuler des jeux de données en fonction des différents modèles hypothétiques à tester. Par exemple, des programmes de simulation en génétique des populations tels que FASTSIMCOAL (Excoffier & Foll 2011) ou MS (Hudson 2002) permettent de générer rapidement un grand nombre de génotypes simulés selon des scénarios démographiques très précisément paramétrables. Au fur et à mesure des simulations, les paramètres du modèle dont on ignore les valeurs sont tirés dans une distribution *a priori* et varient ainsi entre chaque simulation, permettant d'explorer tout l'espace des paramètres. A l'issue de la procédure de simulation, des statistiques résumées sont extraites des génotypes simulés afin de disposer de données réduites à comparer avec les données empiriques. Dans son approche la plus simple, la méthode ABC produit une probabilité postérieure pour chaque modèle en fonction de la part de chacun de ces modèles dans les simulations les plus proches des données réelles (Beaumont 2010; Csilléry *et al.* 2010; Sunnåker *et al.* 2013). De la même façon, la distribution des paramètres variables dans la fraction des données simulées les plus proches des données observées produit une distribution postérieure de chacun des paramètres à estimer. Des approches plus complexes ont été développées par la suite, permettant d'estimer les probabilités postérieures sur la base de méthodes de régression tout en étant moins sensible à la fraction des simulations retenue pour l'estimation (Beaumont 2010; Blum & François 2010). Les méthodes ABC s'avèrent particulièrement robustes pour retracer des histoires démographiques ou de colonisation et peuvent donc fournir un outil puissant pour comprendre les relations qu'entretiennent des populations ou les facteurs déterminant des changements de taille de population (Lopes & Boessenkool 2009).

Toutefois, les variations neutres ne caractérisent que les échanges entre populations, et éventuellement les dynamiques démographiques, mais ne renseignent pas directement sur le potentiel adaptatif des groupes génétiques ainsi définis. L'analyse de **marqueurs soumis à des pressions de sélection variant spatialement** permet de détecter des adaptations locales. Des variations génétiques et phénotypiques sont fréquemment observées le long de gradients environnementaux, révélant des spécialisations écologiques permettant aux populations de maximiser leur fitness dans les conditions environnementales auxquelles elles sont exposées (Savolainen *et al.* 2013). L'existence de ces adaptations peut être déterminante pour la survie des

populations. Elles constituent ainsi un élément important de leur conservation. Elles peuvent également révéler des barrières au flux de gènes précurseurs de spéciation. Dès lors, leur identification au sein de populations naturelles peut constituer un critère essentiel dans la définition d'unités de conservation (Luikart *et al.* 2003).

Les approches moléculaires communément menées pour identifier des locus soumis à adaptation se basent sur des **scans génomiques** dont l'objectif est de mettre en évidence des locus présentant une différenciation entre populations significativement supérieure à l'ensemble du génome (Butlin 2010). Ces méthodes, initiées par le test de Lewontin et Krakauer (1973), se basent sur l'analyse du F_{ST} inter-populations à différents locus afin de déterminer une distribution nulle des F_{ST} au sein du génome. Les locus dont les allèles sont sélectionnés différentiellement en fonction des populations présenteront une différenciation moyenne entre populations supérieure à celle des locus purement neutres. A l'inverse, les locus soumis à sélection équilibrante présenteront une différenciation inférieure aux locus neutres, présumés majoritaires (Beaumont 2005). Lorsqu'un locus *outlier* est lié à un gène aux fonctions connues, il est possible de déterminer plus précisément les causes de l'adaptation locale. De nombreuses méthodes, notamment basées sur des simulations coalescentes de la distribution neutre des variations de fréquences alléliques (Beaumont & Nichols 1996; Beaumont & Balding 2004; Foll & Gaggiotti 2008; Excoffier *et al.* 2009), ont été développées plus récemment et permettent d'identifier plus efficacement les régions du génome soumis à des divergences adaptatives.

Malgré son intérêt, cette approche n'est pas indispensable en tant que telle à la définition de populations d'intérêt pour la conservation. En premier lieu, la non-identification de locus sous sélection n'assure pas que ces derniers n'existent pas au sein des populations concernées, mais seulement que les locus testés ne relèvent pas d'adaptations locales. Au contraire, l'identification d'un isolement à des locus neutres met en évidence la possibilité pour la sélection naturelle de sélectionner des allèles spécifiques en fonction de l'environnement, même si les locus potentiellement impliqués ne sont pas connus. D'autre part, même en l'absence d'adaptations locales identifiées, l'existence d'un isolement génétique vis-à-vis des autres populations caractérise un groupe d'individu ayant une histoire évolutive et un potentiel adaptatif communs, même si cela peut être à court terme, et des caractéristiques génétiques propres au moins en termes de fréquences alléliques. Dès lors, la caractérisation de locus soumis à adaptations locales ne sera pas explorée au cours de cette thèse, laquelle se concentrera sur les variations génétiques neutres entre populations.

1.2.3. Conservation des populations périphériques

Au cours du temps, et tout particulièrement actuellement sous l'action anthropique, les distributions d'espèces se décalent, se contractent ou au contraire s'étendent. Ces changements peuvent refléter l'évolution de la niche écologique ou à l'inverse traduire des changements dans les conditions environnementales qui conduisent une espèce à suivre sa niche dans l'espace (Sexton *et al.* 2009). Dans tous les cas, les populations localisées en **périphérie** de l'aire de distribution, étant situées en limite de leur niche, sont les plus exposées aux changements rapides des conditions écologiques (Brown *et al.* 1996). Elles se trouvent donc à la fois plus facilement menacées par des changements environnementaux trop brutaux, mais également susceptibles d'initier des adaptations à ces changements. Elles constituent dès lors des populations d'intérêt tout particulier en biologie de la conservation, tant pour leur relative fragilité que pour leur importance dans le potentiel évolutif d'une espèce (Lesica & Allendorf 1995). Il existe des hypothèses relatives aux variations écologiques et génétiques attendues au sein d'une aire de distribution qui permettent de prédire les caractéristiques de ces populations et leur réponse aux changements environnementaux.

1.2.3.1. Relations centre-périphérie

La distribution spatiale des effectifs d'une espèce au sein de son aire de répartition conditionne de nombreux patrons de variations macrogéographiques. En effet, une hypothèse biogéographique fondamentale est que l'abondance d'une espèce est maximale au centre de sa distribution et va en décroissant vers la périphérie (Hengeveld & Haeck 1982). L'idée sous-jacente est que la distribution d'une espèce reflète un gradient écologique dont l'optimum constitue le cœur de cette distribution, tandis que les caractéristiques environnementales deviennent de moins en moins favorables vers la périphérie. En conséquence, la survie et la reproduction sont maximales au centre tandis que la fitness diminue en périphérie, avec pour conséquence un gradient de taille de population du centre vers les limites de la répartition de l'espèce associé à un gradient opposé de fragmentation des populations (Brussard 1984). En pratique, cette hypothèse du **centre abondant** est assez rarement constatée empiriquement (Sagarin & Gaines 2002), notamment du fait que les distributions sont la plupart du temps complexes, leur cœur réel ne correspondant pas nécessairement au centre de gravité géométrique (Channell & Lomolino 2000; Sagarin *et al.* 2006). Néanmoins, les hypothèses proposées restent valides lorsque l'on considère non plus le centre

géographique de la distribution mais celui de la niche écologique, et peuvent façonner les patrons de variations génétiques à large échelle au sein de la distribution d'une espèce.

La conséquence génétique la plus couramment admise du gradient d'abondance est que ce gradient conditionne deux paramètres essentiels : la taille efficace de population et le taux de migration, tous deux maximums au centre et décroissants vers la périphérie (Eckert *et al.* 2008). Il en résulte que les populations périphériques, de fait de la dérive génétique plus forte dans des populations de petite taille et isolées, présenteraient une **diversité génétique réduite** et une plus **importante différenciation** génétique que les populations centrales. Ce patron serait renforcé par le fait que les populations marginales sont susceptibles de connaitre des cycles d'extinction/colonisation plus rapides, ainsi que des pressions de sélection importantes favorisant la fixation rapide d'allèles (Hoffmann & Blows 1994).

Cependant, le processus parfaitement opposé peut également être envisagé. En effet, les différences démographiques entre centre et périphérie peuvent entrainer un flux de gènes asymétrique au sein de la distribution. Si la dispersion est aléatoire, les populations centrales abondantes et productives, devraient envoyer plus de migrants vers les plus petites populations périphériques que l'inverse. Cette asymétrie, analogue à un **système île-continent** voir source-puits (Pulliam 1988), aurait alors pour conséquence de maintenir une diversité génétique élevée dans les populations périphériques et d'empêcher leur différenciation. En limitant en conséquence les possibilités d'adaptations locales dans ces populations, ce phénomène serait à l'origine de la limite de l'aire de distribution en elle-même (García-Ramos & Kirkpatrick 1997; Kirkpatrick & Barton 1997). La dérive génétique dans les petites populations périphériques qui réduit la diversité génétique, et le flux de gènes asymétrique qui la maintient, se déroulent *a priori* simultanément bien qu'ils aient des conséquences opposées. Dès lors, la direction dans laquelle les variations génétiques évolueront au sein de la distribution d'une espèce dépend de l'intensité du flux de gènes centre vers périphérie relativement à la dérive, et donc en partie de la distribution des distances de dispersion de l'espèce

1.2.3.2. Menaces sur les petites populations périphériques

Les populations périphériques se retrouvent la plupart du temps dans des environnements suboptimaux relativement à leur niche écologique, et sont souvent plus petites et isolées que les populations centrales. Elles sont ainsi plus naturellement susceptibles aux risques d'extinction dûs à des événements stochastiques qui peuvent mettre en péril rapidement leur persistance (Lesica &

Allendorf 1995). En plus des risques démographiques classiques (susceptibilité aux fluctuations démographiques, effet *Allee* (Courchamp *et al.* 1999)), les petites populations peuvent présenter des risques d'extinction accrus en raison de leurs caractéristiques génétiques (Lande 1993).

La petite taille et l'isolement des populations périphériques favorisent en effet la fixation et la disparition d'allèles par la dérive génétique et donc une réduction de la diversité génétique au sein de ces populations (Reed & Frankham 2003; Charlesworth 2003). Ceci est susceptible d'entrainer des effets délétères dus à la **dépression de consanguinité**. Ce syndrome correspond à une réduction de *fitness* causée par la fixation d'allèles délétères récessifs (Lande 1994) ou un accroissement moyen de l'homozygotie conduisant à la perte de l'avantage hétérozygote à certains locus (Vrijenhoek 1994). Des effets démographiques négatifs, y compris un risque d'extinction accru (Bijlsma 2000), peuvent alors se faire sentir, tout particulièrement lorsque le flux de gènes ne suffit pas à compenser la dérive génétique (Frankham 1996). Un déclin de la *fitness* moyenne causé par des taux de consanguinité élevés est ainsi fréquemment observé tant en captivité (Ralls *et al.* 1988) que dans les populations naturelles (Crnokrak & Roff 1999; Frankham 2009).

En règle générale, on observe fréquemment une corrélation entre le niveau d'hétérozygotie d'une population et l'expression de traits liés à la fitness des individus (*heterozygosity–fitness correlation HFC*) (Chapman *et al.* 2009; Szulkin *et al.* 2010). En fonction des marqueurs génétiques utilisés pour mettre en évidence cette relation, différentes hypothèses peuvent expliquer l'existence d'une **corrélation hétérozygotie-fitness** (Hansson & Westerberg 2002). (i) La première hypothèse, ou effet général (David 1998), stipule que l'hétérozygotie aux locus analysés est révélatrice de l'hétérozygotie moyenne au sein du génome. Dans ce cas, une réduction d'hétérozygotie à un jeu de marqueurs neutres tels qu'un ensemble de microsatellites (Väli *et al.* 2008; Forstmeier & Schielzeth 2012) est directement lié au taux de consanguinité de la population. (ii) Au contraire, l'hypothèse dite de l'effet direct est que les marqueurs utilisés pour mesurer l'hétérozygotie sont exprimés et ont un effet direct sur la fitness, comme par exemple les gènes du CMH (Complexe majeur d'histocompatibilité) (Worley *et al.* 2010). (iii) Enfin, l'hypothèse de l'effet local, en quelque sorte intermédiaire, correspond à des marqueurs en déséquilibre de liaison avec des locus exprimés et agissant sur la fitness (Hansson *et al.* 2001, 2004). Au vu des conséquences possibles sur la survie des populations, la restauration de la diversité génétique dans les populations naturelles en déclin est devenue un enjeu majeur de la biologie de la conservation (Hedrick & Kalinowski 2000; Nichols *et al.* 2001; Hedrick 2004; Frankham 2005).

Les infections parasitaires peuvent représenter un médiateur essentiel des corrélations hétérozygotie-fitness (MacDougall-Shackleton *et al.* 2005; Townsend *et al.* 2009; Voegeli *et al.* 2012). En effet, des facteurs génétiques peuvent avoir un rôle prépondérant dans la probabilité d'infection par un **pathogène** (Frankham *et al.* 2002; Hawley *et al.* 2005). La perte de diversité génétique dans les petites populations, notamment périphériques, peut ainsi entrainer une susceptibilité plus élevée aux pathogènes par rapport aux populations plus « saines » génétiquement (Spielman *et al.* 2004). La disparition d'allèles de résistance spécifiques du pool génique d'une population a pour effet de rendre cette dernière plus vulnérable à la mise en contact de nouveaux parasites (Hedrick 2002). De même, lorsqu'existait un avantage hétérozygote à la résistance à un pathogène, la diminution de diversité génétique au sein d'une population tend à réduire l'hétérozygotie moyenne des individus dans celle-ci et donc la susceptibilité aux pathogènes (MacDougall-Shackleton *et al.* 2005; Evans & Neff 2009). La variabilité génétique des locus directement impliqués dans la réponse immunitaire est à cet égard tout particulièrement importante. L'hypervariabilité du CMH, notamment, a évolué en réponse aux pressions de sélection exercées par les pathogènes, par l'intermédiaire de l'avantage conféré aux allèles rares ou aux hétérozygotes, ou encore à cause des fluctuations spatio-temporelles des populations de pathogènes (Spurgin & Richardson 2010). Il a ainsi été montré chez différentes espèces que la perte de diversité au niveau du CMH pouvait entrainer une susceptibilité accrue aux infections (Westerdahl 2007; Radwan *et al.* 2010). Les infections parasitaires peuvent avoir des conséquences majeures sur la survie des populations naturelles (Altizer *et al.* 2003), allant jusqu'à l'extinction rapide d'espèces peu adaptées à des pathogènes auxquelles elles n'avaient jamais été exposées (Van Riper III *et al.* 1986). Alors que les déplacements humains permettent maintenant de mettre rapidement en contact des populations et des pathogènes non-natifs (Daszak *et al.* 2000), la compréhension des interactions hôte-parasite et le maintien de la diversité génétique dans les populations de taille réduite s'avèrent particulièrement primordiaux.

1.3. LE RALE DES GENETS

Les thématiques développées au cours de cette thèse seront abordées au travers d'une espèce d'oiseau prairial paléarctique, le Râle des genêts.

1.3.1. Biologie et écologie

1.3.1.1. Description générale

Le Râle des genêts (*Crex crex, Linnaeus* 1758) est un oiseau de la famille des *Rallidae*, dans l'ordre des Gruiformes. Une seule autre espèce du genre *Crex* a été décrite, *Crex egregia* ou Râle des prés, espèce prairiale présente toute l'année en Afrique subsaharienne. Le Râle des genêts présente une longueur du corps comprise entre 27 et 30 cm pour une envergure de 46 à 53 cm (Schäffer & Koffijberg 2004). La masse des individus, de 120 à 200 g environ, peut être utilisée pour estimer l'âge des jeunes (Green & Tyler 2005), bien que des variations significatives de la prise de poids en fonction des conditions climatiques soient observées (Tyler & Green 2004). Le dimorphisme sexuel est très peu prononcé chez cette espèce. Mâles comme femelles présentent une coloration brune striée de noir sur le dos, avec un plumage gris sur le ventre (Figure 4). Les ailes sont d'un roux relativement uniforme. Les jeunes naissent couverts d'un duvet noir (Figure 4), puis le plumage devient largement similaire aux adultes après quelques semaines. Rapidement, seule la couleur de l'iris ainsi que la forme des plumes permettent de distinguer les immatures des individus plus âgés (Schäffer & Koffijberg 2004; Green & Tyler 2005). Le plumage est intégralement renouvelé par une mue complète qui débute à la fin de la période reproductrice, soit en juillet pour les mâles dès que le chant cesse, et quelques semaines plus tard pour les femelles, lorsque les jeunes sont âgés d'une dizaine de jours. Une mue partielle intervient également en hiver, avant le départ en migration vers l'aire de reproduction (Schäffer & Koffijberg 2004).

Figure 4: Photographies de Râles des genêts. **A** : Mâle adulte, source: Isle of Man Government, 2006, licence CC BY 2.0 ; **B** : Mâle en train de chanter, source: Debbie Bozkurt©, 2007; **C**: Femelle accompagnée de 4 jeunes, source: Len Eades©, 2011

Très caractéristique, le chant du Râle des genêts (« *krek krek* ») est à l'origine de son nom scientifique (*Crex crex*). Ainsi, les suivis de populations de l'espèce sont réalisés par le comptage systématique des mâles repérés grâce à leur chant (Broyer 2002). Sauf à de rares exceptions près (Ottvall 1999), seul le mâle chante, d'avril à juillet et principalement entre 23h et 3h. Le comportement de chant chez le Râle des genêts est associé à la défense d'un territoire ainsi qu'à la communication intersexuelle (Schäffer 1995). Les caractéristiques du chant permettent de réduire la dégradation du signal lors de la transmission dans la végétation dense, tant à courte qu'à longue distance (Ręk & Osiejuk 2011; Ręk 2013c). Le chant présente des différences interindividuelles suffisamment importantes pour assurer une certaine capacité de discrimination entre individus (May 1998; Peake *et al.* 1998; Ręk & Osiejuk 2011; Budka & Osiejuk 2013a, 2013c). Bien que la stricte identification individuelle reste difficile (Mikkelsen *et al.* 2013), les caractéristiques du chant des mâles peuvent dans certains cas être utilisées pour le suivi des populations et des déplacements au même titre que les méthodes de marquage ou de télémétrie (Peake & McGregor 2001; Mikkelsen *et al.* 2013). Il a été montré qu'il existe des variations géographiques dans le chant du Râle des genêts, et notamment que les individus les plus proches géographiquement ont tendance à avoir un chant plus proche (Peake & McGregor 1999; Budka *et al.* 2014). Cependant, on retrouve au sein d'une même population des variations interannuelles du même ordre de grandeur que les variations inter-populations observées à une année donnée (Budka *et al.* 2014).

1.3.1.2. Habitat et régime alimentaire

L'habitat du Râle des genêts se caractérise essentiellement par la présence de végétation herbacée haute et dense (Figure 5). De ce fait, on le retrouve principalement dans les prairies alluviales, dernier habitat prairial disponible dans de nombreux pays d'Europe de l'ouest (Green *et al.* 1997a). Cependant, ces prairies semi-naturelles exploitées pour la production de foin sont souvent fauchées trop tôt pour permettre le bon déroulement de la reproduction, à moins que des mesures de protection ne soient spécifiquement mises en place (voir partie Conservation). La présence de Râle des genêts est parfois constatée dans des habitats inhabituels et probablement sub-optimaux tels que des champs cultivés ou des jachères, tout particulièrement du fait d'un déplacement des individus suite aux fauches des sites de reproduction (Green *et al.* 1997a). D'autres habitats sont également occupés, telles que des prairies alpines (Brambilla & Pedrini 2011) ou des anciennes zones agricoles abandonnées (Keišs 2005). La structure de la végétation est un élément primordial de l'habitat du Râle des genêts. En effet, l'espèce a besoin d'un couvert végétal suffisant (au moins 20 à 30 cm de haut) mais les zones de prairie trop dense (à cause, par exemple, de l'utilisation de fertilisants) sont évitées (Wettstein *et al.* 2001). Omnivore, le Râle des genêts a un régime alimentaire varié constitué à la fois d'arthropodes (diptères, orthoptères, odonates, coléoptères), de gastéropodes, de nématodes, d'arachnides, mais également de graines et de végétaux verts (Schäffer & Koffijberg 2004). La proportion de chacun de ces groupes dans l'alimentation du Râle peut varier considérablement en fonction de la ressource alimentaire disponible. Par exemple, des gastéropodes ont été identifiés dans plus de 80% des fèces analysées en Ecosse tandis qu'ils n'étaient présents que dans moins de 50% des fèces en Pologne (Schäffer & Koffijberg 2004). Le Râle des genêts semble avoir un régime alimentaire largement opportuniste, ce qui suggère l'importance, non pas de la ressource en tant que telle, mais plutôt de la structure de la végétation disponible, dans la sélection de ses sites de reproduction.

1.3.1.3. Reproduction

La saison de reproduction se déroule théoriquement entre mai et septembre, mais peut varier en fonction des régions et de la date d'arrivée des individus, ainsi que de l'activité humaine qui détruit fréquemment les sites de reproduction tôt dans la saison (voir partie 1.3.2 Conservation). Les mâles commencent à chanter quelques jours après leur arrivée, en choisissant une place de chant caractérisée par une couverture végétale plus dense ou plus haute tels qu'un buisson isolé au milieu de la prairie ou une tache d'herbe dense. Les mâles tendent à se regrouper autour de

zones de chant favorables, et cherchent à exclure leurs congénères de leur site de chant. Ce comportement a pour conséquence de former des agrégats de mâles et à expliquer en grande partie la répartition des individus à échelle fine (Schäffer & Koffijberg 2004). Lorsque le mâle est apparié avec une femelle, son activité de chant cesse. L'évolution du nombre de chanteurs sur un site permet ainsi de suivre la chronologie de la reproduction de l'espèce (Tyler & Green 1996). Mâles comme femelles peuvent s'apparier séquentiellement avec plusieurs partenaires au cours d'une saison (Schäffer & Koffijberg 2004). Après copulation, ils forment un couple monogame pour quelques jours seulement, les mâles repartant en quête d'une nouvelle femelle dès la ponte du premier œuf (Tyler 1996). Environ 10 œufs sont pondus au sol dans un nid sommaire à une fréquence moyenne de 1.2 œufs par jour, après 16 à 19 jours d'incubation (Tyler 1996; Schäffer & Koffijberg 2004). En l'absence de perturbation par les fauches, il a été montré que la survie des jeunes est très bonne et indépendante des conditions climatiques (Tyler *et al.* 1998; Tyler & Green 2004). Lorsque cela est possible, deux pontes sont produites par saison (1.85 en moyenne), bien que les fauches précoces empêchent la plupart du temps la réalisation de la seconde ponte.

Figure 5: Exemple de sites typiques de reproduction du râle des genêts. (gauche : GB, droite, Basses vallées angevines)

1.3.1.4. Répartition et comportement migratoire

On retrouve le Râle des genêts, en période de reproduction (avril – septembre), dans une grande partie du Paléarctique (Figure 6). La distribution de l'espèce est alors limitée au sud à 40°N environ (Balkans, Arménie, Kirghizistan) et au nord à 62°N environ (sud de la Scandinavie). Longitudinalement, on le retrouve de l'Europe de l'ouest à la Sibérie, jusqu'au lac Baïkal approximativement (120°E). Autrefois beaucoup plus continue, l'aire de distribution du Râle des

genêts s'est largement fragmentée en Europe de l'ouest avec la disparition de nombreux habitats favorable depuis le XIX^{ème} siècle (Green & Rayment 1996). L'aire d'hivernage de l'espèce est beaucoup moins bien connue, tout comme les couloirs de migration empruntés. D'après le peu de données disponibles, l'espèce est considérée présente en hivernage dans les savanes d'Afrique du Sud et des pays du sud-est de l'Afrique (Mozambique, Zimbabwe, Tanzanie, voir Figure 6) (Schäffer & Koffijberg 2004). Pourtant, une étude récente remet en cause cette vision, en révélant que des individus observés en Ecosse hivernent également en Afrique de l'ouest, notamment dans des zones de prairie ouvertes au cœur de la forêt congolaise (Green 2013). Bien que la migration soit également peu documentée, on connait l'utilisation du détroit du Nil comme escale migratoire par le Râle des genêts, où il est chassé en grand nombre (Baha El Din *et al.* 1996). On sait également que le détroit de Gibraltar est emprunté par les individus venant ou remontant en Europe de l'ouest (Green 2013), tandis qu'on suppose le passage par la péninsule arabique pour les individus asiatiques (Schäffer & Koffijberg 2004).

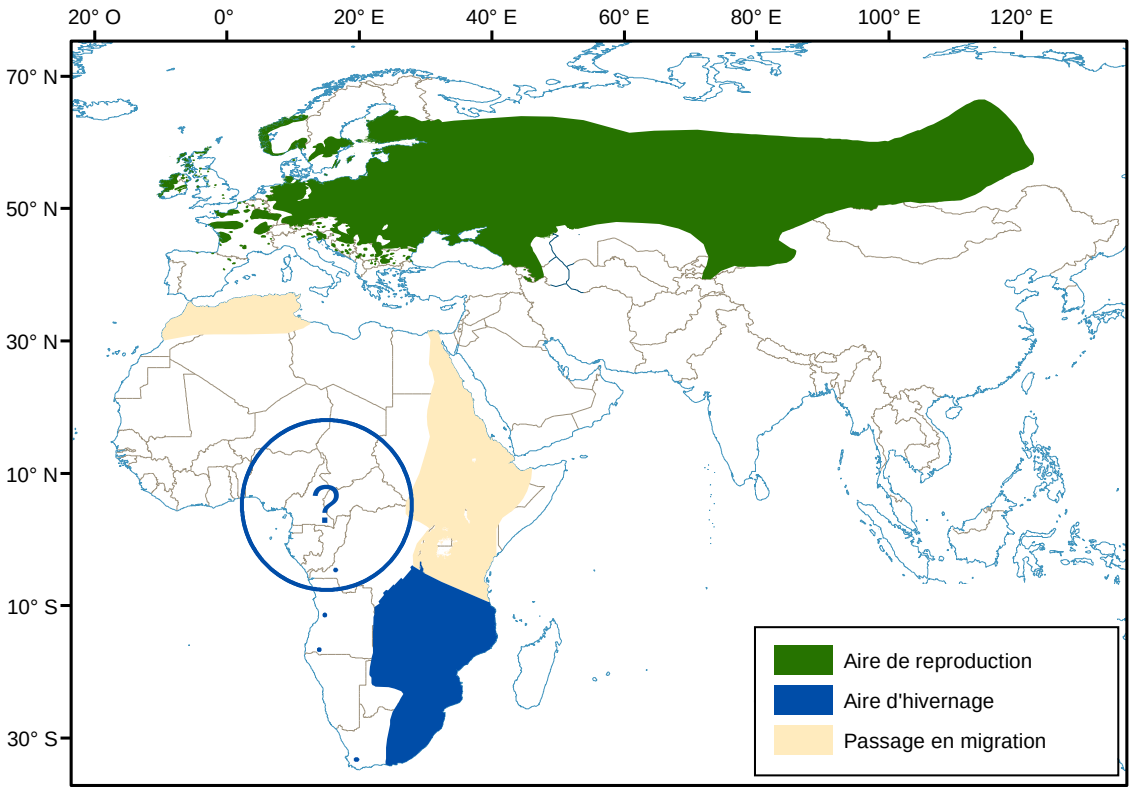


Figure 6: Carte de répartition du Râle des genêts. En vert figure l'aire de distribution en période de reproduction, d'avril à septembre environ selon les régions. En bleu est représentée l'aire d'hivernage traditionnellement connue de l'espèce, bien que des données récentes indiquent un possible hivernage en Afrique occidentale (cercle bleu). Les zones de transit connues lors de la migration sont figurées en beige. Adapté de Birdlife / IUCN.

1.3.2. Conservation

1.3.2.1. Situation globale

La problématique de conservation du Râle des genêts revêt des enjeux très différents selon la zone géographique considérée. Au niveau global, l'espèce est classée en « préoccupation mineure » par l'UICN (Birdlife International 2012), mais la situation est très contrastée entre l'Europe de l'ouest d'un côté, et de l'autre l'Europe de l'est ainsi que la partie asiatique de sa distribution.

La majorité des pays d'Europe occidentale ont ainsi vu leur population de Râle des genêts s'effondrer depuis plusieurs décennies. Plusieurs études font état d'un déclin depuis la fin du XIX^{ème} siècle, notamment en Allemagne, en Angleterre, en Irlande ou au Danemark (Green *et al.* 1997a). Dès 1939, il est montré que l'espèce a disparu de la majorité de son aire de distribution au Royaume-Uni, passant d'une répartition dans l'ensemble du pays au XIX^{ème} siècle à quelques sites anglais et écossais (Norris 1947). La généralisation des comptages de mâles chanteurs durant la deuxième moitié du XXème siècle atteste de la chute considérable des effectifs dans la plupart des pays d'Europe de l'ouest : en Angleterre, un premier recensement chiffré des sites de reproduction en 1970 donne une estimation de 5000 mâles (Sharrock 1976), puis les comptages précis qui se mettent en place par la suite montrent une chute jusqu'à 480 mâles en 1993 (Green 1995). Sur la période 1970–1990, des réduction de population de plus de 50% sont enregistrés en Irlande, aux Pays-Bas, en Norvège, en Pologne, ainsi qu'au Danemark (Green *et al.* 1997a). La même tendance est observée en France (Broyer 1994) (voir section 1.3.2.2 France).

Les causes du déclin de l'espèce en Europe sont bien connues et ont été documentées dès la mise en évidence de celui-ci. La mécanisation de l'activité agricole apparue avec la révolution industrielle, responsable d'une intensification croissante des pratiques, est sans aucun doute le principal facteur responsable de la chute des populations de Râle des genêts (Green & Rayment 1996; Tyler *et al.* 1998). En effet, les prairies alluviales fréquentées par l'espèce sont la plupart du temps exploitées pour la production de foin, activité qui permet d'ailleurs le maintien de cet habitat en limitant l'enfrichement et en le préservant d'une agriculture encore plus destructrice comme la conversion en culture ou en peupleraie. Contrairement aux fauches manuelles, les fauches mécaniques de plus en plus rapides et utilisant des barres de coupe plus longues contribuent à la mortalité directe des jeunes non volants, voire d'adultes, ainsi qu'à la destruction des nids (Schäffer & Green 2001). Cela est particulièrement vrai lorsque la fauche s'effectue de l'extérieur vers le centre de la parcelle, puisque cela contribue à piéger les individus dans la bande

centrale qui est fauchée en dernier (Green 1996; Green *et al.* 1997b). De plus, les méthodes actuelles permettent la fauche rapide de très grandes surfaces de prairies, offrant donc peu de refuges pour les jeunes ayant échappé à la destruction directe (Schäffer & Green 2001). Enfin, les dates de fauches sont particulièrement inadaptées à la phénologie du Râle des genêts, puisqu'elles surviennent au cœur de la période de reproduction de l'espèce (Broyer 1994; Green 1996; Deceuninck *et al.* 1997; Green *et al.* 1997a, 1997b). De ce fait, en plus de la mortalité importante due aux fauches, la deuxième ponte, qui survient normalement en juillet, a très peu de chances d'aboutir du fait de l'absence d'une couverture végétale favorable en été (Broyer 1995). Dans un contexte de survie des adultes très faible (0.18-0.34, Green 1999), la production de jeunes apparait pourtant l'élément central de la persistance des populations.

L'ampleur du déclin des populations de Râle des genêts a justifié la mise en place de mesures de conservation (Schäffer & Weisser 1996) déclinées par des plans d'action européens (Crockford *et al.* 1997; Koffijberg & Schäffer 2006) et nationaux (par exemple Irlande (NPWS 2005) et France (Hennique *et al.* 2013)). Parmi les actions proposées dans le cadre de mesures agri-environnementales figure le retardement des dates de fauche à une période plus favorable, en permettant notamment la ponte, la croissance et l'envol des jeunes avant la destruction de leur habitat (Broyer 1995; Green 1999). Ce retard de fauche entraine une dépréciation de la qualité du foin lorsqu'il est récolté au-delà de la date optimale de maturité de la prairie. En conséquence, ce type de mesures s'accompagne d'un dédommagement financier pour l'exploitant afin de compenser la perte de qualité fourragère et de rendement. Toutefois, le maintien de l'activité de fauche est essentiel pour ces milieux puisqu'il permet de conserver l'habitat prairial, rendant indispensable la mise en place d'un compromis entre viabilité de économique de ces prairies et compatibilité avec la biodiversité. De même, la fauche centrifuge permet de repousser les individus vers l'extérieur de la prairie fauchée au lieu de les emprisonner au centre, augmentant ainsi considérablement le taux de survie (Broyer 1996; Green *et al.* 1997b; Green 1999). Le simple ralentissement de la fauche limite également la mortalité directe des individus (Schäffer & Koffijberg 2004). Enfin, des bandes refuges non fauchées permettent de maintenir un couvert végétal et potentiellement des ressources alimentaires pour les individus après les fauches (Broyer 2003). Bien que le déclin se poursuive dans plusieurs pays d'Europe de l'ouest, le succès de ces mesures de conservation au Royaume-Uni (Schäffer & Koffijberg 2004; O'Brien *et al.* 2006) met en évidence leur potentiel pour enrayer la régression des populations de Râle des genêts.

À l'inverse de la situation décrite en Europe occidentale, le statut de conservation de l'espèce apparaît bien plus favorable en Europe de l'est et en Asie. Initialement, le peu de

données disponibles sur le Râle des genêts hors d'Europe de l'ouest laissait supposer un déclin similaire en Russie et en Europe de l'est. Des réduction de populations sont ainsi reportées en Russie européenne dès les années 30 (Green *et al.* 1997a). Cette baisse généralisée, et la connaissance de l'espèce limitée aux populations d'Europe occidentale en très fort déclin, avaient motivé le classement UICN de l'espèce en « menacé » en 1988. Cependant, au début des années 2000, la nécessité de recenser plus précisément les populations russes et orientales de Râle des genêts est devenue évidente. Un suivi régulier de l'espèce en Russie européenne est ainsi mis en place (Sukhanova & Mischenko 2003), permettant d'obtenir une vision plus précise des densités en Russie et d'estimer les tendances démographiques (Mischenko 2008). Ces données mettent en évidence des densités importantes sur tous les sites suivis, et une extrapolation à l'ensemble de l'habitat prairial disponible en Russie permet de considérer le Râle des genêts comme une espèce beaucoup plus commune et répandue qu'on le croyait auparavant (Broyer *et al.* 2007). De plus, des suivis de populations en Europe de l'est, et notamment en Lettonie, ont permis de constater que la déprise agricole qui a suivi la chute de l'URSS a contribué à l'augmentation des populations de Râles des genêts dans les pays concernés (Keišs 2003, 2005). En effet, des anciennes surfaces agricoles abandonnées constituent maintenant un habitat favorable pour l'espèce. Toutefois, ces milieux s'avèrent être temporaires, puisque leur évolution à moyen terme ne peut être que l'enfrichement total ou la reprise d'une activité agricole. De nouvelles estimations des effectifs en Europe de l'est et en Russie a entrainé le reclassement progressif du classement UICN de l'espèce, de « menacé » en 1988 à « préoccupation mineure » en 2010, en passant par « vulnérable » en 1994 et « quasi menacé » en 2004 (Birdlife International 2012). A l'heure actuelle, on estime que 80% de l'espèce se reproduit en Russie (1 500 000 à 3 000 000 de mâles chanteurs) dans des habitats peu menacés pour le moment (Schäffer & Koffijberg 2004).

1.3.2.2. En France

Les populations de Râle des genêts en France ont connues une évolution similaire à la plupart des pays d'Europe de l'Ouest. Depuis les premières estimations, puis les comptages annuels systématiques des mâles chanteurs, un déclin très prononcé a été observé (Figure 7) (Broyer 1991, 1994). Alors que la population était estimée à 2400 mâles environ en 1976, moins de 400 chanteurs étaient comptabilisés en 2012 (Hennique *et al.* 2013). Alors que la chute semblait stabilisée dans les années 1990, une diminution brutale des effectifs recensés est enregistrée entre 2000 et 2001. De même, le début des années 2000 semblait montrer un maintien des effectifs un peu au-dessus de 500 chanteurs, mais le déclin a rapidement repris après quelques années stables.

La tendance à long terme est tellement à la baisse que, si les effectifs diminuent de façon linéaire, on pourrait s'attendre à l'extinction de l'espèce en France dès 2015 (Figure 7).

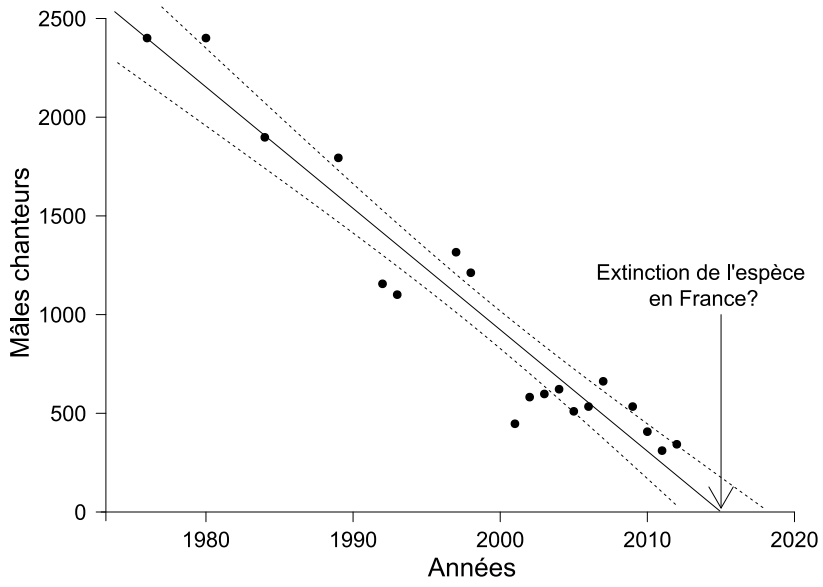


Figure 7: Résultats des comptages de Râles des genêts en France, exprimés en nombre de mâles chanteurs, entre 1976 et 2012. La droite de régression montre l'évolution linéaire du nombre de chanteurs dénombrés en fonction de l'année ± l'intervalle de confiance ($F = 202$, $P < 0.001$, $R^2 = 0.92$). D'après données LPO.

En plus de la très forte baisse des effectifs de Râles de genêts recensés, on observe une contraction de la distribution de l'espèce en France (Figure 8). Même si les données anciennes sont rares, on estimait dans les années 1930 que l'espèce était présent dans l'ensemble des départements métropolitains à l'exception du pourtour méditerranéen (Mayaud 1936). Le Râle des genêts a connu une régression progressive sur le territoire français et une concentration de ses effectifs dans les grandes vallées alluviales du territoire métropolitain: Loire et Basses Vallées Angevines, Charente, Meuse, Saône, Oise (Figure 8). Au fil des ans, la contribution des Basses Vallées Angevines dans les effectifs français a augmenté, faisant de ce site le dernier à disposer d'une population significative (plus de 50% de la population française en 2012).

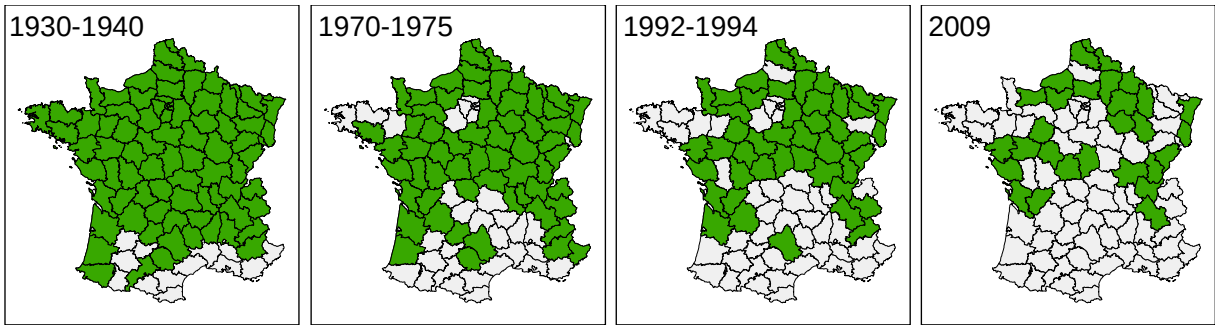


Figure 8: Départements occupés par le Râle des genêts en France (vert) de 1930 à 2009, découpé en 4 périodes. D'après données LPO.

Des mesures agri-environnementales (MAE) ont été mises en place en France à partir du début des années 1990 pour tenter d'enrayer le déclin de l'espèce. Bien que leurs modalités exactes puissent varier en fonction des régions, leur mode d'action est toujours basé sur le retard de la date de fauche contre une indemnisation destinée à compenser la perte de qualité du foin récolté à une date non optimale. Ainsi, dans le site majeur des Basses Vallées Angevine, trois contrats sont proposés aux agriculteurs dans le cadre des mesures agroenvironnementales territorialisées (MAET) : fauche des prairies interdite avant le 20 juin, avant le 10 juillet ou avant le 20 juillet, en échange d'un dédommagement de 193 €, 268 € et 321 € par hectare respectivement (source : chambre d'agriculture du Maine-et-Loire). Selon les régions, ces mesures peuvent être accompagnés de préconisations, voire d'obligations, de faucher à vitesse ralentie et du centre vers l'extérieur de la parcelle. En outre, les dates de contrat peuvent varier même à l'échelle d'un département en fonction des caractéristiques du milieu agricole, mais ces variations sont en grande partie déconnectées de la réalité de la phénologie du Râle. Bien que des améliorations sensibles de la survie étaient enregistrées dans les secteurs contractualisés (Deceuninck *et al.* 1997), l'évaluation de l'efficacité de ces mesures reste ancienne au regard de l'évolution rapide des pratiques, et on constate qu'elles apparaissent insuffisantes pour stopper le déclin du Râle des genêts en France (Figure 8). La fauche simultanée de grandes surfaces prairiales sur un même secteur correspondant aux dates de contrat, la première date de contrat trop précoce (dédiée en fait au maintien de surfaces en herbe mais inadaptée à la phénologie du Râle des genêts), ainsi que la trop rare utilisation de bandes refuges ou d'autres méthodes d'atténuation (faible vitesse de fauche, barre de coupe courte, fauche centrifuge, maintien de zones refuge), font que l'avenir de l'espèce en France apparait fortement compromis à court terme.

Chapitre 1 : Modélisation de l'aire de distribution

2. CHAPITRE 1 : MODELISATION DE L'AIRE DE DISTRIBUTION

Le Râle des genêts, considéré initialement en déclin au niveau global, a vu son statut de conservation IUCN passer de « Menacé » en 1988 à « Préoccupation mineure » en 2010 (Birdlife International 2012) du fait de nouvelles estimations de population en Europe de l'Est et en Russie qui ont mis en évidence des effectifs élevés et des tendances démographiques favorables (Keišs 2004; Mischenko 2008). Cependant, lorsque l'on observe plus précisément la source de ces données, on s'aperçoit qu'un seul expert local est à l'origine de quelques suivis de populations ponctuels (Sukhanova & Mischenko 2003; Broyer *et al.* 2007; Mischenko 2008) dont ont été extrapolées toutes les estimations de tailles de populations et de distribution. Dès lors se pose la question de la fiabilité de ces estimations, à plus forte raison lorsqu'elles déterminent un changement radical du statut de conservation international de l'espèce.

A cet égard, les méthodes de modélisation de distribution fournissent un outil puissant pour prédire la distribution spatiale des zones favorables à une espèce (Guisan & Zimmermann 2000). Toutefois, le déséquilibre considérable des données disponible entre Europe de l'ouest et Russie pose également le problème de la fiabilité de modèles réalisés à partir de données présentant un tel biais. La prise en compte du biais d'échantillonnage dans la réalisation de modèles de distribution est d'autant plus essentielle lorsqu'il s'agit de déterminer le statut de conservation de l'espèce. A ce titre, le Râle des genêts présente une situation très intéressante pour aborder ces questions en raison d'un biais d'échantillonnage très prononcé, caractérisé par une abondance de données inverse de l'abondance des populations. La résolution de ce problème présente un enjeu pour évaluer la robustesse de l'aire de distribution du Râle estimée à dire d'expert, pour une espèce dont le statut a beaucoup évolué en peu d'années.

Ce chapitre développe ainsi une approche de modélisation de distribution du Râle des genêts qui doit permettre d'apporter à la fois des éléments à la connaissance de l'écologie de l'espèce mais également de montrer le potentiel de la modélisation pour confirmer ou infirmer les prédictions à dire d'expert. En effet, les deux articles formant le corps de ce chapitre ont pour objectifs principaux, (i) dans un premier temps d'améliorer la mise en œuvre des procédures de modélisation de distribution utilisant un jeu de données biaisé spatialement, puis (ii) de préciser par une méthode de modélisation la situation globale du Râle des genêts, à confronter aux

données issues de la connaissance experte seule comme utilisée dans l'évaluation des statuts de conservation par les organismes internationaux.

2.1. ARTICLE 1 : CORRECTION DU BIAIS D'ECHANTILLONNAGE

Il est maintenant bien documenté que l'utilisation d'un jeu de données spatialement biaisé dans une procédure de modélisation de distribution peut conduire à des modèles largement erronés lorsque les occurrences disponibles sont également biaisés dans l'espace environnemental (Leitão *et al.* 2011; Bystriakova *et al.* 2012), c'est-à-dire lorsque des pans entiers de la niche écologique de l'espèce sont sur- ou sous-représentés. Bien que le problème du biais d'échantillonnage soit connu et que de nombreuses études utilisent des données d'observation opportunistes (Dennis & Thomas 2000) typiquement affectées par ce type de biais, cet aspect est rarement pris en compte dans les études de modélisation publiées. Pourtant, différentes méthodes ont été développées pour résoudre au mieux le biais (Osborne & Suárez-Seoane 2002; Dudík *et al.* 2005; Phillips 2008; Phillips *et al.* 2009; Rödder & Lötters 2009; Hijmans & Elith 2012), mais leur performance respective n'a été évaluée que très récemment (Syfert *et al.* 2013; Kramer-Schadt *et al.* 2013; Varela *et al.* 2014; Boria *et al.* 2014). Il n'existe à l'heure actuelle pas de consensus sur la meilleure méthode à utiliser dans un contexte de biais d'échantillonnage.

Afin de clarifier la situation et d'identifier la meilleure technique de correction de biais, nous avons mené une évaluation comparée des cinq méthodes habituellement utilisées. Deux jeux de données empiriques et une espèce virtuelle ont servi de base à l'étude, et ont été artificiellement biaisés en fonction des quatre types de biais les plus courants. A partir de ces jeux de données biaisés, cinq méthodes de correction du biais d'échantillonnage ont été appliquées lors d'une procédure de modélisation MAXENT. L'efficacité de la correction a alors été évaluée en comparant les modèles corrigés issus de données biaisées avec les modèles issus des données originales non biaisées. Outre l'AUC, méthode couramment utilisée dans l'évaluation des modèles de distribution, d'autres méthodes d'évaluation de la performance de correction ont été utilisées, en comparant les modèles en fonction de leur recouvrement dans l'espace environnemental et dans l'espace géographique (probabilités de présence relatives et cartes binaires). Cette évaluation cherche à identifier une méthode de correction applicable dans la majorité des types de biais auquel les chercheurs sont confrontés. Elle vise également à déterminer la méthode optimale de

correction du biais en amont de la modélisation que nous mènerons pour le Râle des genêts par la suite (Article 2).

Cette partie théorique, à laquelle ont contribués deux chercheurs de l'université de Göttingen et du Muséum de Bonn respectivement, a fait l'objet d'une publication dans le journal « PloS One » présentée ci-après :

Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MaxEnt using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS ONE*, **9**, e97122.

Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias

Yoan Fourcade[1*], Jan O Engler[2], Dennis Rödder[3], Jean Secondi[1]

[1] LUNAM, Université d'Angers, GECCO (Groupe écologie et conservation des vertébrés), 2 bd Lavoisier, 49045 Angers, France

[2] Department of Wildlife Ecology, University of Göttingen, D-37077 Göttingen, Germany

[3] Zoological Research Museum Alexander Koenig, D-53113 Bonn, Germany

*Corresponding author:

2.1.1. Abstract

MAXENT is now a common species distribution modeling (SDM) tool used by conservation practitioners for predicting the distribution of a species from a set of records and environmental predictors. However, datasets of species occurrence used to train the model are often biased in the geographical space because of unequal sampling effort across the study area. This bias may be a source of strong inaccuracy in the resulting model and could lead to incorrect predictions. Although a number of sampling bias correction methods have been proposed, there is no consensual guideline to account for it. We compared here the performance of five methods of bias correction on three datasets of species occurrence: one "virtual" derived from a land cover map, and two actual datasets for a turtle (*Chrysemys picta*) and a salamander (*Plethodon cylindraceus*). We subjected these datasets to four types of sampling biases corresponding to potential types of empirical biases. We applied five correction methods to the biased samples and compared the outputs of distribution models to unbiased datasets to assess the overall correction performance of each method. The results revealed that the ability of methods to correct the initial sampling bias varied greatly depending on bias type, bias intensity and species. However, the simple systematic sampling of records consistently ranked among the best performing across the range of conditions tested, whereas other methods performed more poorly in most cases. The strong effect of initial conditions on correction performance highlights the need for further research to

develop a step-by-step guideline to account for sampling bias. However, this method seems to be the most efficient in correcting sampling bias and should be advised in most cases.

2.1.2. Introduction

A key issue in ecology and conservation biology is to determine how species are distributed in space. Since extinction risk is associated with range size (Purvis *et al.* 2000), a significant reduction of a species range often determines change in conservation status (see for example IUCN criteria (IUCN 2001; Mace *et al.* 2008)) and prime conservations actions (Rodrigues *et al.* 2006; Harris & Pimm 2008). Likewise, protected areas usually focus on biodiversity hotspots (Myers *et al.* 2000) in order to conserve efficiently as many species as possible (Mittermeier *et al.* 1998; Reid 1998; Moilanen *et al.* 2005). Therefore, conservationists often need precise assessments of species ranges. Beyond simple range description, identifying which main factors limit distributions is essential to efficiently forecast the benefits of conservation management. In order to deal with these questions, several methods of species distribution modeling (SDM), also known as ecological niche modeling (ENM) (Peterson 2006), have been developed since the 1980s (Ferrier 1984).

The principle of SDM is to relate known locations of a species with the environmental characteristics of these locations in order to estimate the response function and contribution of environmental variables (Austin *et al.* 2006), and predict the potential geographical range of a species (Elith & Leathwick 2009). These models estimate the fundamental ecological niche in the environmental space (*i.e.* species response to abiotic environmental factors (Soberón & Peterson 2005)) and project it onto the geographical space to derive the probability of presence for any given area or, depending on the method, the likelihood that specific environmental conditions are suitable for the target species (Elith *et al.* 2011). Distribution models are used by conservation practitioners to estimate the most suitable areas for a species and infer probability of presence in regions where no systematic surveys are available (Elith & Burgman 2002). They can also assess the potential expansion of introduced species in newly colonized areas (Jeschke & Strayer 2008; Jimenez-Valverde *et al.* 2011), estimate the future range of a species under climate change (Jeschke & Strayer 2008; Sinclair *et al.* 2010) or assist in reserve planning (Thorn *et al.* 2009).

Several statistical models exist to predict the distribution of a species (Franklin 2009). Beyond classical regression methods (Resource Selection Function RSF (Manly *et al.* 1993; Boyce & McDonald 1999), Generalized Linear Models GLM (McCullagh & Nelder 1989)), algorithmic

modeling based on machine learning (for example Artificial Neural Networks (Ripley 1996), Maximum Entropy MAXENT (Phillips *et al.* 2004), Classification And Regression Trees CART (Breiman *et al.* 1984)) have become increasingly popular in recent years. Among these, MAXENT has been described as especially efficient to handle complex interactions between response and predictor variables (Elith *et al.* 2006, 2011), and to be little sensitive to small sample sizes (Wisz *et al.* 2008). This, as well as its extreme simplicity of use, has made MAXENT the most widely used SDM algorithm. In December 2013, 1886 citations of the article describing the method (Phillips *et al.* 2006) were reported in Web of Science.

MAXENT modeling, and SDM in general, is now commonly implemented in conservation-oriented studies (Elith & Leathwick 2006). Regional or continent-wide studies are facilitated by the recent availability of global datasets. Environmental layers, such as the global climate variables developed in the WorldClim project (Hijmans *et al.* 2005), offer continuous description of very large areas (Kozak *et al.* 2008). Similarly, the development of open biodiversity databases (see for example the Global Biodiversity Information Facility, GBIF, http://www.gbif.org) increases manifolds the spatial coverage of fieldwork observations that could have been collected by a single project. Such databases usually provide presence-only data that can be handled by modeling methods like MAXENT.

However, datasets derived from opportunistic observations or museum records rather than from planned surveys often exhibit a strong geographic bias (Dennis & Thomas 2000), some areas being visited more often than others because of their accessibility (Kadmon *et al.* 2004) or their naturalistic interest. This unequal survey coverage of a species distribution is often referred as sampling bias, sample selection bias or survey bias. The quality of the model can be strongly affected if entire parts of the environmental space suitable to a species are absent or poorly represented in the survey dataset (Leitão *et al.* 2011; Bystriakova *et al.* 2012), or alternatively, if some areas are overrepresented due to locally high sampling efforts. Several studies questioned the effect of sampling design (Edwards *et al.* 2006), or the biased nature of museum and herbarium datasets (Loiselle *et al.* 2008) on the predictive performance of SDMs. Surprisingly, the issue of quantifying and correcting sampling bias has been poorly addressed despite its crucial importance. Although authors pointed out that the distribution of locations in the geographical and/or ecological space may impact the reliability of the model (Reddy & Davalos 2003; Kadmon *et al.* 2004; Costa *et al.* 2009; Leitão *et al.* 2011; Syfert *et al.* 2013; Beck *et al.* 2013), the potential effect of the sampling bias in the dataset is usually poorly taken into account or not considered at all. However, very different SDM outputs can be generated that lead to

contrasting conclusions whether sampling bias is corrected or not (Fourcade *et al.* 2013), making SDM studies that did not incorporate this issue highly doubtful.

In regard to the considerable influence of sampling bias on SDM prediction ability, Araújo et al. (2006) considered the improvement of sampling designs as one of the five major challenges for future development of SDMs. Several bias correcting methods have been proposed (Osborne & Suárez-Seoane 2002; Dudík *et al.* 2005; Phillips 2008; Phillips *et al.* 2009; Rödder & Lötters 2009; Hijmans & Elith 2012) but they have been rarely used so far. Comparison and evaluation of different methods to correct sampling bias have only been recently carried out and no consensual guideline emerged to solve it. A few recent studies explored the consequences of and potential solutions to correct for sampling bias (by Syfert et al. (2013), Kramer-Schadt (2013), Varela et al. (2014) and Boria et al. (2014)). In spite of their interest, authors investigated a single case study so that it is not possible to evaluate the efficiency of a correction method across species. Second, the empirical bias caused by sampling intensity (Syfert *et al.* 2013) was never tested and no more than two correction techniques have been evaluated at the same time whereas many more have been proposed or used in the literature (Dudík *et al.* 2005; Phillips 2008; Veloz 2009). Therefore, the influence of the nature and intensity of bias on the capacity of various techniques to correct for sampling bias has not been investigated. This remains however a critical issue, especially for users who need robust and reliable SDM predictions such as conservation practitioners.

The goal of this comparative study is to test the effect of bias type, bias intensity, and correction method on MAXENT model performance. Unlike the previous cited studies (Syfert *et al.* 2013; Kramer-Schadt *et al.* 2013; Varela *et al.* 2014; Boria *et al.* 2014) we assessed the performance of five bias correcting methods among the most frequently used under various conditions of bias type and intensity. We used a virtual species to generate four types of sampling biases and three bias intensities, and applied on these biased datasets different corrections. We quantified the relative correction performance across the range of bias conditions and across species. The same framework was also applied on two real datasets. The full workflow on which analyses were based is sketched in Figure 9. Therefore, the present study provides for the first time a comprehensive multi-species evaluation of the most common methods of sampling bias correction under different scenarios of bias and intensities of bias. Intended for conservationists who use MAXENT on a regular basis, we expect this work to provide insights on the selection of the most suitable methods to produce reliable distribution models using biased datasets.

Furthermore, it encourages modelers to develop improvements of techniques to correct for sampling bias suitable for the vast set of modeling methods available.

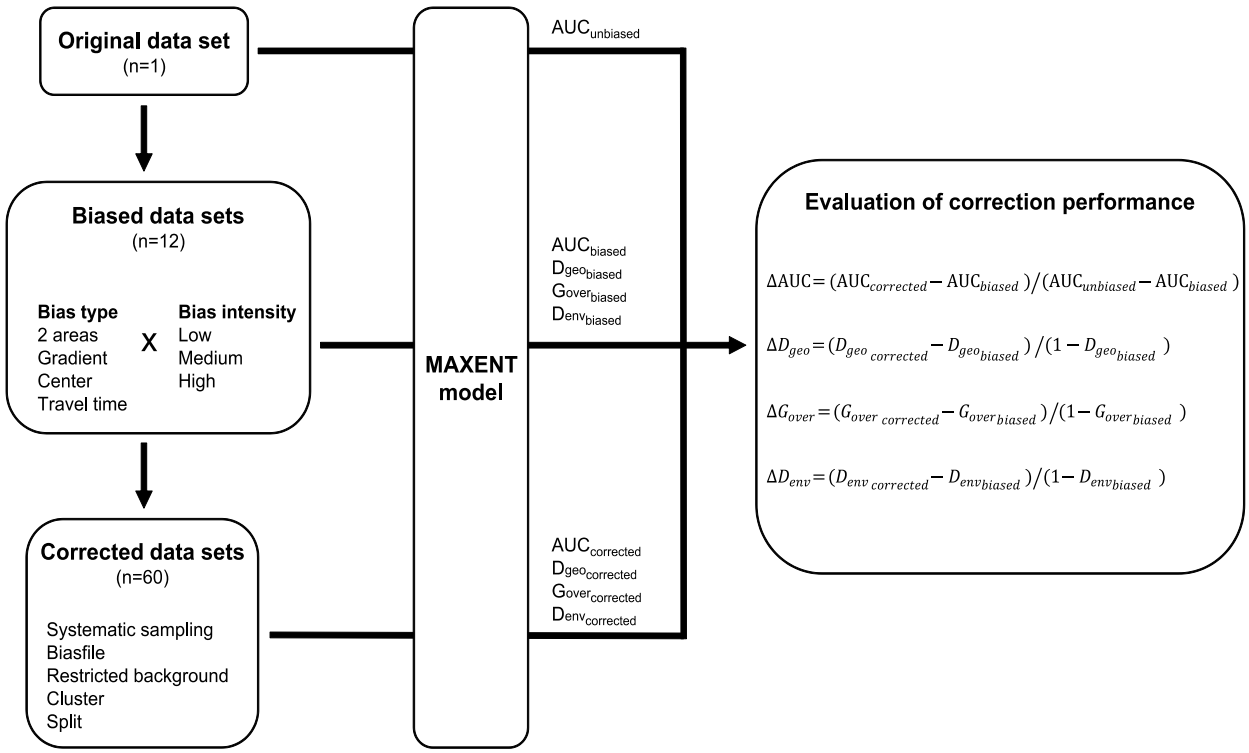


Figure 9: Workflow used in analyses. Original datasets of a virtual and 2 real species were altered to create 12 bias combining 4 bias types and 3 bias intensities. Five methods of sampling bias correction were employed to assess the improvement in the modeled distribution relative to the original distribution using MAXENT. Correction performance was assessed using AUC and 3 measures of overlap between the corrected the original unbiased model.

2.1.3. Materials and Methods

2.1.3.1. Species datasets

In order to obtain a true unbiased dataset, we created a virtual species by randomly sampling a set of points in a given environment determined by a single categorical variable. A similar approach has been used formerly (Kramer-Schadt *et al.* 2013). We extracted 2000 random coordinates from the "Closed to open shrubland" category of the North American Globcover map (Globcover 2009, http://due.esrin.esa.int/globcover (Arino *et al.* 2008)) to generate an unbiased dataset. The geographical extent of the virtual species was chosen to match the scale of the two real species ranges which are also both located in North America. In practice, the virtual species covered the

larger part of western North America, including the lower valleys of the Rocky Mountains, Mojave Desert, Baja California peninsula, and Northern Mexico (Figure 10).

We additionally used the occurrence datasets of two real species (Figure 10). We compiled a set of 1825 occurrences for the painted turtle (*Chrysemys picta*) downloaded from the World Turtle Database (http://emys.geo.orst.edu; accessed on May 2011). The second species was the white-spotted slimy salamander (*Plethodon cylindraceus*) for which we collected a set of 208 observations. A part of the dataset was provided by J. Milanovich and additional records were obtained from the Global Biodiversity Information Facility (http://www.gbif.org, accessed on May 2011). According to our knowledge of the distribution of these species, these original datasets seem relatively unbiased, *i.e.* the distribution of records over space reflects the known spatial distribution of the species.

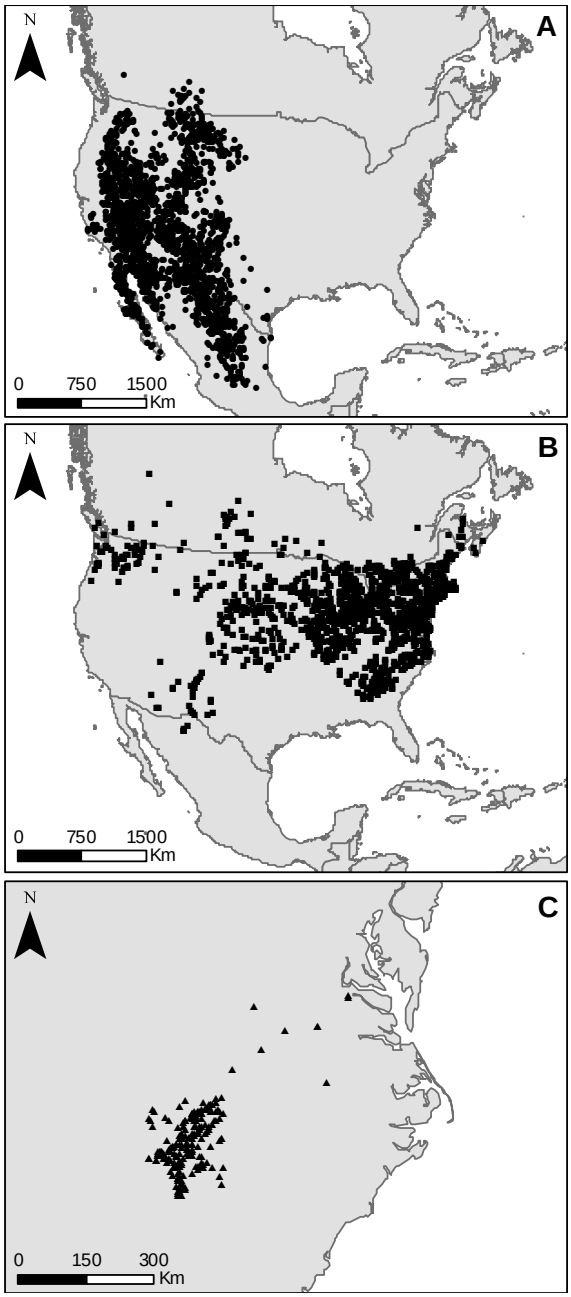


Figure 10: Locations of records used for modeling. (A) Virtual species; (B) *Chrysemys picta*; (C) *Plethodon cylindraceus*

2.1.3.2. Environmental predictors

We used distinct sets of environmental predictors depending on the modeled species. Climatic and topographic grids were downloaded from the WorldClim database (Hijmans *et al.* 2005) (http://www.worldclim.org) at a resolution of 2.5 arc-min (4.63 km at the equator). The global map of land cover provided by the European Spatial Agency was downloaded in its 2009 version (Globcover 2009 (Arino *et al.* 2008), http://due.esrin.esa.int/globcover) and rescaled to fit the 2.5

arc-min resolution of the other variables. Finally, we compiled 5 years of 10-day periods of NDVI (2007-2011), a measure of vegetation productivity derived from multispectral remote-sensing images, downloaded from the SPOT-VEGETATION project (Maisongrande *et al.* 2004) (http://free.vgt.vito.be). We averaged across these 5 years three layers of mean, minimum and maximum annual NDVI.

We removed for each species some highly intercorrelated (correlation coefficient computed by ArcGIS 10; > 0.9 or < -0.9) variables because multicollinearity may violate statistical assumptions and may alter model predictions (Heikkinen & Luoto 2006). The resulting variable sets were composed of 14 predictors (Table 1). Since the geographical distribution of the virtual species and *Chrysemys picta* records covered a large range in North America, we modeled both species across the same geographic area across North America. *Plethodon cylindraceus* occurrences are restricted to a smaller area of Eastern USA. Accordingly, the geographical range of predictors was restricted to a narrower area (Table 1).

Table 1: Environmental predictors included in MAXENT modeling for the virtual species, *Chrysmeys picta* and *Plethodon cylindraceus*. The selected layers are indicated for each species, as well as the extent of modeling (in decimal degrees).

		Virtual species	*Chrysemys picta*	*Plethodon cylindraceus*
Variable				
Altitude		×	×	×
Annual mean temperature		×	×	×
Mean diurnal range		×	×	×
Isothermality		×	×	×
Temperature annual range		×	×	×
Mean temperature of wettest quarter		×	×	×
Annual precipitation		×	×	×
Precipitation seasonality		×	×	×
Precipitation of warmest quarter		×	×	×
Precipitation of coldest quarter		×	×	×
Land cover		×	×	×
Maximum NDVI		×	×	
Mean NDVI		×	×	×
Minimum NDVI		×	×	×
Mean temperature of driest quarter				×
Extent				
Longitude				
	min	-140.00	-140.00	-88.63
	max	-50.00	-50.00	-70.68
Latitude				
	min	20.00	20.00	28.91
	max	75.00	75.00	45.46

2.1.3.3. Generation of sampling bias

The three original datasets were altered to generate four types of bias that might occur when collecting observations (Figure 11). The original datasets were thus subsampled so that the remaining records were biased in the geographical space. We also created three levels of bias intensity, hereafter referred as "low", "medium" and "high" to assess the effect of this parameter on model outputs (Figure 11). For each species, each combination of bias type (4) and bias intensity (3) was replicated 10 times resulting in a total of 360 biased datasets used to model distribution. The four types of sampling bias were generated as follows:

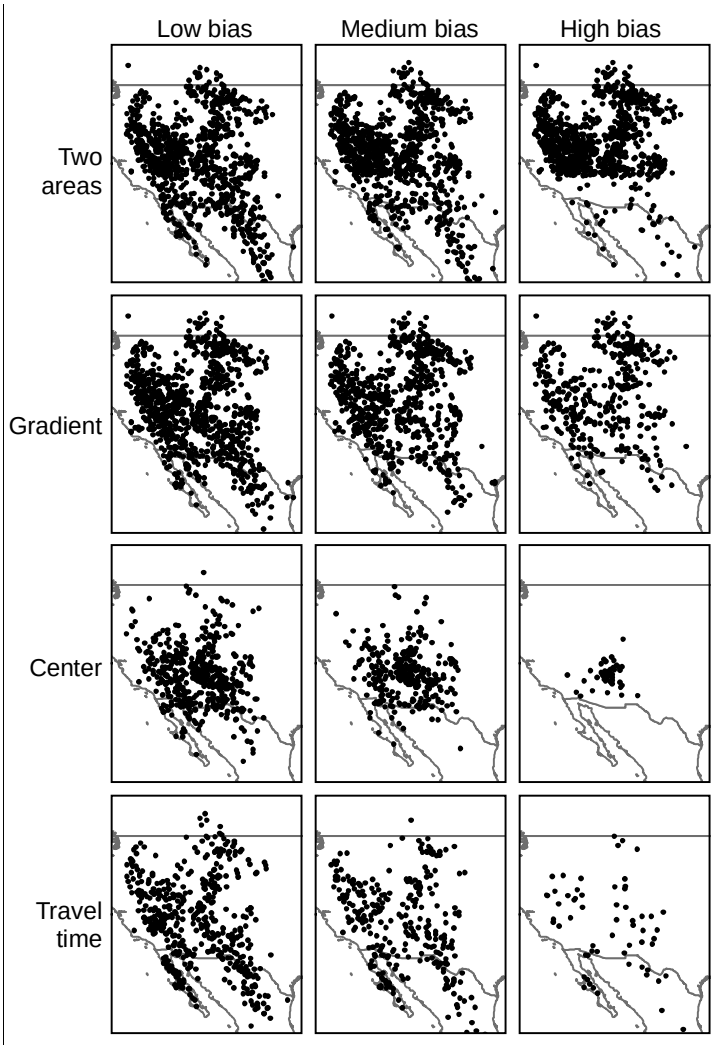


Figure 11: Generation of sampling bias for the virtual species. To generate artificial sampling bias, the original dataset (here the virtual species) was altered into 4 different types of bias (rows), each with 3 intensities (columns).

(1) TWO AREAS - The original dataset was biased such that its northern part exhibited a high density of records and the southern part a low density. This kind of bias is common when a species is systematically monitored in one part of its range and not surveyed in the other, for instance in different countries or groups of countries (Fourcade *et al.* 2013).

(2) GRADIENT - We generated a density gradient of observations decreasing from the north to the south of the range. This bias is close to the first one but here record density changed gradually. Such a bias would not reflect a difference in survey schemes between administrative divisions but a gradual reduction of sampling intensity towards a limit of species range.

(3) CENTER - The density of occurrences gradually decreased from the core of the distribution to the periphery. Such bias mimics cases in which sampling effort is concentrated in the centre of the known range of the species, whereas peripheral areas, potentially less suitable (Sagarin & Gaines 2002), are neglected.

(4) TRAVEL TIME - We used the travel time to the nearest city, using a map produced by the European Commission (Nelson 2008) (available at bioval.jrc.ec.europa.eu/products/gam). This variable integrates both the distance to the city and the presence of road networks. This map was used as a grid of sampling probability, in which probability of keeping a record was highest close to cities and in areas with dense road networks. This bias corresponds to a common situation where most of records are located around cities or along roads (Kadmon *et al.* 2004; McCarthy & Fletcher 2012).

The full details of the generation of sampling bias are given in Supporting information, supplementary material S1.

2.1.3.4. Species distribution modeling

We used for modeling the software MAXENT (Phillips *et al.* 2006), a machine learning algorithm that applies the principle of maximum entropy to predict the potential distribution of species from presence-only data and environmental variables (Phillips *et al.* 2004). Currently, this widely used method is particularly efficient to handle complex interactions between response and predictor variables (Elith *et al.* 2006, 2011), and is little sensitive to small sample sizes (Wisz *et al.* 2008). All models were computed using the version 3.3.3k of MAXENT (http://www.cs.princeton.edu/~schapire/maxent/). Runs were conducted with the default

variable responses settings, and a logistic output format which results in a map of habitat suitability of the species ranging from 0 to 1 per grid cell, wherein the average observation should be close to 0.5 (Elith *et al.* 2011). The models were evaluated by the area under the ROC curve (AUC), and three measures of overlap with the unbiased model (see below section "Model evaluation and statistical analyses").

2.1.3.5. Methods of sampling bias correction

We applied on all our biased datasets five methods of bias correction that have been already published. In order to evaluate their usefulness in real conditions, we used these methods as if the source, shape or strength of the sampling bias was unknown. Therefore, we did not select a correction method according to our knowledge of the bias, as this information is unknown in most empirical studies.

(1) SYSTEMATIC SAMPLING – A subsample of records regularly distributed in the geographical space was selected (Veloz 2009; Hijmans & Elith 2012; Ihlow *et al.* 2012; Boria *et al.* 2014). MAXENT already discards redundant records that occur in a single cell. We removed neighboring occurrences at a coarser resolution than MAXENT does. We created a grid of a defined cell size and randomly sampled one occurrence per grid cell. This subsampling reduces the spatial aggregation of records but does not correct the lack of data due to low sampling effort in some areas. This method could also underestimate the contribution of suitable areas where the high density or records reflects the true ecological value for the species. The resolution of the reference grid was 2 degrees for *Chrysemys picta* and the virtual species, and 0.2 degree for *Plethodon cylindraceus*.

(2) BIAS FILE – This option is implemented in MAXENT. The software can be fed with a bias grid (Dudík *et al.* 2005; Elith *et al.* 2010) that is a sampling probability surface. The cell values reflect sampling effort and give a weight to random background data used for modeling. An ideal way of creating biasfiles would be to represent the actual sampling intensity across the study area. Although it can be roughly estimated by the aggregation of occurrences from closely related species (Phillips *et al.* 2009), in most real modeling situations, this information is lacking. Thus, instead of using our knowledge of the artificially created biases, we produced bias grids by deriving a Gaussian kernel density map of the occurrence locations, rescaled from 1 to 20, following Elith et al. (Elith *et al.* 2010). These maps were implemented in the biasfile option in MAXENT.

(3) RESTRICTED BACKGROUND – MAXENT, as most other presence-pseudoabsence methods, generates a "background" or "pseudo-absence" sample of points (Elith *et al.* 2011). It has been argued that the selection of background points may strongly affect the resulting model (Chefaoui & Lobo 2008; Barve *et al.* 2011; Acevedo *et al.* 2012). By default 10000 pseudo-absences are randomly selected from the whole rectangular study area. This approach was followed for all the other cases, as most SDM studies keep the default MAXENT selection of background points. However, according to Phillips (Phillips 2008), if occurrences are restricted to a fraction of the study area, model performance can be enhanced by drawing the background points from this fraction of the area. The reliability of predictions should be improved when the model is transferred to the rest of the area. Following this recommendation, we randomly sampled 10000 pseudo-absences in buffer areas around occurrences and used them as background samples in MAXENT. Buffer size was a radius of 500 km for the virtual species and *Chrysemys picta*, and 100 km for *Plethodon cylindraceus*.

(4) CLUSTER – Biased datasets typically lead to spatial autocorrelation of records and artificial spatial clusters of observations thus violating the assumption of independence (Dormann *et al.* 2007). This bias can be circumvented by sampling one point per cluster in environmental space (Rödder *et al.* 2009b; Stiels *et al.* 2011; Varela *et al.* 2014). We first performed a principal component analysis (PCA) on the environmental descriptors of occurrences using the "ade4" R package (Dray & Dufour 2007) in order to define independent axes in the environmental space. Then, we ran a cluster analysis based on Euclidean distance on PCA axes space. The resulting dendrogram was used to define a number of classes corresponding to half of the occurrences. One record was randomly sampled per class and the models were run on these subsampled datasets.

(5) SPLIT – When occurrence frequency greatly differs between two areas because of unequal sampling effort, the area can be split in two strata within which coverage probability is more homogeneous, and one model be computed for each stratum. This method has been used for species occurring over a large distribution range and extended environmental gradients (Osborne & Suárez-Seoane 2002; Gonzalez *et al.* 2011; Fourcade *et al.* 2013). We split our biased datasets in a northern and a southern stratum. We combined the model outputs to produce a composite model for the entire range, keeping the highest value in pixels where strata overlapped.

2.1.3.6. Evaluation of correction methods

To estimate the ability of each correction method to recover the information contained in the original unbiased data set, we used 4 criteria that correspond to the interests of different end users of SDMs. We compared the models obtained after applying a bias in the dataset to the original model, and after applying a sampling bias correction method. The original dataset of the virtual species was created to be unbiased. The model computed from it will be thus referred as the unbiased model. Although we do not formally know the actual bias in the *Plethodon* and *Chrysemys* original datasets, we will also refer to the models computed from their original datasets as unbiased models. We should keep in mind that we compare the change in the resulting model, whatever the original bias. The models referred as biased were computed after applying a sampling bias and the corrected models after applying a correction method to the biased dataset.

(1) AUC – The area under the receiver operating curve (ROC), known as the AUC is one of the most common statistics to assess model performance. AUC can be interpreted as the probability that a presence cell have a higher predicted value than a absence cell (or pseudo-absence), both of them being chosen randomly (Elith *et al.* 2006). Although the use of AUC for the evaluation of ecological models has been criticized, especially when calculating against background points rather than true-absences (Lobo *et al.* 2008; Jimenez-Valverde *et al.* 2012), it should be reliable to compare models generated for a single species in the same area and the same predictors.

The calculation of AUC was performed using the R package "PresenceAbsence" (Freeman & Moisen 2008), using as test points the fraction of the original dataset which was excluded to create the biased dataset. The metrics was calculated by the comparison between these test points and either (i) 10000 true absence points sampled outside the range, which corresponds here to the true occupancy for the virtual species, or (ii) 10000 background points randomly sampled in all the study area for the real species. For original models, in the absence of true test points, a mean AUC value was computed using 5 random splits of the dataset, each subsample being used in turn to evaluate the model.

(2) D_{GEO} overlap in the geographical space – Several metrics of niche overlap are available (e.g. *D* (Schoener 1968), modified Hellinger distance *I* (Warren *et al.* 2008) or *BC* (Bray & Curtis 1957)). We used the Schoener's *D* index (Schoener 1968) that has been suggested to be best suited for SDM outputs (Rödder & Engler 2011). This statistics considers the probability distributions across space of the difference in the probability of presence of two species, based on their

respective distribution models. D_{geo} index is ranged from 0 (no overlap) to 1 (complete overlap, identical models).

(3) D_{ENV} overlap in the environmental space – We estimated niche overlap in the environmental space between the unbiased, biased, and corrected models. We used the PCA-env approach described by Broennimann (Broennimann *et al.* 2012) to calculate Schoener's *D* index based on the environmental characteristics of two sets of occurrences. This approach defines the environmental space by the two first axes of a principal component analysis of all the pixels of the study area. The niche overlap is calculated from the smoothed density of occurrences in the environmental space following a kernel density function applied on each dataset. Depending on the correction method used, some models can be based on the same input dataset (*e.g.* the biasfile correction uses the same records as the biased model). In order to compare our SDMs, we applied the PCA-env method on 500 points sampled from the SDM outputs instead of on the input records. We selected these points in the output map using the SDM probability of presence as a sampling probability.

(4) G_{OVER} overlap between binary maps – SDM outputs are often converted to binary maps that are more tractable for conservationists who for instance need to delineate protected areas. In such maps, a pixel is considered as either suitable to the species or not. We used the non-fixed 10th percentile training presence threshold value to generate binary maps as proposed by Liu et al. (Liu *et al.* 2005). We then measured the overlap between biased and corrected models, and unbiased models. This measure strictly measures the geographical overlap whereas the *D* index estimates the overlap of ecological niches in environmental space (D_{env}) or projected onto the geographical space (D_{geo}).

To evaluate the performance of bias correction, we derived new indicators from AUC, D_{geo}, D_{env}, and G_{over} that quantify the improvement of the corrected model to the biased model, standardized by the difference between the unbiased and the biased models. These indices, named respectively ΔAUC, ΔD_{geo}, ΔG_{over} and ΔD_{env} were calculated as follows:

$$\Delta\mathrm{AUC} = (\mathrm{AUC}_{corrected} - \mathrm{AUC}_{biased})/(\mathrm{AUC}_{unbiased} - \mathrm{AUC}_{biased})$$

$$\Delta D_{geo} = (D_{geo}{}_{corrected} - D_{geo}{}_{biased})/(1 - D_{geo}{}_{biased})$$

$$\Delta G_{over} = (G_{over}{}_{corrected} - G_{over}{}_{biased})/(1 - G_{over}{}_{biased})$$

$$\Delta D_{env} = (D_{env}{}_{corrected} - D_{env}{}_{biased})/(1 - D_{env}{}_{biased})$$

These four indices range from -∞ to 1, a positive value indicating that the model was actually corrected (with 1 corresponding to perfect correction, *i.e.* corrected model exactly similar as the unbiased one) whereas a negative value indicates that the correction produced a worse model than the biased one.

2.1.4. Results

2.1.4.1. Effect of bias type and intensity on model outputs

The biased models resulted in a reduction of AUC in all cases and a deviation from the unbiased model for all overlap measures (Figure 12). The deviation varied largely depending on species and bias type though: 0.28–0.91 for D_{geo}, 0.18–0.89 for D_{env}, 0.03–1 for and G_{over}. For the virtual species, the bias type yielding the largest deviation depended on the performance measures considered: "2 areas" for the AUC, "travel time" for D_{geo}, "Center" for D_{env} and "Gradient" for G_{over}, the latter providing the lowest values of overlap with the original unbiased models. However, the differences between unbiased and biased models remained moderate (mean percentage variation: AUC: -1.74%; D_{geo}: -24.41%; D_{env}: -24.62%; G_{over}: -30.07%). All bias types had also globally similar effects in terms of deviation from the unbiased model (Figure 12). For *Chrysemys picta*, all evaluation measures were strongly affected by the "center" bias. AUC decreased more than 5%, and overlaps with the unbiased model ranged from 0.26 to 0.49. In contrast, the *P. cylindraceus* dataset was overall weakly affected by the biases so that the biased models did not lead to noticeable differences with the unbiased model. The strong effect of the "center" bias was visible in *Plethodon cylindraceus* only for the overlap of binary maps (G_{over}). This bias in which only the central zone is sampled may exclude a large part of the original environmental space and lead to very inaccurate SDM outputs. Interestingly, the decrease in AUC performance for all bias types was more pronounced in *C. picta* than the two other datasets even when the values of overlap were in the same range.

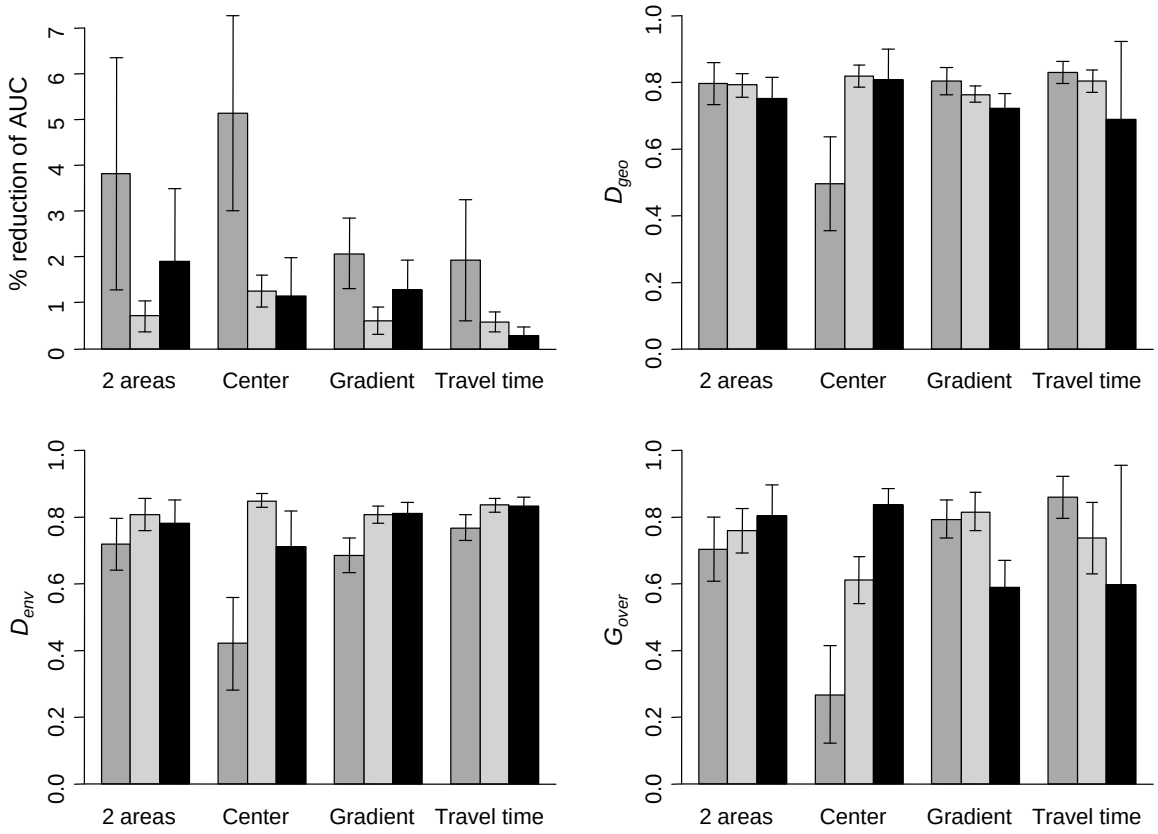


Figure 12: Evaluation indices of biased models across bias intensities. Reduction of AUC between unbiased and biased models (in percentage), Schoener's *D* overlaps between biased and unbiased models computed on SDMs (D_{geo}) and in environmental space (D_{env}), and G_{over} the binary distribution overlap (mean ± SD). Dark grey bars: *Chrysemys picta*, light grey bars: *Plethodon cylindraceus*, black bars: virtual species.

2.1.4.2. Relative performance of correction method

Since we evaluated the performance of correction methods using indices with different sets of assumptions, interpretation may slightly differ with the measure considered. However, as we mainly aimed at comparing SDM outputs, *i.e.* maps of habitat suitability, we primarily focused our interpretations on ΔD_{geo} which truly evaluates the overlap between standard SDM maps. Moreover, the two measures based on Schoener's *D*, in geographic (ΔD_{geo}) and in environmental space (ΔD_{env}), were highly correlated and outputs of bias correction were qualitatively similar (Supporting information, Fig. S1). We discuss here results for ΔD_{geo} only.

Correction performance strongly depended on the species (Table 2). Considering the three species together, less than half (29%) of all combinations (species × bias type × bias intensity ×

correction method) yielded corrected models (following ΔD_{geo}) with more accurate predictions than the biased model. For the virtual species, and considering ΔD_{geo}, 57 % of corrected models (34 out of 60 combinations of bias type, bias intensity and correction method) were more similar to the model generated with the unbiased dataset than the biased model (Table 2). Most cases for which no method was able to provide bias correction were "center" and "travel time" biases, with medium to high intensities. Conversely, only 7% of *P. cylindraceus* models were corrected (4 cases out of 60), while 25% of *C. picta* models were corrected (15 cases out of 60) and offered a better result than the biased model.

Table 2: Mean correction performance across 10 replicates, for each species, bias type, bias intensity and correction method. Positive values, *i.e.* cases where the bias was actually corrected, are shown in bold. For each combination of species and bias (type × intensity), the best method (*i.e.* the one which has the highest value) is highlighted by an asterisk. Correction performance is estimated by three measures: (a) ΔAUC, (b) ΔD_{geo}, (c) ΔG_{over}.

	2 areas			*Center*			*Gradient*			*Travel time*		
	low	medium	high	low	medium	high	low	medium	high	low	medium	high
(a) ΔAUC												
Chrysemys picta												
Biasfile	-0.64	-0.1	-0.01	-0.07	-0.03	**0.25***	-0.93	-0.61	-0.31	-2.09	-0.57	-0.2
Cluster	**0.09**	**0.09**	**0.28**	**0.24**	**0.19**	**0.12**	-0.08	-0.06	**0.01**	-1.12	-0.22*	-0.05*
Restricted background	-0.02	-0.21	-0.16	-0.43	-0.69	-0.85	-0.12	-0.22	-0.16	-0.96	-0.46	-0.62
Split	**0.60***	**0.64***	**0.74***	-0.12	-0.14	**0.1**	**0.10***	**0.02***	**0.14***	-0.94*	-0.28	-0.43
Systematic sampling	-0.85	**0.19**	**0.42**	**0.31***	**0.33***	-0.82	-0.97	-0.58	-0.16	-3.76	-1.17	-0.26
Plethodon cylindraceus												
Biasfile	**0.12**	**0.14**	-0.03	**0.1**	**0.09**	**0.22**	**0.03**	-0.24	-0.11	-0.07	-0.02	**0.02**
Cluster	**0.17**	**0.11**	**0.15**	**0.19**	**0.19**	**0.26**	**0.03**	**0.00***	-0.12	**0.02**	**0.02***	-0.08
Restricted background	-10.03	-10.45	-10.49	-5.36	-4.12	-5.04	-6.19	-10.64	-7.88	-12.74	-9.47	-11.12
Split	**0.03**	**0.01**	**0.2**	-0.12	-0.04	**0.2**	-0.13	-0.09	-0.18	-0.07	-0.07	-0.24
Systematic sampling	**0.31***	**0.32***	**0.42***	**0.51***	**0.51***	**0.50***	**0.13***	-0.09	**0.14***	**0.17***	**0.01**	**0.04***
Virtual species												
Biasfile	-3.3	**0.42**	**0.49**	**0.17**	**0.35***	**0.26**	-0.1	**0.34**	**0.41**	-2.87	-2.40*	-0.27*
Cluster	-0.75	**0.11**	**0.24**	-0.18	-0.15	-0.17	-0.01	**0.09**	**0.07**	-1.82	-3.3	-0.43
Restricted background	-2.09	-0.34	-0.78	-1.97	-1	**0.05**	-1.04	-1.26	-1.5	-9.87	-18.76	-6.43
Split	**0.18***	**0.72***	**0.82***	-0.2	-0.29	-0.17	**0.38***	**0.42***	**0.60***	-1.80*	-2.48	-0.53
Systematic sampling	-3.26	**0.45**	**0.72**	**0.39***	**0.33**	**0.35***	-0.83	-0.08	**0.39**	-2.54	-2.77	-0.69

(b) ΔD_{geo}												
Chrysemys picta												
Biasfile	-0.41	-0.21	**0.03**	**0.3**	**0.28**	**0.31**	-0.49	-0.32	-0.13	-0.72	-0.56	-0.28
Cluster	-0.34	-0.28	-0.13	**0.01**	**0.08**	**0.03**	-0.27	-0.13*	-0.03	-0.18*	-0.14*	0.00*
Restricted background	-0.31	-0.19	-0.11	-0.11	-0.09	-0.12	-0.19*	-0.16	-0.09	-0.28	-0.33	-0.44
Split	-0.24*	**0.07***	**0.37***	-0.14	-0.08	**0.04**	-0.27	-0.23	-0.05	-0.37	-0.28	-0.21
Systematic sampling	-0.47	-0.19	**0.02**	**0.41***	**0.39***	**0.32***	-0.47	-0.25	**0.02***	-0.59	-0.31	-0.07
Plethodon cylindraceus												
Biasfile	-1.17	-1.14	-0.88	-1.1	-0.75	-0.62	-0.96	-0.98	-0.94	-1.2	-0.76	-0.61
Cluster	-0.05*	-0.07*	-0.02*	-0.17	-0.09	-0.03	-0.15	-0.12*	-0.14	-0.01*	-0.06	**0.02**
Restricted background	-3.95	-2.72	-2.27	-4.55	-3.64	-3.08	-2.15	-2.3	-1.95	-3.8	-3.08	-2.54
Split	-0.12	-0.23	-0.73	-0.16	-0.05	-0.02	-0.14	-0.23	-0.26	-0.18	-0.09*	-0.07
Systematic sampling	-0.16	-0.29	-0.19	-0.06*	**0.04***	**0.01***	-0.10*	-0.19	-0.09*	-0.24	-0.1	**0.03***
Virtual species												
Biasfile	**0.14**	**0.46**	**0.09**	**0.61***	-0.64*	-1.35	**0.36**	**0.25**	**0.47**	**0.79***	-0.09*	-0.07*
Cluster	**0.28**	**0.32**	-0.55	**0.2**	-2.1	-2.44	**0.41**	**0.18**	**0.31**	**0.75**	-0.44	-0.27
Restricted background	0	**0.2**	-0.76	-0.07	-2.74	-2.89	**0.25**	-0.22	-0.14	**0.53**	-1.76	-1.4
Split	**0.43**	**0.58***	**0.32***	**0.1**	-2.51	-2.35	**0.49***	**0.25**	**0.41**	**0.72**	-0.64	-0.51
Systematic sampling	**0.32***	**0.51**	**0.14**	**0.41**	-1.35	-0.99*	**0.45**	**0.30***	**0.50***	**0.78**	-0.14	-0.42
(c) ΔG_{over}												
Chrysemys picta												
Biasfile	**0.18**	**0.25**	**0.09**	**0.26**	**0.07**	**0**	-0.10*	-0.05	-0.23	-0.11*	-0.19	-0.2
Cluster	-0.37	-0.14	-0.07	-0.01	**0.06**	**0.01**	-0.38	-0.09	-0.15	-0.55	-0.45	-0.84
Restricted background	-0.12	-0.09	-0.05	**0.01**	0	-0.01	-0.14	-0.13	-0.08	-0.2	-0.32	-0.10*
Split	-0.01	**0.08**	**0.57***	-0.04	-0.05	0	-0.53	-0.56	-0.18	-0.99	-0.9	-1.25
Systematic sampling	**0.33***	**0.40***	**0.31**	**0.65***	**0.41***	**0.34***	-0.25	**0.05***	**0.18***	-0.23	**0.25***	-0.98
Plethodon cylindraceus												
Biasfile	**0.29**	**0.74***	**0.52**	**0.71***	**0.72***	**0.63***	**0.48***	**0.33**	**0.43***	**0.43***	**0.29***	**0.32***
Cluster	**0.29**	**0.3**	**0.23**	**0.39**	**0.14**	**0.11**	**0.08**	**0.39**	**0.15**	**0.26**	-0.23	**0.11**
Restricted background	-0.62	-0.19	-0.45	-0.34	-0.17	-0.21	**0**	-0.35	-0.17	-1.09	-0.79	-0.87
Split	**0.52***	**0.4**	**0.83***	**0.54**	**0.51**	**0.32**	**0.41**	**0.48***	**0.29**	**0.19**	-0.09	-0.03
Systematic sampling	**0.48**	**0.65**	**0.62**	**0.67**	**0.49**	**0.46**	**0.26**	**0.12**	**0.21**	**0.43***	**0.01**	**0.31**
Virtual Species												
Biasfile	**0.41***	**0.77***	-0.31	-0.25	-2.33	-3.09	**0.80***	**0.65***	**0.66**	**0.86***	-0.39	-0.44
Cluster	-0.21	**0.2**	-2.37	-1.13	-3.53	-3.05	**0.61**	**0.29**	**0.4**	**0.8**	-0.73	-0.62
Restricted background	0	**0.34**	-2.21	-1.17	-3.7	-3.17	**0.69**	**0.39**	**0.41**	**0.83**	-0.53	-0.42*
Split	**0.31**	**0.62**	**0.02***	-1.16	-3.68	-3.03	**0.69**	**0.47**	**0.61**	**0.83**	-0.64	-0.57
Systematic sampling	**0.05**	**0.62**	-0.07	**0.02***	-1.37*	-2.01*	**0.62**	**0.42**	**0.73***	**0.86***	**0.17***	-0.72

Regardless of the species, the bias type, and the metrics considered, the restricted background method failed to improve the biased models in almost all tested cases. The other methods performed better but were ranked differently depending on bias type. Systematic sampling performed slightly better and more consistently among the competing methods as

shown by the relative performance of each method across bias types (Figure 13). Although systematic sampling was not always ranked first, it showed very little deviation from the most performing method and performed on average better than the others (for ΔD_{geo}: mean rank $_{\text{Systematic sampling}}$ = 2.11 ± 1.08 SD; mean rank $_{\text{Split}}$ = 2.53 ± 1.08 SD, mean rank $_{\text{Cluster}}$ = 2.613 ± 1.23 SD; mean rank $_{\text{Biasfile}}$ = 3.31 ± 1. 31 SD). In contrast, the restricted background method recorded the least correction (mean rank $_{\text{Restricted background}}$ = 4.44 ± 1.13 SD).

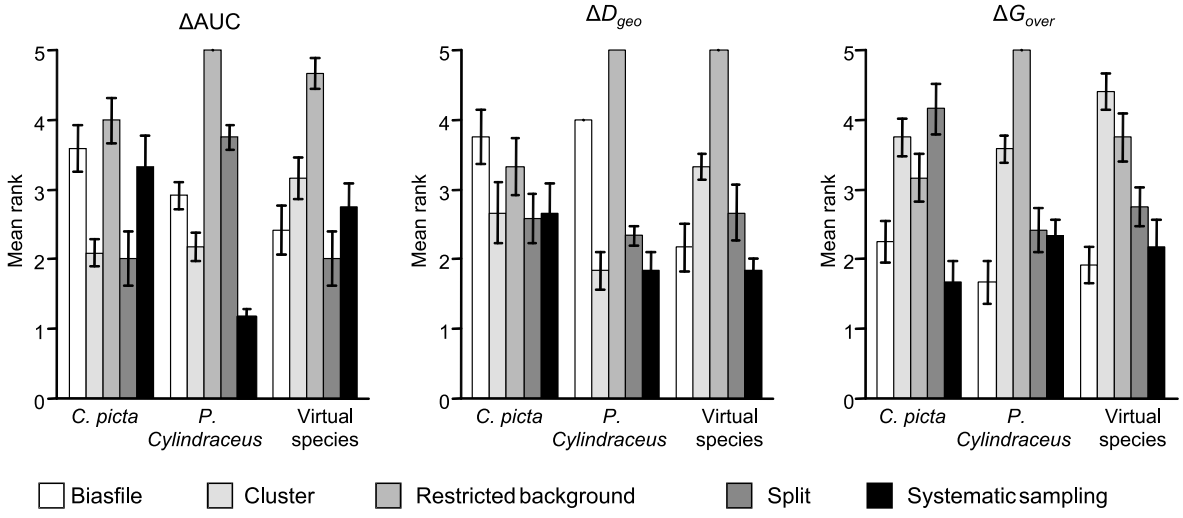


Figure 13: Rank of each method to correct sampling bias. Mean ranks ± standard-error for the performance of each method to correct sampling bias for each species (*Chrysemys picta*: left, *Plethodon cylindraceus*: center, *virtual species*: right), following 3 measures of correction performance: ΔAUC (left), ΔD_{geo} (centre), and ΔG_{over} (right). For each type of bias and bias intensity, the method which results in the most efficient correction is set to 1 whereas the least powerful method is set to 5. The plotted values are the mean rank across the 4 types of bias and 3 intensities.

Overall, and considering only ΔD_{geo}, the systematic sampling method was able to correct the bias ($\Delta D_{geo} > 0$) in 33% of the test cases. This success rate rose to 66% in the case of the virtual species, for which we were able to compare to a true unbiased model. However, the biasfile corrected the initial bias in 23% of test cases. The cluster and split method were both efficient in 23% of cases while only 6% of cases were corrected by the restricted background method.

Interestingly, relative performance between methods was consistent across metrics (Figure 13). The restricted background method was always the least performing one in terms of ΔAUC, ΔD_{geo} and ΔG_{over} (but for *C. picta* and ΔD_{geo}, for which it is ranked 4/5). The systematic sampling method was among the most performing methods. However, even if systematic sampling was overall the most efficient method across species in terms of ΔD_{geo}, it was outperformed by the

split or biasfile methods in some cases for the virtual species, or by the cluster method under some combinations of bias × intensity for the two real species (Table 2). However, when the systematic sampling was unable to resolve the bias, this latter was most often equally poorly corrected by any of the methods tested.

2.1.5. Discussion

As an unexpected first finding, we noticed that the range of AUC values obtained for biased and corrected models remained high even for models with the strongest biases. The decrease in AUC observed after applying the bias was moderate, less than 2% on average, across species and bias type. Moreover, the AUC values of the biased models were almost always over 0.8 or 0.9, which would classify the models as "good" or "very good" (Araújo et al. (2005) adapted from Swets (1988)). Together with other studies (Lobo *et al.* 2008; Jimenez-Valverde *et al.* 2012; Jimenez-Valverde 2014) our results highlight that this measure may poorly reflect model accuracy. Therefore, studies that focus solely on the AUC value should interpret their results with caution. AUC may be a good statistical measure of discrimination ability, but it often fails to quantify the ecological realism of modeled distribution (Lobo *et al.* 2008; Jimenez-Valverde *et al.* 2012; Jimenez-Valverde 2014) especially when estimated from presence-only data. Because we have a reference model, we will mainly focus on the overlap indices with the unbiased model as a measure of predictive accuracy performance.

Contrary to previous studies investigating sampling bias correction in SDM that focused on a few methods and simple biases (Syfert *et al.* 2013; Kramer-Schadt *et al.* 2013; Varela *et al.* 2014; Boria *et al.* 2014), we reviewed here five different ways to deal with sampling bias and used both real and virtual datasets under various bias scenarios. We also considered bias intensity that has been to our knowledge never assessed and proved to be of as a high concern as the type of bias. Moreover, instead of relying only on classical measures of SDM performance such as AUC (as used in Syfert et al. (2013), Varela et al. (2014) and Boria et al; (2014)) or omission/commission error (as used in Kramer-Schadt et al. (2013)), we evaluated the correction performance by directly comparing the SDM outputs. Therefore, we actually assessed the ability of the tested methods to recover the unbiased model, which is the expected behavior of an efficient sampling bias correction. In addition, rather than basing our conclusions on island species (Syfert *et al.* 2013; Kramer-Schadt *et al.* 2013; Boria *et al.* 2014), we used continental

species whose distributions are clearly shaped by climate and not by a geographically bound space.

Our results clearly evidence that the different methods of sampling bias correction tested here may have very variable efficiency depending on the modeling conditions (biases type and correction method). Interestingly, the correction may have a positive effect, and actually contributes to correct the bias; nonetheless, in some cases it may produce a poorest model than the biased model. These results suggest that the problem of sampling bias in species distribution modeling has probably multiple answers depending on the context. We especially emphasize that the type and intensity of bias influence the ability of various methods to resolve the initial bias.

However, correction methods did not perform equally across the various conditions. The less efficient method restricted the spatial extent of the background whereas in other methods, the background points were selected from the whole available environment (*i.e.* randomly drawn from the area covered by the environmental grid files). Surprisingly, this method is often used and have been contributed to improve SDM performance in some cases (Phillips 2008; Phillips *et al.* 2009). However, as suggested by Thuiller et al. (2004) and Vanderwal et al. (2009), excessively restricting the geographical extent of pseudo-absences to a narrow area or selecting them from a too large area reduces model accuracy. Background selection may greatly influence the resulting model as it determines the underlying assumptions of the model to use (Chefaoui & Lobo 2008). Therefore, this step should be undertaken with caution. The size of the buffer used for background selection also greatly influences model performance. For instance, AUC often increases with the size of the study area because it contributes to include background points that have environmental characteristics greatly distant from the species requirement, resulting in artificial increase of SDM validation (Barve *et al.* 2011). The selection of the training area should therefore be strictly relevant to the ecology of the species and the objective of the study. A relevant selection of the training area (the geographic region in which background points are selected) should reflect the geographical space accessible to the species over a given time period (Barve *et al.* 2011). It may thus be essential to carry out a rigorous investigation of the optimal geographic distance between the set of occurrences used to train the model and background points. It has to be both optimal for model training and biologically meaningful. The interpretation of the modeled distribution must also be engaged carefully as it may reflect the fundamental niche or the true occupied range, and often a position between both.

Regarding the high variability in correction performance of the different methods depending on various factors, it is difficult to propose a universal guideline to solving sampling

bias. It might be advisable to evaluate first several types of correction. The final choice of correction method would be then based on their effect in classical SDM evaluation metrics (*e.g.* AUC, Kappa, True Skill Statistics TSS) and possibly the adequacy of output maps to *a priori* knowledge of the species distribution. A first useful step might also be to evaluate bias type to design and select the most appropriate correction. For instance, the split method makes sense only if the most and less sampled areas are at least roughly known. Most of the time, sampling bias is only inferred by the known sampling effort or empirical knowledge of the species distribution. The true severity and shape of this bias is almost always lacking. However, in some cases, bias can be evaluated by comparing the geographic distribution of the available occurrences to known sources of bias. An estimation of sampling probabilities across the study area can provide insights on the potential bias that may affect the collected observations and help in further choice of the correction method. The known characteristics of the modeled species may also condition the strategy to use. A species with an expected very large geographical and/or environmental range should benefit from being split into two or more partitions that are combined afterward (Osborne & Suárez-Seoane 2002). Finally, the five methods tested here are not necessarily exclusive. For instance, it is possible to split a large dataset and apply systematic sampling to each dataset in species with broad distributions (Fourcade *et al.* 2013) or to apply the biasfile method after using first another correction method

Nevertheless, we found that only systematic sampling constantly performed well irrespective of species and bias type. Interestingly enough, it happened to be the simplest and most obvious way to solve the geographic bias. Beside, this method can be quickly applied to any dataset, even if the nature of the bias is unknown. Kramer-Schadt et al. (2013) and Boria et al. (2014) also identified this technique, which they named spatial filtering, as the most effective (but note that Varela et al. (2014) found that environmental filtering, equivalent to our cluster method, may provide better results). We might keep in mind that this method may have a few drawbacks though. Subsampling the complete dataset may alter the distribution of occurrences in the environmental space and exclude some portion of environmental space from the input records. This problem may be circumvented by adjusting the grid size but this may not be possible when the sample size of occurrences (presences) is too low. On the contrary, smoothing the distribution of occurrences may lead to overestimating the probability presence in marginal areas. The grid cell size used to sample the occurrences may also be a source of problem; it must be large enough to resolve the bias but not too large to result in a strong loss in resolution. A fine adjustment of the parameters of the method (*e.g.* the resolution of systematic sampling that

provides the optimal trade-off between sampling bias correction and information reduction) is thus necessary to maximize its performance.

Our results suggest that the difference in performance between the systematic sampling method and the optimal method for the dataset under investigation is slight. Since systematic sampling seems to be robust enough to differentiate sampling bias and species, we suggest that this could be the method selected first if no further attempts to correct sampling bias are to be made. This method provides a quick and efficient way to improve SDM from a biased dataset and is likely to be relevant most of the time. Even if the elaboration of a step-by-step framework would be ideal to assist in the definition of a reliable strategy of bias correction, we highlight that most SDMs studies, at least those that use MAXENT as modeling algorithm, would highly benefit from this simple tweaking of input occurrences. Regarding the major impacts that conservation actions are expected to provide (Adams *et al.* 2004; Hoffmann *et al.* 2010), they must be based as much as possible on irrefutable models. Therefore, it is essential to develop simple methods intended for conservation practitioners, such as techniques of sampling bias correction, to build robust distribution models. In this regard, we encourage MAXENT users to carefully take sampling bias into account, and to use systematic sampling of their input occurrences as a quick and simple resolution of bias.

2.1.6. Acknowledgements

This study was supported by a grant to YF and JS from Plan Loire Grandeur Nature and ERDF. JOE was kindly supported within the German Federal Environmental Foundation fellowship program.

2.1.7. Supporting information

Supplementary material S1: Details of the generation of sampling bias.

(1) Two areas – The original set of species occurrences was divided into a northern set and a southern set following the median latitude of available records. The northern set was defined as the area of high density of records whereas the southern set was defined as the area of low density of records. In order to create the biased density, we randomly removed occurrences in each subset, keeping a higher proportion of records in the northern set. For the high bias, 95% of

records were kept in the north and 5% in the south. For the medium and low biases, the ratio of observation was respectively 80% / 20% and 70% / 30%. The two datasets were thereafter merged to be used in modeling. As a result, the number of occurrences was the same for each bias intensity, corresponding to the half of original records.

(2) Gradient – We generated a sampling probability function depending on latitude, ranging from 1 for the northernmost point to 0 for the southernmost point. We randomly selected occurrences from the original dataset following this probability function. The low bias was generated following a linear sampling probability function while the medium and high biases followed an exponential function. The intensity of bias was modulated according to the following equations, where *y* is the latitude:

Low intensity: $\mathrm{p} = 1 - \left(\frac{\max(y) - y}{\max(y) - \min(y)}\right)$

Medium intensity: $\mathrm{p} = 1 - \left(\frac{\max(e^{xy}) - e^{xy}}{\max(e^{xy}) - \min(e^{xy})}\right)$ where $x = 0.1$

High intensity: $\mathrm{p} = 1 - \left(\frac{\max(e^{xy}) - e^{xy}}{\max(e^{xy}) - \min(e^{xy})}\right)$ where $x = 0.2$

(3) Center - The records were selected from sampling probability functions depending on the distance to the geographical centroid of occurrences. Here, the probability was highest at the centroid of the original set of points and decreased towards the periphery, according to the following equation, where *d* is the distance to centroid:

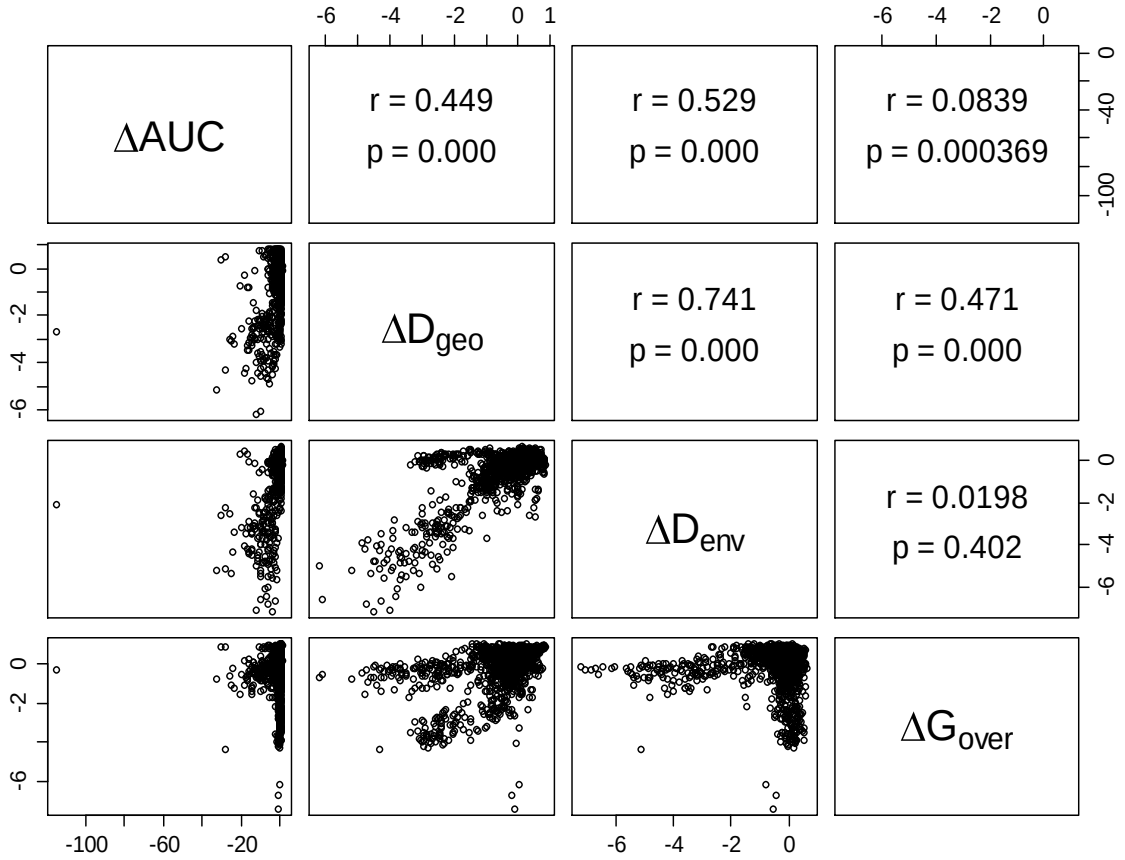
$$\mathrm{p} = 1 - \left(\frac{\max(e^{xd}) - e^{xd}}{\max(e^{xd}) - \min(e^{xd})}\right)$$

The intensity of bias was modulated by varying the value of x. Low bias: $x = 1/3$, medium bias: $x = 1/5$, high bias: $x = 1$

(4) Travel time - For each occurrence point, we extracted the corresponding travel time (extracted from the map produced by the European Commission (Nelson 2008) available at bioval.jrc.ec.europa.eu/products/gam/) and used this value as a sampling probability weight. The

three bias intensities were generated by sampling respectively 25%, 15% and 5% of original records for the virtual species and *Chrysemys picta*. Because of the lower number of available records, we sampled 50%, 30% and 20% of original dataset for *Plethodon cylindraceus*.

Figure S1: Correlations between measures of correction performance. Lower panel: correlation scatterplots, upper panel: Pearson's correlation coefficient *r* and associated p-value.



2.1.8. Conclusion Article 1

Au vu des résultats obtenus, on constate une assez grande variabilité dans la performance de correction, en fonction de l'espèce modélisée, du type de biais, de son intensité et de la méthode retenue. Il est donc sans doute hasardeux de définir une méthode universelle, mais il est malgré tout possible de proposer quelques recommandations.

En effet, parmi les méthodes testées, la restriction des pseudo-absences autour des points de présences (*restricted background*) (Phillips *et al.* 2009) semble particulièrement inefficace, bien qu'il puisse s'agir d'une étape indispensable. Il est ainsi recommandé de choisir en *background* uniquement les zones que l'espèce est susceptible d'atteindre (ensemble M dans le diagramme BAM, voir Introduction). Tandis que l'utilisation d'un fichier de description du biais (*biasfile*) et la séparation en deux sous-modèles (*split*) ont une performance variable, il est à noter que le sous-échantillonnage environnemental (*cluster*) a une efficacité presque toujours inférieure au sous-échantillonnage spatial (*systematic sampling*).

Cette dernière méthode de correction du biais semble avoir la meilleure performance moyenne. Cette approche se trouve être la plus simple à mettre en œuvre parmi toutes celles évaluées. Elle consiste simplement à définir une grille régulière et à ne retenir qu'une seule occurrence par cellule afin de dissoudre les agrégats de points qui peuvent survenir lorsqu'une même zone est sur-échantillonnée par rapport aux autres. Toutefois, le sous-échantillonnage entraine nécessairement une perte d'information lorsque la résolution de la grille est trop large et lorsque les agrégats de points correspondent à une réalité écologique. Il convient donc de définir le meilleur compromis entre la correction du biais et l'information à retenir.

Ce même résultat a également été identifié par plusieurs études menées parallèlement à celle-ci (Kramer-Schadt *et al.* 2013; Boria *et al.* 2014). Il semble que, en l'absence de renseignement précis sur la nature du biais, cette méthode puisse être conseillée dans la plupart des cas. Elle doit permettre, par une manipulation très simple du jeu de données initial, de résoudre relativement efficacement le biais d'échantillonnage.

2.2. ARTICLE 2 : MODELISATION DE LA DISTRIBUTION DU RALE DES GENETS

L'étude théorique relative à la correction du biais d'échantillonnage dans les modèles de distribution nous permet maintenant d'envisager des moyens efficaces de modéliser la distribution du Râle des genêts. Cette approche de modélisation est motivée par l'apparente pauvreté des données qui déterminent la distribution et les effectifs actuellement considérés dans l'évaluation de la conservation de l'espèce (Sukhanova & Mischenko 2003; Mischenko 2008).

En règle générale, les aires de répartitions incluses dans les évaluations des statuts de conservations par l'IUCN sont encore souvent issues d'opinions d'experts. Pourtant, les tailles et contraction d'aires de distribution font parties des critères d'évaluation essentiels utilisés par l'IUCN (IUCN 2001; Mace *et al.* 2008). Or, malgré l'intérêt des avis d'experts dans de nombreux domaines en écologie et conservation (Martin *et al.* 2012), notamment lorsque les données empiriques sont en trop faible quantité, ceux-ci sont connus pour être sujet à de nombreux biais potentiels (Iglecia *et al.* 2012; McBride *et al.* 2012). Dès lors, lorsqu'il existe des méthodes prédictives plus objectives, comme dans la modélisation d'aires de distributions, celles-ci semblent devoir être privilégiées.

Nous avons donc confronté dans le cas du Râle des genêts la carte de distribution distribuée par l'IUCN avec une distribution modélisée sous MAXENT. Afin de prendre en compte le biais d'échantillonnage considérable existant dans les données de présence, un sous-échantillonnage spatial des données, couplé à une modélisation scindée en deux parties (est et ouest) ont été mis en place. La distribution IUCN a été comparée avec les distributions modélisées (avec ou sans prise en compte du biais) par des indices de recouvrement de niche. De plus, les estimations de tailles de populations par pays ont été mis en relation avec la somme des indices de qualité d'habitat tirés du modèles (ou probabilités de présence relatives) et avec l'intensité de l'agriculture. Bien que non prévus pour cet usage, il a en effet été montré que les modèles de distribution permettent parfois de prédire les abondances (VanDerWal *et al.* 2009b; Brambilla & Ficetola 2012). Cette approche a pour objectif de mettre en évidence l'apport de la modélisation de distribution dans l'évaluation des statuts de conservations et de clarifier la distribution réelle du Râle des genêts.

Cette partie a fait l'objet d'une collaboration avec un autre doctorant du laboratoire et, comme l'article 1, avec des chercheurs allemands. Les résultats ont été publiés dans un article paru dans le journal « Biological Conservation », présenté ci-après :

Fourcade Y, Engler JO, Besnard AG, Rödder D, Secondi J (2013) Confronting expert-based and modelled distributions for species with uncertain conservation status: a case study from the Corncrake (Crex crex). *Biological Conservation*, **167**, 161–171.

Confronting expert-based and modelled distributions for species with uncertain conservation status: a case study from the corncrake (*Crex crex*).

Yoan Fourcade[1*], Jan O. Engler[2,3] , Aurélien G. Besnard[4], Dennis Rödder[2], Jean Secondi[1]

[1] LUNAM, Université d'Angers, GECCO (Groupe écologie et conservation des vertébrés), F-49045 Angers, France

[2] Zoological Research Museum Alexander Koenig, D-53113 Bonn, Germany

[3] Department of Forest Zoology and Forest Conservation, University of Göttingen, D-37077 Göttingen, Germany

*Corresponding author

2.2.1. Abstract

The Red List classification of IUCN has become one of the most important evaluations of threats that affect biodiversity at the species level. However, many estimations of species range, one essential factor in the Red List classification, are derived from expert-based assessments that sometimes lack empirical evidence. Our study focused on the corncrake (*Crex crex*), a grassland Palaearctic bird whose conservation status has been revised recently following some new assessments of range and population size. However, the amount of data that form the basis of this reclassification appears weak compared to the large area involved. We used a method of species distribution modelling (MAXENT) to predict the corncrake range and confronted it to the expert-based map. We resolved the huge geographic bias in the distribution of presence points by using a relevant method of sampling bias correction. We found a rather similar distribution with the IUCN estimated range, although less widespread. We also highlighted a relationship between habitat suitability computed by the model and population estimates per country when the effect of agriculture intensity is taken into account. This result supports the current expert-based estimates of corncrake distribution and emphasizes that a relevant modelling strategy should be able to predict the distribution of a species even from a biased dataset. IUCN estimates of species' ranges would certainly benefit from a model-based approach in addition to expert and field controls.

2.2.2. Introduction

The conservation status of a species determines the implementation of environmental policy and national or local target actions. Therefore, it is crucial to obtain baseline information, which is as reliable as possible (Sutherland *et al.* 2004). Although objective criteria are available to assign a status level, assessment reliability frequently depends on estimates of distribution and population size. These can be rough or of very unequal quality across a species range. The problem is acute for species distributed over many countries where sampling effort is highly unequal. The worst case happens when a nation where estimate precision is low hosts a large fraction of the world population. The uneven distribution of information quality can thus strongly bias estimates of range or population size which may in turn affect the conservation status assignment. If the assessed conservation status is lower than the actual status, action may be delayed or not considered at all but if the status is too high, unnecessary actions may divert resources for higher priority species or habitats. Thus, it is important to confront different approaches used to assess the worldwide distribution of species.

Currently, one of the main international organisations to assess the threats to wildlife is the International Union for Conservation of Nature (IUCN) whose Red List classifies species according to their extinction risk (IUCN 2001). Although its assignments about species status have no formal legal value, they are often used by national or intranational authorities or NGOs as a reference to prioritize conservation actions. Despite the distribution gaps and sampling biases that its assessments may reflect, mostly due to lack of data or outdated information, IUCN remains the only organization to centralize such large-scale information relevant to biodiversity conservation. The estimation of species range and population trends forms the basic information for IUCN conservation assessments. Among the five criteria of the Red list classification of threatened species (criteria A to E), criterion B is directly dependent on range size. Criterion D also considers the fragmentation of range associated with a reduced population size (IUCN 2001; Mace *et al.* 2008). The estimations of range size are expressed as the "Extent of occurrence" (EOO) or the "Area of occupancy" (AOO). These two measures of species range are derived from the locations of known occurrences of the species. The most common application of EOO and AOO definitions is to draw a convex hull polygon around all the known locations of the species (Burgman & Fox 2003). The estimated range is then used in the assignment of a species to one of the three threatened categories (Critically Endangered, Endangered, and Vulnerable). For instance, following criterion B, a species is classified as "Vulnerable" if its extent of occurrence is estimated to be less than 20000 km² and if its population or range present signs

of decline or extreme fluctuations. Different thresholds in range size estimates are used to assign a category (IUCN 2001, criterion B).

The accuracy of EOO or AOO, as well as population size estimates, are crucial to the definition of many conservation programs that directly stem from the IUCN assessments (IUCN 2005; Rodrigues *et al.* 2006). However, a species' status is often based on mixed data, systematic surveys and more fragmented data collected and analyzed by local or national experts (Hayward 2009). This recourse to expert-knowledge is widely used to provide additional information when the amount of empirical data is low. Experts may be used in various ways, from the direct supply of data to the definition of priors for Bayesian modelling (Perera *et al.* 2012). However, the quality of expert-based information may vary substantially depending on personal beliefs and experiences and on the way information is elicited (Martin *et al.* 2012). Furthermore, for species distributed across several countries, the availability and quality of quantitative data may substantially vary too. As a consequence, uncertainty accumulates over the assessment process resulting in population sizes with large error estimates which may often be provided along with basic distribution maps representing envelopes of possible presence.

Therefore, it is important to confront expert-based assessments with other methodology that reduces the number of steps cumulating errors. The reliability of range estimates can be improved by the mean of statistical models that predict the occurrence of a species within a geographical area (Cassini 2011). Two distinct statistical approaches can be used (Breiman 2001b). On the one hand, one can model species' response functions to predictors through classical regressions, such as in Resource Selection Function (RSF, Boyce and L. McDonald, 1999; Manly et al., 1993) or Generalized Linear Models (GLM, McCullagh and Nelder, 1989). On the other hand, various algorithms relying on machine learning have been developed to model ecological niches and project it onto its potential geographic distribution. Among them, the most popular are Artificial Neural Networks (Ripley 1996), Maximum Entropy (MAXENT, Phillips et al., 2004), or those based on Classification And Regression Trees (CART, Breiman et al., 1984), such as Random Forest (Breiman 2001a), Multiple Adaptive Regression Splines (Friedman 1991) or Boosted Regression Trees (Ridgeway 1999). Contrary to classic statistical methods, machine learning methods are able to deal with complex and non-linear relationships between predictors and can be regarded as more suitable for ecological modelling even if their interpretation is less straightforward.

These methods, collectively known as species distribution modelling (SDM), are now widely used to predict habitat suitability of a species across space and time (Guisan &

Zimmermann 2000; Franklin 2009; Elith & Leathwick 2009; Peterson *et al.* 2011). The principle of SDM is to link locations of occurrences to environmental predictors such as climate, topography or land cover to map the species' potential distribution or probability of a given site to be suitable for it (Guisan & Zimmermann 2000). SDMs produce qualitative maps so that the proportion of areas of high probability could be contrasted with areas of low probability and, ideally, the sum of probabilities or suitability should reflect the species population size (VanDerWal *et al.* 2009b; Brambilla & Ficetola 2012). Such models can provide predictions of a species' range, providing that model assumptions are met, *i.e.* relevant predictors are chosen, species records are precisely located and constitute a non-biased sample relative to the species' real distribution. Moreover, SDM have many applications in conservation prioritization since they estimate the most suitable areas for a species, as well as potential range shifts under climate change or the possible expansion of invasive species (Jeschke & Strayer 2008). Conservation status assessment should benefit from this approach, and the possibility to use SDMs to assess EOO is already included in IUCN guidelines (IUCN Standards and Petitions Subcommittee 2010).

The present study aims at confronting expert-based range proposed by Birdlife, for the Corncrake (*Crex crex*) and outputs from species distribution modelling. Birdlife, a British non-governmental organization, provides assessments about the distribution and threats of bird species for IUCN. This approach should be able to estimate the accuracy of the widespread distribution of the species which form the basis of the recent IUCN reclassification to "Least Concern". However, the coverage of survey between the western and the eastern part of the distribution is very unbalanced (see Methods) which is likely to affect the resulting model. Geographical sampling bias is known to be detrimental to the modelling process when it reflects sampling bias in the environmental space (Grand *et al.* 2007; Costa *et al.* 2009; Leitão *et al.* 2011; Bystriakova *et al.* 2012). Such bias has often been described but it is still rarely taken into account. Several methods of sampling bias correction have been proposed (*e.g.* Dudík et al. 2005; Phillips et al. 2009) but there is no consensual guideline to reduce it (Fourcade *et al.* 2014). In order to improve model accuracy, we applied a sampling bias correction method specifically designed for our case study. We then compared the output maps of both corrected and uncorrected models to the range estimate provided by Birdlife. We also tested the consistency between model outputs and population estimates at the state level.

2.2.3. Material and methods

2.2.3.1. The model species

The Corncrake (*Crex crex*) is a Palaearctic grassland bird whose conservation status has been revised in the recent years. The species has dramatically declined in most Western European countries over the past decades (Green *et al.* 1997a) and has been maintained owing to conservation measures. The intensification of agricultural practices caused large habitat loss and drastically reduced brood survival in the remaining suitable habitats because of the earlier mowing of meadows (Green *et al.* 1997b; Tyler *et al.* 1998), which greatly affected western populations (Green & Stowe 1993; Green & Rayment 1996). These well identified threats have motivated the classification of the species as "Vulnerable" in early 90s by Birdlife International (Birdlife International 2012). During the first decade of the 21st century, new assessments of population size in former USSR have changed the earlier view on corncrake distribution and population trends (Keišs 2003; Sukhanova & Mischenko 2003; Mischenko 2008). These studies concluded that the species was more widely distributed in Russia and Asia than previously thought. They also revealed that the changes in agricultural land use that followed the fall of USSR have promoted the development of uncultivated meadows and abandoned arable lands which have turned into favourable habitats for corncrakes (Keišs 2005). As a result, the IUCN conservation status of the species was reclassified from "Vulnerable" to "Near Threatened" in 2004 and to "Least Concern" in 2010 (Birdlife International 2012), the lowest category of threat. This reclassification turned the corncrake from a globally threatened species to a species of minor preoccupation.

Nonetheless, the current status of the species has been evaluated from limited information and non-independent sources. Besides, only few survey sites were conducted in Russia to infer the estimated range (from Western Europe to Baikal Lake approximately) and population size (2 – 3 million pairs) (Schäffer & Koffijberg 2004; Birdlife International 2012). The main part of the estimated range in Asian Russia, as well as Kazakhstan or Western China, has been poorly investigated so far. On the contrary, in many EU countries most individuals are censused on a year or a multiyear basis. This high level of uncertainty in the eastern part of the range may question the reliability of these estimates for the worldwide population. These are of major importance given the threats for the species in the western part of its range.

2.2.3.2. Occurrence data

A set of presence data of the Corncrake was collected through many sources (Appendix B, Table B.1). Most of occurrences consisted on precise coordinates of singing males collected during annual systematic surveys organized by local or national organizations involved in corncrake conservation. As a result, long-term and unbiased datasets of observations were available in many Western European countries. The remaining data came from scientific publications, online databases or naturalist reports. Some less precise observations were georeferenced according to locality names. We only included observations recorded during the breeding season (from 15th April to the end of June) to avoid the presence of migrating birds in the dataset. Depending on the dataset provider, the occurrences were recorded from 1984 to 2011 (Appendix B, Table B.1).

Contrary to Western Europe, most of Eastern European or Asian countries do not survey this species. Consequently, information quality about the location of singing males during the breeding period greatly varies across the species' range. Furthermore, the dataset exhibits a strong geographical sampling bias (Figure 14). More than 99% of observations are located between -8° and 32° longitude whereas the remaining 1% lies between 32° and 105° longitude, *i.e.* mainly in Russia and Kazakhstan that are considered to host the largest parts of the world population and range (BirdLife International 2004; Koffijberg & Schäffer 2005). A sampling bias in the geographical space will not result in a biased distribution model if the occurrences reflect the actual species distribution in the environmental space (Kadmon *et al.* 2004). However, considering the continental scale of our study, it is very likely that the strong geographic bias overestimates the contribution of European climates over the conditions encountered in Asia ,which part is strongly underrepresented in the dataset (Appendix B, Figure B.1 left) (See section 2.4 for bias correction). A total of 41790 occurrences were collected. Some records where duplicated, revealing multiple observations of a same bird, or several birds located in a single place. After removing duplicates (multiple points with same coordinates), the remaining sample size of records was 35744. Because the modelling method uses only one occurrence per environmental grid cell, we also removed duplicate observations occurring in a single grid cell. The final dataset available for the modelling process consisted of 7605 species records.

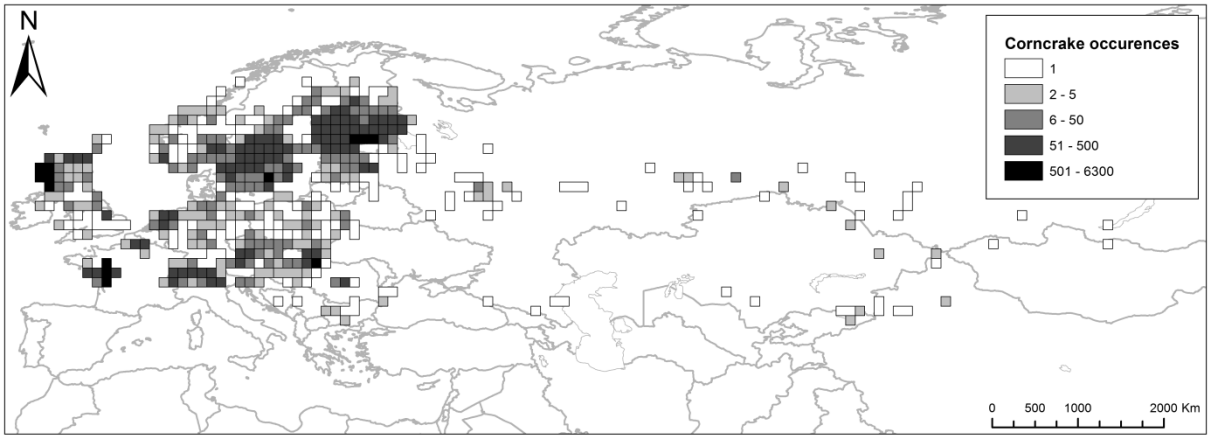


Figure 14: Number of corncrake occurrences collected per 1° grid cell. The colour is function of the number of presence points in the cell ranging from white (1 observation per cell) to black (>500 observations per cell, up to 6300). The distribution of points was highly biased in space, with more than 99% of observations located in the western part of the expected range (up to 30° longitude).

2.2.3.3. Species ecology and environmental predictors

The Corncrake arrives at its breeding grounds from April or May and stays until mid-September. It mainly occurs in fertile tall grasslands such as floodplain meadows or steppe (Schäffer & Koffijberg 2004). We used a first set of environmental predictors that represent spring climate, *i.e.* the species requirements on its arrival, and annual climate, which determines the long term habitat structure (forest, steppe). We downloaded climatic and topographic grids from the WorldClim database (Hijmans et al. 2005, http://www.worldclim.org) at a resolution of 2.5 arc-min. We also used a land cover map from the Globcover 2009 project (Arino et al. 2008, http://due.esrin.esa.int/globcover) and rescaled the grid to fit the 2.5 arc-min resolution of the other variables. Finally, we compiled 5 years (2007 – 2011) of ten-day periods of NDVI data, a measure of vegetation productivity derived from multispectral remote-sensing images, downloaded from the SPOT-VEGETATION project (Maisongrande et al. 2004, http://free.vgt.vito.be). Taking its migrating nature into account, we assumed that the distance to its wintering range could restrain its breeding area. We thus computed a grid of distance to the wintering zone using the estimated area provided by Birdlife (Birdlife International 2012). Highly inter-correlated variables (correlation coefficient computed by ArcGIS 10; > 0.9 or < -0.9) were removed (Heikkinen & Luoto 2006), resulting in a set of 18 variables (Table 3).

Table 3: Environmental predictors used for modelling.

Environmental predictors	Initial resolution	Resolution used for SDM	Provider	Period	Source
Altitude	3 arc-seconds ; vertical error < 16m	2.5 arc-min	NASA SRTM[1]	2000	Shuttle Radar Topography Mission
Annual mean temperature	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	Interpolated using data from GHCN, FAO, WMO, CIAT, R-Hydronet and others
Mean diurnal range	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Isothermality	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Maximum temperature of driest quarter	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Temperature annual range	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Mean temperature of wettest quarter	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Annual precipitation	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Precipitation of wettest month	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Precipitation of driest month	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Precipitation seasonality	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Precipitation of coldest quarter	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Mean precipitation of April and May	30 arc-sec	2.5 arc-min	Worldclim[2]	1950 - 2000	
Land cover	300 m	2.5 arc-min	ESA Globcover[3]	2009	ENVISAT satellite mission
Maximum NDVI for the period April to June	1 km	2.5 arc-min	SPOT-VEGETATION[4]	2007-2011	SPOT4 satellite
Minimum NDVI for the period April to June	1 km	2.5 arc-min	SPOT-VEGETATION[4]	2007-2011	
Vegetation growth (NDVI of June minus NDVI of April)	1 km	2.5 arc-min	SPOT-VEGETATION[4]	2007-2011	
Distance to wintering area		2.5 arc-min	Layer of distance computed from Birdlife international range map		Expert-based map of African wintering range

2.2.3.4. Bias correction

The main source of geographical bias in the dataset is the high variation in sampling effort between Western Europe (high) in one hand, and Russia and Asia (low) in the other hand. To correct this bias, we split the dataset in two geographical units before computing the model. Even without this extreme sampling bias, such a strategy may be a good way to model species with a large distribution (Osborne & Suárez-Seoane 2002). In order to remove the bias that may remain in these subsets, we carried out a systematic sampling of occurrences within each dataset. We expect this method to perform relatively better than others in a variety of conditions (Fourcade *et al.* 2014). A grid with a resolution of 1.5° was created and we randomly selected one occurrence

per grid cell. A western model was then computed including only the occurrences up to 30° longitude whereas an eastern model was computed with the remaining occurrences from 30° to 105° longitude. The final corrected model was a combination of western and eastern model in which the highest value was kept in overlapping pixels.

To evaluate the effect of bias correction, we also computed a distribution model without any correction of the sampling bias including all the 7605 occurrences available. This model is expected to be of poor quality given the huge geographic bias of observations. However, it is a good way to compare our corrected model to a basic model run without taking the bias into account.

2.2.3.5. Species distribution modelling

The three species distribution models (Birdlife model, uncorrected model and corrected model) were carried out with the software MAXENT (Phillips *et al.* 2006). It is currently one of the most commonly used SDM method and its popularity has considerably increased over the past years owing to its robustness to complex scenarios (Elith *et al.* 2006) and to small sample sizes (Wisz *et al.* 2008). It is a machine learning algorithm that applies the principle of maximum entropy, in a way that is mathematically equivalent to Poisson regression (Renner & Warton 2013), to predict potential distribution of species from presence-only data and environmental variables (Phillips *et al.* 2004; Elith *et al.* 2011; Merow *et al.* 2013). The models were computed with the version 3.3.3k of MAXENT (http://www.cs.princeton.edu/~schapire/maxent/) with the default features settings. The logistic output format produced a map of probability of presence ranging from 0 to 1 per grid cell. As the Corncrake is a long-distance migrant, we had no a priori expectation about its dispersal capacity across Eurasia. We therefore let MAXENT select background (10000 locations) randomly across the whole study area. The evaluation of models was performed by 10-fold cross-validation.

2.2.3.6. Birdlife model

In order to evaluate congruence of model-based and expert-based approaches, we computed the overlap between the corrected or the uncorrected SDMs and the envelope provided by the IUCN. The latter distribution was developed from a set of 2000 locations randomly sampled in

the Corncrake range provided by Birdlife (Birdlife International 2012) and adopted by the IUCN. These occurrences were then included in a distribution model computed with the default MAXENT settings. Although this model does not exactly represent the EOO proposed by the IUCN, we considered it as a good method to convert the Birdlife range estimate to a species distribution model (considering not just the realized but also the potential distribution). When applying a threshold that maximizes sensitivity and specificity (here 0.38), the SDM matched very well the Birdlife expert-based envelope (Figure 15).

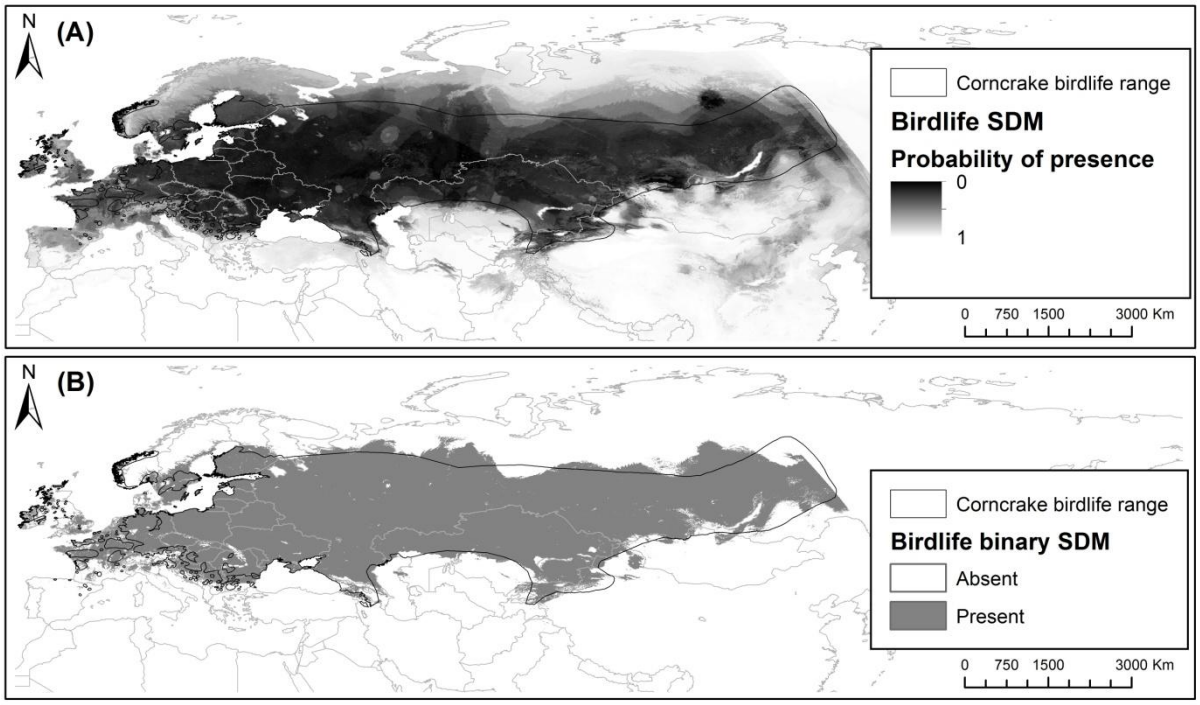


Figure 15: Species distribution model derived from 2000 points sampled in the Birdlife estimated range, without threshold (A) and converted to a binary map following the threshold that maximizes sensitivity and specificity (overlap with Birdlife envelope = 0.81) (B).

2.2.3.7. Evaluation and comparison of distribution models

SDM performance is commonly assessed by the area under the receiver operating curve (ROC), known as the AUC (Hanley & Mcneil 1982). AUC ranges from 0 (predictions exactly inverse to the actual values) to 1 (model with optimal discrimination abilities), a random model given an AUC value of 0.5. We reported this measure because it has been extensively used in SDM studies but it has been argued that it is a poor estimate of the accuracy of the predicted distribution (Lobo *et al.* 2008; Jimenez-Valverde *et al.* 2012). The AUC values reported for both Birdlife and

uncorrected model were calculated directly by MAXENT whereas the AUC of the corrected model was computed with 'ROCR' R package (Sing *et al.* 2012) after combination of the two models. We also computed a corrected version of AUC (cAUC) which removes spatial sorting bias by pairwise-distance sampling and is calibrated by a spatial null model (Hijmans 2012). Finally, we reported the True-Skill Statistics (TSS), a threshold-dependent measure of model accuracy (Allouche *et al.* 2006), using the 'PresenceAbsence' R package (Freeman & Moisen 2008).

We compared the three distribution models computed using three measure of overlap:

D_{SDM}: the Schoener's *D* index (Schoener 1968), is a classical and reliable (Rödder & Engler 2011) measure of niche overlap widely used in ecological studies and SDM applications in particular. It is computed from the model output maps. This measure ranges from 0 (no overlap) to 1 (complete overlap) and is derived from the difference in probability distributions over space produced between two SDMs. Prior to analyzes, the suitability values inferior to the threshold that maximises specificity and sensitivity were set to 0 to remove noise that could inflate the estimate of D_{SDM}.

D_{env}: is another computation of Schoener's *D* index, following the PCA-env approach of Broennimann (2011). This method computes the overlap from occurrence data in environmental space instead of model projections correcting for the structure of the available climate space.

Finally, because many conservation actions require a simple prediction based on predicted presence/absence of the species, we computed a measure of overlap between binary maps. We converted the logistical grid produced by MAXENT to a binary map following the threshold that maximize the sensitivity and the specificity. We then calculated the overlap between our binary models and the reference model.

2.2.3.8. Relationship between distribution model and population size estimations

Although there is no direct relationship between occurrence and abundance, this link often exists between occupancy area and abundance. The area occupied by a species at a given scale, such as country-scale, is often correlated to its abundance in this territory (Gaston *et al.* 2000). As a consequence, and despite the fact that SDM is not intended to estimate population size, we may

expect the spatial distribution of suitability to correlate with the species abundance (VanDerWal *et al.* 2009b; Brambilla & Ficetola 2012). Such a relationship would provide a cross-validation of our model, in addition to the analysis of the distribution itself.

In order to test whether our model outputs matched nationwide estimates of population size we fitted a linear model describing the relationship between the sum of probabilities of presence per country produced by the corrected SDM and the mean population size estimation per country. We only used the corrected model, considering that it was the least biased. The population data came from the last available estimations of corncrake populations (BirdLife International 2004; Schäffer & Koffijberg 2004), which provided a range between minimum and maximum estimated population size (early 2000s estimates), and we used the mean number of singing males estimated per country. The sums of probabilities of presence and the estimated population were standardized by the country area. Population estimates are lacking for some countries that do not monitor the species at all, notably Kazakhstan. We therefore included in our analysis data from the 38 countries for which estimates were available. Corncrake populations have declined because of agriculture intensification in Western Europe whereas populations remained stable or increased in Eastern Europe and Asia where this process was less pronounced or has been reversed during the last decades (Green *et al.* 1997a; Keišs 2003). Thus, we included a measure of agriculture intensity as a covariable in the model. As a proxy, we selected agriculture machinery, expressed as the number of tractors per 100 km² of arable land for the year 2007 (FAO 2007). To account for the large range of variables and to linearize the relationship, the sum of probabilities, population estimate and density of tractors were log-transformed. We carried out model selection following an information-theoretic (IT) approach (Burnham and Anderson, 2002) to assess the relative role of agriculture and habitat (SDM) on population sizes. Although IT model selection have been criticized (Guthery *et al.* 2005), it should be a reliable tool for selecting models based on more than one variable (Stephens *et al.* 2005), providing that appropriate hypotheses and interpretation are followed. Since IT methods can lead to overfitted models when too many useless variables are included (Guthery *et al.* 2005), we tested here only two variables that are likely to be somehow linked with corncrake populations. Likewise, IT methods give no information on prediction ability of models (Guthery *et al.* 2005) but base model selection on the relative loss of information. We thus also based our interpretation on models' coefficient of determination (R^2) to estimate how well models fit our data. In order to avoid possible misinterpretation due to the use of a single criterion, we reported results for both AICc (Akaike Information Criterion corrected for finite sample size) (Burnham and Anderson, 2002) and BIC (Bayesian Information Criterion) (Schwarz 1978).

2.2.4. Results

2.2.4.1. Model-based distribution vs. Expert-based range

The reference Birdlife model matched well with Birdlife range estimate (overlap with binary map: 0.81) (Figure 15) and had a fair AUC (0.79). Its cAUC and TSS were the lowest (0.57 and 0.63) (for models evaluation metrics, see Appendix B, Table B.2), which is not surprising knowing that it derived from non-true presence points and cannot be considered biologically meaningful. The distribution model derived from the whole dataset had a higher AUC (0.85), which would be classified as good (Araújo et al. 2005, adapted from Swets 1988), and higher cAUC and TSS (0.61 and 0.71). However, a visual inspection of the output map revealed that the predicted distribution strongly departed from the expert-based distribution provided by Birdlife (Figure 16). High probabilities of presence were strongly skewed towards Western Europe compared to Birdlife EOO. The three measures of overlap also revealed a large deviation from the Birdlife model. The niche overlap computed between SDMs was low (D_{SDM} = 0.28), as well as the environmental overlap (D_{env} = 0.32) (Figure 17). The binary overlap was even lower (0.18) which highlighted a very poor match between the uncorrected model and the Birdlife estimation.

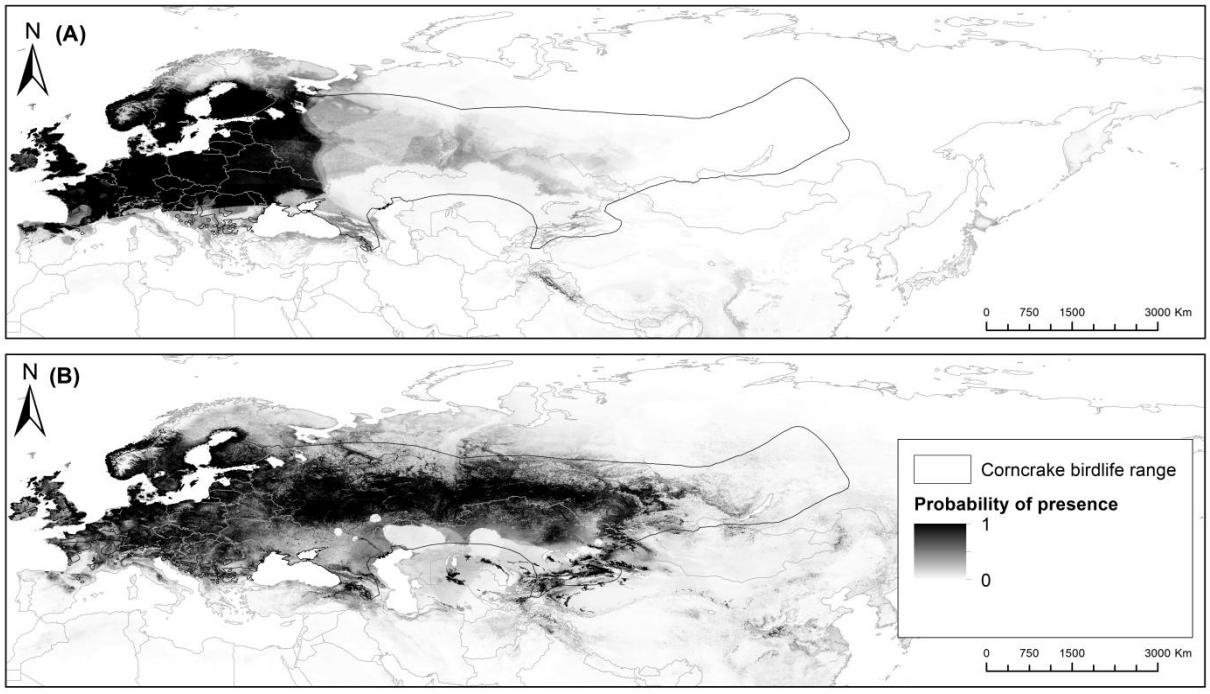


Figure 16: Corncrake species distribution models, without correction (A) and with sampling bias correction based on the combination of two sub-models (Western and Eastern) and systematic subsampling of occurrences (B).

The SDM derived from the combination of eastern and western model, after systematic sampling of records had a higher performance than the global model (AUC = 0.89, cAUC = 0.74, TSS = 0.73). It produced a very different range than the uncorrected model. Contrary to the latter, the predicted distribution visually covered a similar range to the Birdlife EOO (Figure 3). The three measures of overlap also indicated a better match with the Birdlife model (D_{SDM} = 0.60, D_{env} = 0.50 and Binary overlap = 0.40) (Figure 17). Compared to the full set of corncrake records, our approach smoothed the distribution of occurrences in environmental space. It also successfully corrected for the different weights of eastern and western occurrences in the complete dataset (Appendix B, Figure B.1).

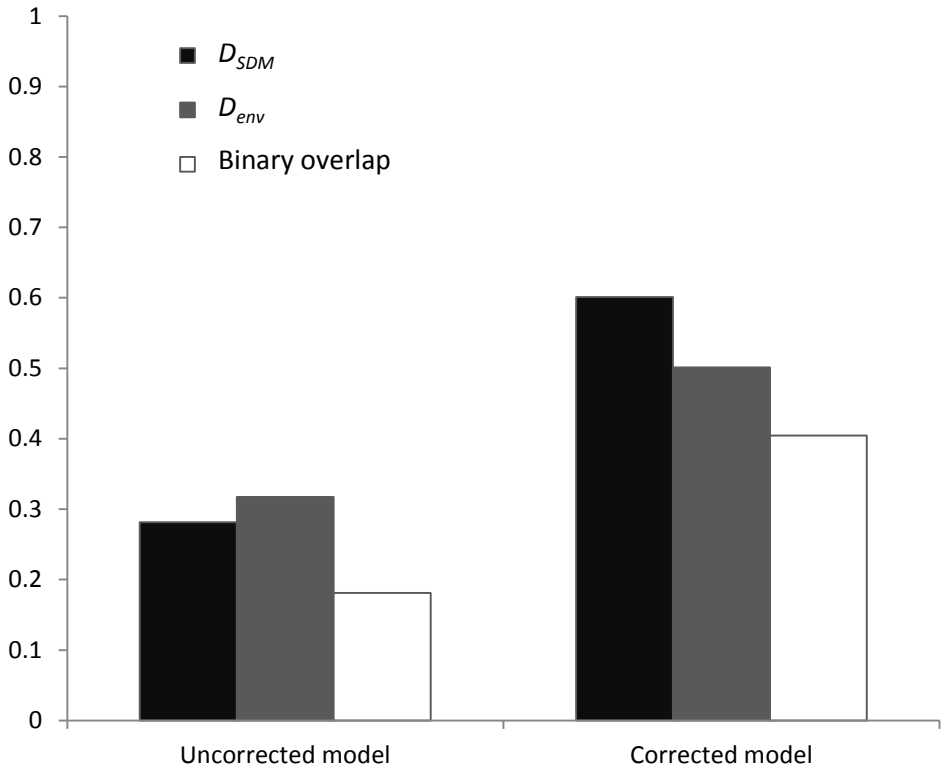


Figure 17: Overlap between the null model (based on Birdlife estimated range) and the uncorrected model (left) or the corrected model (right). The overlap was measured using Schoener's *D* index calculated on SDMs and in the environmental space, and the overlap between binary maps.

2.2.4.2. Predicted probabilities of presences vs. estimated abundance

We selected the complete linear model (Appendix B, Table B.3) that included both the sum of

probabilities of presences and agricultural intensity (AICc = 158.5 ; ΔAIC with 2nd best model: 10.7; AIC weight = 0.99, BIC = 163.9 ; ΔBIC with 2nd best model: 9.6; BIC weight = 0.98). This model best explained the mean estimated corncrake population size per country and accounted for more than 40% of total variation (intercept: 0.04217; log (Sum probabilities presence / Country area): 2.154; log (Tractors / 100 km²): -1.085, $R^2 = 0.42$). The effect of habitat suitability or agriculture intensity alone accounted for only 17 and 16% of variation in population sizes respectively. The probabilities of presence per country derived from the corrected SDM were positively related to the nationwide population sizes and negatively related to agriculture intensity (Figure 18). Thus, our corrected SDM provided outputs that were very consistent with Birdlife population size estimates.

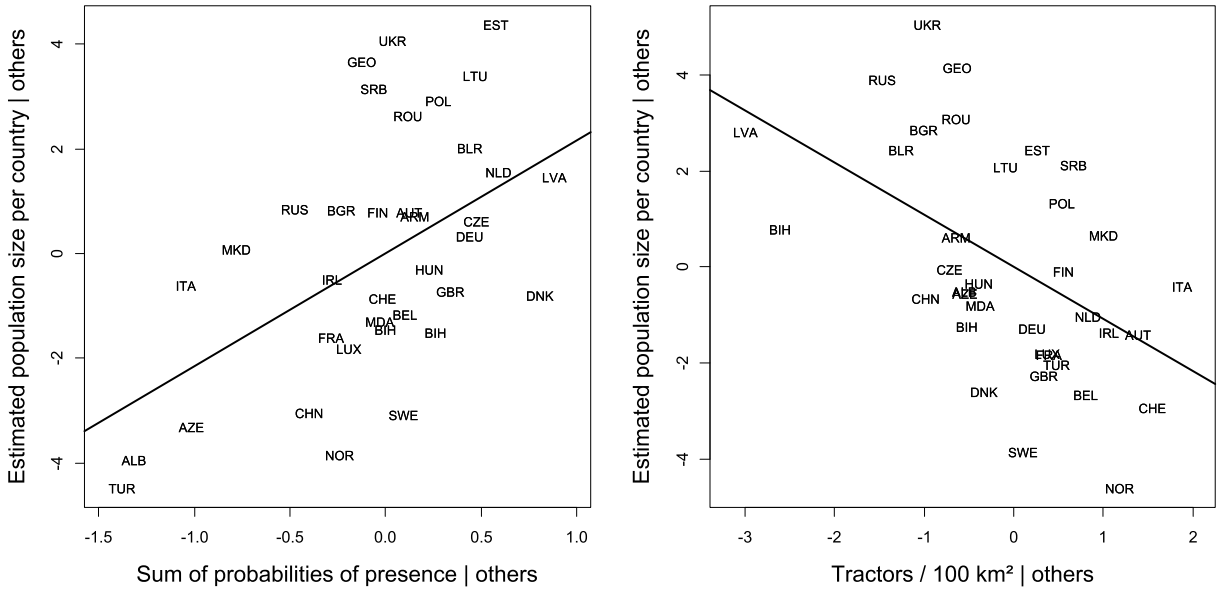


Figure 18: Partial regression plots of Birdlife population size estimations per country vs. sum of SDM probabilities of presence (left) and number of tractors per 100 sq. km of arable land (right). The three variables were log-transformed and partial residuals were plotted. The sums of probabilities of presence and the estimated population sizes were standardized by country area. Country names were coded following the ISO 3166-1 alpha-3 code (Appendix B, Table B.4).

2.2.5. Discussion

2.2.5.1. Comparison between model-based and expert-based distributions

We built a preliminary distribution model without taking into account the strong sampling bias in the dataset. The map of predictions greatly differed from the Birdlife range. Without further

analyses, the uncorrected model predicted a distribution centred on Europe. This would suggest that the corncrake distribution in Russia is highly overestimated which strikingly contrasts with the expert-based assessment. However, despite the lack of information available in this part of corncrake range, there is some evidence of the existence of high densities in Russia. Indeed, the few sites studied across Russia revealed large densities and population sizes (Sukhanova & Mischenko 2003), which is not predicted by the uncorrected model. This questions the reliability of this model which only predicts low suitability after 30° longitude.

On the contrary, after correcting the sampling bias, the potential distribution was much more congruent with the Birdlife range. It exhibited a higher overlap and visually presented a better match with the reference model. However, even if the corrected model can be considered as the best one, the overlap with Birdlife range remained moderate. This was especially noticeable when considering the overlap between binary maps, which is lower than 0.5. Indeed, the potential distribution seems narrower in Asia than the Birdlife estimate suggests. The model indicates a narrow strip of high suitability across Russia up to 90° longitude whereas Birdlife considers the species present up to Baikal Lake. The model may underestimate the eastern distribution because of the low number of available records. The shape of the expert-based range in its easternmost part is likely to correspond to the local knowledge of the species presence which was incorporated in the expert evaluation. However, by nature such data is impossible to include in statistical models like SDMs. Depending on the algorithm and its ability to project the modelled distribution in an unsampled environment, the modelling approach may failed to detect the presence of a species in areas with incomplete coverage. Species' migration and dispersal abilities are also likely to constrain distribution (Allouche *et al.* 2008), which would especially affect SDM accuracy at the range edge.

On the contrary, in Western Europe the modelled distribution is larger than estimated by expert-based assessments. This is probably a consequence of the decline of the species in this part of its range which is not reflected by the model. SDM indeed estimate the potential habitats but cannot take into account the deviations from this theoretical distribution unless relevant predictors such as *e.g.* anthropocentric actions and biological interactions are included in the analysis. The overestimation of IUCN range estimations has already been pointed out and seems to be common (Jetz *et al.* 2008; Herzog *et al.* 2012). It has been suggested that, rather than modelling the entire distribution of a species, the more efficient approach might be to model species distribution at a local scale. Small-scale models could thus be more likely to take local conservation concerns into account (Jiménez-Alfaro *et al.* 2012), but may fail to accurately predict

future distribution under climate change (Barbet-Massin *et al.* 2010). Another approach is to build a consensus model from the weighted combination of several individual models based on different statistical techniques (Araújo and New, 2007). By doing so, one may expect to reduce uncertainty, smooth individual errors and improve model precision. This approach may provide more robust predictions for conservation practitioners, although analysing the reasons for divergence between models is not straightforward and requires a deep understanding of very different and complex statistical methods.

2.2.5.2. Effect of sampling bias correction

A relevant sampling bias correction strategy should improve significantly the SDM output. Although we cannot ensure that the corrected model had a perfect fit to the true distribution, our results suggest that we produced a more accurate model than the one created without taking sampling bias into account. Several methods of bias correction have been proposed in previous studies such as including a grid of sampling probability in MAXENT modelling (Elith *et al.* 2010) or sampling one record per environmental cluster (Rödder *et al.* 2009b). However, there has been no clear guideline to build non biased SDM so far (but see Kramer-Schadt et al., 2013).

We designed here a specific strategy adjusted to our species. The large range of the Corncrake and the easy identification of the sampling bias made our case study ideal for splitting the dataset in Western and Eastern subsets. At this scale, the distribution of occurrences encompasses a large range of climates which makes relevant the building of two models in which the distribution of the species is driven by different predictors (Appendix A). Coupled with to the split approach, we used the systematic sampling of occurrences within each dataset. This simple method dissolve the spatial aggregation of records and, in most cases, also ensures that the occurrences are not clustered in the environmental space (Appendix B, Figure B.1).

The strong deviation between the uncorrected and the corrected models highlights the crucial importance of the prior evaluation of sampling bias and the design of a relevant correction strategy. Not considering this step might underestimate the actual range of a species, which would be highly detrimental in a conservation perspective. Along with the biased training dataset, the lack of independent test data might distort the interpretation of these models. A rigorous model assessment would imply the use of data collected independently from the training data. The collection of data *a posteriori*, such as field expeditions to confirm or infirm the presence of a species in area where the model predicts high suitability may be a way to test model

performance (Fielding & Bell 1997). However, such model assessment would be highly costly and would be most of time out of the reach. Another solution, which combines statistical modelling and expert knowledge, is to involve experts in model assessment. The model can be submitted to one or more experts who have enough knowledge about the species to evaluate the plausibility of the model.

2.2.5.3. Relationship probabilities of presence / abundance and effect of agriculture

We observed a positive correlation between population size estimates and the sum of probabilities of presence. Our distribution model was therefore able to discriminate between countries with high and low corncrake abundance. Some countries exhibit a notable deviation from this relationship: western countries generally hosted smaller populations than expected by the model whereas several Eastern European countries had higher population sizes. This deviation from the SDM / abundance relationship was efficiently corrected by agriculture intensity. The difference in agricultural intensity between high income western European countries and former soviet countries influenced the pattern of deviation from the correlation. Therefore, we observed a reasonable congruence between SDM outputs and Birdlife population size estimates. We also corroborated the effects of agricultural practices on corncrake presence. However, the model residuals still showed a distinction between Western and Eastern countries. Some other factors may explain this pattern, such as a variation in model accuracy due to the high differences in data quality.

SDM predictions are considered as indicators of environmental suitability (Elith & Leathwick 2009), and it is thus expected that they represent the ability for a species to establish and maintain feral populations in a given area. However, there is little empirical evidence that probabilities of presence computed by SDMs are actually correlated with local abundance or density. It has been shown that SDM predictions can be linked to population density but not in a simple linear relationship (Tôrres *et al.* 2012) because contrary to abundance, SDM output is ranged from 0 to 1. SDM can be interpreted as a proxy for local carrying capacity (VanDerWal *et al.* 2009b). Although maximum local abundance should have a positive relationship with suitability, many factors may affect local abundance even in case of high environmental suitability (*e.g.* when microhabitat conditions becomes more important rather than environmental suitability on a macro scale; Filz et al. 2013). The expected pattern would thus be a triangular correlation

between suitability and observed density (Tôrres *et al.* 2012). We found support for the existence of such a link in the present study. However, the relationship is likely to be improved by additional covariables that reflect local factors affecting species presence.

2.2.5.4. Implications for corncrake conservation

According to the corrected model, that we consider acceptable, we are able to confirm most of IUCN/Birdlife assessments about the Corncrake range. Our model highlights the existence of a large patch of suitable habitat in Russia and central Asian republics. This area is estimated to host the largest part of corncrake populations (between 1.5 and 3 million pairs estimated in Russia on a global population of 1.7-4.8 million breeding pairs), and the model seems to be in accordance with this assumption. This supports the widespread distribution of the species in the eastern part of its range that has conditioned the recent re-evaluation of its conservation status.

Although the SDM approach appears to corroborate the Birdlife assessments about the large distribution and population size of corncrakes, one has to take this conclusion with caution. First, the output map of SDM is only a spatial projection of realized niche computed from the input occurrences and predictors but not necessarily the current occupied range (Elith & Leathwick 2009). Indeed, if some factors such as biotic interactions or anthropogenic threats are not reflected by the predictors used during the modelling process, the actual species distribution may significantly deviate from the modelled one. This is visible in our model which remarkably predicts the presence of the Corncrake in Western Europe, *e.g.* in UK or France, in areas from where the species has been extirpated by agriculture intensification (Green *et al.* 1997a), increasing the fragmentation of the breeding range. Without further analyses at the local scale, we are not able to assess the contribution of agriculture and human activity in the Asian part of corncrake range that may affect the true distribution of the species. This contribution of economic development to species' local extinctions, and especially agriculture, is widely encountered (Czech & Krausman 1997). For instance, the recent accession of Poland to European Union has permitted the expansion of intensive agriculture, which had a substantial impact on farmland bird populations (Sanderson *et al.* 2013). Moreover, the model does not entirely consider the dispersal capacity of the species, even if we tried to account for it by including the migration distance as an environmental predictor (which is ranked third important variable in the eastern SDM). Our SDM indeed predicts high suitability areas in far-east Asia, when the species is known to be absent, for historical and dispersal reasons. The potential species' distribution in migratory

pathways and wintering range is also ignored from the SDM approach since it only predicts the suitable breeding range. However, winter survival and migratory conditions may have a crucial role in the decline of threatened avian populations (Ockendon *et al.* 2012).

Even so, given the strong accordance between the model-based and expert-based distributions, we can consider that the actual distribution of the species is unlikely to largely differ from our model. Nevertheless, the Corncrake remains a high conservation priority in some parts of its range, and would benefit from monitoring in the eastern part of its range. Except a significant recovery in UK, populations are still in a tenuous state in several western European countries. Moreover, as well as many bird species, climate change is likely to affect corncrake distribution in various ways. Firstly, a range shift towards northern latitudes is expected in response to changes of climatic conditions (Chen *et al.* 2011). Secondly, global warming may modify agricultural land-use (Olesen & Bindi 2002) and, for instance, disturb areas currently spared by intensive agriculture. Modelling the future distribution of the species under climate change can help to anticipate future conservation issues (Matthews *et al.* 2011). However, the accuracy of such SDM remains highly dependent on our ability to model future climatic conditions and its economic drivers.

Furthermore, the current trends in agricultural changes in former USSR are in favour of an upcoming decline of grassland birds (Kamp *et al.* 2011). The increase of agricultural exploitation of steppes may greatly affect the availability of favourable habitats in the years to come. Our results tend to confirm that central Asia is the core of the Corncrake range. Forthcoming conservation issues are likely to take place there, especially because of the current economic development of these regions (Czech 2000). A rigorous monitoring of the species and its potential threats in this part of its range would allow anticipating future events and may lead to reconsider its global status.

2.2.6. Conclusion: role of expert knowledge and SDM predictions

The value of expert knowledge and SDM predictions are both to be considered in conservation and are likely to provide benefits at different steps of assessment. On the one hand, predictions, such as those of SDMs or demographic models, may unravel unexpected conservation issues. On the other hand, experts may alert stakeholders of a threat and be the starting point of further assessment, or evaluate a model output through their knowledge of the species ecology. Model-

based assessment of species distribution provides an evaluation of species range which is free from the potential subjective point-of-view or overconfidence of experts (Burgman et al., 2011). The respective role of modelling and expert judgment will highly depend on the amount of available data (Drew & Perera 2011). When the amount of information is very low, resorting to expert knowledge, professional ecologist or traditional expert (Berkes 2004) is likely to estimate a species' range more accurately than a mathematical model, whereas SDMs should provide less biased range estimates when the input dataset is adequate (see for instance Bustamante and Seoane, 2004; Pearce et al., 2001). However, this study highlights the value of SDM approach even with highly biased datasets. Regardless of the method used to predict the distribution and conservation status of a species, the actual effectiveness of a conservation action program often relies on the way recommendations are implemented by authorities or practitioners.

2.2.7. Acknowledgements

The authors would like to thank all the people, organizations and websites that provided corncrake records: A. Mishenko, artportalen.se, Birdlife International, Birdlife Suomi, C. Moga, Conservatoire d'espaces naturels de Picardie, elurikkus.ut.ee, Fugle og Natur, Global Biodiversity Information Facility, J. Kamp, K. Koffijberg, L. Bozic, Ligue pour la Protection des Oiseaux (LPO) France, LPO Anjou, LPO Loire-Atlantique, M. Brambilla, M. Flade, Ma Ming, naturgucker.de, O. Keiss, Office National de la Chasse et de la Faune Sauvage, Royal Society for Protection of Birds, S. Boldogh, Schweizer Vogelschutz, worldbirds.org, xeno-canto.org. YF, AGB and JS received support from the Plan Loire Grandeur Nature and the European Regional Development Fund (ERDF). JOE received support within the German Federal Environmental Foundation fellowship programme.

2.2.8. Appendices

Appendix A

Table A.1: Relative contribution of each predictor for western and eastern model in the corrected distribution model. In bold are shown the three most important variables for each model.

Variable	Percent contribution	
	Western model	Eastern model
Altitude	13.2	1
Annual Mean Temperature	7.6	4.4
Annual Precipitation	0.3	0.8
Distance to wintering area	5	**19.7**
Isothermality	2.4	1.7
Land cover	2.4	**21.7**
Maximum Temperature of Warmest Month	**14.7**	0
Maximum NDVI	1.3	0.8
Mean Diurnal Range	1	0
Mean Temperature of Wettest Quarter	1.2	2
Minimum NDVI	2.9	0
Precipitation April to May	0.4	0.3
Precipitation of Coldest Quarter	0.3	0.5
Precipitation of Driest Month	**22.9**	6.8
Precipitation of Wettest Month	0.1	0.1
Precipitation Seasonality	0.3	0.9
Temperature Annual Range	**21.1**	1.7
Vegetation growth (NDVI of June - NDVI of April)	3.2	**37.6**

Appendix B. Supplementary material

Table B.1: Sources and years of corncrake occurrence data. When the observations correspond to data already used in an article, the reference is indicated by an exponent number and detailed below.

Provider	Years	Number of occurrences
A. Mishenko[123]	1995 - 2003	31
artsobservasjoner.no	2006 - 2010	524
artportalen.se	2006 - 2010	5183
avibase.bsc-eoc.org	2009 - 2011	27
birding.hu	2011	61
BirdLife Suomi	2000 - 2010	7961
C. Moga[4]	2010	111
Conservatoire d'espaces naturels de Picardie	1993 - 2010	379
ebird.org	2011	1
elurikkus.ut.ee	2004 - 2011	348
FugleogNatur	2007 - 2011	11
Global Biodiversity Information Facility (GBIF)	1999 - 2011	1174
J. Kamp	2004 - 2012	25
K. Koffijberg	2007 - 2010	707
L. Bozic[5]	2004	22
Ligue pour la Protection des Oiseaux (LPO) France	1994 - 2009	589
LPO Anjou	1984 - 2009	4999
LPO Loire-Atlantique	1999 - 2011	191
M. Brambilla	1996 - 2008	656
M. Flade	2001	5
Ma Ming[6]	1998 - 2000	9
naturgucker.de	1985 - 2011	161
O. Keiss[7]	1984 - 2004	20
Office National de la Chasse et de la Faune Sauvage	2002 - 2009	167
Royal Society for Protection of Birds	1993 - 2010	14832
S. Boldogh[8]	1997 - 2011	1703
Schweizer Vogelschutz	1996 - 2010	1655
worldbirds.org	2000 - 2011	229
xeno-canto.org	2007 - 2011	9

Articles from which corncrake observations were obtained:

[1]Mischenko, A. Corncrake *Crex crex* monitoring in European Russia in 2002-2003: a pilot study. Revista Catalana d'Ornitologia 65–70 (2008).

[2]Sukhanova, O. & Mischenko, A. Monitoring Corncrake *Crex crex* numbers in European Russia: the first stage. Ornis Hungarica 12, 135–141 (2003).

[3]Mischenko, A. & Sukhanova, O. The Corncrake (*Crex crex*) in Russia (European Part). Proceedings International Corncrake Workshop 1998 77–82 (1999).

[4]Moga, C. I., Hartel, T. & Öllerer, K. Status, microhabitat use and distribution of the corncrake *Crex crex* in a Southern Transylvanian rural landscape, Romania. North-Western Journal of Zoology 6, 63–70 (2010).

[5]Bozic, L. Breeding distribution and population size of Corncrake *Crex crex* in Slovenia in 2004. Acrocephalus 26, 171–179 (2005).

[6]Ming, M. & Qishan, W. New records of Corncrake *Crex crex* in Xinjiang, China. Forktail 18, 158–158 (2002).

[7]Keišs, O. Impact of changes in agricultural land use on the Corncrake *Crex crex* population in Latvia. Acta Universitatis Latviensis 691, 93–109 (2005).

[8]Boldogh, S., Szegedi, Z., Szentgyörgyi, P. & Petrovics, Z. Distribution, population size and conservation status of Corncrake *Crex crex* in North-east Hungary, 1997 – 2006. Vogelwelt 130, 153–158 (2009).

Table B.2: Evaluation metrics of the 3 models computed: Area under the ROC curve (AUC), Hijman's corrected AUC (cAUC) and True Skill Statistics (TSS).

	AUC	**cAUC**	**TSS**
Birdlife model	0.79	0.57	0.63
Uncorrected model	0.85	0.61	0.71
Corrected model	0.89	0.74	0.73

Table B.3 (a) Table of model selection following Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) and (b) coefficients and significance of factors of the selected model (Population Size ~ ΣProbability of presence + Tractors density + intercept). Population size per country and sum of SDM probabilities of presence are standardized by country area, and both factors and response variable are log transformed.

(a)

Intercept	ΣProb. of presence	Tractors	df	logLik	AICc	ΔAICc	AIC weight	BIC	ΔBIC	BIC weight	R^2
(A) Pop. Size ~ ΣProb. of presence + Tractors density + intercept											
0.0422	2.154	-1.085	4	-74.63	158.5	0.00	0.991	163.6	0.00	0.983	0.42
(B) Pop. Size ~ ΣProb. of presence + intercept											
-4.0410	1.667		3	-81.23	169.2	10.66	0.005	173.2	9.62	0.008	0.17
(C) Pop. Size ~ Tractors density + intercept											
9.2290		-0.8323	3	-81.40	169.5	11.00	0.004	173.5	9.96	0.007	0.16
(D) Pop. Size ~ intercept											
4.2410			2	-84.51	173.4	14.84	0.001	176.2	12.60	0.002	0.00

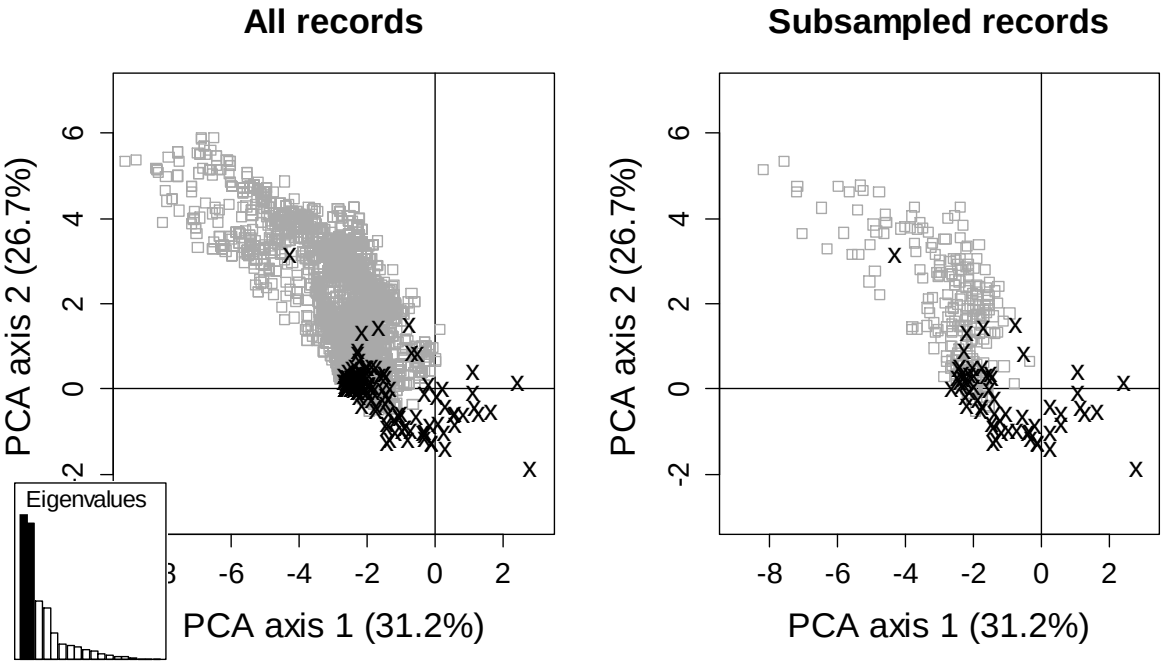
(b)

	Coefficient	Std. Error	p
ΣProb. of presence	2.1539	0.5549	0.00047
Tractors density	-1.0851	0.2838	0.00055
intercept	0.0422	2.9074	0.98851

Table B.4: ISO 3166-1 alpha-3 code used to shorten country names.

Country name	ISO 3166-1 alpha-3 code
Albania	ALB
Armenia	ARM
Austria	AUT
Azerbaijan	AZE
Belarus	BLR
Belgium	BEL
Bosnia & Herzegovina	BIH
Bulgaria	BGR
China	CHN
Croatia	BIH
Czech Republic	CZE
Denmark	DNK
Estonia	EST
Finland	FIN
France	FRA
Georgia	GEO
Germany	DEU
Hungary	HUN
Ireland	IRL
Italy	ITA
Latvia	LVA
Liechtenstein	LIE
Lithuania	LTU
Luxembourg	LUX
Macedonia	MKD
Moldova	MDA
Netherlands	NLD
Norway	NOR
Poland	POL
Romania	ROU
Russia	RUS
Serbia	SRB
Slovakia	SVK
Slovenia	SVN
Sweden	SWE
Switzerland	CHE
Turkey	TUR
Ukraine	UKR
United Kingdom	GBR

Figure B.1: Scatterplot of corncrake occurrences on the two first axes of a principal component analysis calibrated on the environmental conditions of the whole study area. Left: all occurrences, right: subsample of occurrences used in the corrected model (grey squares: western points (long < 30°), black crosses: eastern points (long > 30°))



2.2.9. Conclusion Article 2

Le manque de données collectées dans la partie orientale de la distribution du Râle des genêts faisait craindre de trop grandes approximations de la part de l'expert responsable de l'estimation de l'aire de reproduction de l'espèce. Pourtant, après correction du biais, la procédure de modélisation montre une bonne adéquation avec la distribution IUCN estimée à dire d'expert. Nos résultats confirment donc la large distribution du Râle des genêts à travers une grande part du Paléarctique. Cette distribution apparait continue dans sa partie orientale, où se concentrent les habitats les plus favorables, tandis que les sites d'Europe de l'ouest apparaissent plus fragmentés. On constate également qu'il existe une relation positive entre la modélisation et les estimations de tailles de populations lorsque l'intensité de l'agriculture est prise en compte, ce qui tend à montrer le double effet des conditions écologiques naturelles (habitat modélisé) et de l'activité agricole sur les populations de l'espèce.

Ce travail met ainsi en lumière le double intérêt des approches de modélisation et d'utilisation de la connaissance experte pour l'évaluation des statuts de conservation d'espèce. Ici, les deux approches permettent de délimiter approximativement la même distribution et d'estimer de la même façon les tailles de populations sous l'effet de l'activité humaine. La méthodologie développée ici montre la possibilité de conduire une modélisation de distribution efficace même à partir d'un jeu de données extrêmement biaisé, si tant est qu'une procédure de réduction du biais pertinente est appliquée.

Chapitre 2 : Structure spatiale des populations à l'échelle continentale

3. CHAPITRE 2 : STRUCTURE SPATIALE DES POPULATIONS A L'ECHELLE CONTINENTALE

A l'heure actuelle, on ignore largement la structuration spatiale des populations de Râles des genêts. Une certaine fidélité au site est connue chez cette espèce mais dans le même temps, des mouvements à longue distance sont documentés en période de reproduction. Ce chapitre a pour objectif de déterminer le degré d'isolement des populations de Râle des genêts en Europe, et notamment de comprendre les échanges existants entre les sites en déclin d'Europe occidentale et les sites abondants présents à l'est. Deux approches différentes ont été menées pour répondre à ces questions, via la caractérisation des variations du chant des mâles dans un premier temps, puis via une approche de génétique des populations.

3.1. ARTICLE 3 : VARIATIONS GEOGRAPHIQUES DU CHANT DES MÂLES EN EUROPE

La plupart des espèces d'oiseaux possèdent un vaste répertoire de vocalisations leur permettant de communiquer avec leurs congénères. Ce trait a essentiellement évolué sous l'action de la sélection sexuelle (Read & Weary 1990) et est utilisé notamment pour délimiter un territoire ou attirer l'attention de partenaires sexuels potentiels (Bradbury & Vehrencamp 2011). Les espèces pour lesquelles le chant n'est pas appris au cours du développement et pour lesquelles il n'existe donc pas de transmission culturelle présentent l'intérêt que seuls la génétique (Catchpole & Slater 2008) et les conditions écologiques locales expliquent les caractéristiques du chant, sans influence d'un quelconque processus d'apprentissage. Dès lors, la structure spatiale du chant de ces espèces est susceptible de refléter directement la structure génétique des populations. Etudier les variations géographiques et temporelles du chant dans ces populations peut alors permettre de déterminer des paramètres tels que la fidélité au site ou les migrations inter-populations.

Le Râle des genêts présente un chant stéréotypé constitué de deux syllabes séparées par une pause, répété toute la nuit afin de défendre un territoire de reproduction et d'attirer les femelles (Schäffer 1995; Ręk & Osiejuk 2010). Les paramètres du chant sont largement influencés par les interactions sociales, tout spécialement lorsque la présence d'un mâle compétiteur à

proximité induit des interactions agressives (Ręk & Osiejuk 2010; Budka & Osiejuk 2013a), et peuvent permettre de discriminer les individus (Mikkelsen *et al.* 2013). Il a été montré que les vocalisations du Râle varient entre populations (Peake & McGregor 1999) mais également au cours d'une saison (Osiejuk *et al.* 2004), tout du moins en ce qui concerne la durée des syllabes et l'intervalle entre elles.

Un total de 352 mâles a été enregistré dans 8 sites européens localisés en France, Norvège, Pologne et République Tchèque. Sept paramètres du chant ont été relevés afin de préciser les variations géographiques du chant des mâles Râles des genêts. On a notamment cherché à déterminer s'il existe des différences entre populations en fonction de la distance géographique qui les sépare ou si on retrouve un *pattern* géographique plus complexe. Dans deux sites, on disposait de deux années d'échantillonnage (2010 et 2011). Les différences interannuelles de chant ont donc pu être analysées dans ces deux populations dans le but de déterminer si les caractéristiques du chant observées dans une population sont stables entre deux saisons de reproduction.

Cette partie, issue d'une collaboration avec l'université Adam Mickiewicz de Poznań (Pologne), a fait l'objet d'une publication en anglais dans le journal « Journal of Avian Biology » présentée ci-après :

Budka M, Mikkelsen G, Turcokova L, Fourcade Y, Dale S, Osiejuk TS (2014) Macrogeographic variations in the call of the corncrake Crex crex. *Journal of Avian Biology*, **45**, 65–74.

Macrogeographic variation in the call of the corncrake *Crex crex*

Michał Budka[1*], Gunvar Mikkelsen[2], Lucia Turčoková[3], Yoan Fourcade[4], Svein Dale[2] and Tomasz S. Osiejuk[1]

[1] Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61614 Poznan, Poland

[2] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, NO-1432 Aas, Norway

[3] Laboratory of Ornithology, Palacký University, Tř. Svobody 26, 777 21 Olomouc, Czech Republic

[4] LUNAM, Université d'Angers, GECCO (Groupe écologie et conservation des vertébrés), 2 bd Lavoisier, 49045 Angers, France.

*Corresponding author

3.1.1. Abstract

This study was conducted to characterise macrogeographic variation in the vocalisation of the corncrake *Crex crex*, a bird species with a non-learned and highly stereotyped call. We also examined: 1) whether call characteristics remained stable across successive breeding seasons within two of the study populations and 2) whether call similarity was related to distance between populations.

Recordings of 352 males from eight populations were analysed. The analyses focused on variation in 1) temporal characteristics (duration of syllables and intervals, duration of the intervals between consecutive maximal amplitude peaks within syllables, called pulse-to-pulse duration (PPD)), and 2) spectral characteristics (minimal and maximal frequency, frequencies below which 25%, 50% and 75% acoustic energy of signal is distributed). We found significant differences in most of the temporal and all of the spectral characteristics between populations.

No differences were found in PPD. Significant interannual differences in spectral characteristics were found in both of the populations examined, whereas differences in temporal characteristics were only observed in one population. In general, geographic variation in calls showed clinal distance-dependence, where distant populations showed larger differences in call than neighbouring populations. Our results show that geographic variation in corncrake calls may be very dynamic in the short term and that within-population variation may occur on the same scale as between-population variation. This finding is surprising because call characteristics in non-learners are essentially inherited, and genetic transmission should be very slow. We suggest that the social interactions between males and/or the specific dispersal patterns of this species and the low site fidelity of adult and young birds may be responsible for such pattern.

3.1.2. Introduction

Birds are one of the most spectacularly vocalising groups of animals and exhibit enormous variation of vocal signals both among and within species (Catchpole & Slater 2008). This variation is also reflected in geographical differences in signals, which often vary on both micro and macrogeographical scale (Mundinger 1982). These two extremes are generally defined by the distance between individuals and the dispersal capability of a species (Nottebohm 1969; Mundinger 1982).

Regardless of the spatial scale at which differences in birds' vocalisations are analysed, most studies have been restricted to bird species that learn their structure of songs during ontogenesis, including songbirds (Oscines), parrots (Psittaciformes), and hummingbirds (Trochillidae) (reviewed in Podos & Warren 2007). Relatively few studies have considered geographic variation in vocalisation of non-learning species (Bretagnolle 1996; Galeotti *et al.* 1996; Appleby & Redpath 1997; Bretagnolle & Genevois 1997; Mager *et al.* 2007; Podos & Warren 2007; Odom & Mennill 2012). The main difference between learning and non-learning species is that both genetic (innate) and cultural (learned) components are responsible for song variation in the first group, whereas in the second vocal signals are essentially inherited (Catchpole & Slater 2008). However, innate vocalisations can be flexible and individuals can use them differently or can change some call characteristics such as: duration, intensity, rhythm or pitch of the sound. Nevertheless, these changes in vocalisation cannot be defined as a cultural transmission in the same sense as learning of song in *Oscines* or speech in human. In general, the occurrence of song variation in non-learners should largely be the result of evolutionary and

ecological factors such as: genetic drift, sexual selection, social interactions, morphological features of the vocal apparatus, habitat vegetative structure, or climate conditions, whereas in learners, additional song learning process must always be taken into account, because the cultural transmission of vocal characteristics is strongly developed in this group.

In non-learning species the geographical pattern of vocalisations should mainly reflect genetic differences between populations, and these differences should increase with increasing distance or population isolation. Therefore, one of two spatial patterns of geographic call variation may be expected. First, differences between neighbouring populations may increase continuously and in tandem with the distance between populations as a consequence of high connectivity among neighbouring populations and limited connectivity between distant populations (Isler *et al.* 2005). Second, call variation may be diffuse, with large differences occurring among both neighbouring and distant populations as a consequence of local genetic differences (Bretagnolle & Genevois 1997; Peake & McGregor 1999; Lein 2008; Odom & Mennill 2012).

The corncrake *Crex crex* is a highly secretive, non-learning bird species (Brenowitz 1991). Its breeding area extends from western Europe to western Siberia, and the birds winter in southern Africa (Cramp & Simmons 1980). During the breeding season, corncrakes inhabit mostly wet meadows with dense vegetation. Males are vocally active at night and make a characteristic loud cracking call (Cramp & Simmons 1980; Green *et al.* 1997a). The cracking call is the functional equivalent of a songbird song, since it is used in mate attraction (Cramp & Simmons 1980; Schäffer 1995) and rival deterrence (Ręk & Osiejuk 2010). In contrast to many songbird species, corncrakes demonstrate no repertoire variation, and corncrake call organisation is highly stereotyped. The call consists of two syllables separated by an interval, and it is repeated in a stereotypical manner thousands of times over the course of the night (Green *et al.* 1997a) (Figure 19). The syllables are structurally toneless and usually consist of 14 – 22 repeated pulses. Furthermore, individual birds have characteristic and invariable pulse-to-pulse duration (PPD) patterns, where PPD is the duration of the interval between consecutive maximal amplitude peaks (Peake *et al.* 1998). Therefore, researchers often use PPD to recognize individual corncrake males by call (Terry *et al.* 2005; Mikkelsen *et al.* 2013).

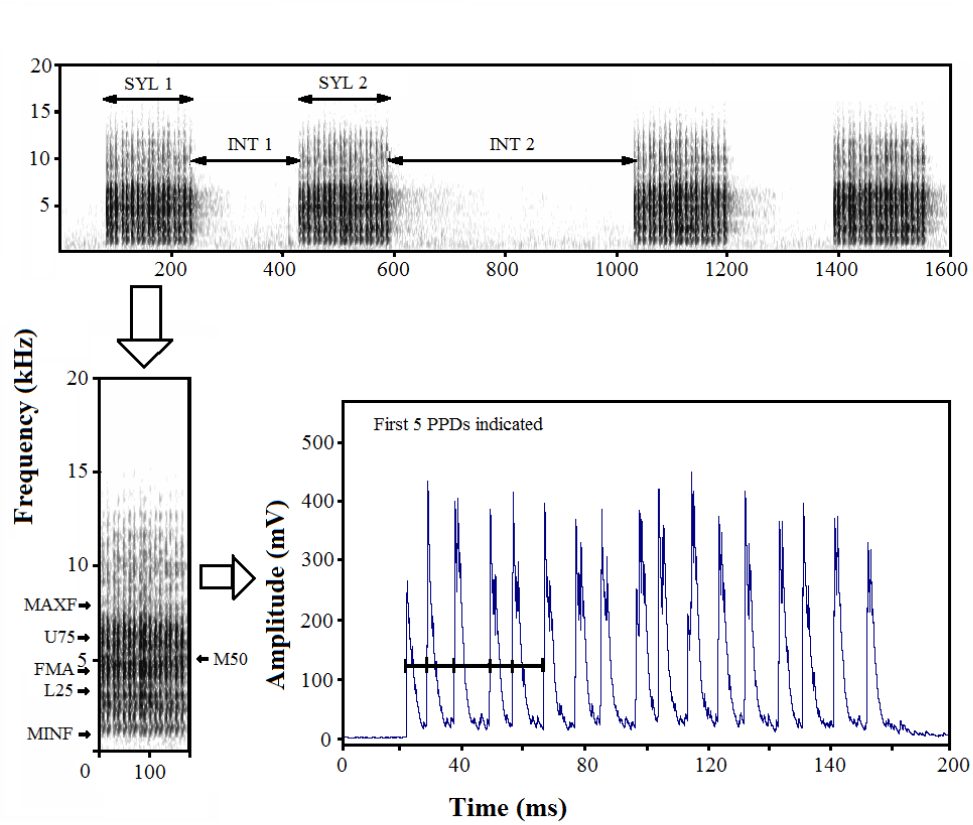


Figure 19: Illustration showing the corncrake calls characteristics measured. Abbreviations are as follows: SYL1 and SYL2 – duration of the first and the second syllable; INT1 and INT2 – duration of the within- and between-call interval; L25, M50, and U75 – frequency below which 25%, 50%, and 75% of the total signal energy is distributed; MINF – minimum frequency; MAXF – maximum frequency. Each syllable usually consists of 14 – 22 repeated maximal amplitude peaks. Number of pulses within first syllable (NP SYL1) is lesser than within second syllable (NP SYL2). Rhythm of calling is defined as a: RHYTHM = INT2 / (SYL1 + INT1 + SYL2).

Geographic variation in the corncrake call was studied by Peake and McGregor (Peake & McGregor 1999), who found that populations differed in both the duration of syllables and the interval between them. However, a later study on intraseasonal variation in calls revealed that individuals showed varied syllable and interval duration over the course of a single season (Osiejuk *et al.* 2004). Both syllables decreased slightly in the course of the season, whereas intervals showed U-shape curve, with the lowest values in the middle of the season and the longest in the beginning and in the end of the season (Osiejuk *et al.* 2004). Intraseasonal differences between extreme values in duration of both syllables and interval between them were small (ca 15 – 16 ms), whereas differences in duration of between-call interval were larger (up to

155 ms; Osiejuk *et al.* 2004). Therefore, variation in syllable and interval duration that is assumed to be geographical may actually reflect intraseasonal differences; to properly determine the source of variation, it is important to take into consideration the time at which the recordings were made. Moreover, Ręk and Osiejuk (2010; 2011) have shown that the rhythm of cracking, which is affected by duration of syllables and intervals (Figure 19), is a type of a conventional signal, and the rhythm and between-call interval are established during male interactions and are stabilized by a receiver retaliation rule. During aggressive interactions, males increase between-call interval and rhythm of calling (Budka & Osiejuk 2013a), and intermittent calling males (high value of rhythm) are more aggressive than monotonous calling males (low value of rhythm) (Ręk & Osiejuk 2010). Therefore, the social interactions can have important consequences for geographical variation in vocalizations of corncrake, and theoretically, different patterns of call characteristics may arise independently in various populations.

In this study, we focused on macrogeographic variation in male corncrake calls among populations that are separated by hundreds of kilometres. We investigated three components of corncrake call: 1) variation in syllables and intervals duration, and proportion between them resulting calling rhythm; 2) variation in pulse distribution within syllable; 3) variation in the distribution of acoustic energy within the vocal frequency range. These call components should presumably exhibit different patterns of variation as a consequence of their functions in the birds' life and the different mechanisms shaping their variation. Additionally, to test more specific hypotheses, we examined: 1) whether call characteristics were consistent across two different breeding seasons in two of the study populations and 2) whether call characteristic similarity was related to the distance between populations in order to determine whether geographic isolation plays a role in call variation.

3.1.3. Methods

3.1.3.1. Studied populations and call recording

Corncrake's calls were recorded in the Czech Republic, France, Norway and Poland during the 2009-2012 breeding seasons (Figure 20). Recordings were made at night (22:00-04.00, local time) between May 24 and July 17. The distance between microphone and bird varied from a few to about 20 m. In France, recordings were made using Sony DAT TCD-D8 recorder and an EMU 4535 electret condenser with EM 700 shotgun microphone. The rest of recordings were made using Marantz PMD 620 recorders and Sennheiser ME67 directional microphones with K6

powering modules. The individual birds we recorded were not marked. Thus, to avoid recording the same male twice, we attempted to record all the males within each suitable habitat patch during a single night.

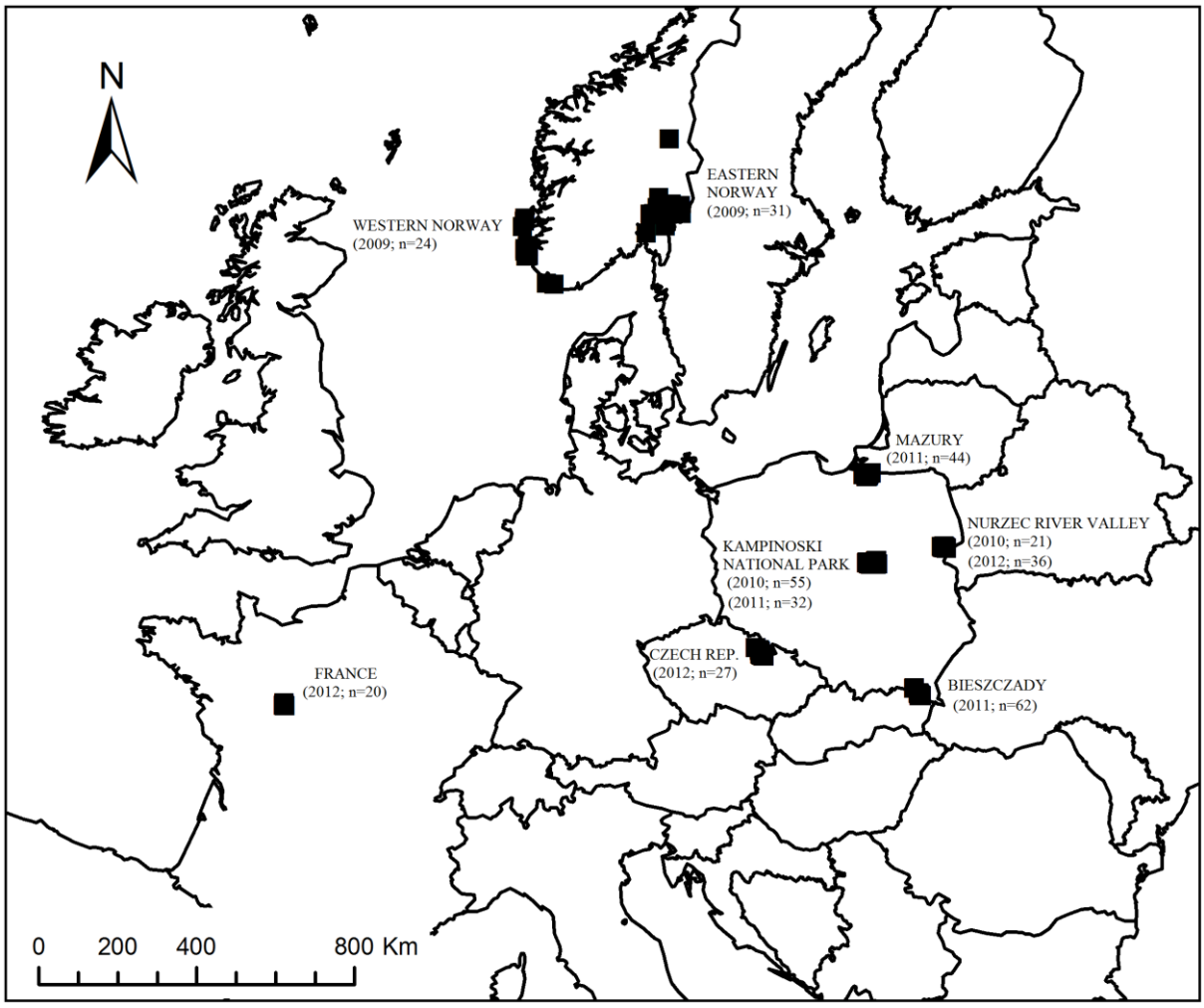


Figure 20: Map showing areas in which male corncrake calls were recorded. The year of recording and the number of recorded birds are given.

We collected recordings from 352 corncrake males in eight different populations: one in France (n = 20), two in Norway (western Norway, n = 24; eastern Norway, n = 31), one in the Czech Republic (n = 27), and four in Poland (Bieszczady, n = 62; Kampinoski National Park, n = 87; Nurzec River Valley, n = 57; and Mazury, n = 44). In two populations, recordings were obtained in two different years: in 2010 (n = 55) and 2011 (n = 32) in Kampinoski National Park and in 2010 (n = 21) and 2012 (n = 36) in the Nurzec River Valley. However, in our main

analyses, we used the year with the larger number of birds sampled (i.e. 2010 for Kampinoski National Park and 2012 for the Nurzec River Valley). Distances between populations ranged from 184 to 1710 km. The exact locations of the calling males were determined using GPS receivers. All GPS data were transferred from the receivers to a PC workstation and then calibrated using the WGS 84 datum. ArcGIS ver. 9.3 was used to calculate the distances between populations.

3.1.3.2. Call analysis

In the populations in Norway and France, calls were recorded at a 48.0 kHz/16 bit sampling rate. In the other populations, a 44.1 kHz/16 bit sampling rate was used. Therefore, all recordings were calibrated so that they had the same digital quality (44.1 kHz/16 bit sampling) before we analysed the calls. For each male, we chose 20 calls. The call was defined as a first and second syllable, interval between syllables and interval between calls. We always analysed the first 20 consecutive calls that did not have significant background noises. As a result, the sample was essentially random. The sound files were analysed using the following settings in Avisoft SASLab Pro ver. 5.2.04 software: FFT length = 512, Frame = 25%, Window = Hamming, and Temporal Overlap = 87.50%. The resulting spectrogram had a 448 Hz bandwidth at a sampling frequency of 86 Hz and a time resolution of 1.45 ms (Specht 2007). Using the above settings, we measured the following temporal characteristics: SYL1 and SYL2 – duration (ms) of the first and the second syllable; INT1 and INT 2 – duration (ms) of the within-call and between-call interval. Then we calculated calling rhythm (RHYTHM = INT2/(SYL + INT1 + SYL2)), which reflects whether calling is monotonous (low values of RHYTHM) or intermittent (high values of RHYTHM) and is a signal of aggression and/or male quality (Osiejuk *et al.* 2004; Ręk & Osiejuk 2010).

To analyse the PPD structure and the number of pulses (NP) within syllables, we used the 'pulse train analysis' function in Avisoft SASLab Pro. Before PPD measurements were made, the FIR time-domain filter (500 Hz; high pass setting) was used to remove low-frequency noises from all sound files. In the pulse train analysis, we used the 'rectification + exponential decay' method to make our PPD measurements. We also initially used the following settings: time constant = 1 ms, threshold = 0.10 V, hysteresis = 10 dB, and start end threshold = -8 dB. However, for a few sound files, we had to decrease the threshold and set hysteresis between 9

and 12 dB to correctly detect all the pulses. Each measurement of pulse distribution was visually checked to ensure that all the pulses had been detected.

Variation in the distribution of acoustic energy within the vocal frequency range was measured using a one-dimensional function called 'Amplitude spectrum (linear)' and employing a Hamming evaluation window implemented in Avisoft SasLab Pro. We measured the following spectral characteristics: L25, M50 and U75 – frequency (Hz) below which respectively 25%, 50% and 75% of the total signal energy is distributed; MINF – minimum frequency (Hz); MAXF – maximum frequency (Hz) at which syllable amplitude falls below – 20 dB (relative to the maximum amplitude). To remove background noises, the spectral characteristics option 'total' was on and the minimum frequency range was limited to 0.5 kHz. This value was used in accordance with the guidelines provided in Osiejuk and Olech (2004). All the measurements are illustrated in Figure 19.

3.1.3.3. Statistical analysis and data selection

We tested for differences in call characteristics among the eight European populations (299 males in main analyses) and between years in two of the study populations (Nurzec River Valley 2010 vs 2012 and Kampinoski National Park 2010 vs 2011). First, we calculated the mean values of all the characteristics using the 20 calls from each male. The mean values were used in subsequent analyses. Multivariate analysis of variance (MANOVA) was used to compare differences in the temporal characteristics of calls among populations and between years. Call characteristics were entered into the models as dependent variables and population or year was entered as a fixed factor. Intercept was included in the models. All models had a full factorial design and were run using the type III sum of squares. Because the temporal call characteristics of individual corncrakes, especially INT2 and RHYTHM, can vary over a single season (Osiejuk *et al.* 2004), we first determined whether date was correlated with any temporal call characteristics, both within each population and across the combined dataset. We found no significant relationships (linear and nonlinear regression models; in all cases $p>0.05$). Consequently, we did not use time as a covariate in the models examining geographical variation in corncrake calls.

Call characteristics describing energy distribution in the frequency domain (i.e. L25, M50, U75, MINF and MAXF) were highly and significantly correlated between the first and second

syllables (Pearson correlation: in all cases $r > 0.96$, $p < 0.0001$). Therefore, we only used the characteristics of the first syllable in the analyses. Because L25, MINF, and MAXF variances were not equal across groups (Levene's test of equality of error variances; $p < 0.05$), we performed five separate Kruskal–Wallis tests. The Bonferroni correction (Holm' s method) was used to control for the multiple comparisons (Holm 1979).

Geographic variation in PPD structure was examined using a randomization test we designed. In our study, the lowest number of pulses was 12 within the first syllable and 14 within the second syllable. However, the correlation between the first 11 PPDs of the first and second syllables was high (Pearson $r > 0.96$; $p < 0.0001$). Therefore, in the analyses, we used only the first 11 PPDs of the first syllable. We tested whether pairs of males from the same population in the same year had more or less similar PPD structure than pairs of males belonging to different populations. To do this, we first calculated squared Euclidean distances between all pairs of males within each population. We then compared them with values obtained in the same way from randomly reshuffled pairs of males from different populations.

Additionally, to check whether differences in population call characteristics were distance-dependent, we ran a simple exact Mantel test in Rundom Pro 3.14 software (Jadwiszczak 2009). First, we calculated the mean values of temporal (SYL1, INT1, SYL2, INT2, NP SYL1, NP SYL2 and RHYTHM) and spectral (first syllable's L25, M50, U75, MINF and MAXF) call characteristics for each population. Then, we performed the exact Mantel tests (number of permutations = 40 320; matrix size = 8 × 8), where we compared a matrix of geographic distances between populations (central latitude and longitude coordinates, in km) and matrices of the differences in each call characteristic between populations (Euclidean distance). To confirm that geographic calls was distance-dependent, we performed four series of Mantel tests (in each test eight populations were considered), in which all possible combinations of different recording years in Kampinoski National Park and Nurzec River Valley were considered. The Bonferroni correction (Holm's method) was used to control experiment-wide type I error (Holm 1979). All variables were normally distributed (Kolmogorov–Smirnov Z test; $p > 0.05$). All statistical analyses were performed using IBM SPSS Statistics software ver. 21 unless otherwise stated. All p-values are two-tailed.

3.1.4. Results

3.1.4.1. Variation of temporal call parameters

Multivariate general linear model showed that temporal characteristics of corncrake calls varied significantly among populations (Pillai's Trace test: $F_{49, 2037}$=1.860; p<0.0001). Differences were found in SYL1, SYL2, INT2, NP SYL1, NP SYL2 (Table 4, Figure 21). Th is variation was clearly distance-dependent when 2010 recordings from Kampinoski National Park and 2012 recordings from Nurzec River Valley were included in the models. The Mantel-tests also showed significant positive correlations between the distance separating populations and the degree of difference in the temporal characteristics of their calls: neighbouring populations had more similar call than distant populations. However, when we used other combinations of recording years, either only the number of pulses were significantly different or no significant differences in any of the temporal characteristics were found (Table 5).

Table 4: The results of the multivariate analysis of variance of between-population differences in the temporal characteristics of corncrake calls. In the model, SYL1, INT1, SYL2, INT2, NP SYL1, NP SYL2 and RHYTHM were included as dependent variables and population was included as a fixed factor. Abbreviations of variables are defined in Figure 19.

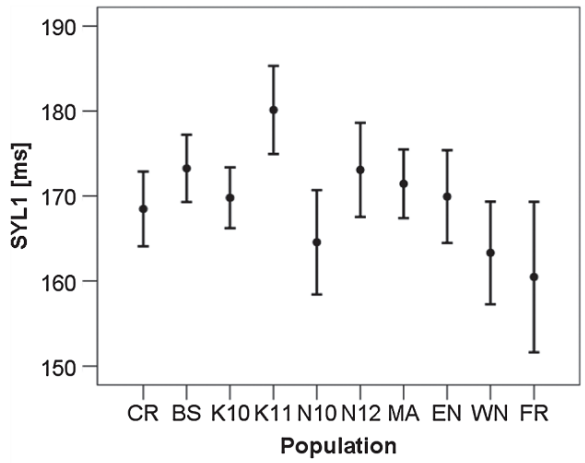
Effect	F	df	P
Multivariate Pillai's Trace test			
Intercept	36962	7,285	0.0001
Population	1.860	49, 2037	0.0001
Between-Subject Effects			
SYL 1	2.681	7	0.011
INT 1	1.387	7	0.210
SYL 2	4.035	7	0.001
INT 2	2.472	7	0.018
RHYTHM	1.992	7	0.056
NP SYL1	2.495	7	0.017
NP SYL2	3.847	7	0.001

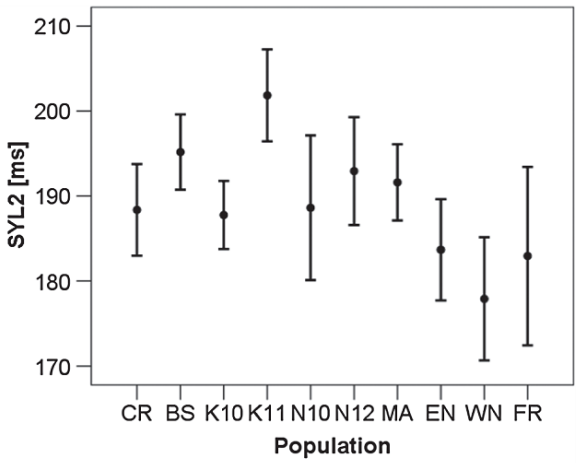
Table 5: Results of the exact Mantel-tests. The matrices of the distances between populations (in km) and the matrices of dissimilarity in male call characteristics (Euclidean distances) were compared. Because calls were recorded in two different years in two of the populations, we analysed all four possible combinations of years. Abbreviations are as follows: KP10 or KP11 tests in which recordings from Kampinoski National Park from 2010 or 2011, respectively, were used; NV10 or NV12 – tests in which recordings from Nurzec River Valley from 2010 or 2012, respectively, were used. High, positive r values mean that neighbouring populations were more similar in their call parameters than distant populations. ∗ indicates the result was significant after the Bonferroni correction (Holm's method) was applied. Abbreviations of variables are defined in Figure 19.

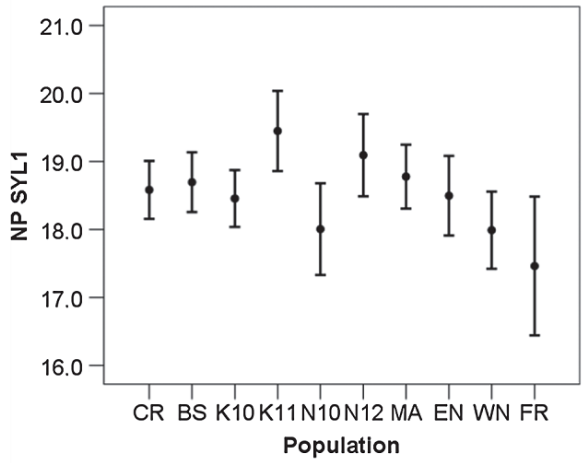
	KP10&NV12		KP10&NV10		KP11&NV12		KP11&NV10	
Variant of test	*r*	p	*r*	p	*r*	p	*r*	p
Temporal characteristics								
SYL1	0.67	0.001*	0.38	0.052	0.45	0.017	0.09	0.630
INT1	0.27	0.153	0.07	0.737	0.31	0.114	0.10	0.603
SYL2	0.53	0.003*	0.47	0.010	0.38	0.050	0.24	0.220
INT2	0.20	0.295	-0.01	0.982	0.12	0.550	0.04	0.841
RHYTHM	0.48	0.010*	0.08	0.687	0.42	0.030	0.09	0.644
NP SYL1	0.67	0.001*	0.44	0.022	0.59	0.001*	0.22	0.271
NP SYL2	0.51	0.004*	0.54	0.003*	0.53	0.005*	0.41	0.032
Frequency characteristics								
L25	0.71	0.001*	0.57	0.003*	0.59	0.001*	0.41	0.029
M50	0.50	0.009*	0.22	0.256	0.31	0.109	0.05	0.816
U75	0.78	0.001*	0.50	0.011*	0.21	0.295	0.05	0.816
MINF	0.51	0.006*	0.51	0.006*	0.65	0.001*	0.65	0.001*
MAXF	0.63	0.001*	0.74	0.001*	0.19	0.323	0.22	0.257

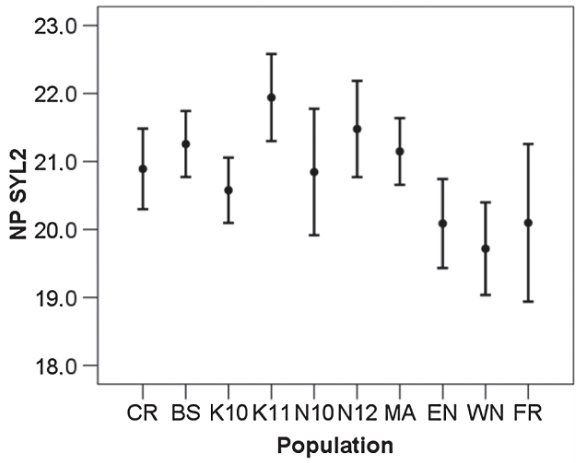
We also found significant interannual differences in the temporal characteristics of calls in Kampinoski National Park (Pillai's Trace test: $F_{7,79}=3.691$, $p=0.002$) , but not in the Nurzec River Valley (Pillai's Trace test: $F_{7,51}=1.925$, $p=0.085$). Between 2010 and 2011, bird calls in the Kampinoski National Park population significantly differed in SYL1 ($F_1=11.573$, $p=001$), SYL 2 ($F_1=18.077$, $p<0.001$), INT2 ($F_1=6.237$, $p<0.014$), NP SYL1 ($F_1=7.985$, $p=0.006$), and NP SYL2 ($F_1=11.932$, $p=0.001$); however, no differences in INT1 ($F_1=0.334$, $p=0.565$) and RHYTHM ($F_1=2.018$, $p=0.159$) were found.

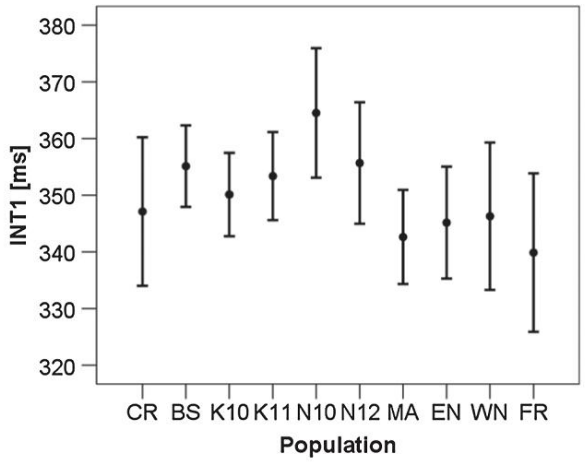
We compared dissimilarity in PPD structure (expressed in squared Euclidean distances) between pairs of males from the same population and pairs of randomly selected males from different populations (6234 pairs of each). We used the first 11 PPDs from the first syllable in our analysis and found no significant difference in PPD structure between pairs of males from the same and different populations (Mann–Whitney U-test; $Z = 1.148$, $p = 0.251$).

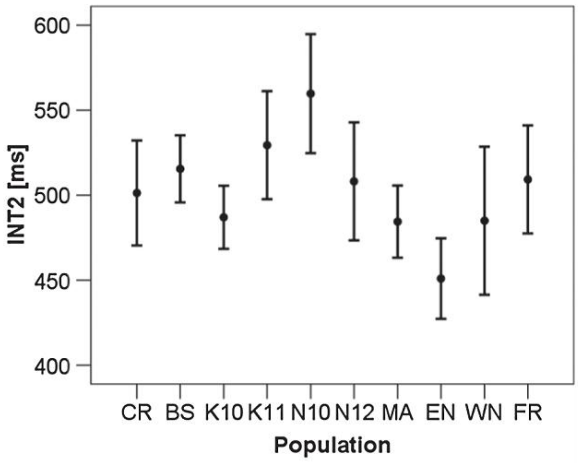

SYL1 [ms]
190
180
170
160
150
CR BS K10 K11 N10 N12 MA EN WN FR
Population


SYL2 [ms]
210
200
190
180
170
CR BS K10 K11 N10 N12 MA EN WN FR
Population


NP SYL1
21.0
20.0
19.0
18.0
17.0
16.0
CR BS K10 K11 N10 N12 MA EN WN FR
Population


NP SYL2
23.0
22.0
21.0
20.0
19.0
18.0
CR BS K10 K11 N10 N12 MA EN WN FR
Population


INT1 [ms]
380
370
360
350
340
330
320
CR BS K10 K11 N10 N12 MA EN WN FR
Population


INT2 [ms]
600
550
500
450
400
CR BS K10 K11 N10 N12 MA EN WN FR
Population

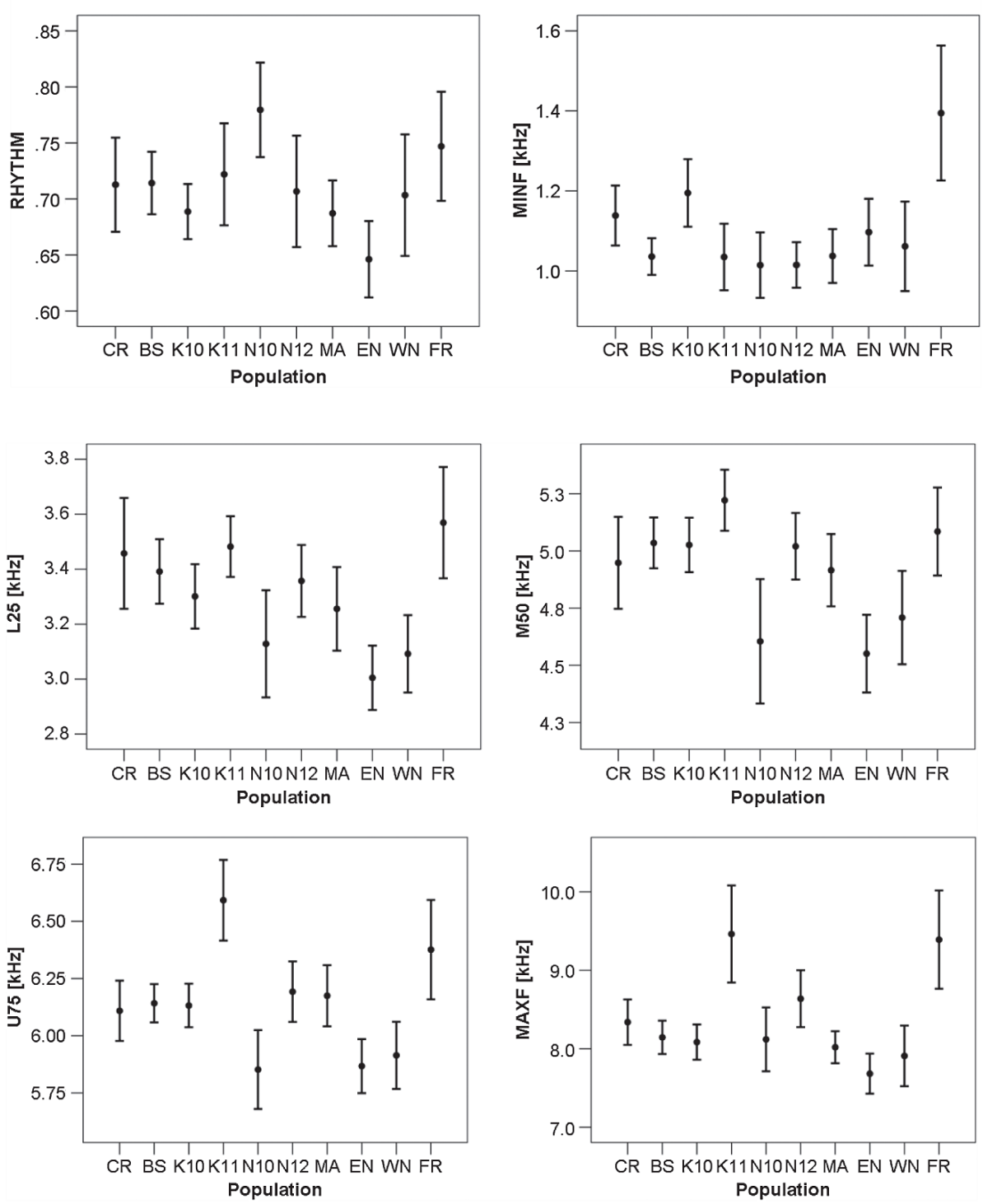


Figure 21: Differences in corncrake calls between- and within-populations. Plots show temporal and frequency call characteristics. Abbreviations of variables are defined in Figure 19. Points are the mean population values, and error bars show the 95% confidence intervals of the mean. The population abbreviations are as follows: CR – Czech Republic; BS – Bieszczady, Poland; K10 – 2010 recordings from Kampinoski National Park, Poland; K11 – 2011 recordings from Kampinoski National Park, Poland; N10 – 2010 recordings from Nurzec River Valley, Poland; N12 – 2012 recordings from Nurzec River Valley, Poland; MA – Mazury, Poland; EN – eastern Norway; WN – western Norway; FR – France.

3.1.4.2. Variation in energy distribution in the frequency domain

We found that energy distribution in frequency domain of corncrake male calls varied among populations. Kruskal–Wallis tests showed significant differences ($p<0.0001$; df=7; significant after Bonferroni correction) for all the examined characteristics: L25 ($\chi^2=32.45$), M50 ($\chi^2=33.34$), U75 ($\chi^2=37.07$), MINF ($\chi^2=34.54$) and MAXF ($\chi^2=43.20$). Th e differences in MINF between populations increased with increasing distance for all possible combinations of recording years. The Mantel tests also showed that L25, M50, U75 and MAXF were positively and significantly distance dependent, but not for all the combinations of years (Table 5). We also found that the spectral characteristics of calls differed significantly between the years in both studied populations. In Kampinoski National Park we found significant differences in MINF, L25, U75 and MAXF among years, whereas in Nurzec River Valley interannual differences were observed only in M50 and U75 (Table 6, Figure 21).

Table 6: Differences in the energy distribution in frequency domain between years in two of the study populations: Kampinoski National Park (2010, n = 55; 2011, n = 32) and Nurzec River Valley (2010, n = 21; 2012, n = 36). Mean values (kHz) and the results of the Mann–Whitney tests are given. ∗ indicates a difference was significant after the Bonferroni correction (Holm's method) was applied. Abbreviations of variables are defined in Figure 19.

	Kampinoski National Park				Nurzec River Valley			
	2010 Mean	2011 Mean	*Z*	p	2010 Mean	2012 Mean	*Z*	p
L25	3301	3482	-2.245	0.025*	3128	3369	-1.916	0.055
M50	5026	5221	-1.778	0.075	4605	5035	-2.660	0.008*
U75	6132	6592	-4.146	0.0001*	5852	6200	-3.325	0.001*
MINF	1195	1035	-2.438	0.015*	1015	1042	-0.657	0.511
MAXF	8086	8593	-4.040	0.0001*	8121	8637	-1.916	0.055

3.1.5. Discussion

Our results clearly reveal the existence of macrogeographical call variation in the corncrake, which is a representative of a poorly studied group of non-learning birds. We found differences in temporal and spectral call characteristics between populations. However, we did not find differences in PPD structure, in spite of the fact that this call characteristic is stable throughout an individual's life (Peake *et al.* 1998) and is probably the result of the anatomical structure of the

vocal apparatus. Thus, we confirm the earlier findings of Peake and McGregor (1999), which showed significant differences in SYL1 and SYL2 between corncrake populations. However, our populations also differed significantly in the number of pulses within both syllables, whereas these call characteristics were not significantly different in the previous study (Peake and McGregor 1999). Furthermore, Peake and McGregor (1999) performed a discriminant function analysis on the first 10 PPDs, SYL1, SYL2, NP SYL1 and NP SYL2; their analysis correctly attributed 54–88% of individuals to their population of origin. In our study, in contrast, INT1 and the first 11 PPDs did not differ significantly between populations. Taken together, these findings suggest that variation in calls between populations may change over time. Moreover, variation between successive years within a population may be similar to that between different populations (Figure 21). Such short-term variation in non-learners may be related to dispersal patterns and the low site fidelity of adults (Mikkelsen *et al.* 2013) and young birds (Green 1999), where different birds return to the same locality in the next year. However, there are no field data available at present to unambiguously test this hypothesis.

Within the breeding season, corncrake males send, receive and respond to acoustic signals. During such social interactions with other males within a population, individuals are able to fine-tune their calls. Especially INT2 and RHYTHM can be changed within a very short time (Rek & Osiejuk 2013). Moreover, males are able to comprehend a new pattern of RHYTHM, if the change of the intruder's RHYTHM is related with the change of his behaviour. After such interaction, males acquire a new signalling strategy and begin signalling aggressive motivation using the new association (Ręk 2013a). Therefore, interannual and between populations variation in INT2 and RHYTHM may be shaped during male-male interactions, and may be stabilized by receiver retaliation rule (Bradbury & Vehrencamp 2011). If so, a geographical pattern in INT2 and RHYTHM that is inconsistent and unstable across successive years may be expected. In this way, geographic and interannual variation in temporal characteristics in a non-learners call might be explained by social interactions.

Temporal and spectral call characteristics have also been found to differ between populations of other non-learning species, such as the blue petrel *Halobaena caerulea* (Bretagnolle & Genevois 1997), the common loon *Gavia immer* (Mager *et al.* 2007), the tawny owl *Strix aluco* (Appleby & Redpath 1997), the barred owl *Strix varia* (Odom & Mennill 2012), and the variable antshrike *Thamnophilus caerulescens* (Isler *et al.* 2005). It remains unknown, however, which factors and mechanisms affect variation in the vocalisation of these non-learner species. The first and the most obvious source of such variation would be genetic differences between different populations. We might expect that, in a variety of genetically transmitted traits, different

populations would exhibit differentiation related to intervening geographic distance and/or evolutionary divergence time (Caizergues *et al.* 2003). On this basis, we would expect to find the same differentiation in corncrake calls: birds from more distant populations or populations that have been isolated for longer time periods should exhibit larger differences in call characteristics than birds within a population or from neighbouring populations (Harrison & Hastings 1996; Oyler-McCance *et al.* 1999; Sosa-López *et al.* 2013). In our study, we found that geographic variation in calls was clearly distance dependent when recordings collected in Kampinoski National Park in 2010 and in Nurzec River Valley in 2012 were used in the analyses. However, when different combinations of recording years were used, this geographical pattern was not so obvious (Table 6). Only MINF increased with increasing distance between populations for all possible combinations of years. However, it is worth noting that, in all the significant results, more distant populations showed larger differences in the temporal and spectral characteristics of their calls. This specific pattern of corncrake call variation, where neighbouring populations are not dramatically different and within-population variation is high, suggests that birds frequently move between populations or that the factors influencing call variation in neighbouring populations are not very different.

Alternatively, individuals occurring in different habitats may differ in their vocalisations because of acoustic adaptation to particular environments (Morton 1975; Hansen 1979; Ippi *et al.* 2011). Therefore, we might expect corncrakes to emit signals with a frequency spectrum that results in less attenuation and degradation during transmission. In fact, many studies have shown that habitat characteristics affect bird vocalisations. Most studies compare structurally different habitats, such as woods vs open areas or natural habitats vs cities with urban noise (Slabbekoorn & Smith 2002; Nicholls & Goldizen 2006; van Dongen & Mulder 2006; Warren *et al.* 2006). However, Cosens and Falls (1984) also showed that differences in song attenuation and degradation exist between marshes and grasslands – both of which are habitats occupied by corncrakes. This hypothesis is theoretically possible in corncrake when differences between populations are considered, but rather unlikely to explanation interannual differences in call within population, because habitats were invariable in successive years in studied by us populations.

The distribution of energy in the frequency domain, and especially at lower signal frequencies, is strongly related to the body size of the sender. It is known that larger animals can produce lower frequency signals (Fletcher 2004; Bradbury & Vehrencamp 2011). Body size co-varies with some anatomical and physiological traits, such as trachea length, syrinx size, and vocal tract resonance, which can directly affect the signal produced (Lambrechts 1996; Fitch 1999).

Thus, geographic patterns in the energy distribution of corncrake calls may be an indirect effect of body size, as in the silvereye *Zosterops lateralis* (Potvin 2013) and common loon (Mager *et al.* 2007), especially since differences in body size between some corncrake populations have been documented (Cramp & Simmons 1980; Schaffer 1999; Keišs *et al.* 2004). Moreover, Keiss et al. (2004) found a significant correlation between wing length and the habitat type occupied by corncrakes, and showed that optimal patches of habitats were inhabited by larger males. If differences in male corncrake size are reflected in the acoustic energy distribution of their signals, then we might expect that larger males in optimal habitats would produce lower frequency signals than smaller males in suboptimal habitats. In that case, habitat differentiation between populations should be reflected in the differentiation of call characteristics, whereas a specific dispersion and a low side fidelity of birds, when various size males return to the population in the next year, may explain interannual variation within population.

The geographical variation seen in corncrake calls seems to be more similar to the clinal variation observed in the common loon, with low differences in call between neighbouring populations and with higher differences between distant populations (Mager *et al.* 2007), than to the haphazard variation observed in the barred owl, where high differences in call occur both between neighbouring and distant populations (Odom & Mennill 2012). We have proposed a few possible hypotheses above that may explain geographical variation in the calls of nonlearning birds. The main difference between learning and non-learning species arises from the factors responsible for song/call variation. In the first group cultural transmission must be always considered. Despite this, the evolutionary mechanisms responsible for differences in call characteristics should be the same in learners and nonlearners. Our study suggests that it is necessary to take into account species ' dispersal capacity, site fidelity, genetic differences, body size, and habitat features to fully understand geographical in the calls of non-learning bird species. In corncrake, a macrogeographic variation occurs in temporal and spectral call characteristics. However, significant annual differences are also observed within population. This means that the pattern of geographic variation in call may be very dynamic, and birds' dispersion and/or social interactions may cause high between-seasonal changes.

3.1.6. Acknowledgements

We thank Paweł Ręk, Sylwester Solaniuk, Krzysztof Jurczak and Thorstein Holtskog for help in the field. The study was partially funded by a grant from the Dean of the Faculty of Biology at Adam Mickiewicz Univ. (GDWB-02/2011).

3.1.7. Conclusion Article 3

Les résultats obtenus montrent qu'il existe des variations géographiques significatives du chant des mâles Râles des genêts à l'échelle européenne. Ces variations sont corrélées avec la distance euclidienne séparant les populations considérées. Toutefois, les variations entre populations ne sont pas présentes pour l'ensemble des paramètres du chant étudiés. De façon intéressante, on observe que les différences inter-annuelles sur un même site peuvent être aussi importantes que les différences entre populations distantes géographiquement.

Les variations macro-géographiques observées peuvent tout d'abord suggérer, si les variations du chant sont déterminées purement génétiquement, à un phénomène d'isolement par la distance où la proximité génétique entre populations est fonction de la distance entre elles. Dans ce cas, les différences de chant entre années consécutives seraient dues à une faible fidélité au site entre saisons de reproductions où à une forte dispersion des jeunes. Toutefois, on sait que le chant des mâles est largement influencé par les interactions compétitrices entre mâles rivaux. Dès lors, la structure du chant pourrait se mettre en place en début de saison, puis se stabiliser au fur et à mesure des interactions. Ceci aboutirait à une structuration spatiale du chant des mâles entre populations, instable entre saisons de reproductions, entièrement déterminée par les interactions sociales de la saison en cours. Alternativement, on peut supposer que les conditions écologiques du site de reproduction peuvent influencer les caractéristiques du chant en modulant sa propagation ou en déterminant la croissance des individus, de laquelle peut dépendre certains paramètres du chant. En conclusion, on constate chez le Râle des genêts une dynamique spatiale et temporelle dans le chant des mâles susceptible d'être influencée, de façon non exclusive, par la dispersion des individus, les interactions entre mâles et les conditions environnementales.

3.2. ARTICLE 4 : STRUCTURE GENETIQUE SPATIALE DES POPULATIONS ET FLUX DE GENE

Identifier des groupes cohérents génétiquement et estimer les relations entre ces groupes (isolement, flux de gènes) permet de répondre à différentes questions pertinentes en biologie de la conservation. Tout d'abord, le détermination d'unités évolutives distinctes (ESU, Ryder 1986) ou d'unités de gestion (MU, Moritz 1994) offre l'opportunité de définir des actions de conservations ciblées sur des entités présentant une histoire évolutive commune, des adaptations locales ou un potentiel évolutif à protéger (Hedrick 2001). L'estimation des flux de gènes informe sur la dispersion des individus et la migration inter-populations. La description des variations spatiales de diversité génétique met en lumière les menaces potentielles posées par la dépression de consanguinité ou la perte de potentiel adaptatif dans les populations à faible effectif ou en déclin (Hedrick & Kalinowski 2000). D'autre part, l'analyse des caractéristiques génétiques des populations peut permettre de reconstituer les dynamiques démographiques historiques ayant affecté ces dernières (Luikart *et al.* 1998).

Les théories biogéographiques fournissent des prédictions quant à la structure génétique et les variations de diversité génétique attendues au sein d'une aire de distribution. Bien que ce modèle ait été largement critiqué (Sagarin & Gaines 2002), on considère généralement que le cœur de la distribution d'une espèce est constitué de populations abondantes et bien connectées tandis que les populations périphériques sont de petite taille et plus isolées (Brussard 1984; Brown 1984). En termes génétiques, on s'attend alors, si le flux de gènes est relativement faible, à avoir des populations périphériques plus structurées et à plus faible diversité génétique que les populations centrales (Hoffmann & Blows 1994). Les populations marginales seraient alors plus sensibles aux changements environnementaux (petite taille, potentiel adaptatif limité, dépression de consanguinité, etc.) et donc plus facilement sujettes à des risques d'extinction (Lesica & Allendorf 1995). Pourtant, si les capacités de dispersion de l'espèce sont suffisantes, le flux de gènes du centre vers la périphérie peut contrecarrer les effets de la dérive génétique dans ces petites populations périphériques et homogénéiser la diversité génétique entre populations centrales et marginales (Kirkpatrick & Barton 1997).

Dans cet article, les patrons de diversité et de structure génétiques ont été analysés dans les populations européennes de Râle des genêts. Cette espèce présente des populations à effectifs

importants en Europe de l'est (Russie, Pologne, Lettonie, Ukraine) tandis que les populations d'Europe occidentale (France, Royaume-Uni) et d'Europe du nord (Suède, Danemark) sont de taille largement plus réduite (Schäffer & Koffijberg 2004). On se retrouve alors, à l'échelle européenne, dans un contexte démographique tel que celui décrit précédemment. Cette distribution des populations est, au moins en partie, causée par l'activité agricole qui a drastiquement réduite les populations d'Europe de l'ouest tandis que l'Europe de l'est était relativement épargnée, tout particulièrement depuis la chute de l'URSS (Green & Rayment 1996; Green *et al.* 1997a; Keišs 2003). L'hypothèse d'une différenciation accrue et d'une perte de diversité génétique dans les populations périphériques a ainsi été testée afin de déterminer (i) les relations entre populations centrales encore en bon état de conservation et les petites populations périphériques menacées d'Europe de l'ouest, (ii) si, en plus des mauvaises tendances démographiques observées, des conséquences génétiques sont à craindre dans ces populations (dépression de consanguinité). Afin de mieux comprendre les dynamiques opérant à l'échelle de la distribution du Râle, on a également analysé la signature génétique des changements démographiques intervenus dans ces populations, que l'on a confrontée aux tendances démographique locales décrites par les suivis de populations.

Pour ce faire, 15 populations de Râles des genêts ont été échantillonnées à travers l'Europe, de la façade atlantique jusqu'à la Russie occidentale, et de la Roumanie à la Suède. Environ 500 individus ont été génotypés à l'aide de 15 marqueurs microsatellites, afin (i) de calculer la diversité génétique présente dans chaque population échantillonnée, (ii) de caractériser la structure génétique des populations de Râles des genêts à l'échelle européenne, en utilisant un large jeu d'analyses (F_{ST}, STRUCTURE, analyse discriminante, test de l'isolement par la distance), (iii) de tester le scénario démographique le plus probable dans chacun des sites (population décroissante, croissante ou constante) par une méthode d'*Approximate Bayesian Computation*. Les résultats attendus doivent permettre de mieux comprendre les dynamiques structurant les variations génétiques d'une espèce à l'échelle continentale dans un contexte de perturbations anthropiques variables au sein de l'aire de distribution, tout en fournissant des données essentielles pour la conservation du Râle des genêts.

Cette partie, menée grâce à une collaboration avec David Richardson de l'Université d'East Anglia (Norwich, GB), a fait l'objet d'un manuscrit rédigé en anglais et soumis pour publication.

Range dynamics at the continental scale: testing the central-marginal hypothesis in the Corncrake, a species with a complex conservation status

Yoan Fourcade[1, 2], David S Richardson[2 *], Oskars Keišs[3], Michał Budka[4], Rhys E Green[5, 6], Sergei Fokin[7], Jean Secondi[1*]

[1] Université d'Angers, GECCO, 49045 Angers, France

[2] Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

[3] Laboratory of Ornithology, Institute of Biology, University of Latvia, LV-2169 Salaspils, Latvia

[4] Department of Behavioural Ecology, Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, 61614 Poznan, Poland

[5] Department of Zoology, Conservation Science Group, Downing Street, Cambridge CB2 3EJ, United Kingdom

[6] The Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, United Kingdom

[7] State informational-analytical centre of game animals and environment ("Centrokhot control"), Moscow, Russia

* Joint senior and corresponding authors

3.2.1. Abstract

Aim: Understanding variation in genetic diversity and gene flow across a species range, especially in relationship to core and marginal populations, has always been a central question in biogeography. Such understanding is critical in conservation biology to determine the genetic status and viability of small peripheral populations. In this study, we investigated pan-European-scale patterns of genetic structure and demography in the Corncrake *Crex crex*, a grassland bird species distributed according to a core-periphery pattern.

Location: Europe

Methods: We genotyped 496 individuals at 15 microsatellites loci sampled from 15 European populations of corncrake. We assessed population structure and variation in genetic diversity. We also estimated the population demographic trends and the direction of gene flow between eastern and western sites by approximate Bayesian computation (ABC).

Results: We found no evidence of a decrease in genetic diversity, or an increase in population differentiation, in peripheral populations as predicted under the classical central-marginal hypothesis. Instead, analyses revealed low genetic structure across the entire range with high levels of gene flow between all sites. However, we did find some evidence that the westernmost populations were, to a limited extent, differentiated from the rest of the European population. Demographic trends inferred by ABC concurred with known local trends, namely that population numbers have decreased in western Europe and remained constant across eastern Europe. Results also suggest unidirectional gene flow from eastern populations to western populations.

Main Conclusions: We conclude that the contrasting demographic regimes between eastern and western populations probably causes asymmetric gene flow that buffers small peripheral populations against diversity loss, but also progressively erases past genetic structure. Furthermore, our study highlights that ABC methods provide a valuable tool to unravel demographic processes that may be undetectable by classical surveys.

3.2.2. Introduction

Characterising geographic patterns of genetic diversity and structure at the range scale is a central issue in biogeography and conservation biology. Understanding how processes such as gene flow and source-sink dynamics determine the distribution of genetic diversity across a range enable us to evaluate the threat posed to specific populations by inbreeding depression and/or the loss of adaptive potential (Hedrick & Kalinowski 2000). It is also a crucial step in identifying Evolutionarily Significant Units (ESU, Ryder 1986) and thus designing informed conservation action plans. However, levels of genetic variation, and the processes that drive them, require molecular analyses as they cannot be assessed using classical species surveys (Noss 1990). When this genetic information is lacking general models of range dynamics may be accepted by default, which may be misleading.

Under the central-marginal model, also referred in terms of demography as the abundant centre model (Hengeveld & Haeck 1982; Brussard 1984; Brown 1984), a given species' abundance is expected to be higher at the range core (*i.e.* the area of ecological optimum), and less abundant and more isolated at the periphery as environmental conditions gradually depart from the ecological optimum. This biogeographical model has implications for genetic variation at the range-scale (Eckert *et al.* 2008) and the evolution of species' range (Hoffmann & Blows 1994; Kirkpatrick & Barton 1997). It implies that populations at the range core have higher effective population size and produce more migrants than the smaller, peripheral populations. Under this model genetic drift, only partially compensated by limited gene flow from the core area, results in lower local genetic diversity in the peripheral populations and increases their differentiation (Hoffmann & Blows 1994; Lesica & Allendorf 1995; Eckert *et al.* 2008). Consequently, marginal populations are expected to be demographically and genetically more sensitive to environmental changes - either stochastic or directional - and more prone to extinction (Lesica & Allendorf 1995; Channell & Lomolino 2000). Even though the central-marginal model is widely accepted, the hypothesis is frequently challenged by empirical and theoretical studies (Sagarin & Gaines 2002; Sagarin *et al.* 2006; Samis & Eckert 2007). For example, Sagarin & Gaines (2002) found that only 39% of empirical studies showed the expected decline of abundance from the range centre to the margins, while a decline in genetic diversity towards the periphery was detected in only 64% of the studies reviewed by Eckert *et al.* (2008).

Alternative models have been proposed which predict different genetic structure across the range scale. First of all, the central-marginal model itself can generate opposite patterns. If core populations are large and peripheral populations are small, the core population will send more migrants to neighbouring populations than the marginal populations, resulting in an asymmetric gene flow from core to periphery (Kirkpatrick & Barton 1997). This asymmetric gene flow is analogous to a source-sink (Pulliam 1988), or island-continent model (Slatkin 1987). Homogenisation of genetic diversity and weak structure at the range scale is expected if the effect of this asymmetric gene flow is greater than the combined effects of drift and selection at the range margins. Another hypothesis is that strong and/or fluctuating selection pressures at the range periphery may drive higher level of genetic diversity in marginal populations than in central populations where environmental conditions are more stable (Fisher 1930). Finally, Kark *et al.* (2008) suggested that a peak of genetic diversity is expected to arise in sub-peripheral populations which experience both a fluctuating environment, but also relatively high levels of gene flow from the range core.

We used the Corncrake (*Crex crex*) as a model species to study gene flow between core and peripheral populations because its distribution and demography is typically consistent with the abundant centre model. Recent land use changes across Europe have strongly affected the population size and distribution of this species in parts of its range (Green *et al.* 1997a). Spatially heterogeneous constraints on populations affect many species at the range scale and may disturb or even interrupt range dynamics. Consequently, current patterns and processes may not be identifiable by classical surveys alone. In the corncrake the extensive population monitoring undertaken in many European countries allows the survey-based demographic trends to be compared against the historical demography inferred using genetic data. The availability of such fine-scale demographic data provides an exciting opportunity to determine if apparent local trends, which usually drive conservation actions, concur with the continental-scale demographic landscape.

Specifically we tested whether the corncrake follows the central-marginal model or whether the demographic imbalance between eastern and western population generates a net gene flow towards the periphery that homogenises populations across the range. So far, apart from an unpublished PhD thesis (Wettstein 2003), no study has investigated the genetic structure of Corncrakes, and other methods (*e.g.* monitoring returning individuals) do not provide adequate amounts of data to determine dispersal patterns, connectivity between sites, and distinct evolutionary significant units (Ryder 1986). We used a suite of microsatellite markers to assess genetic diversity and structure across the European range of the corncrake. Approximate Bayesian computation (ABC) (Beaumont *et al.* 2002) was used to estimate Corncrake historical demography at the population scale in order to assess fine-scale spatial variation in demographic trends across Europe. We then compared this genetically inferred demographic landscape to the local demographic trends estimated from surveys.

3.2.3. Methods

3.2.3.1. Study species and sample collection

The Corncrake is a migratory bird that breeds on grasslands across the Palearctic (Schäffer & Koffijberg 2004). Ecological niche modelling (Fourcade *et al.* 2013) and expert field knowledge (Schäffer & Koffijberg 2004) suggest that the species' range core is located in Russia and eastern Europe, while favourable habitats are scarcer and more fragmented in western Europe. Changes in anthropogenic activities, *e.g.* the intensification of agricultural practices, have contributed to

creating large demographic differences between eastern Europe and central Russia, where the species is abundant, and western Europe, where the species has declined severely. For instance, a massive population decline and range contraction have been reported in France (Deceuninck *et al.* 2011) and Great Britain (Green & Gibbons 2000) during the last decades. The situation in eastern Europe and Asia, which includes 90% of the world's corncrake population (Schäffer & Koffijberg 2004) is fundamentally different. There the impact of agriculture intensification during the 20th century is difficult to assess, but was probably less important than in western Europe. Indeed recent surveys of Russian and eastern European populations highlight the positive effect on corncrake populations of agricultural abandonment after the fall of USSR (Keišs 2005; Mischenko 2008). We focus on the European part of the corncrake range, from Scotland to western Russia. The current distribution across this part of its breeding range includes a core area (eastern Europe) in which corncrakes are relatively abundant and evenly distributed, surrounded by several smaller populations in the North (Sweden), West (Scotland, France) and South (Romania, Italia) of the range. In the studied area, demographic differences between the core and the periphery are, at least partly, driven by human activity.

We collected 496 corncrake samples from 15 locations across Europe (Figure 22). DNA sampling was conducted in 2011 and/or 2012 thanks to the collaboration of local ringers. Samples were collected from May to July, in order to avoid the capture of migrating birds. Individuals were attracted at night using playback of conspecific male calls during the peak of calling activity. Birds were caught using a mistnet or large dipnet and released immediately after sampling. Because of the playback-assisted capture method only males were sampled, which is typical for this species (Green 1999). Depending on the local legislation and experience of the people involved in sampling, different sources of DNA were collected. In France, Germany, Italy, Hungary, Poland (all sites), Czech Republic, Latvia, Belarus and Russia (20 samples out of 32), ca. 50 µl of blood was collected from the brachial vein and stored in absolute ethanol. In Scotland buccal swabs served as a source of DNA, whereas feathers were collected from Romania, Sweden (all sites), and Russia (12 samples out of 32). Population sample sizes ranged from 7 to 66 (Figure 22). In all sites but Hungary, Romania and Russia the mass of the bird was recorded to the nearest 1 g. Although within-species body mass variations can be linked to individual conditions (Schulte-Hostedde *et al.* 2005) or environmental factors (Lima 1986; Ashton 2002), a previous study has shown that the mass of male corncrakes varies consistently across European populations (Keišs *et al.* 2004). We compared genetic structure and variations of body mass to evaluate the congruence between genetic differentiation and morphological variations.

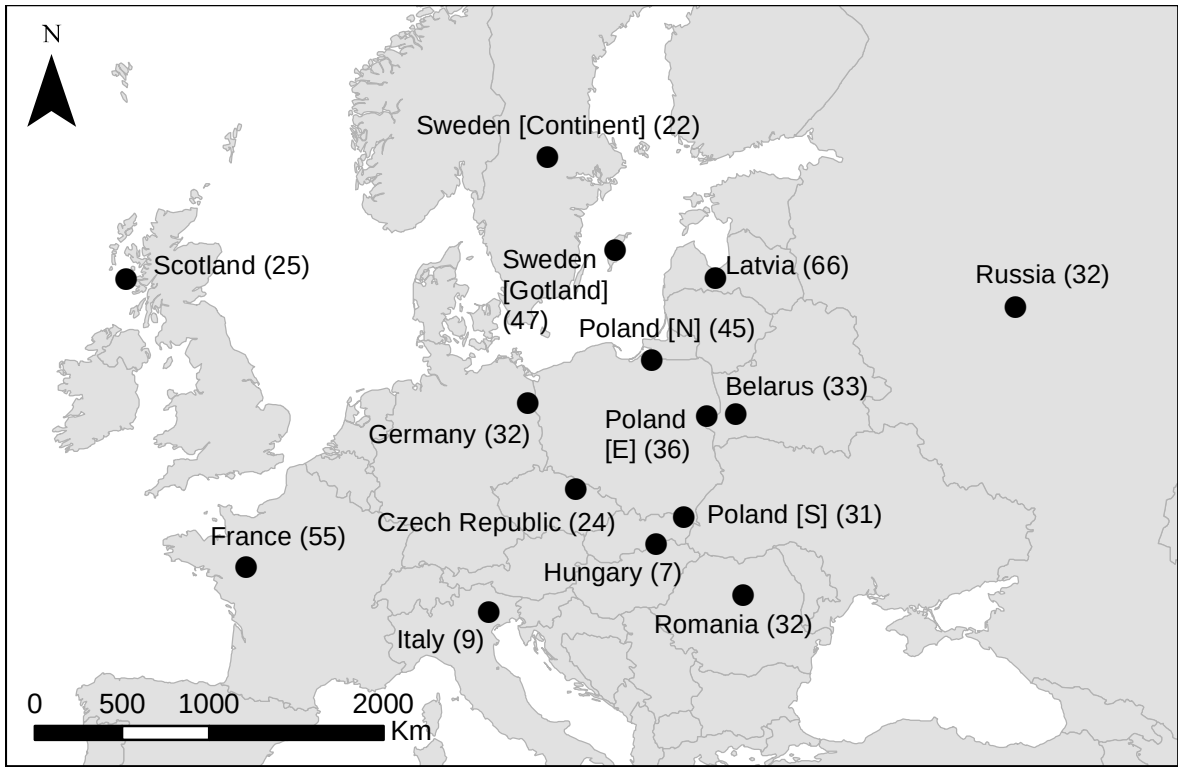


Figure 22: Map of sampling localities of corncrake across Europe. The name of each site is figured along with the number of samples collected in brackets.

3.2.3.2. Microsatellite genotyping

We extracted genomic DNA from each sample following a method of salt extraction (Richardson *et al.* 2001). All individuals were genotyped at 15 microsatellite markers of which eight had been previously designed for the corncrake: *Crex1*, *Crex2*, *Crex6*, *Crex7*, *Crex8*, *Crex9*, *Crex11*, *Crex12* (Gautschi *et al.* 2002). The seven other markers are conserved across a large range of bird families: *CAM18* (Dawson *et al.* 2013), *TG02-120*, *TG04-12*, *TG04-12a*, *TG04-41*, *TG05-30* and *TG012-15* (Dawson *et al.* 2010). We amplified markers by Polymerase Chain Reaction (PCR) in three multiplexes (Supporting information, Table S1), using 1 μL of Qiagen Multiplex MasterMix, 1 μL of DNA (dried in the tube, ca. 15 ng) and 1 μL of 5μM primer mix (Kenta *et al.* 2008). The PCRs were run under the following conditions: an initial step at 95°C for 15 min, followed by 40 cycles of 94°C during 30 s (denaturation), 56.6°C during 90 s (annealing) and 72°C during 60 s (elongation). The final stage consisted of 30 min at 60°C. Amplified fragments were mixed with a solution of formamide and GeneScan 500 ROX Size Standard (Applied Biosystems) and separated by micro-capillary electrophoresis. Alleles were subsequently scored using GeneMapper v3.7 software (Applied Biosystems).

Deviations from Hardy-Weinberg and linkage equilibria were estimated for each pair of populations and loci using the R (R Development Core Team 2012) package "adegenet" (Jombart 2008) and the GENEPOP software (Rousset 2008) respectively. Significance levels were adjusted using Bonferroni correction for multiple comparisons. The proportion of null alleles in the dataset and its influence on G_{ST} estimation was assessed using the FreeNA package (Chapuis & Estoup 2007). To test for the potential effect of any null alleles in the dataset, we ran a STRUCTURE analysis after exclusion of the marker which exhibited the highest rate of null alleles, using the same parameters as for the main analysis.

We computed multilocus standard genetic diversity statistics for each population. Observed (H_o), expected heterozygosity (H_e), and rarefied allelic richness (*rAR*) using the R package "Hierfstat" (Goudet 2005). We assessed the effect of geography on genetic diversity by testing the correlation between the genetic indices and longitude and latitude. Single-locus observed and expected heterozygosity were also computed at each locus for each population (Supporting information, Table S3).

3.2.3.3. Population structure

Population structure was first examined using two measures of pairwise genetic differentiation: Nei's G_{ST} (Nei 1973), the extension of F_{ST} for multi-allelic loci, and Jost's D (Jost 2008), both computed using "DEMEtics" R package (Gerlach *et al.* 2010). Significance was estimated based on 1000 permutations. Differentiation was considered as significant for p-values <0.05 after Bonferroni correction. We also estimated the contribution of within individual and within and among population variance on global genetic variation using an analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) computed using Arlequin 3.5 (Excoffier & Lischer 2010).

Isolation by distance, *i.e.* an increase in pairwise genetic differentiation with geographic distance, was tested using both population and individual level data. Significance was tested using Mantel tests with 5000 permutations comparing pairwise geographic distance and; (i) pairwise population differentiation $G_{ST}/1\text{-}G_{ST}$ and (ii) pairwise individual genetic distance $\hat{a}$ (Rousset 2000) computed using the software SPAGeDi (Hardy & Vekemans 2002).

We also tested for the presence of genetic structure using the software STRUCTURE 2.3.4 (Pritchard *et al.* 2000) which uses a Bayesian approach to assign individuals to genetic clusters based on allele frequencies (full detail in Supporting Information, Appendix S1). We

varied the number of *K* clusters from 1 to 15 (the number of populations sampled). The most probable number of clusters was subsequently determined using both the likelihood of *K* and the change of likelihood between two values of *K* (⊿*K*) following Evanno *et al.* (2005).

We estimated genetic clustering of our samples using the method of Discriminant Analysis of Principal Components (DAPC, Jombart *et al.* 2010) implemented in the "adegenet" R package (Jombart 2008). This approach does not assume any migration model or prior based on sampling location, but aims to identify synthetic variables that distinguish between groups while minimizing within-group variation. We assessed the most likely number of clusters using Bayesian Information Criterion (BIC).

3.2.3.4. Demographic trends and gene flow

We used Approximate Bayesian Computation (ABC) to assess demographic trends in the sampling sites (Beaumont *et al.* 2002). The ABC approach estimates parameters in absence of computable likelihood functions by comparing empirical observations to simulated data. It first generates a large set of simulated data using parameters randomly drawn from prior distributions. Observed and simulated data are then reduced to a set of summary statistics. The posterior probability of models and parameters are estimated using the fraction of simulated models whose summary statistics are closest to those of observed data (Beaumont 2010; Csilléry *et al.* 2010).

We tested whether the changes in census size reported by national surveys (Green *et al.* 1997a; Koffijberg & Schäffer 2006) were reflected in the genetic data. For each population three scenarios of demographic variation over time (constant, decreasing and increasing effective population size) were tested. We used ABC Toolbox (Wegmann *et al.* 2010) to sample parameters in our prior distributions and coalescent simulations were computed using Fastsimcoal (Excoffier & Foll 2011) under the three demographic models. The posterior probabilities of models were then evaluated using the 'abc' R package (Csilléry *et al.* 2012) under the neural network approach, which has been shown to be less sensitive to tolerance rate and correlation between summary statistics than regression-based methods (Blum & François 2010).

The direction of longitudinal gene flow was assessed using another set of simulations in the same ABC framework. Estimates of gene flow may be affected by the fact that genetic structuring across all sampled populations is unclear and that some populations showed evidence of declining size over time (see Results section and Figures 2 and 4). Therefore, in order to

simplify our model, we focused on gene flow between two pairs of distant populations, France–Russia and Scotland–Latvia respectively. We determined the posterior probabilities of three models: a reference model in which the two populations exchange migrants symmetrically and two models with a unidirectional gene flow (from west to east and from east to west respectively). The full details of the ABC methodology are given in Supporting Information, Appendix S1.

3.2.4. Results

3.2.4.1. Genetic diversity

The 15 microsatellites genotyped were all polymorphic and highly variable, with the number of alleles ranging from 9 to 34. The corncrake specific markers were more variable than the cross species utility markers (mean allele number: 26 vs. 12 respectively). Across 225 tests (15 populations*15 loci), 26 showed a significant departure from Hardy-Weinberg equilibrium, but the same loci or populations were not consistently affected (Supporting information, Table S3). Similarly, GENEPOP revealed no significant deviation from linkage disequilibrium after Bonferroni correction. The proportion of null alleles was moderate (mean null allele frequency over loci = 0.039, SD = 0.031), ranging from 0.011 (*Crex12*) to 0.118 (*Crex11*). Moreover, the mean G_{ST} estimated after correction for null alleles (0.009, 95% CI = 0.006-0.013) was similar to the value calculated without taking null alleles into account (0.008, 95% CI = 0.005-0.012). Given this, and since the presence of null alleles would have little impact on Bayesian genetic clustering anyway (Carlsson 2008), we kept all markers in further analyses.

All populations were similar in terms of genetic diversity (Table 7); allelic richness ranged from 3.86–4.68, observed heterozygosity (H_o), from 0.63–0.75, and expected heterozygosity (H_e), from 0.70–0.77. Mean rarefied allelic richness was 4.42 (3.86–4.42). None of these components of genetic diversity exhibited a significant geographic pattern (Spearman correlation tests with longitude and latitude, all *P*-values >0.08) (Supporting information, Table S4).

Table 7: Basic genetic diversity statistics calculated for each sampled location. *n*: number of individuals genotyped, H_o: observed heterozygosity, H_e: expected heterozygosity or gene diversity, *rAR*: rarefied allelic richness.

Population	*n*	H_o	H_e	*rAR*
Scotland	25	0.64	0.7	3.99
France	55	0.66	0.75	4.49
Italy	9	0.64	0.69	3.86
Germany	32	0.63	0.74	4.54
Sweden (continent)	22	0.64	0.72	4.33
Czech Republic	24	0.75	0.76	4.56
Sweden (Gotland)	47	0.65	0.73	4.43
Poland (north)	45	0.68	0.72	4.43
Hungary	7	0.73	0.77	4.68
Poland (south)	31	0.7	0.73	4.48
Poland (east)	36	0.68	0.73	4.47
Latvia	66	0.7	0.77	4.67
Belarus	33	0.63	0.74	4.5
Romania	32	0.71	0.73	4.43
Russia	32	0.71	0.74	4.48

3.2.4.2. Population structure

Despite the considerable power of the highly variable marker set no clear pattern of isolation by distance was detected (Supporting information, Figure S1), either at the population level (mantel test, correlation = 0.300, $P = 0.112$), or at the individual level (mantel test, correlation = 0.039, $P = 0.078$). Pairs of individuals or populations separated by large geographic distances did not exhibit higher genetic differentiation than those located close to each other.

Pairwise G_{ST} and D did not show significant differentiation between populations after Bonferroni correction (*P*-values from 0.103 to 1.00) (Supporting information, Table S5). Both indices exhibited very low values (mean ± SD; $G_{ST} = 0.008 \pm 0.008$; $D = 0.062 \pm 0.060$). However, the highest values constantly involved the same two populations: G_{ST} was higher than 1% in 28 pairwise comparisons, 14 involving Scotland and 14 involving Italy, including the G_{ST} between Scotland and Italy which was the highest in the dataset (0.033). Similarly, the highest values of D always involved Scotland and Italy. The AMOVA analysis revealed that within individual variation represents the vast majority of global genetic variance (93.4%) while variation among population contribute to only 0.44% of total variation (Supporting information, Table S6).

Following the ΔK method (Evanno *et al.* 2005), the Bayesian clustering analysis performed by STRUCTURE retained four significant genetic clusters (Figure 23). However, the

likelihoods from $K = 1$ to $K = 4$ were very close, indicating that the support for $K = 4$ against 1, 2 or 3 was very low. Assuming $K = 4$, individual estimated memberships highlighted reasonably high support for a Scottish cluster, since almost all birds sampled in Scotland had a probability > 0.7 of belonging to the same cluster. French and Italian populations appeared to be grouped in a second cluster, suggesting the possible existence of a southwestern European cluster. However, membership coefficients for this group indicated a probability of belonging to several genetic clusters, thus revealing weak differentiation. All eastern European populations were roughly similar, with individuals mainly assigned to two other clusters. Romania was the only site where more than 10% of individuals were assigned to the Scottish cluster. Details of the mean membership coefficients per sampling population are available in Supporting information, Table S7. The genetic differentiation of French and Scottish birds from most continental populations seems to be in accordance with the morphological data which indicate that these individuals are on average heavier, especially those from Scotland (Figure 24). The difference in morphology of the French/Italian cluster is not as clearly pronounced as the German birds appear to have similar morphological characteristics as the French ones, while Italian corncrakes are more similar to the eastern European ones (Figure 24).

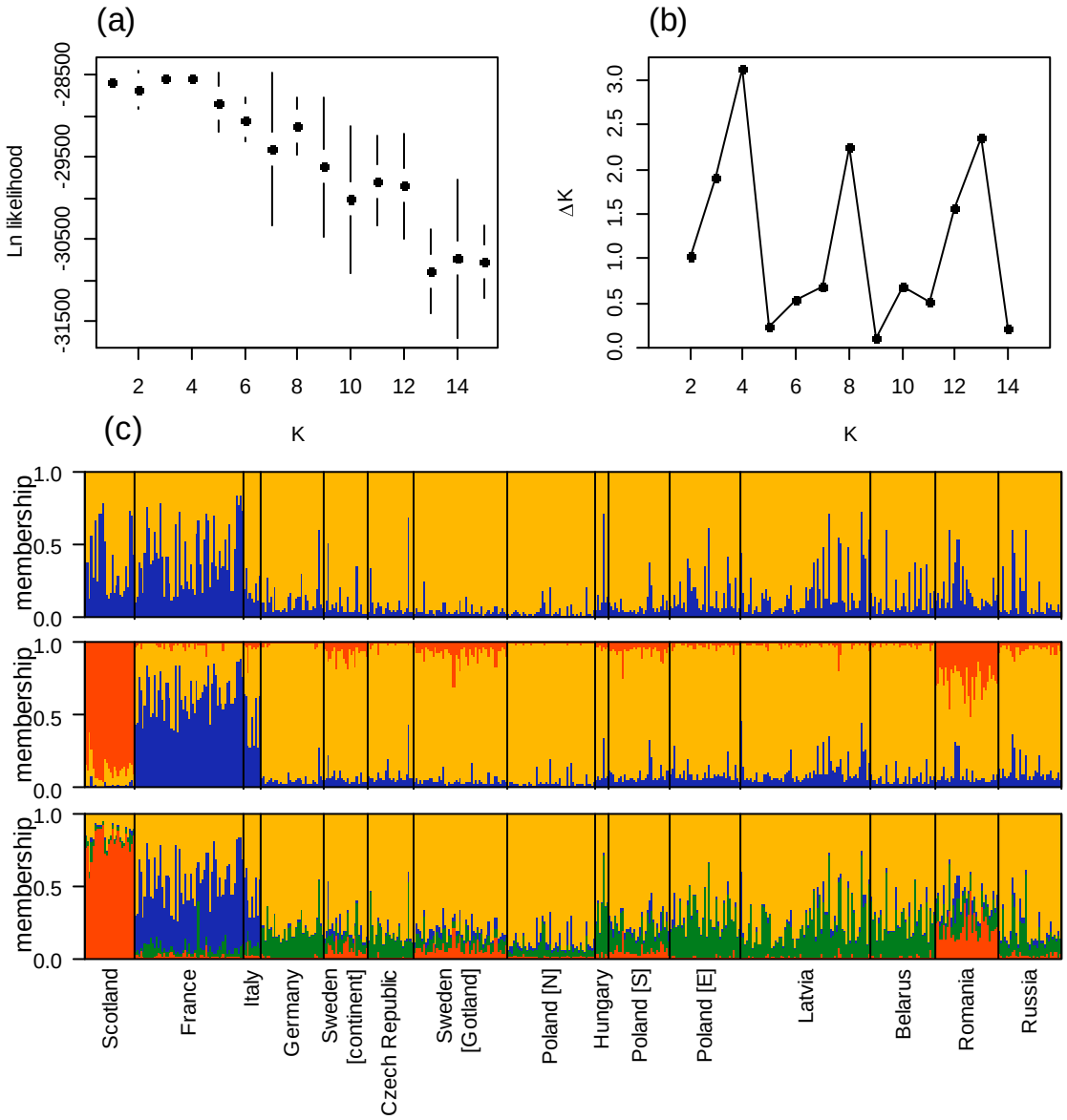


Figure 23: Genetic structure among European corncrake populations based on the Bayesian clustering algorithm STRUCTURE. (a) Ln likelihood with confidence intervals of the ten replicates (b) ΔK for each value of K. The highest peak of ΔK and Ln likelihood at $K = 4$ indicates support for four genetic clusters. (c) Bar plot of individual membership to each for $K = 2$, $K = 3$ and $K = 4$. Sampling sites are separated by vertical bars and plotted according to their longitude.

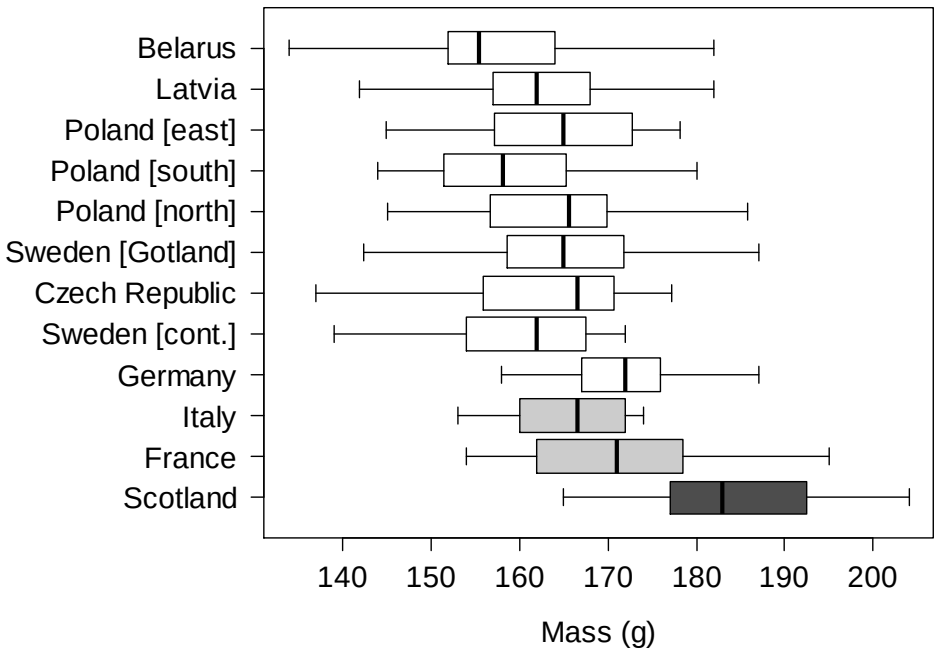


Figure 24: Boxplot of corncrake mass (g) in 12 sampled locations. The central boxes show the median and 25 and 75% quartiles, while the whiskers show the extreme values except those that exceed 1.5 times interquartile. The greyscale colours correspond to the three main groups inferred by STRUCTURE. Statistical differences inferred by Wilcoxon test: Scotland vs. any other group: P always < 0.01; France vs. Poland, Belarus ($P < 0.01$) and Latvia ($P = 0.04$); Germany vs. Belarus ($P < 0.01$); Sweden vs. Belarus ($P < 0.01$).

Eighty principal components (PC) were kept in the discriminant analysis of principal components (DAPC) to retain more than 80% of the total variation, indicating that genetic variation could be hardly partitioned. Following Bayesian information criterion, the optimal number of genetic clusters was five (Supporting information, Figure S2). However, these five clusters did not match the geographic distribution of samples and were mixed between individuals from multiple origins. DAPC was thus unable to identify a reliable population structure in our dataset.

3.2.4.3. Approximate Bayesian computation and estimation of demographic history and gene flow

Using cross-validation, we showed that the three models of demographic trends or gene flow could be discriminated between (Supporting information, Table S8). The whole dataset indicated

a scenario of decreasing effective population size (Table 8) with high confidence (posterior probability of the 'decreasing' model for all data pooled together = 0.98). Populations considered separately gave different results (Figure 25, Table 2). All western European sites (Scotland, France, Italy and Germany) clearly supported a decreasing demographic model (Table 8). In contrast, the southern and easternmost populations (Hungary, Romania, Belarus and Russia) supported a demographic scenario of constant effective population size. Among the other sites, four were assigned to the decreasing model (Sweden (Gotland), Poland (north), Poland (east) and Latvia) and three to the constant model (Czech Republic, Sweden (continent) and Poland (south)) (Table 8). In all analyses, the model of increasing population size always had a null posterior probability, indicating very strong support for the rejection of this demographic scenario in all corncrake populations.

Table 8: Posterior probability of the three demographic models – for each population and for all data pooled together – inferred from the neural network method. In bold is shown the highest posterior probability. The last column shows local demographic trends inferred from population surveys. Data come from Schäffer & Koffijberg (2004) unless stated otherwise.

	Decreasing	Constant	Increasing	Local trend
All data	**0.98**	0.02	0.00	
Scotland	**1.00**	0.00	0.00	Decreasing / increasing[a]
France	**0.85**	0.15	0.00	Decreasing
Italy	**1.00**	0.00	0.00	Decreasing
Germany	**0.98**	0.02	0.00	Increasing ?[b]
Sweden (continent)	0.36	**0.64**	0.00	Increasing[c]
Czech Republic	0.12	**0.88**	0.00	Increasing
Sweden (Gotland)	**0.91**	0.09	0.00	Decreasing[d]
Hungary	0.24	**0.76**	0.00	Fluctuating
Poland (north)	**0.73**	0.27	0.00	
Poland (south)	0.46	**0.54**	0.00	Increasing
Poland (east)	**1.00**	0.00	0.00	
Latvia	**0.91**	0.09	0.00	Increasing
Belarus	0.26	**0.74**	0.00	Constant
Romania	0.11	**0.89**	0.00	Increasing
Russia	0.40	**0.60**	0.00	Fluctuating

[a] Long-term decrease (Green 1995) followed by recent recovery (O'Brien *et al.* 2006)
[b] Schäffer & Koffijberg (2004) indicate an increasing population but Busche (1994) indicates a declining population in Northern Germany
[c] Berg & Gustafson (2007)
[d] Green *et al.* (1997a)

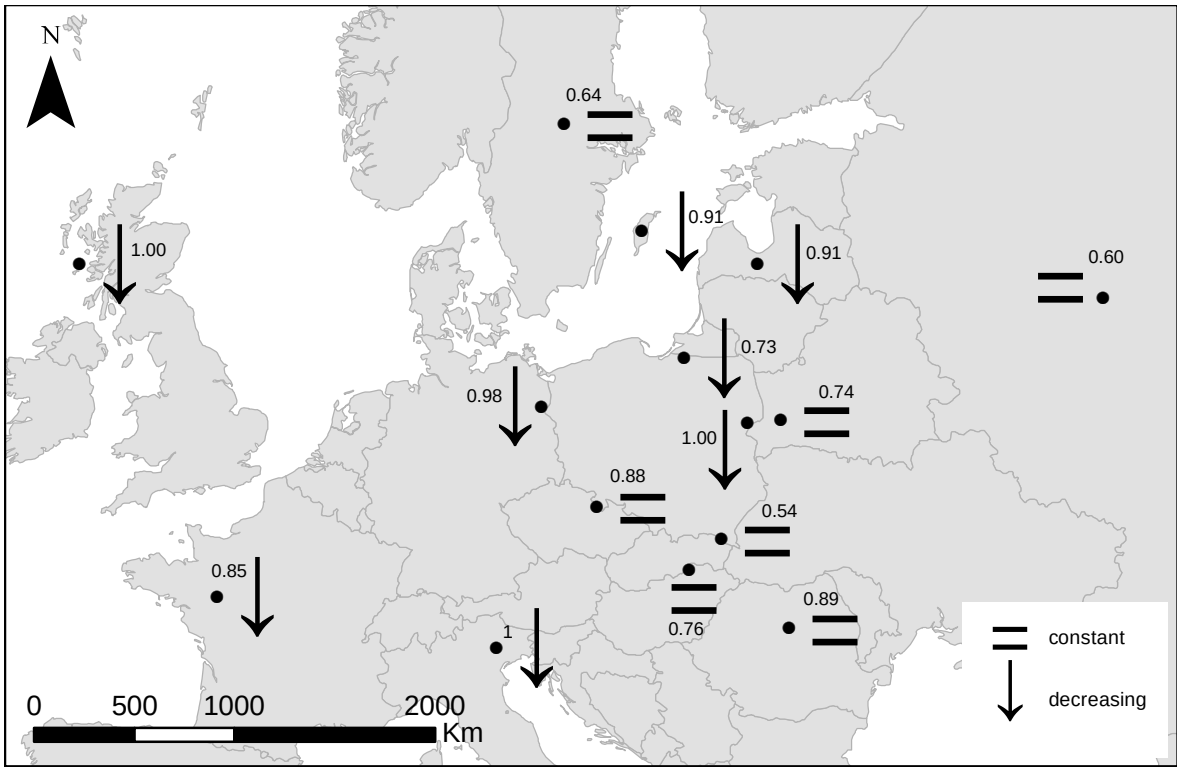


Figure 25: Map of demographic scenarios selected by Approximate Bayesian Computation at each sampling site. The posterior probability of the selected scenario under the neural network approach is given next to the corresponding symbol.

The simulations of gene flow between France and Russia resulted in a strong support for the model of unidirectional gene flow from Russia to France (posterior probability = 0.90, Table 9). The result was more ambiguous for the second pair of western/eastern population (Scotland – Latvia), but seemed to reveal the same pattern of asymmetric gene flow from east to west (Table 9). The support for this model was lower than for France and Russia (posterior probability = 0.53) but was the highest among the three simulated scenarios (west to east: 0.47, symmetric: 0.00).

Table 9: Posterior probability of the three models of gene flow for the two pairs of populations population (France–Russia, Scotland–Latvia), inferred from the neural network method. In bold and underlined is shown the highest posterior probability.

	France - Russia	Scotland - Latvia
West to East	0.06	0.47
Symmetric	0.04	0.00
East to West	**0.90**	**0.53**

3.2.5. Discussion

Our results revealed that there was only very limited genetic structure among European populations of the corncrake, as evidenced by the G_{ST} and D values, as well as the AMOVA and DAPC analyses. However, the STRUCTURE analysis, using the sampling locations as priors for its estimation, provided some evidence that a small number of slightly differentiated separate clusters exist (1-4). A finer analysis of the STRUCTURE output revealed that European corncrake populations may be best depicted by 3 major groups. The first group encompasses all eastern European populations, while two groups, France and Italy on one hand, and Scotland on the other hand, showed evidence of differentiation. Other peripheral populations, *i.e.* the most northern (Sweden) and southern (Romania) sites lacked even these low levels of differentiation from the core. ABC analyses revealed that an asymmetric gene flow from eastern to western populations may have contributed to the observed weak genetic structure and to the maintenance of high genetic diversity in all populations.

At first sight, two contradictory scenarios may explain the apparent genetic structure despite the inferred high levels of gene flow. First, this pattern may result from very recent, or ongoing, isolation of the western populations from an originally panmictic European population. We know from historical data that corncrake was still common and widespread in Europe in the early 20th century (Green *et al.* 1997a) and we can assume that favourable habitats were highly connected. However, habitat fragmentation (Donald *et al.* 2001; Tockner & Stanford 2002) and genetic drift may have since caused the limited population differentiation. The second possibility is that corncrake populations were structured in the past. Assuming a constant migration rate, the differences in population sizes between eastern and western populations would automatically result in a higher number of effective migrants leaving the core populations for the smaller peripheral populations than vice versa. In agreement with this hypothesis, our analyses supported the scenario of an asymmetric gene flow from east to west. Thus, differences in demographic regimes may have gradually erased most, if not all, of any initial difference in allele frequencies between Scotland, France, and eastern Europe, resulting in the observed pattern of high gene flow with low, but still apparent, genetic structure.

The detection of (weakly) genetically differentiated sub-units in western Europe is consistent with other data. For example, previous biometrical analyses found that French and British corncrakes were heavier than eastern European birds (Keišs *et al.* 2004; Schäffer & Koffijberg 2004), a pattern that we also observed (Figure 24). Furthermore, a recent study detected geographic variation in male calls across Europe, but also high inter-annual variations

(Budka *et al.* 2014). These patterns could plausibly be the result of a scenario where there is limited genetic structuring but with some interpopulation migration. Finally, an ongoing study of corncrake migration (Green 2013) has revealed that Scottish birds winter in western Africa, whereas previously all corncrakes were believed to winter in eastern and southern Africa (Schäffer & Koffijberg 2004). This finding may indicate that western and eastern European populations have distinct wintering areas. Taken altogether this evidence, even discounting the potential environmentally determined mass data, supports the genetic evidence and suggests that there may be some limited differentiation between the western populations (France and, especially, Scotland) and the more continental populations.

Although patterns of genetic variation at the range scale, and related hypothesis *e.g.* the central-marginal hypothesis, have been studied for many years, there are still various unresolved issues and uncertainties regarding theoretical expectations (Sagarin *et al.* 2006; Eckert *et al.* 2008). Under the central-marginal hypothesis, differentiation of peripheral populations is expected as a result of higher dispersion rate and drift in the smaller marginal populations (Sagarin & Gaines 2002; Eckert *et al.* 2008). Instead, in our study on the corncrake the genetic and morphological data suggest a longitudinal differentiation between western and eastern populations rather than a pure core-periphery system. However, our analyses show that considerable gene flow now occurs between corncrake populations. Given this, it is possible that original pattern of genetic structure in the corncrake may have been erased as a result of changes caused by anthropogenic factors over the last decades. We thus cannot exclude the hypothesis that original genetic structure better reflected a typical core-periphery system.

In our study, we did not detect a reduction in genetic diversity in peripheral populations as expected under the central-marginal hypothesis (Eckert *et al.* 2008). Indeed, all measures of genetic diversity remained notably high across the entire European range (Table 1). The high level of gene flow inferred from this study suggests that all European corncrake populations are interconnected. It also suggests that the threatened populations of western Europe are sustained by birds from core eastern populations – providing suitable habitat is restored. As a consequence, this dispersal of individuals to the peripheral populations should allow the maintenance of genetic diversity within and among populations across the corncrake's European range. Since low genetic diversity is generally associated with a reduction of fitness (Beardmore 1983; Charlesworth & Charlesworth 1987), this effect should be positive in terms of avoiding the risks usually linked to inbreeding depression and loss of adaptive potential (Frankham 2005) in the most reduced populations. These dynamics should enable corncrake populations to persist in western Europe

sites with declining populations where their fate has appeared uncertain. However, the gradual replacement of western European birds through the source-sink dynamics may, potentially, lead to a loss of local adaptation (Kawecki & Ebert 2004) and therefore increase long-term extinction risk.

Our Approximate Bayesian Computation analyses of demography supported a model of decreasing population size when all populations were included as one (the global analysis), but revealed a more complex pattern at a smaller scale. A model of demographic decline was supported for all western European populations (Scotland, France, Germany and Italy) which corroborates the trends identified from the national surveys which tend to indicate a decrease since the late 19th century (Green & Gibbons 2000). A scenario of constant population size was selected for for some of the most southern or eastern sampling sites. (Czech Republic, Hungary, Romania, South Poland, Belarus and Russia) where recent population surveys also suggest that corncrake populations have remained roughly stable, or even increased (Bürger *et al.* 1998; Keišs 2003; Sukhanova & Mischenko 2003; Schäffer & Koffijberg 2004). More surprising is that a scenario of decreasing population size was identified in Latvia and two Polish populations, despite no survey-based evidence of declining corncrake numbers in these populations. On the contrary, recent agricultural decline in former communist countries appears to have favoured population expansion (Keišs 2003, 2005). However, human activities may have negatively affected these populations during the Soviet period (Tucker *et al.* 1994), leaving a genetic signature that is still detectable in the current corncrake populations. Importantly, these results indicate that the ABC method is able to identify population trends that are not detected by classical surveys, such as historical declines or trends masked by fluctuations in local population numbers. Although the time period reflected by ABC analyses remains uncertain, such methods may be particularly useful for species whose behaviour makes the accurate detection of population trends difficult. For example, species with high dispersal ability such as the corncrake (Mikkelsen *et al.* 2013) may undertake long-distance movements during the breeding season to avoid unsuitable conditions, therefore causing large annual fluctuations in population sizes recorded at specific sites..

Approximate Bayesian computation (ABC) (Beaumont *et al.* 2002) provides an innovative simulation-based tool which is now widely used to distinguish between demographic scenarios (Bertorelle *et al.* 2010). However, this method has mostly been used to infer broad-scale demographic history such as colonisation patterns (Boubou *et al.* 2012; Duchen *et al.* 2013; Fontaine *et al.* 2013), or variations in effective population size at the metapopulation level

(Robinson *et al.* 2012). Here we show that ABC methods can be used to identify discrepancies between population trends observed in surveys and variation in effective population size. By doing so, we may gain insight about the actual connectivity between sampling locations and thus improve the design of future conservation actions. Developing similar approaches in other species would certainly provide insights into range dynamics in large continental landmasses like the Palearctic. The Palearctic region has been recently subject to large scale changes that differ in nature and intensity in different regions. The ABC approach would provide opportunities to better understand the dynamics of ranges and refine predictions about future distributions at the continental scale. In this regard, ABC methods could be a valuable tool to assist in national or international conservation action plans (Lopes & Boessenkool 2009).

3.2.6. Acknowledgments

The authors wish to thank all people who helped in corncrake DNA sampling: Attila Sandor, Sandor Boldogh, Peter Olsson, Peter Sverev, Paolo Vasca, Patrik Rhönnstad, Susanne Arbeiter, Jochen Bellebaum, Joachim Sadlik and especially people from LPO Anjou: Gilles Mourgaud, Emmanual Séchet and Edouard Beslot. We also thank Christophe Lemaire for his useful suggestions on ABC analyses and Karl Phillips for his help in genotyping. This study was funded by a grant from Plan Loire Grandeur Nature and European Regional Development Fund.

3.2.7. Supporting information

Appendix S1: Detailed methods of genetic analyses

Population structure

STRUCTURE analyses were run assuming an admixture model with correlated allele frequencies (Falush *et al.* 2003) and sample location as a prior (Hubisz *et al.* 2009). Runs were conducted with 100 000 burn-in steps following by 500 000 iterations. For each value of K (1–15), 10 replicates were computed. The likelihood of K and the ΔK parameter (Evanno *et al.* 2005) were investigated using STRUCTURE HARVESTER (Earl & vonHoldt 2011) and "CorrSieve" R package (Campana *et al.* 2011). STRUCTURE results were averaged across the 10 replicates with CLUMPP 1.1.2 (Jakobsson & Rosenberg 2007).

Approximate Bayesian computation

Three demographic models (constant, decreasing and increasing effective population size over time) were defined to test for the demographic trend in each sampled population. The three scenarios differed only by their growth rate: no growth rate was defined for the "constant" scenario whereas the "decreasing" and "increasing" scenarios included a positive and a negative growth rate respectively (see Supporting information, Table S2 for prior distributions). The demographic history inferred by ABC was then compared to local trends known from population surveys.

We estimated the direction of gene flow between the populations of corncrake in France and Russia, and Scotland and Latvia using three models of gene flow: symmetric, west to east and east to west. These scenarios simulated the genetic characteristics of the two populations of interest and the migration between them. The two populations were simulated using the parameters inferred by the previous analysis of demographic trends. For the two models of unidirectional migration, migration rate was set to 0 in one direction. For the symmetric model, migration rate was equal in both directions. The full detail of priors and summary statistics is given in Table S2.

We used ABC Toolbox (Wegmann *et al.* 2010) to sample parameters in our prior distributions and coalescent simulations were computed using Fastsimcoal (Excoffier & Foll 2011). One million (demographic trend analysis) or 500 000 (gene flow analysis) simulations of each scenario were computed for each population or pair of populations. We simulated 15 microsatellite markers and the number of individuals corresponding to the sample size of each location. After each iteration, summary statistics (Supporting information, Table S2) were calculated by "arlsumstat", a command-line tool in the software Arlequin (Excoffier & Lischer 2010). The ability to discriminate between the three scenarios was assessed using a cross-validation approach implemented in the "abc" R package (Csilléry *et al.* 2012). We ran 100 cross-validations for each model using a multinomial logistic regression method with 10% tolerance rate. At each iteration one simulation was used as a pseudo-observation and classified into one of the three models. We reported the percentage of correctly and incorrectly assigned pseudo-observed data. The posterior probabilities of models were then evaluated to determine which scenario best matched our data. We used the neural network method with a 25% tolerance rate, *i.e.* the 250 000 simulations that are the closest to observed data were used to estimate posterior probabilities. In all analyses the neural network algorithm used 10 networks and 5 hidden layers and parameters were logit transformed to prevent extrapolation outside prior distributions.

Table S1: Genetic diversity statistics calculated for each microsatellite locus. N_a: number of alleles per locus, H_o: observed heterozygosity, H_e: expected heterozygosity or gene diversity, F_{IS}: Wright's inbreeding coefficient, G_{ST} and D: two estimators of population differentiation. The grouping of loci into multiplexes for genotyping and the fluorescent labelling used (dye) are indicated in 2nd and 3rd columns.

Locus	Multiplex	N_A	H_o	H_e	F_{IS}	F_{ST}	*Jost's D*
Crex1	3	24	0.69	0.86	0.191	0.007	0.069
Crex2	3	24	0.82	0.90	0.083	0.009	0.097
Crex6	1	34	0.87	0.91	0.039	0.022	0.240
Crex7	1	21	0.83	0.87	0.046	0.024	0.184
Crex8	1	24	0.83	0.90	0.078	0.003	0.041
Crex9	1	21	0.87	0.91	0.038	0.006	0.070
Crex11	2	33	0.89	0.92	0.031	0.019	0.236
Crex12	2	24	0.60	0.85	0.292	0.006	0.072
CAM18	2	16	0.62	0.60	-0.025	0.013	0.021
TG02-120	3	11	0.32	0.43	0.249	0.040	0.038
TG04-012	3	9	0.65	0.66	0.008	0.008	0.017
TG04-012a	2	11	0.24	0.26	0.071	0.035	0.014
TG04-041	1	12	0.72	0.75	0.031	0.002	0.009
TG05-30	3	9	0.58	0.54	-0.072	0.009	0.009
TG012-015	1	13	0.58	0.67	0.132	0.003	0.013

Table S2: The parameters used in the ABC simulations of, A) demographic trends and, B) gene flow between France and Russia. (a) Prior distributions and (b) summary statistics computed at each iteration and used to compare empirical and simulated datatsets.

A) Demographic trends

(a) Simulation parameters	Shape of prior distribution	Lower limit	Upper limit
Effective population size (log transformed)	Uniform	1	6
Growth rate			
"Decreasing" model	Uniform	0	0.05
"Increasing" model	Uniform	0	-0.05
Mutation rate	Uniform	0.00001	0.0001
Gamma parameter of microsatellite mutation	Uniform	8	15

(b) Summary statistics
Mean number of alleles over loci
Standard-deviation of number of alleles over loci
Mean heterozygosity over loci
Standard-deviation of heterozygosity over loci
Mean Garza-Williamson statistic over loci
Standard-deviation of Garza-Williamson statistic over loci
Mean allelic range over loci
Standard-deviation of allelic range over loci

Table S2: (B)

B) Gene flow

(a) Simulation parameters	Shape of prior distribution	Lower limit	Upper limit
Migration rate*	Uniform	0	0.1
Mutation rate	Uniform	0.00001	0.0001
Gamma parameter of microsatellite mutation	Uniform	8	15
(b) Summary statistics			
Mean number of alleles over loci output for each population			
Standard-deviation over loci of the number of allele for each population			
Mean number of alleles over loci and population			
Standard-deviation over populations of the mean number of alleles			
Mean total number of alleles over loci:			
Mean heterozygosity over loci output for each population			
Standard-deviation over loci of the heterozygosity for each population			
Mean heterozygosity over loci and population			
Standard-deviation of the mean heterozygosity over populations			
Mean total heterozygosity			
Mean Garza-Williamson statistic over loci for each pop			
Standard-deviation over loci of the Garza-Williamson index for each population			
Mean Garza-Williamson statistic over loci and pops			
Standard-deviation over pops of the mean Garza-Williamson statistic			
Mean Garza-Williamson index computed over a poll of all populations			
Mean over loci of the modified Garza-Williamson index for each population			
Standard-deviation over loci of the modified Garza-Williamson index for each population			
Mean over loci and populations of the modified Garza-Williamson index			
Standard-deviation over populations of the mean modified Garza-Williamson index			
Mean allelic range over loci for each pop			
Standard-deviation over loci of allelic range for each population			
Mean allelic range over loci and pops			
Standard-deviation over pops of the mean allelic range			
Mean total allelic range over loci and pops			
Pairwise F_{ST}			
Mean number of differences between pairs of populations			
Mean delta mu-square (square difference in mean allele length between pairs of populations) over loci			

*Symmetric migration for the "symmetric" model, unidirectional migration for the "west to east" and "east to west" models

Table S3: Genetic diversity for each of the 15 loci and 15 populations. The observed heterozygosity (H_o), expected heterozygosity (H_e) and number of alleles (N_A) are indicated. Statistically significant departures from Hardy-Weinberg equilibrium based on 1000 permutations are indicated in bold and underlined.

Pop	n		CAM18	Crex1	Crex11	Crex12	Crex2	Crex6	Crex7	Crex8	Crex9	TG012-015	TG02-120	TG04-012	TG04-012a	TG04-041	TG05-30
Germany	32	H_o	0.62	0.56	**0.80**	**0.48**	0.81	0.87	0.78	0.72	0.77	**0.50**	0.33	0.56	0.09	0.81	0.67
		H_e	0.70	0.88	0.93	0.87	0.90	0.90	0.89	0.88	0.92	0.63	0.49	0.64	0.18	0.79	0.58
		N_A	9	15	19	13	15	21	13	14	16	6	6	7	3	8	3
Belarus	33	H_o	0.63	0.64	0.91	**0.50**	0.91	0.75	0.69	**0.67**	0.79	**0.19**	0.66	0.79	0.16	0.56	0.55
		H_e	0.57	0.87	0.96	0.90	0.91	0.90	0.88	0.92	0.92	0.56	0.53	0.72	0.18	0.75	0.50
		N_A	6	14	23	14	14	19	12	16	17	5	6	5	3	7	3
France	55	H_o	**0.69**	**0.63**	0.93	**0.58**	0.91	0.79	0.78	0.89	0.84	**0.34**	0.42	**0.44**	0.40	0.74	0.47
		H_e	0.73	0.84	0.91	0.88	0.91	0.89	0.88	0.89	0.90	0.63	0.49	0.58	0.43	0.73	0.60
		N_A	11	20	19	14	18	16	14	16	12	6	8	5	6	6	8
Hungary	7	H_o	0.75	0.86	1.00	0.67	0.86	0.86	1.00	1.00	1.00	0.50	0.20	0.86	0.17	0.86	0.43
		H_e	0.58	0.92	0.92	0.90	0.96	0.93	0.93	0.92	0.90	0.71	0.80	0.65	0.17	0.77	0.48
		N_A	3	9	8	6	11	9	10	9	8	3	4	4	2	5	3
Italy	9	H_o	0.44	0.89	0.78	0.50	0.78	0.89	0.78	0.67	0.89	0.63	0.22	0.78	0.11	0.44	0.78
		H_e	0.64	0.81	0.92	0.70	0.81	0.87	0.71	0.96	0.85	0.62	0.63	0.64	0.11	0.68	0.49
		N_A	4	6	11	6	7	8	5	11	8	4	3	3	2	5	2
Latvia	66	H_o	0.68	**0.62**	0.92	0.68	0.71	0.86	0.91	0.86	0.86	0.74	0.47	**0.64**	0.24	0.76	**0.47**
		H_e	0.66	0.90	0.94	0.85	0.91	0.91	0.89	0.90	0.93	0.75	0.58	0.67	0.30	0.77	0.57
		N_A	10	17	25	16	19	26	15	18	18	6	8	7	8	9	6
Poland (east)	36	H_o	**0.58**	**0.64**	0.97	0.63	0.86	0.89	0.86	0.83	0.92	0.55	0.08	0.72	0.50	0.66	0.56
		H_e	0.65	0.85	0.94	0.83	0.91	0.94	0.92	0.90	0.89	0.69	0.13	0.65	0.40	0.71	0.52
		N_A	10	18	23	14	18	24	17	15	16	6	4	5	3	7	3
Czech Republic	24	H_o	0.71	0.71	0.88	0.68	0.83	1.00	0.92	0.88	0.88	0.83	0.29	0.71	0.63	0.75	0.50
		H_e	0.68	0.85	0.95	0.89	0.90	0.92	0.90	0.91	0.92	0.72	0.26	0.71	0.51	0.73	0.54
		N_A	7	9	22	10	14	15	12	15	13	8	3	6	3	6	3
Poland (north)	45	H_o	0.42	0.67	0.89	0.60	0.87	0.89	0.89	0.96	0.89	0.62	0.33	0.69	0.11	0.69	0.62
		H_e	0.45	0.85	0.94	0.86	0.89	0.92	0.87	0.92	0.92	0.67	0.39	0.67	0.11	0.77	0.58
		N_A	7	14	24	16	18	20	12	16	17	6	7	5	4	8	4
Poland (south)	31	H_o	0.45	0.77	0.97	0.61	0.84	0.90	0.87	0.90	0.90	0.74	0.45	0.52	0.23	0.81	0.61
		H_e	0.39	0.91	0.91	0.87	0.88	0.92	0.91	0.88	0.91	0.62	0.50	0.71	0.21	0.78	0.56
		N_A	5	15	19	12	15	18	14	14	16	6	7	6	4	8	5
Russia	32	H_o	**0.56**	0.81	0.94	**0.50**	0.94	0.94	0.88	0.84	0.97	0.68	**0.35**	**0.81**	0.19	0.66	0.57
		H_e	0.56	0.85	0.94	0.82	0.91	0.92	0.89	0.88	0.93	0.64	0.49	0.71	0.23	0.75	0.57
		N_A	8	11	23	10	16	22	14	12	15	7	6	6	4	7	3
Romania	32	H_o	0.74	0.72	0.94	0.69	0.93	0.95	0.84	0.85	0.81	0.76	0.27	0.48	0.25	0.72	0.69
		H_e	0.70	0.87	0.95	0.85	0.92	0.89	0.88	0.90	0.88	0.71	0.24	0.62	0.29	0.72	0.58
		N_A	7	16	20	13	17	14	13	14	13	7	4	5	7	7	3

Sweden (cont.)	22	H_o	0.70	0.68	0.91	0.55	0.73	0.86	0.82	0.76	1.00	0.50	0.18	0.55	0.05	0.81	**0.57**
		H_e	0.58	0.82	0.95	0.83	0.89	0.93	0.81	0.88	0.94	0.71	0.25	0.64	0.22	0.75	0.54
		N_A	4	12	20	8	14	21	8	12	16	7	4	5	4	6	3
Sweden (Gotland)	47	H_o	0.64	0.57	0.83	**0.48**	0.80	0.91	0.81	0.85	0.83	0.60	0.23	**0.64**	0.21	0.85	0.55
		H_e	0.56	0.82	0.95	0.87	0.90	0.92	0.89	0.90	0.90	0.66	0.42	0.65	0.21	0.77	0.54
		N_A	8	12	21	14	17	25	15	17	15	7	6	6	4	11	3
United-Kingdom	25	H_o	0.63	0.64	0.70	0.89	0.57	0.73	0.68	0.80	0.71	**0.52**	**0.38**	**0.63**	0.29	0.76	0.72
		H_e	0.57	0.79	0.68	0.85	0.88	0.87	0.83	0.90	0.85	0.70	0.36	0.59	0.34	0.76	0.51
		N_A	7	11	8	9	11	12	10	11	10	6	5	5	6	7	2

Table S4: Correlation between genetic diversity (observed heterozygosity H_o, gene diversity H_e, rarefied allelic richness *rAR* and F_{IS}) per population and geography (longitude and latitude). Values shown are the Spearman correlation coefficient ϱ. None of the values are significant.

	longitude	latitude
H_o	0.47	-0.26
H_e	0.31	-0.15
rAR	0.32	-0.11
F_{IS}	-0.29	0.39

Table S5: Pairwise differentiation between populations. No value significantly differs from 0 (*P* always > 0.05 after Bonferroni correction, based on 1000 permutations). Above diagonal: G_{ST}, below diagonal: Jost's *D*.

	Germany	Belarus	France	Hungary	Italy	Latvia	Poland (east)	Czech Republic	Poland (north)	Poland (south)	Russia	Romania	Sweden (cont.)	Sweden (Gotland)	Scotland
Germany	---	0.001	0.005	0.003	0.027	0.001	0.008	0.007	0.002	0.002	0.001	0.005	0.001	-0.001	0.015
Belarus	0.017	---	0.007	-0.001	0.022	0.002	0.008	0.007	0.001	0.001	0.000	0.007	0.003	0.000	0.016
France	0.044	0.069	---	0.009	0.024	0.005	0.009	0.006	0.009	0.008	0.005	0.007	0.006	0.005	0.015
Hungary	0.034	0.027	0.062	---	0.019	0.002	0.007	0.005	0.002	0.003	0.001	0.006	0.004	0.001	0.014
Italy	0.179	0.165	0.159	0.184	---	0.019	0.030	0.028	0.022	0.019	0.020	0.023	0.028	0.024	0.033
Latvia	0.004	0.023	0.048	0.013	0.141	---	0.008	0.005	0.003	0.003	0.001	0.006	0.003	0.001	0.015
Poland (east)	0.026	0.034	0.063	0.004	0.142	0.021	---	0.001	0.008	0.010	0.008	0.006	0.005	0.006	0.014
Czech Republic	0.014	0.016	0.040	-0.005	0.170	0.011	0.001	---	0.008	0.008	0.006	0.006	0.005	0.004	0.017
Poland (north)	0.006	0.021	0.060	0.041	0.133	0.014	0.018	0.021	---	0.001	0.001	0.007	0.002	0.000	0.015
Poland (south)	0.001	0.027	0.064	0.033	0.130	0.010	0.040	0.021	0.021	---	0.001	0.008	0.004	0.001	0.012
Russia	0.016	0.027	0.050	0.046	0.144	0.018	0.044	0.036	0.020	0.021	---	0.005	0.002	0.000	0.013
Romania	0.079	0.065	0.093	0.021	0.182	0.073	0.080	0.080	0.076	0.076	0.070	---	0.003	0.002	0.010
Sweden (cont.)	-0.004	0.020	0.039	0.058	0.155	0.009	0.030	0.010	0.009	0.008	0.015	0.052	---	0.002	0.012
Sweden (Gotland)	0.001	0.009	0.039	0.017	0.160	0.015	0.014	0.008	0.009	0.009	0.009	0.042	-0.002	---	0.011
Scotland	0.159	0.177	0.153	0.126	0.236	0.165	0.149	0.182	0.169	0.116	0.137	0.090	0.120	0.123	---

Table S6: Analysis of molecular variance (AMOVA) partitioning molecular variation between within individual, within populations and among populations variation.

Source of variation	d.f.	Sum of squares	Variance components	Percent variation
Among populations	14	99.48	0.02	0.44
Within populations	481	2684.96	0.33	6.22
Within individuals	496	2443.00	4.93	93.34
Total	991	5227.44	5.28	

Table S7: Individual membership coefficients averaged by sampling population, *i.e.* mean probability of belonging to one of the four clusters inferred by STRUCTURE. For each population the first cluster, in terms of mean membership, is in bold and the second one is underlined.

Sampling population	Mean estimated membership			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
UK	**0.794**	0.012	0.045	<u>0.150</u>
France	0.017	**0.469**	0.062	<u>0.452</u>
Italy	0.033	<u>0.341</u>	0.070	**0.556**
Germany	0.010	0.015	<u>0.190</u>	**0.785**
Sweden [Continent]	0.071	0.056	<u>0.098</u>	**0.776**
Czech Republic	0.009	0.023	<u>0.166</u>	**0.803**
Sweden [Gotland]	0.072	0.049	<u>0.074</u>	**0.805**
Poland [north]	0.009	0.049	<u>0.067</u>	**0.875**
Hungary	0.028	0.018	<u>0.342</u>	**0.611**
Poland [south]	0.056	0.049	<u>0.147</u>	**0.749**
Poland [east]	0.013	0.029	<u>0.266</u>	**0.693**
Latvia	0.011	0.036	<u>0.229</u>	**0.724**
Belarus	0.015	0.031	<u>0.241</u>	**0.713**
Romania	<u>0.215</u>	0.059	0.134	**0.592**
Russia	0.024	0.046	<u>0.140</u>	**0.790**

Table S8: Confusion matrix of models in Approximate Bayesian Computation, inferred by 100 cross-validations. (A) Demographic trends (B) Gene flow. The table shows the average across the 15 models populations (A) or between the two pairs of populations (B).

A

	Constant	Decreasing	Increasing
Constant	**0.73**	0.06	0.20
Decreasing	0.06	**0.86**	0.07
Increasing	0.20	0.07	**0.71**

B

	East to West	Symmetric	West to East
East to West	**0.92**	0.08	0.00
Symmetric	0.01	**0.88**	0.12
West to East	0.00	0.29	**0.71**

Figure S1: Isolation by distance (IBD) at, A) the individual level, B) the population level. Pairwise genetic differentiation A) $G_{ST}/1\text{-}G_{ST}$, B) Rousset's $\hat{a}$, is plotted against geographic distance (in km).

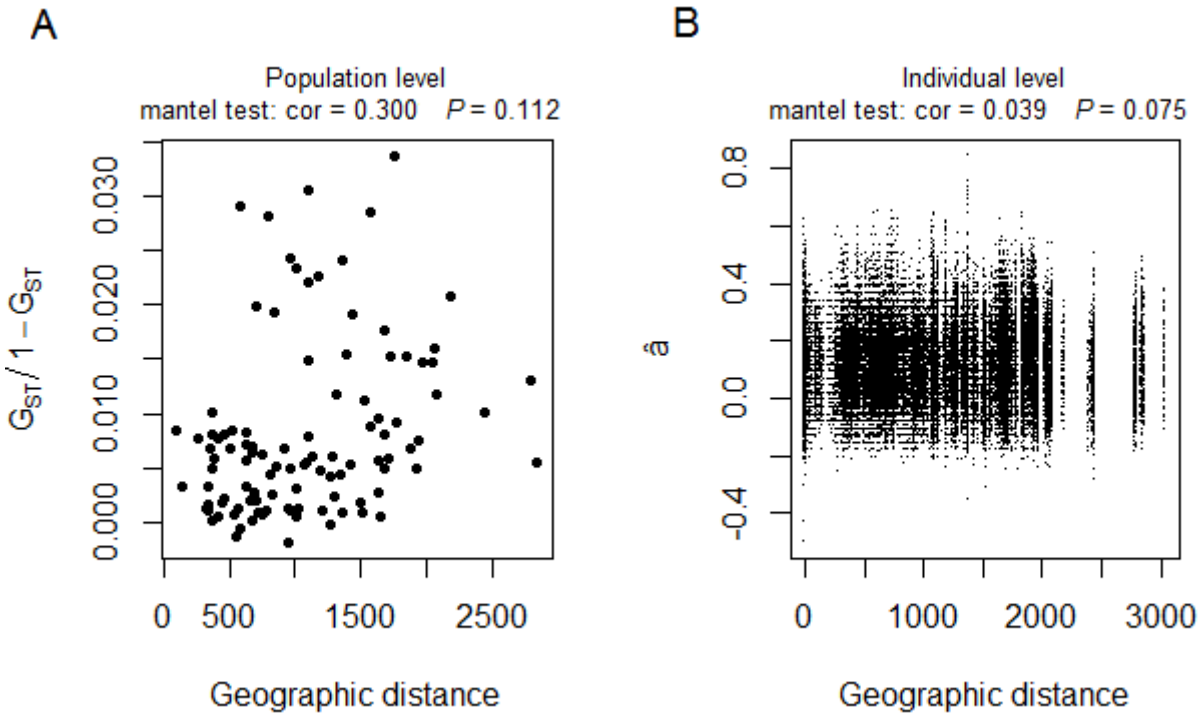
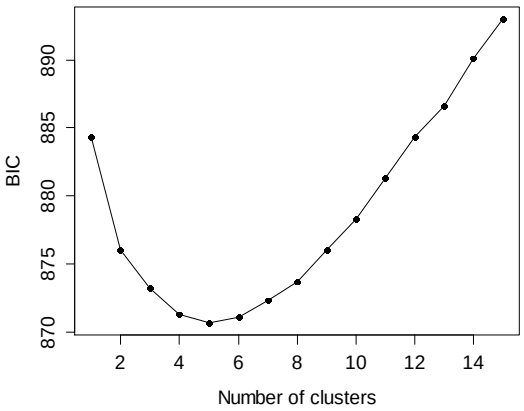
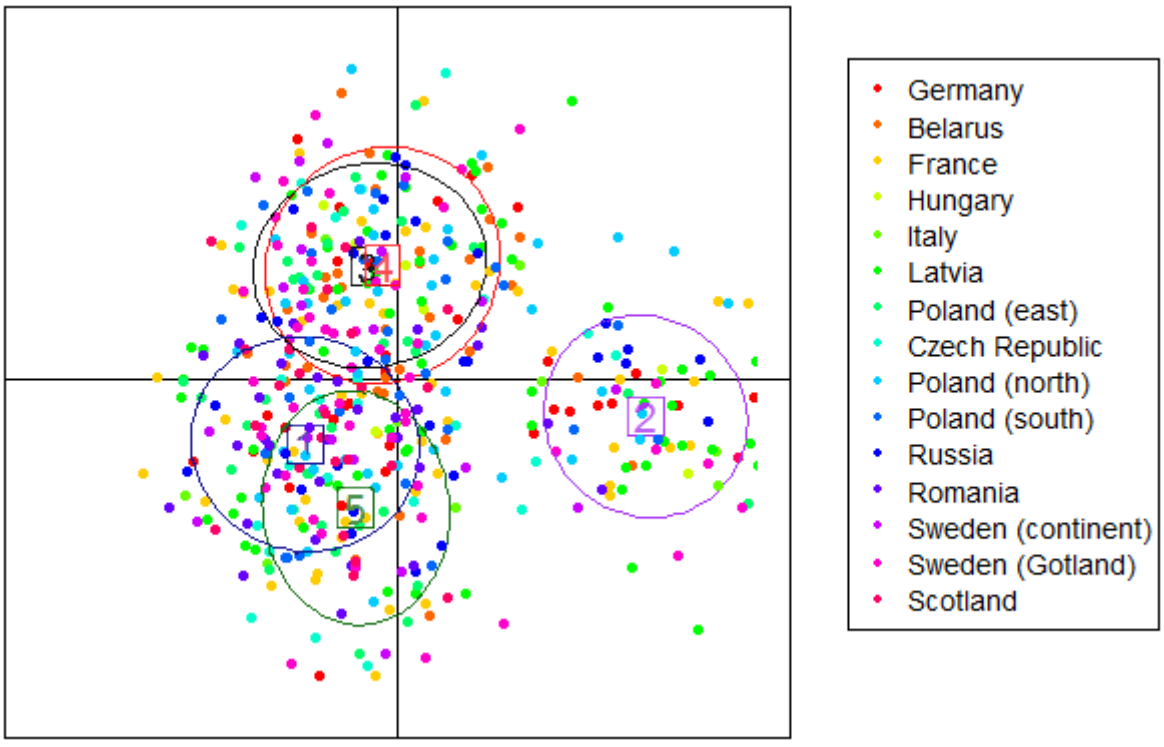


Figure S2: Inference of genetic clusters following Discriminant Analysis of Principal Components (DAPC). A) Selection of the optimal number of clusters with Bayesian Information Criterion (BIC); here BIC is maximum at 5 clusters. B) Scatterplot of the first two axes of the discriminant analysis. The five clusters are highlighted by ellipses and colours indicate the population of origin of each individual. Here the genetic clusters inferred by DAPC do not match the sampling sites.

A



B



3.2.8. Conclusion Article 4

On a pu observer l'existence d'une faible structure génétique, qui tend à différencier les populations française d'une part, et écossaise d'autre part, d'un ensemble relativement homogène constitué par les autres sites échantillonnés. Toutefois, en fonction des analyses, cette structure peut apparaitre non significative, montrant bien la faiblesse de cette différenciation. Dans le même temps, il s'avère que toutes les populations, y compris celles qui ont subis une très forte réduction d'effectif dans les dernières années, présentent une diversité génétique élevée. Malgré cela, l'analyse ABC permet de retrouver la majeure partie des tendances démographiques connues : populations décroissantes en Europe de l'ouest, populations constantes dans la majeure partie des sites d'Europe de l'est.

Le test de la direction du flux de gènes entre populations éloignées (est vs ouest) qui a été mené *via* une procédure ABC permet de mieux comprendre ces résultats pouvant apparaître au premier abord ambigus. En effet, les analyses soutiennent l'hypothèse d'un flux de gènes majoritairement dirigé depuis les populations d'Europe de l'est vers les populations d'Europe de l'ouest. Ceci s'explique par le déséquilibre démographique existant entre les populations européennes de Râles des genêts, avec des effectifs très abondants en Europe de l'est tandis que les populations de l'ouest sont très réduites. Dès lors, à taux de migration constant, un plus grand pourcentage de migrants venus de l'est se retrouvera dans les populations occidentales, entrainant un flux de gènes asymétrique est vers ouest. Certains indices, tels qu'une forte suspicion de voies migratoires différentes, laissent penser que les populations occidentales pouvaient présenter une différenciation par rapport aux populations d'Europe de l'est. Cette dispersion asymétrique sur l'aire de reproduction peut alors avoir contribué à masquer cette structure génétique en homogénéisant l'ensemble des populations européennes tout en maintenant une diversité génétique élevée.

Chapitre 3 : Conditions écologiques et variations du statut parasitaire des populations à l'échelle continentale

4. CHAPITRE 3 : CONDITIONS ECOLOGIQUES ET VARIATIONS DU STATUT PARASITAIRE DES POPULATIONS A L'ECHELLE CONTINENTALE

4.1. ARTICLE 5 : VARIATIONS DE L'INFECTION PARASITAIRE A L'ECHELLE EUROPEENNE

Les maladies induites par les pathogènes peuvent représenter une menace sérieuse pour les populations naturelles (McCallum & Dobson 1995; Pounds *et al.* 2006), tout particulièrement celles soumises à de nouveaux pathogènes auxquels elles ne sont pas adaptées (Woolhouse *et al.* 2005). La menace est d'autant plus grande que les populations sont de petite taille et donc plus sensibles aux fluctuations démographiques qui peuvent conduire à l'extinction (Saccheri *et al.* 1998; Bijlsma 2000). De nombreux facteurs, tant écologiques que génétiques, influencent la distribution des pathogènes au sein d'une aire de distribution. Les caractéristiques génétiques de l'hôte peuvent ainsi déterminer la probabilité d'infection (Frankham *et al.* 2002). Les populations ayant subi une perte de diversité génétique présentent souvent une réponse immunitaire amoindrie (Spielman *et al.* 2004), et sont ainsi plus sensibles aux pathogènes du fait de la perte d'allèles de résistance ou d'un avantage hétérozygote (Hedrick 2002; MacDougall-Shackleton *et al.* 2005). D'autre part, des facteurs écologiques (Morgenstern 1982; Schrag & Wiener 1995) tels que la densité des hôtes (Ebert *et al.* 2000), des vecteurs (Trape *et al.* 1992), la connectivité entre populations (McCallum & Dobson 1995) ou encore les caractéristiques de l'habitat (climat, etc.) (Hoshen & Morse 2004) jouent un rôle dans la transmission des pathogènes. Les variations spatiales de ces différents facteurs vont donc déterminer la distribution de la prévalence, c'est-à-dire la part d'individus infectés dans une population.

Concernant le Râle des genêts, on présupposait que le déclin important constaté dans les populations d'Europe de l'ouest, ainsi que leur position périphérique, aurait conduit à une baisse de diversité génétique dans ces populations. Dans ce contexte, on s'attendait à ce que celles-ci présentent une susceptibilité aux pathogènes plus importante que les populations abondantes d'Europe de l'est. Or, le chapitre précédent a mis en évidence un flux de gènes de l'est vers l'ouest qui contribue à homogénéiser les populations et permet le maintien de la diversité génétique. Dans ces conditions, les variations de diversité génétique entre populations sont extrêmement réduites et on s'attend à ce qu'elles ne puissent pas générer des différences de prévalence des pathogènes à travers l'Europe. Toutefois, ce contexte présente une opportunité

intéressante pour étudier les facteurs écologiques responsables d'éventuelles variations de prévalence à l'échelle continentale, puisque les facteurs génétiques sont ici contrôlés. Dans cette optique, on a analysé les patrons d'infection par la malaria aviaire, un type de parasite aviaire très commun présent chez une grande diversité d'hôtes et largement répandu mondialement (Valkiūnas 2005), chez le Râle des genêts. Les variations de prévalence entre sites européens ont ainsi été mises en relation avec différents facteurs écologiques afin d'identifier des déterminants non-génétiques de l'infection par la malaria chez cette espèce.

Parmi les quinze populations européennes de Râle des genêts échantillonnées pour l'analyse génétique, neuf d'entre elles ont été utilisées pour l'analyse des malarias. Il s'agit de celles pour lesquelles des échantillons de sang étaient disponibles, un pré-requis indispensable pour la détection des parasites sanguins. Une méthode moléculaire (Waldenström *et al.* 2004) a permis l'identification des individus infectés et la caractérisation des souches de malarias. Dans un premier temps, on a vérifié que les caractéristiques génétiques de l'hôte n'influencent pas le pattern d'infection observé : la probabilité pour un individu d'être infecté a ainsi été mise en relation avec la diversité génétique individuelle, puis on a analysé la relation entre diversité génétique à l'échelle populationnelle et prévalence. Sachant qu'on ne s'attend à rencontrer aucune relation significative entre génétique et infection, on a analysé dans un deuxième temps l'effet de différents facteurs écologiques, notamment le climat, la taille des populations hôtes et l'intensité agricole, sur les variations de prévalence en malaria aviaire en Europe. Cette étude doit permettre d'évaluer comment des variables écologiques peuvent déterminer les patrons d'infection parasitaires à l'échelle continentale, dans un contexte de conditions génétiques contrôlées.

Cette partie, menée grâce à une collaboration avec David Richardson de l'Université d'East Anglia (Norwich, GB), a fait l'objet d'une publication dans le journal « Evolutionary Applications » présentée ci-après :

Fourcade Y, Keišs O, Richardson DS, Secondi J (2014) Continental-scale patterns of pathogen prevalence: a case study on the corncrake. *Evolutionary Applications*. **In press**

Continental-scale patterns of pathogen prevalence: a case study on the corncrake

Yoan Fourcade[1,2], Oskars Keišs[3], David S Richardson[2*], Jean Secondi[1*]

[1] Université d'Angers, GECCO, 49045 Angers, France

[2] Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

[3] Laboratory of Ornithology, Institute of Biology, University of Latvia, Salaspils LV-2169, Latvia

[*] Equal contribution as senior and corresponding authors

4.1.1. Abstract

Pathogen infections can represent a substantial threat to wild populations, especially those already limited in size. To determine how much variation in the pathogens observed among fragmented populations is caused by ecological factors, one needs to examine systems where host genetic diversity is consistent among the populations, thus controlling for any potentially confounding genetic effects. Here, we report geographic variation in haemosporidian infection among European populations of corncrake. This species now occurs in fragmented populations, but there is little genetic structure and equally high levels of genetic diversity among these populations. We observed a longitudinal gradient of prevalence from western to eastern Europe negatively correlated with national agricultural yield, but positively correlated with corncrake census population sizes when only the most widespread lineage is considered. This likely reveals a possible impact of local agriculture intensity, which reduced host population densities in Western Europe and, potentially, insect vector abundance, thus reducing the transmission of pathogens. We conclude that in the corncrake system, where metapopulation dynamics resulted in variations in local census population sizes, but not in the genetic impoverishment of these populations, anthropogenic activity has led to a reduction of host populations and pathogen prevalence.

4.1.2. Introduction

Pathogens affect host fitness in various ways, including through loss of fecundity and reductions in survival (Lanciani 1975; Smith *et al.* 2009), and are thus a major driver of evolutionary

dynamics (Altizer *et al.* 2003). The deleterious effects of pathogens can also be a serious threat to any population (McCallum & Dobson 1995; Pounds *et al.* 2006; Martel *et al.* 2013) but especially to small populations that already experience elevated extinction risk due to demographic and genetic processes (Saccheri *et al.* 1998; Bijlsma 2000; O'Grady *et al.* 2006; Wright *et al.* 2007). For example, extinction probability is negatively related to population size because of the increasing impact of stochastic environmental events and epizootic infections with decreasing size (Lande 1988). Understanding what factors determine pathogen prevalence is therefore also important to conservation biology (Daszak *et al.* 2000).

Various ecological parameters influence pathogen infection (Morgenstern 1982; Schrag & Wiener 1995; Plowright *et al.* 2008). Density-dependent transmission (Dietz 1988; McCallum *et al.* 2001) has been shown to be responsible for pathogen dynamics in a vast range of host species (see for example, Burdon & Chilvers 1982; Jaffee *et al.* 1992; Ebert *et al.* 2000; Hochachka & Dhondt 2000). Together with the density of hosts, the density of vectors may determine infection probability of vector-transmitted pathogens (Trape *et al.* 1992; Pinto *et al.* 2000; Sol *et al.* 2000). Likewise, habitat fragmentation affects pathogen transmission (McCallum & Dobson 2002; Horan *et al.* 2008), as pathogens spread more rapidly between well-connected habitat patches. Therefore, we may expect habitat quality (driving local carrying capacity) and habitat connectivity (driving colonization/extinction rate and dispersal between populations) to determine the density of hosts and/or vectors. As a consequence, these factors would influence the rate of pathogen transmission within and among populations, and therefore, pathogen prevalence.

Host genetic characteristics also contribute to variation in pathogen distribution across a species range (Frankham *et al.* 2002; Hawley *et al.* 2005). Small host populations with depleted genetic diversity appear to be particularly susceptible to pathogens (Spielman *et al.* 2004) as a result of various genetic factors, including the loss of individual heterozygote advantage (MacDougall-Shackleton *et al.* 2005; Evans & Neff 2009) and/or the lack of specific alleles conferring resistance within the population level (Hedrick 2002). A negative relationship between host genetic diversity and prevalence is expected if prevalence reliably reflects (*i.e.* is positively correlated to) susceptibility. However, the opposite pattern may be observed if only genetically diverse individual survive infection. So although it is difficult to determine, *a priori*, the most likely pattern of correlation between pathogen susceptibility and observed infection, it is clear that host genetic diversity – at the population or individual level – can be an important driver of pathogen infection dynamics (Hedrick 2002; Altizer *et al.* 2003).

Understanding the relative contribution of genetic and ecological factors as drivers of pathogen distribution is a challenging issue. Range-scale studies offer the opportunity to analyse variation in pathogen infection across gradients of ecological conditions and host genetic diversity However, species with high dispersal capacity and low genetic structuring will provide particularly good systems, in which to investigate the effect of ecological factors on pathogen prevalence as gene flow will homogenise genetic diversity across their range, thus controlling for the potentially confounding effects of host genetic factors.

Avian malaria, here defined as infection by *Plasmodium* or related genera *Haemoproteus* and *Leucocytozoon* protozoans (Martinsen *et al.* 2008), has been shown to impact individual survival (Beier *et al.* 1981; la Puente *et al.* 2010) and reproductive success (Kilpatrick *et al.* 2006a; Knowles *et al.* 2010). Such haemosporidian parasites infect almost all bird species ever tested (Valkiūnas 2005), with various levels of pathogen–host specificity (Bensch *et al.* 2000; Cumming *et al.* 2013). Parasites of the genera *Plasmodium* and *Haemoproteus* are transmitted via mosquitoes belonging to the family *Culicidae*, while *Leucocytozoon*'s vectors are mainly flies of the family *Simuliidae* (Valkiūnas 2005). The transmission of avian haemosporidian parasites is mostly thought to occur during spring and summer in temperate climates (Atkinson 2008), but can also occur in tropical climates, such as the African wintering grounds of migrant bird species (Loiseau *et al.* 2012). Molecular methods now allow the rapid and efficient screening of these infections, as well as the identification of the parasite lineages involved (Bensch *et al.* 2000; Waldenström *et al.* 2004; Hellgren *et al.* 2004). Thus avian malaria has become a model of host-parasite interactions and their impact on host evolution, ecology and conservation (Westerdahl *et al.* 2005; Njabo *et al.* 2011; Asghar *et al.* 2011). Various studies have explored the effect of host genetic diversity on haemosporidian infection status in birds (MacDougall-Shackleton *et al.* 2005; Ortego *et al.* 2007). Infection patterns have also been linked to ecological factors at a relative fine scale, such as altitude (Marzal & Albayrak 2012), distance to water (Wood *et al.* 2007), food availability (Knowles *et al.* 2011), host density (Isaksson *et al.* 2013; Lachish *et al.* 2013) or other habitat characteristics (Lachish *et al.* 2013; Gonzalez-Quevedo *et al.* 2014). However, the contribution of ecological factors on the variation in haemosporidian prevalence at larger, continental scale has received little attention.

The corncrake (*Crex crex*), is a widely distributed bird species that breeds in grassland habitats from western Europe to Siberia (Schäffer & Koffijberg 2004). Its conservation status differs greatly across different regions of its range. In the westernmost areas, agriculture intensification has resulted in the degradation of habitat suitability and, consequently, population

fragmentation, thus leading to a decreasing gradient in population census size from eastern to western Europe (Green & Rayment 1996; Green *et al.* 1997a; Birdlife International 2013; Fourcade *et al.* 2013). Interestingly, spatial genetic structure is weak across the European range, and gene flow from the eastern to the western sites appears to maintains high genetic diversity in all populations (Fourcade et al., *submitted*). This species, as well as many farmland bird species in Europe (Donald *et al.* 2001), has seen its distribution and population trends shaped by anthropogenic activity during the last century. Such disturbance, occurring over a large geographic scale and an extended period, may have disrupted previous host-parasites dynamics and could thus pose overlooked threats to these already declining populations. Therefore, analysing the current patterns of pathogen infections and their ecological drivers seems essential to efficiently anticipate long-term conservation actions.

Here we investigated the geographic pattern of haemosporidian infection (as a model of a widespread pathogen), in relation to ecological factors across the corncrake's European breeding range. Infection status, and the identity of infecting parasites lineages, was determined for all individuals across populations using molecular screening (Hellgren *et al.* 2004). In order to test whether host genetic diversity influences malaria prevalence despite the very low interpopulation variation in this parameter, we first verified that prevalence was uncorrelated with estimates of genetic diversity calculated using a suite of microsatellite markers. Second, we tested the effects of various ecological factors, including climate, host population size (census compared to effective population size) and mean agricultural yields on malaria prevalence. We discuss the implications our results have in regards to understanding the large scale structuring of pathogen faunas within animal populations and, more specifically, what implications this may have for corncrake conservation.

4.1.3. Material and Methods

4.1.3.1. Study species and sample collection

The corncrake (*Crex crex*) is a migratory bird that breeds in the Palearctic, from western Europe to Baikal Lake, and winters in southeast Africa. On its breeding ground, it occurs mainly in natural or semi-natural grasslands such as floodplain meadows, alpine grasslands or steppes (Schäffer & Koffijberg 2004). We sampled nine European populations (Table 10) following the longitudinal demographic gradient that occurs in Europe. Blood samples from 354 corncrakes were collected in 2011 and 2012 during the peak breeding period (May – July). Between 11 pm

and 3 am birds were attracted using playback of conspecific male calls and captured with a dipnet or by hand. This method captures males only. Small (ca. 25 µl) blood samples were collected from the brachial vein and stored in absolute ethanol. Each bird was ringed before being released to avoid resampling the same individual within or between years.

4.1.3.2. Haemosporidian parasites screening

DNA was first extracted following a salt extraction protocol (Richardson *et al.* 2001). Haemosporidian infection was detected using a nested PCR (Waldenström *et al.* 2004; Hellgren *et al.* 2004). A first PCR amplifies a 570 bp fragment of the cytochrome *b* gene of species belonging to the genera *Plasmodium*, *Haemoproteus* and *Leucocytozoon*, using the primers HaemNF1 and HaemNR3 (Hellgren *et al.* 2004). Two different PCRs were then run on an aliquot of the first reaction to amplify a shorter fragment of DNA within the first amplicon. The primers HaemF and HaemR2 (Bensch *et al.* 2000) were used to amplify a 477 bp fragment of *Haemoproteus* or *Plasmodium* while the primers HaemFL and HaemR2L (Hellgren *et al.* 2004) were used to amplify a 475 bp fragment of *Leucocytozoon*s.

The first PCR was run in a volume of 10 µL containing 1 µL of extracted DNA (approximately 10 ng/µL), 5 µL of Qiagen TopTaq, 0.4 µL of each primer (initial concentration: 10 mM) and 3.2 µL of pure water. The reaction was performed according to the following conditions: after incubation at 96°C during 3 min, 20 cycles of 20 sec at 94°C, 30 sec at 50°C and 45 sec at 72°C, following by a final incubation at 72°C for 10 min and 20°C for 5 min. The second reaction used 1 µL of PCR product from the first reaction, with the same proportion of reagents. The first and final incubations were similar to the first PCR but the cyclic reaction was as follows: 40 cycles of 30 sec at 94°C, 45 sec at 49°C with *Plasmodium*/*Haemoproteus* primers, or 57°C with *Leucocytozoon* primers, and 45 sec at 72°C. The final amplification was visualized on a 2% agarose gel using Ethidium bromide in order to identify infected birds. Positive and negative controls (using either a known infected sample from another bird species, or 1 µL H_2O respectively) were included in all PCR reactions and on the agarose plates. Each sample was run twice to ensure the detection of infected birds and reduce false negatives. When there was inconsistency between two runs, a third screening was run to ensure the correct assignment of infection status. Only individuals that gave positive results in two runs were counted as being infected.

All positive PCR products were sequenced on an ABI 3730 XL sequencer. Sequences were aligned using BioEdit (Hall 1999) and ClustalW (Thompson *et al.* 1994). We compared the sequences to homologous sequences deposited in the National Centre for Biotechnology Information (NCBI) Genbank (Benson *et al.* 2005) and MalAvi (Bensch *et al.* 2009) databases to identify already known lineages. Exact matches with already published sequences were labelled according to the name of the known strain. When a sequence was already referred to by different names, we chose to keep the first published name. Sequences that differed by 1 bp or more were assigned a new name following the guidelines suggested by Bensch *et al.* (2009): the abbreviated scientific name of the host species (here CRECRE) followed by a number. The phylogenetic relationships between lineages is given in Supporting information, Figure S1, following the protocol described in Supporting information, Appendix S1.

4.1.3.3. Microsatellite genotyping, genetic diversity and effective population size

Each DNA sample was genotyped at 15 microsatellite loci. Eight highly polymorphic markers had been specifically designed for corncrake: *Crex1*, *Crex2*, *Crex6*, *Crex7*, *Crex8*, *Crex9*, *Crex11*, *Crex12* (Gautschi *et al.* 2002) whereas the other markers were identified as being conserved across a large range of bird species: *CAM18* (Dawson *et al.* 2013), *TG02-120*, *TG04-12*, *TG04-12a*, *TG04-41*, *TG05-30* and *TG012-15* (Dawson *et al.* 2010). Full details of the genotyping method and genetic statistics of the markers are given in (Fourcade *et al., submitted*) and Supplementary information, Table S1.

We computed three common estimates of individual multilocus heterozygosity, using "Rhh" R package (Alho *et al.* 2010): the standardized heterozygosity *stH* (Coltman *et al.* 1999), the internal relatedness *Ir* (Amos *et al.* 2001) and the homozygosity by locus *Hl* index (Aparicio *et al.* 2006). We also estimated population-level heterozygosity and genetic diversity using the following measures, computed with "HIERFSTAT" R package (Goudet 2005): observed heterozygosity *Ho*, gene diversity or expected heterozygosity *He*, rarefied allelic richness *Ar* and the inbreeding coefficient F_{IS}. The effective population size (N_e) was calculated for each sampling site using an Approximate Bayesian Computation (ABC) (Beaumont *et al.* 2002) approach. We used simulations already computed to investigate the demographic history of corncrake across Europe (Fourcade et al. *submitted*) using the framework implemented in the "abc" R package (Csilléry *et al.* 2012). The full details of N_e calculation are given in Supporting Information, Appendix 2.

4.1.3.4. Statistical analyses

We assessed the effect of individual measures of genetic diversity on infection probability using binomial regressions. We computed generalized linear mixed models (GLMMs) with population identity as random effect using the "lme4" R package (Bates *et al.* 2014). We used linear regressions to test the relationships between haemosporidian prevalence and the three measures of population-level genetic diversity: *Ho*, *Ar* and F_{IS}.

We then investigated the effect of three main categories of ecological factors on the variation of malaria prevalence:

(i) Climate: We obtained climatic variables from the WorldClim project (Hijmans *et al.* 2005), downloaded at a 2.5 arc-min resolution (www.worldclim.org). The original database contained 19 variables but, as some of them were highly redundant, we selected the subset of eight predictors that described the spatio-temporal variations of temperature and rainfall across the study area: the annual mean temperature (Bio1), the maximum temperature of the warmest month (Bio5), the minimum temperature of the coldest month (Bio6), the temperature annual range (Bio7), the annual precipitation (Bio12), the precipitation of the wettest month (Bio13), the precipitation of the driest month (Bio14) and the precipitation seasonality (Bio15). Since they remained strongly intercorrelated, we performed a principal component analysis (PCA) on these eight climatic grids and used the first axis, which accounted for 50.2% of the total climatic variation in the study area, as a predictor variable. This component mostly depicted the west-east longitudinal gradient from the oceanic to the continental climate (Supporting information, Figure S1). In order to take into account fine-scale variability, we extracted the mean climatic value in a 50 km buffer around each sampling site.

(ii) Agriculture intensity: The mean wheat yields per country (2012 data) were downloaded from FAOSTAT (Food and Agriculture Organization of the United Nations, http://faostat.fao.org/, accessed on 11/03/2014) and was used as a proxy for the level of agriculture intensification across Europe.

(iii) Host population size: We included in our analyses two measures of the corncrake population size, 1) inferred by the national census population sizes of corncrake, obtained from Schäffer & Koffijberg (2004), and 2) the effective population sizes N_e calculated here from genetic data.

Despite the fact that we retained only four potentially informative variables, it is worth noting that they remained correlated (Variance inflation factors VIF: climate: 3.66, census size: 4.73, effective size: 1.23, yield: 5.09). Therefore, after testing for a relationship between each predictor and prevalence using linear regressions, we carried out model selection based on the corrected Akaike information criterion (AICc) (Burnham & Anderson 2002) to determine the variables, or combination of variables, that best explained the observed patterns of prevalence. Model selection was carried out using the "MuMIn" R package (Barton 2013). We carried out the analyses described above for all malaria lineages pooled together, and for SW2 alone, the most common and widespread lineage we detected (see Results section). Additionally, we assessed the linear relationship between haemosporidian lineage richness and the four variables included above.

4.1.4. Results

4.1.4.1. Haemosporidian prevalence and distribution of lineages

We found no evidence of cross-sample contamination or failed amplification based on the negative and positive controls. Observed overall prevalence across all populations was 10% (36/354 birds). Prevalence varied considerably among populations across Europe (Range = 0–30%, $\chi^2 = 18.41$, $P = 0.018$) exhibiting a spatial gradient from south-west (France, 3.3% prevalence) to north-east (Russia, 30% prevalence) (Figure 26, linear regression against longitude: $F_{1,7} = 13.00$, *adjusted* $R^2 = 0.60$, $P = 0.01$, linear regression against latitude: $F_{1,7} = 1.06$, *adjusted* $R^2 = 0.01$, $P = 0.34$).

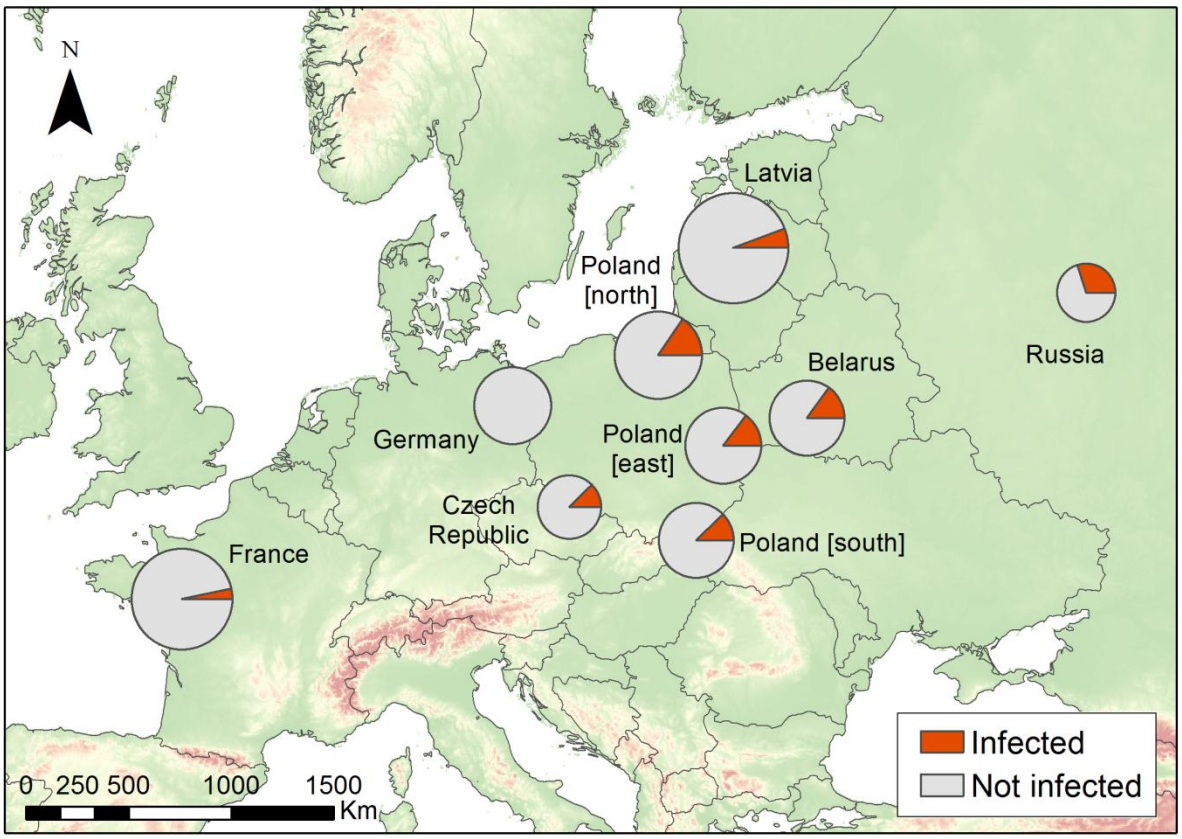


Figure 26: Geographic distribution of malaria prevalence per population across nine European populations of corncrake (*Crex crex*). The size of each circle is function of the number of samples from that location (minimum: Russia, 20 samples; maximum: Latvia, 71 samples).

Ten different lineages of haemosporidian parasites were detected (Table 10): seven *Plasmodium*, two *Leucocytozoon*, and one *Haemoproteus* lineage (Supporting information, Figure S2). One bird was found to be infected by both a *Leucocytozoon* strain and a *Plasmodium* strain. Another four Polish birds showed evidence of mixed infection with both the *Plasmodium* strain SW2 and a previously undescribed haplotype that was 1 bp different (CRECRE1; GeneBank accession number *KJ783457*). This new lineage was confirmed by the repeated amplification and sequencing of the original DNA sample.

Among the ten haemosporidian strains detected, one *Plasmodium* lineage (SW2) occurred in 71% (25/36) of infected corncrakes (Table 10). This haplotype was restricted to the six easternmost populations (Poland, Latvia, Belarus, and Russia) with an average prevalence of 11.6% across these locations. SW5 was found only in Russia, infecting 2 birds. Regarding the western populations, France was characterised by a single lineage found only at this site: ACCTAC01. In the Czech Republic – the westernmost site after France in which haemosporidian parasite was detected – a total of three lineages were found. Two of these

lineages, WA42 and RTSR1, occurred only in the Czech Republic while the lineage SYBOR10 was found here and also in populations further east.

Table 10: Number of infected corncrakes, and prevalence per haemosporidian lineage, for the nine sampling sites across Europe. Sampling sites are ordered from west to east. Genebank accession numbers are provided behind each lineage name.

Location	Long	Lat	Sample size	Infected (prevalence)	Number of positive infections per malaria lineage per population (prevalence)									
					ACCTAC01[a]	SYBOR10[a]	WA42[a]	RTSR1[a]	SW2[a]	CRECRE1[a]	SW5[a]	WW2[b]	SYBOR08[c]	CIAE02[c]
					EU810700	*DQ368390*	*EU810615*	*AF495568*	*AF495572*	*KJ783457*	*AF495574*	*AY831755*	*DQ847239*	*EF607287*
France	-0.51	47.58	60	2 (0.03)	2 (0.03)									
Germany	14.30	53.05	34	0 (0.00)										
Czech Republic	16.49	50.24	24	3 (0.13)		1 (0.04)	1 (0.04)	1 (0.04)						
Poland [north]	20.40	54.31	45	7 (0.16)					6 (0.13)	2 (0.04)		1 (0.02)		1 (0.02)
Poland [south]	22.06	49.29	33	4 (0.12)					4 (0.12)					
Poland [east]	23.23	52.59	34	5 (0.15)		1 (0.03)			4 (0.12)	2 (0.06)				
Latvia	23.67	56.71	71	4 (0.06)					4 (0.06)					
Belarus	24.73	52.66	33	5 (0.15)					4 (0.12)				1 (0.03)	
Russia	39.16	55.87	20	6 (0.30)					3 (0.15)		2 (0.10)			1 (0.05)
Total (mean prevalence)				36 (0.10)	2 (0.01)	2 (0.01)	1 (0.00)	1 (0.00)	25 (0.07)	4 (0.01)	2 (0.01)	1 (0.00)	1 (0.00)	2 (0.01)

4.1.4.2. Relationship between haemosporidian prevalence and genetic diversity

Following a binomial GLMM procedure, we did not find any effect of standardized heterozygosity (*stH*) on infection probability (Wald $Z = 0.51$, $P = 0.61$). No relationship was detected for the two other predictors either: internal relatedness *Ir* (Wald $Z = -1.15$, $P = 0.61$), homozygosity by locus *Hl* (Wald $Z = -1.02$, $P = 0.31$). Similarly, we found no effect of genetic estimators of diversity on infection probability when considering SW2 lineage only (all $P > 0.5$).

Observed heterozygosity (*Ho)* varied between 0.63 and 0.75 among populations, but was not related with haemosporidian prevalence (all lineages: $F_{1,7} = 0.64$, *adjusted* $R^2 = -0.05$, $P = 0.45$, SW2: $F_{1,7} = 0.003$, *adjusted* $R^2 = -0.14$, $P = 0.96$). Similarly, little variation among populations was observed in allelic richness (*Ar*: 8.95–9.78), gene diversity (*He*: 0.72–0.77) and F_{IS} (0.00-0.17), and none of these measures was correlated with haemosporidian prevalence, either for all lineages or for SW2 only (all $P > 0.1$).

4.1.4.3. Estimation of effective population size

Overall, the ABC analysis indicated a mean effective population size across all populations of 117204 ± 65853 (Supporting information, Table S3, minimum: mode $N_{e\_Poland\ (East)} = 50976$, 95% CI: 25787–364012; maximum: mode $N_{e\_Germany} = 277179$, 95% CI: 123777–732928). The estimation of N_e for the whole dataset was higher than for each population separately (mode $N_{e\_all\text{-}data} = 385833$, 95% CI: 85225–744614), and remained within a plausible range given the estimated European corncrake population size of 2.6–4 million birds (Schäffer & Koffijberg 2004; Birdlife International 2013). N_e did not exhibit any longitudinal or latitudinal pattern (longitude: $F_{1,7} = 0.003$, *adjusted* $R^2 = -0.14$, $P = 0.96$; latitude: $F_{1,7} = 0.0008$, *adjusted* $R^2 = -0.14$, $P = 0.98$). Census and effective population size estimated per sampling site were not correlated (effective size vs. census size: $F_{1,7} = 0.17$, *adjusted* $R^2 = -0.12$, $P = 0.70$) (Supporting information, Table S3).

4.1.4.4. Relationship between haemosporidian prevalence/richness and ecological factors

We found that total haemosporidian prevalence exhibited a significant negative relationship with climate ($F_{1,7}$ = 7.54, *adjusted* R^2 = 0.45, P = 0.03), and a positive relationship with agricultural yield ($F_{1,7}$ = 29.91, *adjusted* R^2 = 0.78, P < 0.001) and corncrake census size ($F_{1,7}$ = 14.48, *adjusted* R^2 = 0.63, P < 0.001), but not with effective population size ($F_{1,7}$ = 0.33, *adjusted* R^2 = -0.09, P = 0.58). Among these variables, the model selection procedure identified agricultural yield as the most important factor influencing total haemosporidian prevalence (Table 11 and Figure 27A). All other models greatly departed from this one regarding ΔAICc (difference with 2nd best model = 4.87), showing that the other predictors poorly explained the observed variation of prevalence compared to yield.

Considering only SW2 prevalence, a similar positive relationship was found with corncrake census size ($F_{1,7}$ = 27.30, *adjusted* R^2 = 0.76, P = 0.001) and agricultural yield ($F_{1,7}$ = 11.84, *adjusted* R^2 = 0.55, P = 0.01), but the regression with the climate principle component was not significant anymore ($F_{1,7}$ = 3.84, *adjusted* R^2 = 0.26, P = 0.09). Again, the relationship with corncrake effective population size was not significant ($F_{1,7}$ = 1.48, *adjusted* R^2 = 0.06, P = 0.26). However, here, the best model explaining SW2 prevalence included only corncrake population census size (ΔAICc with 2nd best model = 2.61) (Table 11 and Figure 27B). In this case, agricultural yield, which was ranked first for total prevalence, appeared only in third position (ΔAICc with best model = 5.40).

No significant linear relationship was identified between lineage richness and the four predictors tested (all P > 0.05). The relationship between richness and agricultural yield approached significance though ($F_{1,7}$ = 4.79, *df* = 7, *adjusted* R^2 = 0.32, P = 0.06).

Table 11: Results of model selection by AICc. Linear models linking haemosporidian infection and ecological predictors, for all lineages and for SW2 lineage only, are ranked by AICc. For visual convenience, only models that had an AICc weight > 0.01 are shown. Yield is the mean wheat yield per country as provided by the FAO. The climate variable is a synthetic climatic predictor extracted from a PCA on the *Bioclim* dataset (Hijmans *et al.* 2005). Census and effective sizes are corncrake population size inferred respectively from field surveys (Schäffer & Koffijberg 2004) and genetic analyses.

	Adj. R^2	*F*	df	AICc	ΔAICc	AICc weight
All lineages						
Yield	0.78	29.91	3	-23.50	0.00	0.83
Census size	0.63	14.48	3	-18.60	4.87	0.07
Yield + Census size	0.75	13.30	4	-16.60	6.93	0.03
Yield + Effective size	0.75	12.96	4	-16.40	7.12	0.02
Yield + Climate	0.75	12.82	4	-16.30	7.20	0.02
Climate	0.45	7.54	3	-15.10	8.39	0.01
SW2						
Census size	0.77	27.33	3	-28.60	0.00	0.69
Census size + Effective size	0.84	21.51	4	-26.00	2.61	0.19
Yield	0.58	11.84	3	-23.30	5.40	0.05
Census size + Climate	0.77	14.77	4	-23.10	5.50	0.04
Census size + Yield	0.73	11.78	4	-21.50	7.16	0.02

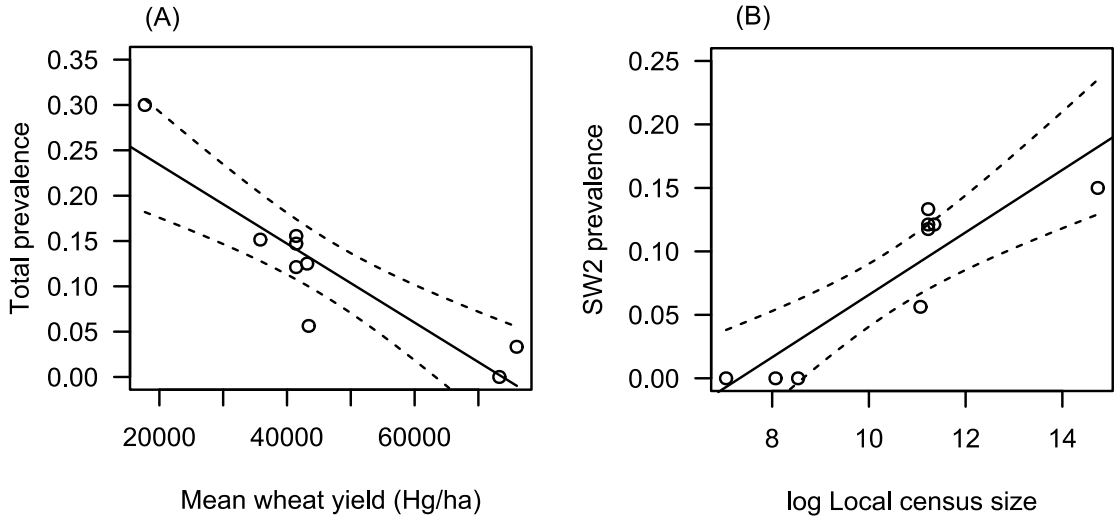


Figure 27: Haemosporidian prevalence in nine European populations of corncrake plotted against (A) agricultural intensity approximated by the mean wheat yield per country (in Hg/ha) for all haemosporidian lineages pooled and (B) corncrake local census population size for the most widespread lineage only (SW2).

4.1.5. Discussion

Mean prevalence of haemosporidian infection across the European range of the corncrake was ca. 10%, which is relatively low compared to other bird species. For example, an analysis of blood parasites across 74 passerine species revealed an average prevalence of 26% (Scheuerlein & Ricklefs 2004). Similarly, a 39% prevalence was found among 50 bird species sampled in Dominican Republic (Latta & Ricklefs 2010). However, in the corncrake, haemosporidian prevalence showed a strong geographical gradient, increasing from western to eastern Europe. Interestingly, the prevalence of easternmost populations was consistent with the average value given above, whereas western populations appear to be almost free of these parasites. In the corncrake, where individual or population heterozygosity had no effect on haemosporidian infection, prevalence was strongly related with agriculture yield per country. However, when only the most widespread lineage SW2 was considered, the most important factor explaining prevalence was local corncrake census size.

The lack of relationships between haemosporidian prevalence and host genetic diversity is consistent with our predictions. As a consequence of high gene flow, no loss of genetic diversity occurred in the threatened westernmost populations. Indeed, genetic diversity varied little between populations (H_o: 0.63–0.75) and the estimates of effective population size provided by the ABC analysis were totally unrelated to the survey-based population estimates. Therefore, corncrake genetic characteristics cannot explain the spatial variation in haemosporidian prevalence. Since genetic diversity differs so little between populations, ecological factors must account for the marked spatial variation of haemosporidian prevalence across the corncrake range. A likely explanation is that haemosporidian prevalence is driven by vector density (Trape *et al.* 1992; Loaiza & Miller 2013). This hypothesis is supported by the negative relationship between haemosporidian prevalence and agricultural yield. Differences of vector density may be caused by variation in natural environmental conditions or in the intensity of human disturbance. The massive drainage of wetlands (Brinson & Malvárez 2002), and intensive use of pesticides in farmland (Geiger *et al.* 2010) across western Europe, may have reduced the number of vectors, either by directly reducing vector populations, and/or indirectly by reducing the size of other host bird populations (Donald *et al.* 2001; Stoate *et al.* 2009). It has been shown that agriculture intensification can lead to a decline of *Diptera* abundance (Wickramasinghe et al., 2004; Paquette et al., 2013). In contrast, in an island system, Gonzalez-Quevedo *et al.* (2014) showed that anthropogenic activity, specifically the creation of water reservoirs and poultry farms, can increase avian malarial infection within a natural bird population. In Europe, agricultural practices

show a gradient of intensity from west to east which may affect vector fitness and, as a consequence, have generated the gradient of haemosporidian prevalence in corncrake populations that we observe. The reasons for such large-scale variation in the density of malaria vectors have never been investigated. Human-driven changes of the environment operate at the ecosystem scale and it seems likely that both vector and host densities have experienced the same gradient of alteration in the last decades. Although our results did not provide direct evidence, they appear to support the hypothesis that agricultural intensity has affected pathogen communities.

Although at the scale of the whole haemosporidian community, the intensity of agriculture appeared to be the main driver of prevalence, it is noticeable that, when we focused on a single malaria lineage (here SW2), haemosporidian prevalence was highly correlated with the gradient in host census sizes across Europe. Classically, host density is a key factor that determines parasite transmission (Dietz 1988), including in malaria (Lachish *et al.* 2011; Isaksson *et al.* 2013). It could account for the observed variations of prevalence at the scale of the SW2 lineage. In our sampling, most infected birds carried this very generalist haemosporidian lineage. It has been described as *Plasmodium homonucleophilum* (Ilgūnas *et al.* 2013), and has been identified in numerous bird species, including sedge warbler *Acrocephalus schoenobaenus* (Waldenström *et al.* 2002), great tit *Parus major* (Beadell *et al.* 2006), and tawny owl *Strix aluco* (Krone *et al.* 2008). Therefore, its transmission relies on a range of hosts, and does not depend on corncrake only, which at first sight limits the impact that corncrake density alone should have on its prevalence. Nevertheless, the observed gradient of corncrake population size along the gradient of agriculture intensity is likely to exist in many bird species affected by agricultural practices (Donald *et al.* 2001), so the overall pool of host species may exhibit the same pattern, thus influencing parasite transmission. Moreover, corncrake males tend to aggregate on specific calling sites during the breeding season (Budka & Osiejuk 2013b; Ręk 2013b) and such behaviour certainly favours density-dependent pathogen transmission. Furthermore, although we do not have direct measures of local density, the large populations of corncrakes in eastern Europe should result in much higher within-patch local densities, or higher densities of such breeding areas, than in western Europe, both of which would facilitate transmission of haemosporidian parasites. Moreover the large populations in eastern Europe may provide a reservoir of chronically infected birds that contributes to the maintenance of relatively high prevalence.

The identity of haemosporidian lineages provides some alternative explanations for the observed pattern. Indeed, most infected birds in eastern Europe were carriers of SW2 while this

lineage was absent from the western sites. This generalist lineage was already identified in several western locations (for example United-Kingdom (Szöllősi *et al.* 2011) or Portugal (Ventim *et al.* 2012)) as well as eastern European countries e.g. Romania (Svoboda *et al.* 2009) and Russia (Ilgūnas *et al.* 2013)). Clearly its range is not restricted to eastern Europe. Therefore, the low prevalence in western sites may explain why the SW2 lineage was not detected there. Nevertheless, these results raise questions about the geographic structure of haemosporidian lineages across the corncrake range. Its distribution may be explained by the use of alternate migration routes and/or wintering areas (Rintamäki & Ojanen 1998; Wirth *et al.* 2005; Durrant *et al.* 2008). There is some data to support this hypothesis. We found evidence that the French and the Scottish population (the latter was not sampled in a way that allowed for disease screening) differ genetically and morphologically from the rest of the Europe corncrakes (Fourcade *et al. submitted*). Similarly, recent data about corncrake migration suggest that birds breeding in Britain may use a different migration pathway than more eastern populations (Green 2013). If the French birds also follow this alternative western migration route, and providing the haemosporidian infections are acquired in wintering grounds, this may explain why this population differs so clearly in terms of the genetic identity and prevalence of pathogens found there. However, this issue remains rather speculative and needs further investigation. Indeed, most haemosporidian strains identified here have already been found in migratory hosts, both in Africa and Europe. For example, ACCTAC01, the *Plasmodium* lineage found in France, has also been identified in resident African species, such as the African Goshawk *Accipiter tachiro*, showing that infection may occur in Africa. In contrast, the widespread SW2 lineage has been found in a non-migrant European species, the tawny owl *Strix aluco* (Krone *et al.* 2008), showing that this parasite can be acquired in the corncrake's breeding grounds. Although infections sites are unknown in the present case, the clear longitudinal pattern of prevalence that we observed in Europe suggests that it depends on factors occurring in the breeding area. Furthermore, there is no explanation why processes occurring in winter would determine the relationships between prevalence and agriculture intensity in Europe.

We predicted, and confirmed, that host genetic diversity would not be driving patterns of pathogen prevalence in the corncrake system because gene flow maintains equally high diversity level across the European range. Therefore, the large variation in haemosporidian prevalence observed must be explained by ecological factors. The longitudinal gradient of haemosporidian parasites prevalence correlated with wheat yields, used here as a proxy for agriculture intensity. Focusing on a single lineage, the most important variable driving prevalence was host population size, but again this factor is directly linked to agriculture activity which contributes to the gradient

of corncrake population sizes. A likely explanation is that agriculture intensification in western Europe has led to reduced infection by strongly limiting both vector and host density. A practical consequence is that infection by haemosporidians – or other pathogens borne by insect vectors and/or where transmission is density-dependent – should not be a major threat to the viability of these small bird populations. Our results also suggest that the massive decline of corncrake in western Europe can be largely imputed to agriculture practices, and not to other neglected factors such as pathogens. Thus, efficient conservation actions could be largely inspired by those applied in United-Kingdom – based on the management of mowing practices – as they managed to halt the decrease of the species, and eventually to recover a significant corncrake population (O'Brien *et al.* 2006).

As already stated, the areas of low haemosporidian prevalence may indicate a deterioration of grassland ecosystems with an extirpation of most insect vectors or a disruption of parasitic cycles. At the European scale, agricultural intensity has been shown to be linked to a decline of arthropod communities in farmland landscapes (Hendrickx *et al.* 2007; Féon *et al.* 2010). As a global decrease of insect populations is observed (Dunn 2005; Conrad *et al.* 2006), managing insect populations is becoming a major issue since their decline directly affects ecosystem services such as pollination (Potts *et al.* 2010). Therefore, efforts should be made to implement conservation strategies that maintain both biodiversity and functional relationships like host-parasite interactions. In this regard, parasites screening in birds hosts may serve in monitoring insect populations and functional interactions, and may thus provide wider insights into biodiversity conservation in agricultural landscapes. More generally, our study system allowed us to assess the effect of large scale ecological factors on prevalence patterns. Further continental-wide studies are needed that provide insights about the relative contribution of extrinsic (ecological) and intrinsic (genetic) factors on pathogen prevalence. These may not only provide ecological and evolutionary understanding of pathogen dynamics but may also improve the design of conservation strategies for wild populations potentially threatened by pathogens (De Castro & Bolker 2004; Smith *et al.* 2009). They may also help to predict the spread of zoonotic diseases carried by migrating animals (see examples for avian influenza, (Reed & Meece 2003; Gilbert *et al.* 2006; Kilpatrick *et al.* 2006b)).

4.1.6. Acknowledgments:

The authors would like to thank all people who helped in the collection of corncrake samples: Michal Budka, Serguei Fokin, Peter Zverev, Susanne Arbeiter, Joachim Sadlik, Edouard Beslot, Gilles Mourgaud, and Emmanuel Séchet. We are also grateful to Tom Jolin and Dave Wright for their contribution to malaria screening. This study was founded by a grant from *Plan Loire Grandeur Nature*, *Région Pays-de-Loire*, *Angers Loire Métropole*, *Conseil Général Maine-et-Loire*, *Direction Régionale de l'Environnement, de l'Aménagement et du Logement Pays-de-Loire* (DREAL) and European Regional Development Fund.

4.1.7. Supporting information

Appendix S1: Phylogenetic analyses of malaria lineages detected in corncrake

We built a tree describing the phylogenetic relationships between the detected malaria lineages. We first assessed the model of sequence evolution with jModelTest 2.1.3 (Darriba et al., 2012). Following the AICc scores, we chose the GTR + gamma model of nucleotide substitution for further analyses. The phylogenetic tree was constructed using the Bayesian method implemented in MrBayes 3.2.1 (Ronquist and Huelsenbeck, 2003). We ran the analysis with 4 simultaneous Markov chains during 5 million generations sampled every 100 generations, and discarding 25% of the trees as burn-in. As an outgroup we included a cytochrome b sequence of *Plasmodium falciparum,* identified in gorillas (*Gorilla gorilla*) (Genbank accession number: GU045311.1). We used FigTree 1.4 (http://tree.bio.ed.ac.uk/software/figtree) to plot the resulting tree and reported the Bayesian support of tree nodes as computed by MrBayes (Figure S2).

Appendix S2: Estimation of effective population size by Approximate Bayesian Computation

We used an Approximate Bayesian Computation (ABC) (Beaumont *et al.* 2002) method previously implemented to investigate the demographic history of corncrake across Europe (Fourcade et al. *submitted*). For each sampling location, we simulated data using "ABCtoolbox" (Wegmann *et al.* 2010) to sample parameters in our prior distributions and "Fastsimcoal" (Excoffier & Foll 2011) for data simulation, under three simple demographic scenarios: decreasing, constant and increasing effective population size. One million simulations were conducted per demographic model and per population, and the most probable scenario was

selected following the neural network approach implemented in the "abc" R package (Csilléry *et al.* 2012) using a 25% tolerance rate. Among the nine sampled populations, five were assigned to the model of decreasing effective population size and four to the constant model with a high posterior probability (0.80 ± 0.16 SD) (Fourcade et al. *submitted*).

Using the models selected in this previous work (Table S2), posterior probabilities of parameters were assessed following the neural network method. Effective/census population size ratio is usually around 0.10 in wild populations (Frankham 1995). Since the European corncrake population is estimated at ca. 3,000,000 birds (Schäffer & Koffijberg 2004), the prior distribution for effective population size was set between 1 and 1,000,000 individuals to cover all plausible values. The prior was log-transformed to allow better estimation of small sizes, and set uniformly between 1 and 6. Tolerance rates were set to 25%, *i.e.* the 250,000 simulations that are the closest to observed data were used to estimate parameters. In all analyses, neural network algorithms used 10 networks and five hidden layers. Parameters were logit-transformed to prevent extrapolation outside prior distributions.

Table S1: Grouping of loci into multiplexes, fluorescent labelling, and genetic diversity statistics calculated for each microsatellite locus. N_A: number of alleles per locus, H_o: observed heterozygosity, H_e: expected heterozygosity or gene diversity, F_{IS}: Wright's inbreeding coefficient, G_{ST} and D: two estimators of population differentiation. From Fourcade *et al.* (submitted).

Locus	Multiplex	Dye	N_A	H_o	H_e	F_{IS}	G_{ST}	*Jost's D*
Crex1	3	NED	24	0.693	0.857	0.191	0.007	0.069
Crex2	3	FAM	24	0.823	0.898	0.083	0.009	0.097
Crex6	1	HEX	34	0.874	0.909	0.039	0.022	0.240
Crex7	1	FAM	21	0.834	0.874	0.046	0.024	0.184
Crex8	1	NED	24	0.832	0.902	0.078	0.003	0.041
Crex9	1	NED	21	0.871	0.906	0.038	0.006	0.070
Crex11	2	NED	33	0.890	0.919	0.031	0.019	0.236
Crex12	2	FAM	24	0.603	0.852	0.292	0.006	0.072
CAM18	2	HEX	16	0.616	0.601	-0.025	0.013	0.021
TG02-120	3	FAM	11	0.325	0.432	0.249	0.040	0.038
TG04-012	3	HEX	9	0.653	0.659	0.008	0.008	0.017
TG04-012a	2	FAM	11	0.241	0.260	0.071	0.035	0.014
TG04-041	1	HEX	12	0.725	0.748	0.031	0.002	0.009
TG05-30	3	HEX	9	0.583	0.544	-0.072	0.009	0.009
TG012-015	1	HEX	13	0.580	0.668	0.132	0.003	0.013

Table S2: Posterior probability of demographic models in each sampling site, and for all data pooled together, inferred by the ABC analysis, according to Fourcade et al. (*submitted*). The selected model is shown in bold and has been used in N_e calculation.

	Posterior probability		
	decreasing	constant	increasing
All data	**0.98**	0.02	0.00
France	**0.85**	0.15	0.00
Germany	**0.98**	0.02	0.00
Czech Republic	0.12	**0.88**	0.00
Poland (north)	**0.73**	0.27	0.00
Poland (south)	0.46	**0.54**	0.00
Poland (east)	**1.00**	0.00	0.00
Latvia	**0.91**	0.09	0.00
Belarus	0.26	**0.74**	0.00
Russia	0.40	**0.60**	0.00

Table S3: Estimates of effective population size (mode of posterior distribution and 95% confidence intervals) inferred from Approximate Bayesian Computing, and mean local census size inferred by field surveys, from Schäffer & Koffijberg (2004) (minimum – maximum estimations), for all populations pooled and for each sampling population separately. In order to provide a fine estimation of small sizes, effective population sizes were estimated on a logarithmic scale (see Figure S2 for full posterior probabilities), the values shown are thus back transformed.

	Effective population size		Census size	
All Populations	385833	(85225-744614)	1650000	(1300000-2000000) [1]
France	94812	(42983-244066)	1150	(1102-1198)
Germany	277179	(123777-732928)	5100	(4000-6200)
Czech Republic	155779	(91755-739527)	3200	(3000-3400)
Poland (north)	66101	(20111-446583)	75000	(60000-90000)
Poland (south)	137631	(83894-723401)	75000	(60000-90000)
Poland (east)	50976	(25787-364012)	75000	(60000-90000)
Latvia	111538	(61370-475166)	64000	(52000-76000)
Belarus	131327	(79565-742067)	85000	(50000-120000)
Russia	150069	(93545-735899)	2500000	(2000000-3000000) [2]

[1] Estimation for the whole European population, from Birdlife International (2012)
[2] Estimation for European Russia

Figure S1: Synthetic climatic predictor, obtained from the first axis of a PCA performed on a set of eight bioclimatic variables from the *Worldclim* project (Hijmans *et al.* 2005).

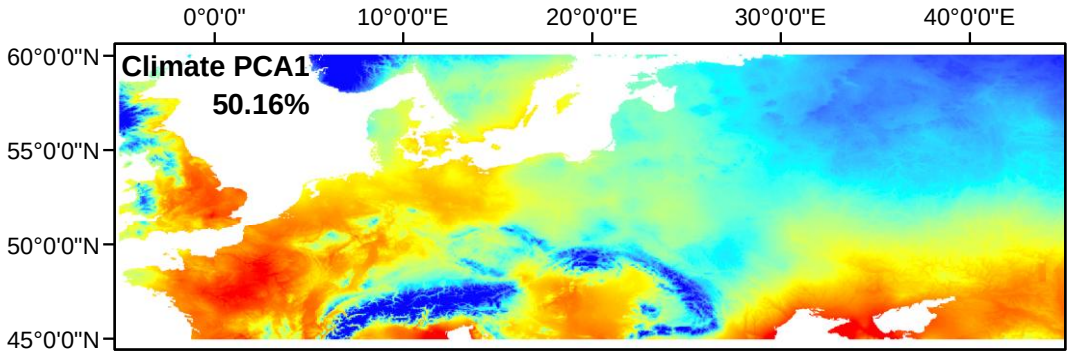
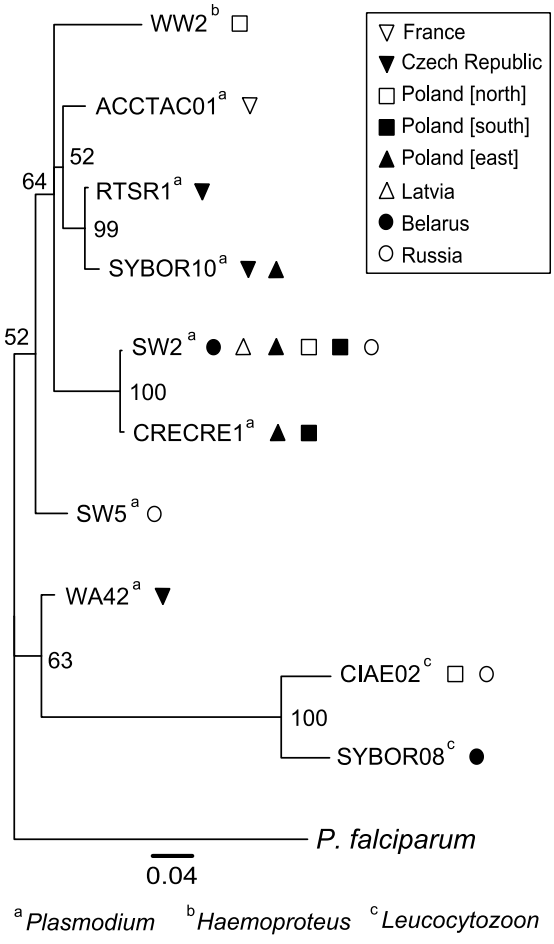


Figure S2: Phylogenetic tree of the ten malaria lineages detected in eight corncrake populations (no malaria infection was detected in Germany) over Europe. *Plasmodium falciparum* is included as an outgroup. The Bayesian support of nodes is placed next to the nodes.



4.1.8. Conclusion Article 5

Les résultats obtenus mettent en évidence un gradient de prévalence allant des populations de l'ouest peu touchées par la malaria (3% en France) aux sites les plus orientaux pouvant rassembler jusqu'à 30% d'individus infectés (Russie). Bien que, comme attendu, les faibles différences de diversité génétique entre populations ne permettent pas d'expliquer le gradient de prévalence observé, il existe une relation très forte entre prévalence et rendements agricoles. Lorsque l'on analyse seulement la prévalence en SW2, la souche de malaria majoritaire dans notre échantillonnage, on observe que la variable expliquant le mieux les variations de prévalence est cette fois-ci la taille des populations hôtes de Râles des genêts, laquelle est encore une fois largement corrélée à l'activité agricole.

Deux hypothèses, non exclusives, permettent d'expliquer cette relation entre malaria et intensité agricole. Tout d'abord, il est possible que l'activité agricole plus intensive en Europe de l'ouest soit responsable d'un déclin des insectes vecteurs de malaria dans ces régions. En effet, une chute des populations d'insectes est observée dans de nombreux pays d'Europe occidentale, en relation avec l'intensification agricole (Conrad *et al.* 2006; Hendrickx *et al.* 2007; Féon *et al.* 2010). L'assèchement des zones humides (Brinson & Malvárez 2002) et l'utilisation intensive de pesticides sont notamment susceptibles d'affecter les populations de moustiques, porteurs de malaria aviaire. En conséquence, il est probable qu'un gradient de densité de vecteurs de malaria existe à travers l'Europe, responsable du gradient de prévalence observé chez le Râle. De la même façon, on sait que les variations d'intensité agricole sont responsables d'un gradient de tailles de populations de Râle des genêts (Green & Rayment 1996). Sachant que la transmission de parasites est souvent densité-dépendante (Dietz 1988), les simples différences de tailles de populations hôtes (Râle mais aussi potentiellement d'autres oiseaux porteurs de malaria souffrant de l'intensification de l'activité agricole) peuvent avoir causé ce gradient longitudinal de prévalence en malaria aviaire à travers l'Europe.

Ce chapitre a ainsi mis en évidence que, même si la malaria aviaire ne semble pas en mesure de poser une menace substantielle sur le Râle des genêts (prévalences relativement faibles), elle révèle l'effet des pratiques agricoles sur les communautés de pathogènes et de vecteurs. En plus de provoquer le déclin visible des espèces d'oiseaux inféodées au milieu agricole et prairial, l'intensification agricole semble également affecter profondément les relations fonctionnelles existant dans ces écosystèmes telles que les interactions hôtes-pathogènes.

Chapitre 4 : Distribution locale et espèce parapluie

5. CHAPITRE 4 : DISTRIBUTION LOCALE ET TEST DU CONCEPT D'ESPECE PARAPLUIE

5.1. ARTICLE 6 : TEST DU CONCEPT DE L'ESPECE PARAPLUIE VIS-A-VIS DES PASSEREAUX PRAIRIAUX EN PAYS-DE-LA-LOIRE

Alors que le déclin global de la biodiversité se poursuit à un rythme soutenu, les budgets alloués aux actions de conservation demeurent largement contraints. Afin de répondre au défi posé par cette situation, développer des approches optimisant le rapport coût-efficacité des actions de conservation devient essentiel. Ainsi, les approches se focalisant sur une seule espèce en tant que « représentante » de son écosystème permettent de résoudre la complexité inhérente à tout système biologique (Simberloff 1998; Caro & O'Doherty 1999). Parmi celles-ci, les espèces indicatrices sont des taxons dont le suivi permet d'avoir une vue d'ensemble simplifiée de la santé d'un écosystème (Noss 1990). Les espèces emblématiques (*flagship species*) sont quant à elles des espèces bien connues du grand public ou des décideurs du fait de leur importance culturelle, et qui peuvent donc attirer l'attention et justifier plus facilement la mise en place d'actions de conservation (Western 1987). Ces espèces sont typiquement des animaux de grande taille bien implantés dans l'imaginaire collectif, auxquels le public peut s'identifier. Cependant, la protection de ces espèces peut s'avérer être d'une portée réduite en dehors de leur propre conservation. Le concept d'espèce parapluie (*umbrella species*) définit alors une espèce dont la protection va assurer la conservation de l'ensemble (ou au moins d'un grand nombre) des espèces vivant dans le même habitat (Roberge & Angelstam 2004). L'efficacité des mesures de conservation basées sur des espèces parapluies ont été fréquemment évaluées *a posteriori*, par exemple en comparant l'abondance ou les tendances démographiques d'espèces censées bénéficier de la protection d'une espèce parapluie (Caro 2003; Roberge & Angelstam 2004; Ozaki *et al.* 2006; Branton & Richardson 2011). En revanche, l'évaluation *a priori* des capacités d'une espèce à agir comme espèce parapluie bénéficie rarement de quantifications rigoureuses. Ce chapitre vise à proposer une méthode d'estimation de l'efficacité d'une espèce parapluie potentielle. Cette méthode sera testée dans le cadre du Râle des genêts et des passereaux prairiaux en Pays-de-Loire.

En France, la politique de conservation du Râle des genêts est largement liée à son statut à la fois d'espèce emblème et d'espèce parapluie. En effet, l'implémentation des mesures agri-environnementales en Basses Vallées Angevines et dans les quelques autres noyaux de population

français est rendu possible notamment grâce au travail d'information sur l'espèce effectué par les acteurs locaux. L'intérêt porté au Râle des genêts par une partie du public et de la profession agricole contribue à la possibilité de mettre en œuvre de mesures contraignantes. Toutefois, alors que les populations continuent de décliner malgré celles-ci, la pertinence du maintien des mesures peut être remise en question, tout particulièrement dans certains sites historiques qui n'accueillent plus d'individus depuis plusieurs années. De plus, les chapitres précédents ont montré que la population française de Râle des genêts ne représente qu'un fragment périphérique mineur de l'espèce quasiment sans caractéristique génétique propre, du moins à l'heure actuelle, et sans doute peu touché par les pathogènes. Dès lors, à l'heure où les contraintes budgétaires fortes obligent à cibler les actions de conservation sur les taxons les plus sensibles, l'énorme effort de conservation du Râle en France pourrait paraître disproportionné, tout spécialement pour une espèce non menacée au niveau global. Cependant, les mesures agri-environnementales, même si elles sont définies par rapport au Râle des genêts, sont considérées comme bénéficiant à tout l'écosystème prairial, faisant du Râle une espèce parapluie. Parmi ces espèces potentiellement bénéficiaires figurent les passereaux prairiaux qui exploitent le même type d'habitat, parmi lesquels le Tarier des prés qui connait également un déclin marqué dans de nombreux pays (Tucker *et al.* 1994). Pourtant, l'utilité de ces mesures centrées sur le Râle des genêts pour la protection des autres oiseaux prairiaux n'a jamais été réellement testée dans les zones concernées.

Une méthodologie combinant trois approches distinctes mais complémentaires a été mise en place pour évaluer le potentiel d'une espèce en tant qu'espèce parapluie. Celle-ci a été testée pour estimer les capacités du Râle des genêts à agir comme une espèce parapluie pour la conservation des passereaux prairiaux en Pays-de-Loire. Dans un premier temps, des variables environnementales ont été choisies et construites afin de représenter de façon pertinente l'habitat des espèces d'intérêts. Le recouvrement des niches écologiques des différentes espèces a alors été quantifié, ainsi que le recouvrement spatial des distributions locales, estimées par une procédure de modélisation de distribution. Enfin, une analyse de la sélection d'habitat liée aux variables paysagères, notamment an ce qui concerne l'évitement des haies, a été menée afin d'évaluer les différences dans l'utilisation de l'habitat à échelle très fine par les différentes espèces. Cette approche doit permettre à la fois de proposer une méthode d'évaluation de l'efficacité potentielle d'une espèce parapluie présumée mais également de confirmer ou d'infirmer le fait que les mesures de conservation Râle-centrées peuvent bénéficier aux passereaux prairiaux.

Ce travail a fait l'objet d'un manuscrit soumis à *Diversity and Distributions* :

Evaluating the potential effectiveness of umbrella species: an approach using species distribution modelling and landscape analysis

Yoan Fourcade*, Aurélien G. Besnard*, Jean Secondi

Université d'Angers, GECCO, 49045 Angers, France

* Equal contribution

5.1.1. Abstract

Aim: The concept of umbrella species has emerged to assist the implementation of simple conservation actions. The rationale is that concentrating resources on the protection of an umbrella species contributes to protect a suite of species and ecological processes belonging to the same ecosystem. The environmental requirement and distribution of the umbrella species should thus overlap those of the group of targeted species. We aimed to build an analytical framework to assess the potential of species to act as an umbrella species, combining techniques of species distribution modelling and more classical habitat selection analysis on fine-scale landscape features.

Location: Western France

Methods: In western France, the conservation of several large grassland floodplains relies on agri-environmental schemes targeting one single bird species, the Corncrake. It is considered as an umbrella species but no real assessment has been carried out so far. We estimated overlaps in the environmental and the geographical space between the main bird species breeding in these grasslands, including the Corncrake and four passerines.

Results: We found that Corncrake distribution partially overlapped the distribution of all other species, and that overall no species would be able to provide enough coverage to be considered as a good umbrella species for the others. Moreover, landscape analyses revealed a much stronger avoidance of hedgerows for passerines than for Corncrake, which also indicate differences in microhabitat use.

Main conclusions: Our study highlights the potential of species distribution modelling to select umbrella species. Such a diagnostic can be effectively completed by analyses of micro-habitat

selection which can reveal interspecific differences at a too fine scale to be detectable by the modelling approach.

5.1.2. Introduction

Implementing efficient management actions for the conservation of ecosytems is a major challenge. This goal is difficult to achieve owing to the complexity of biological diversity even over small areas (Blaustein & Kiesecker 2002). Multi-group monitoring and analysing interactions between processes (ecological or anthropic) is highly resource consuming and difficult to get funded. An alternative option when not all factors and interactions are identified or understood is to focus conservation actions on a single species to protect an ecosystem (Simberloff 1998). Resorting to simple proxies is easier to justify and implement than multi-level approaches. In this regard, various concepts of surrogate single species have been used (Caro & O'Doherty 1999). Flagship species are used to raise public awareness by highlighting negative effects caused by human practices, and ease the acceptance of constraining solutions (Western 1987). Indicator species are simple shortcuts to monitor population trends of several species, taxonomic richness, or the environmental health of a given ecosystem (Noss 1990). It soon appeared advisable to use more operational approaches. In this respect, the definition of the umbrella species concept has been developed (Frankel & Soulé 1981).

The simplest definition of an umbrella species is that its conservation confers protection to a large number of naturally co-occurring species in a given ecosystem (Roberge & Angelstam 2004). Identifying the right species is not straightforward though. An umbrella species mainly requires to be distributed over an area large enough to encompass the largest fraction of the range of as many as possible beneficiary species or groups to be protected (Wilcox 1984). Ideally, its distribution would efficiently delineate a protected area (Caro & O'Doherty 1999). This definition can be extended to any ecological requirements, especially those that are threatened in a given ecosystem (Lambeck 1997), to maintain ecological functions needed by the other species. Thus, a typical umbrella species is a specialist of the target habitat or ecosystem (Caro & O'Doherty 1999) that should have broad enough ecological requirements to overlap those of other species of interest (Seddon & Leech 2008). The umbrella species concept has proven useful in some specific cases but evaluation of its operational efficiency gave equivocal results (Branton & Richardson 2011). By definition, coexisting species have different ecological requirements (Grinnell 1917). For this reason, some studies proposed to consider several species instead of a single one (Lambeck 1997). Threshold values for each environmental parameter of interest are

then defined by the minimal value of the most sensitive species for this parameter. In spite of its limits, the umbrella species concept offers considerable benefits. Since resources available for conservation are limited, it is often a priority to implement sustainable approaches that minimize resources spent and allow long term monitoring. Umbrella species meet this goal but selecting the right species remains challenging and no risk free. Erroneous selection may lead to inadequate public policies if too a restricted fraction of Biodiversity benefits from the protection scheme (Roberge & Angelstam 2004). Therefore, evaluation of the umbrella species as a concept and assessment of its operational efficiency is critical. Recent methodological developments may help to address this issue.

Recently, conservation prioritisation has taken advantage of species distribution modelling (SDM) or ecological niche modelling (ENM) tools. Many conservation-oriented applications have used SDM techniques to predict the expansion of invasive species, estimate range shifts under climate change, or assist in reserve planning (Guisan *et al.* 2013). Although these techniques appear as valuable tool to refine or evaluate the choice of umbrella species, they have been used in that context by very few authors (Estrada *et al.* 2011). However, the resolution of SDM-based suitability maps is often too low to quantify landscape features used for birds to settle territories. Area sensitivity or the proximity to barriers may be detected only at fine geographical scales. Typically, avoidance of wooded edges (Keyel *et al.* 2013) or man-made structures (Wallander *et al.* 2006) by grassland birds occurs below 50 m. Therefore, analysing more precisely the relationships between landscape and bird occurrences may reveal differences in microhabitat use left undetected by classical modelling approaches.

Grassland habitats are one of the most endangered habitat worldwide (Hoekstra *et al.* 2005) and host many threatened species, especially among birds (Fuller *et al.* 1995; Donald *et al.* 2001). The collapse of grassland birds is mainly caused by agricultural intensification (Donald *et al.* 2001). In Western Europe, the Corncrake *Crex crex* has long been seen as a flagship species owing to its dramatic decline over the last decades (Green *et al.* 1997a) which highlighted the need for more protection of grassland ecosystems. It has been eventually considered as an umbrella species and, in France, it has been used to implement agri-environmental schemes (AES) on floodplain meadows. These measures were expected to benefit a larger number of species, including passerines, but the ability of the Corncrake to shield grassland passerines has never been evaluated. More importantly, the actual ecological and spatial overlaps with the other birds have never been studied.

We developed a multi-approach methodology to assess the potential of a putative umbrella species (Corncrake here) to efficiently protect other species of community. Using

specifically designed environmental predictors, we quantified overlaps in the environmental and the geographical space between individual species and the other species. Because landscape structure is one of the most important environmental parameters for grassland birds (Fonderflick *et al.* 2013), we finally assessed interspecific overlaps in habitat selection for two critical landscape features, i.e. meadows and hedgerows.

5.1.3. Methods

5.1.3.1. Study area and species sampling

The study area covers floodplain grasslands on either side of the lower 200 km reach of the Loire River and its main tributaries which was spared from river massive embanking. The main land use is extensive hay meadows (39 009 ha, *i.e.* 75.4 % of total area), generally harvested once a year and grazed afterwards by cattle at low density. Meadows are separated by hedgerows, at various densities, that fragment grassland habitat. Five species dominate the bird community, four passerines (Whinchat *Saxicola rubetra*, Yellow Wagtail *Motacilla flava*, Corn Bunting *Emberiza calandra,* Reed Bunting *Emberiza schoeniclus*), and one rail (Corncrake *Crex crex*). The latter species has undergone a massive decline in most Western European countries (Green *et al.* 1997a), including France. In 2011, around 85% of the remaining French population bred in the study area (Deceuninck *et al.* 2011). AES measures implemented to protect birds mostly consist in delayed mowing after completion of breeding by the Corncrake.

For passerines, we carried out day survey on meadows using binoculars and spotting scopes, and georeferenced all bird occurrences. Almost all grasslands of the study area were visited once so that no sampling bias is expected. Data were mainly collected during the 2009 and 2010 breeding seasons. Additional surveys were carried out in 2011 and 2013. We used corncrake data from the 2006 national survey, the most recent and complete dataset available, which was obtained from *LPO France*. Data were collected by local ornithologists who surveyed the study area by night. Overall we obtained 1602 occurrences of grassland birds: 595 Whinchats, 182 Corn Buntings, 273 Yellow Wagtails, 212 Reed Buntings and 340 Corncrakes.

5.1.3.2. Measurement of environmental overlap between species

We selected 7 environmental predictors to model distributions that were specifically designed for this study area and theses species, and mapped as GIS layers of 250 m resolution. Predictors fell

into four groups: climate (temperature and precipitation), landscape (hedgerow density and proportion of meadow), soil wetness (Topographic Wetness Index TWI and flooding susceptibility derived from the Normalized Difference Vegetation Index NDVI during winter months) and vegetation (Enhanced Vegetation Index EVI during spring months) (Table 12). Methodological details for the selection of predictors and building of GIS layers are given in Appendix S1 in Supporting information.

Table 12: Source of raw data and treatment used to compute the seven environmental variables describing the habitat of the study area

Variable	Source of raw data	Treatment
Climate		
Precipitation	Météo France weather stations	Spatial interpolation of mean March-June data (1980-2009)
Temperature	Météo France weather stations	Spatial interpolation of mean March-June data (1980-2009)
Landscape		
Meadow Percentage	CORELA[1] and digitized from aerial images	Proportion of meadow per 250-m pixels
Hedgerow density	CORELA[1] and digitized from aerial images	Hedgerow length / surface per 250-m pixels
Wetness		
TWI[2]	Topographic layer from BD Alti, IGN[3]	FD8 multiple flow direction TWI algorithm[4]
Winter NDVI[5]	MODIS[6] vegetation indices (MOD13Q1), NASA	Mean value of 5 contrasted flooding levels (winter 2002-2003)[7]
Vegetation		
Spring EVI[8]	MODIS[6] vegetation indices (MOD13Q1), NASA	Mean May EVI across 6 years (2004, 2005, 2008, 2009, 2010,2011)

[1] Conservatoire Régional des Rives de la Loire et de ses Affluents, France
[2] Topographic Wetness Index
[3] Institut national de l'information géographique et forestière, France
[4] see Freeman (1991)
[5] Normalized Difference Vegetation Index
[6] Moderate-Resolution Imaging Spectroradiometer, satellite imaging from the American National Aeronautics and Space Administration (NASA)
[7] Dates of images: 2nd half November 2002, 1st half January 2003, 2nd half January 2003, 1st half February 2003, 2nd half February 2003
[8] Enhanced Vegetation Index

To quantify available habitat, we computed a Principal Component Analysis (PCA) on all pixels of the study area using the values of all predictors. Occurrences were thereafter projected on the environmental space defined by the first two PCA axes to quantify (i) habitat selection of individual species and (ii) the interspecific differences in habitat selection. Significance of habitat selection was assessed by calculating the significance of the marginality of each species, *i.e.* the squared length of the vector connecting the centroid of the available environment to the centroid of the utilized habitat (Hirzel *et al.* 2002). It was computed using the randomization procedure (10 000 permutations) implemented in the "*adehabitatHS*" R package (Calenge 2006).

Environmental overlap between species pair was quantified using the PCA-ENV approach of Broennimann (2012). This method allows the computation of a niche overlap index derived from Schoener's *D* (1968), which ranges from 0 (no overlap) to 1 (identical niches). The environmental space was defined by a grid of 100×100 pixels resolution produced from the ordination scores of the two first axes of the former PCA calibrated on all the pixels of the study area. Niche overlap was then calculated from the smoothed density of occurrences in the environmental space, produced by a kernel density function. We also computed overlaps in the environmental space between one species and the four others merged together. All PCA analyses were performed by the "*ade 4*" R package (Dray & Dufour 2007).

5.1.3.3. Measurement of geographical overlap between species

We built species distribution models (SDMs) to describe the potential distribution of species, using the species occurrences and environmental predictors described previously. We used version 3.3.3k of MAXENT, a machine learning algorithm which predicts the potential distribution of species from presence-only data (Phillips *et al.* 2006). Features settings were set to linear, quadratic and product to obtain interpretable response curves. Bird occurrences were used to train the models and background points were randomly selected across the whole study area (10 000 locations). The whole study area has been surveyed so that we do not expect any sampling bias to affect model outputs (Fourcade *et al.* 2014), and we did not attempt to correct for such a bias.

Model evaluation was performed by calculating the area under the receiver operating curve (AUC), sensitivity, specificity and True Skill Statistics (TSS) using 100 bootstrap replicates. Presence points were randomly split 100 times into a training set (70% of the whole dataset) and a test set (the remaining 30%). For each replicate, a SDM was computed using the training set, and evaluation metrics were obtained using test set, thereafter reported as mean ± SD of 100 replicates. We used the "*PresenceAbsence*" R package (Freeman & Moisen 2008) to compute all evaluation metrics. Variable contributions were assessed by a jackknife procedure which measures at each bootstrap replication the test gain obtained by using a single variable alone. All other analyses were performed using models computed with all occurrences.

Autocorrelation between training and test datasets are known to inflate evaluation measures of SDMs (Veloz 2009). In the bootstrap procedure which consists in splitting the original dataset into training and test points, the two datasets cannot be considered as really independent. Therefore, it is of valuable interest to evaluate the predictions of distribution

models using test points as independent as possible from the training set. We could use data from the survey of the Corncrake population in 1991 (421 occurrences) when its local distribution was more extended. As the ecological variables used here are likely to be changing over time, we could not build an "historical" model to compare with the current distribution of Corncrake. However, we reported for the Corncrake model the same evaluation metrics as described above using these occurrences as test points. Although such evaluation is not available for passerines, it should provide an additional assessment of the robustness of our model

In order to estimate the potential of our study species to perform as an umbrella species, we quantified the spatial overlaps between local distributions predicted by SDMs. Modelled maps were first converted to binary maps using the 10^{th} percentile training presence threshold (Liu *et al.* 2005). We then calculated the percentage, in surface area, of each species' range covered by the modelled range of each other species. Aggregated range maps (including all species but the focal one) were also generated to assess the ability of each single-species SDM to cover the range of the other species.

5.1.3.4. Measurement of fine scale habitat selection

In order to detect differences in microhabitat use at a finer scale than the SDM resolution, we also analysed habitat selection in relation to small-scale landscape features. We first created a dataset of points randomly distributed across the study area (n=1000). Then, in order to describe landscape around each bird occurrence and random points, we created buffers of 6 radius sizes from 50m to 1000m. We extracted from these buffers the percentage of surface occupied by grassland and the density of hedgerow expressed in m/ha. Data extraction was achieved using "*rgeos*" R package (Bivand & Rundel 2013).

We described more precisely the influence of hedgerow proximity on bird presence. First, we extracted for each location the distance between each bird location and the nearest hedgerow, and binned values in 20-m distance classes to obtain bird frequency as a function of distance to hedgerow. The same procedure was repeated for the 1000 random points to get reference values for the available landscape. Raw data for bird occurrences and random points were transformed as a proportion of total occurrences. Second, we calculated for each distance class and for each species the deviation between the observed proportion and the expected proportion of occurrences estimated by the random points. We then modelled the relationship between the deviation and the distance to hedgerow by a non linear regression – a damped sine wave shaped function – with the following equation: $Difference\ vs.\ random = A \times \exp(-\gamma \times distance) \times \cos(\omega \times distance + \varphi) - B$. The unknown coefficients were estimated using the nonlinear

least squares estimate function implemented in R (R Development Core Team 2012). The distance under which the modelled birds' occurrences proportion was lower than the expected was estimated using the "*investr*" R package (Greenwell 2013).

5.1.4. Results

5.1.4.1. Environmental overlap

The two first PCA axes were used to characterise the available environment and accounted respectively for 40% and 18% of the total variance. Visual inspection of birds occurrences projected on the available environmental space of the PCA revealed clear patterns of habitat selection (Figure 28) as occurrence distributions were shifted compared to the centre of the environmental space. This finding was confirmed by the high significance of the marginality vector for the 5 species (all $P < 0.001$). Interspecific differences such as the larger environmental space occupied by the Yellow Wagtail and the shift towards higher values on the Y axis for the Reed Bunting can be observed (Figure 28). Visual assessment is confirmed by Schoener's *D* index of niche overlap (Table S1 in Supporting information). Pairwise comparisons between species ranged from 0.54 (Corncrake *vs.* Reed Bunting) to 0.82 (Whinchat *vs.* Corn Bunting). Analysis of niche overlap between one species and the other species merged datasets revealed that the Whinchat had the highest overlap in environmental space with the other species (0.89). On the contrary, overlap in the environmental space between the Corncrake and the other species was only 0.69 (Figure 30 and Table S1).

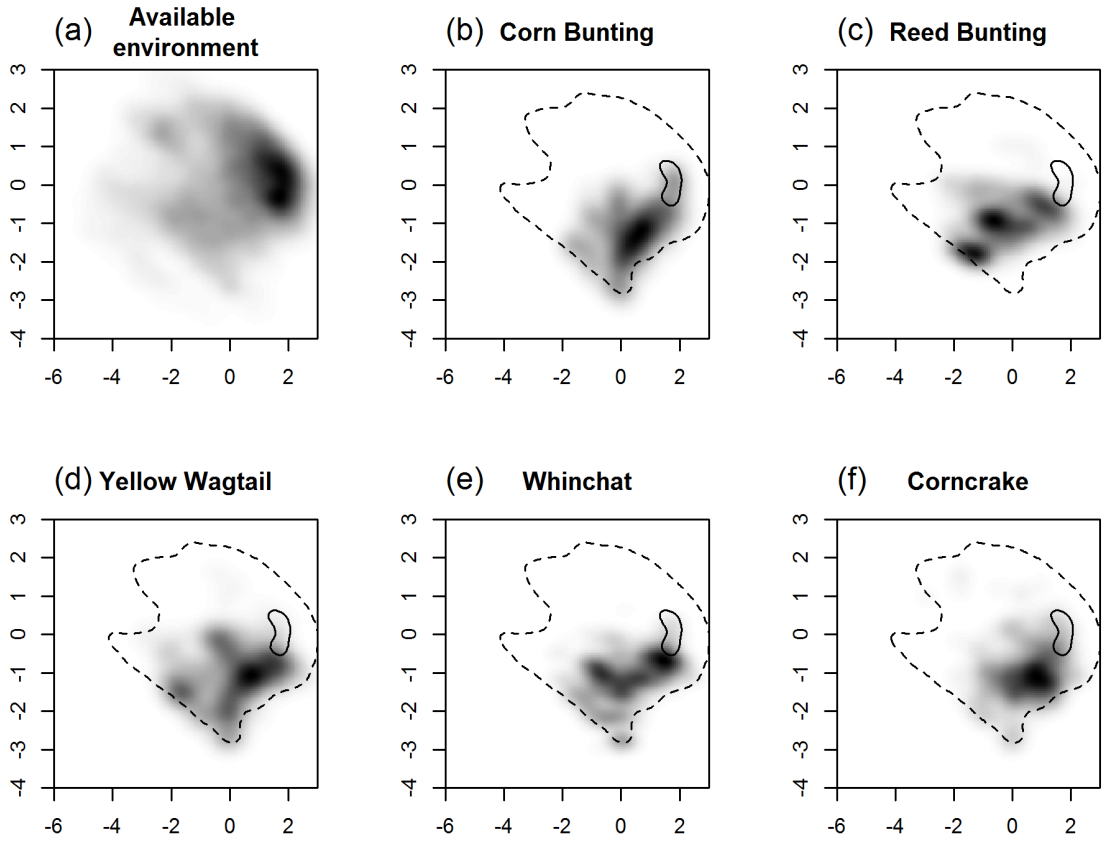


Figure 28: Density of occurrences in the environmental space for each species relative to the available environment. A principal component analysis (PCA) is computed on all pixels of the study area using the 7 environmental variables used afterward for species distribution modelling. (a) points density of the whole environment plotted in grey scale and (b – f) in solid and dotted lines. The occurrences density of the 5 bird species is projected on the PCA plane and plotted in grey scale (b – f).

5.1.4.2. Species distribution modelling

TSS values for all species distribution models (Figure 29) ranged from 0.60 and 0.75, and AUC values ranged from 0.89 to 0.94 (Table S2 in Supporting information). Models can therefore be classified as good to excellent (Araújo *et al.* 2005). Similarly, sensitivity values (0.87–0.89) and specificity values (0.73–0.87) of all SDMs indicated a good fit and differed very little between species. Evaluation measures of the "historical" Corncrake dataset (Table S2) were close to the bootstrap evaluation obtained with the "present" data (TSS = 0.59, sensitivity = 0.82, specificity = 0.77, AUC = 0.88) and revealed a good predictive ability.

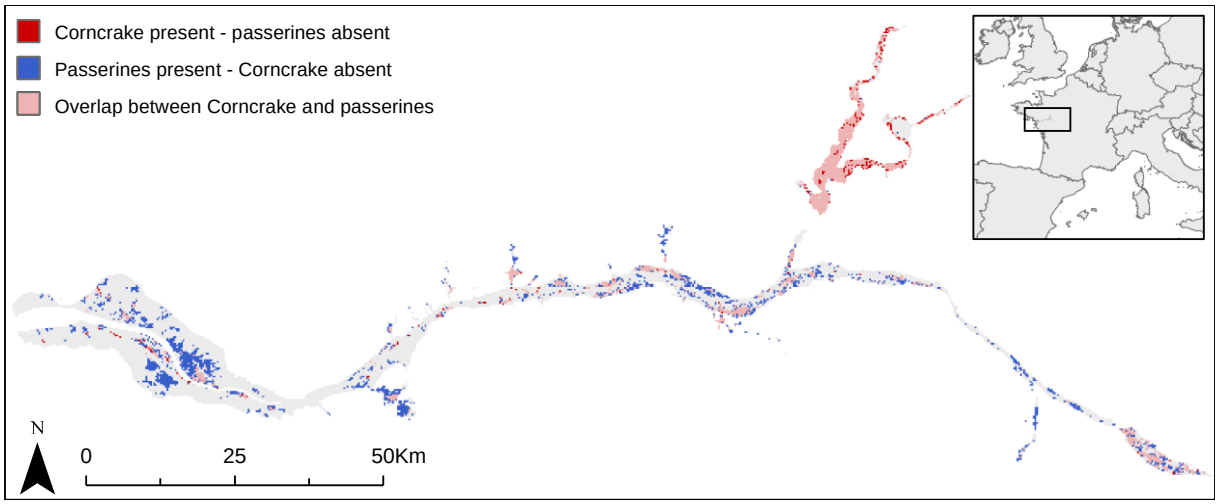


Figure 29: Map showing the similarity between Corncrake modelled distribution and the aggregated model of passerines distribution. The insert shows the location of the study area in western France.

For all passerines SDMs, the environmental predictor with the highest contribution was EVI (mean test gain = 0.82 ± 0.21 SD, Figure 31), followed by percentage of meadow (mean test gain = 0.58 ± 0.13 SD) and precipitations (mean test gain = 0.32 ± 0.11 SD). The next most important variables were hedgerow density (mean test gain = 0.22 ± 0.11 SD) and flooding susceptibility (winter NDVI) (mean test gain = 0.18 ± 0.22 SD). However, the latter had a much higher contribution in the Reed Bunting model (mean test gain = 0.52 ± 0.17 SD) than for the other passerines. The other indicator of wetness (TWI) had little importance in the models (mean test gain = 0.07 ± 0.06 SD respectively), while the temperature had an almost negligible contribution (mean test gain = 0.01 ± 0.03 SD). The Corncrake SDM was slightly different. The most important variable was again the vegetation (EVI, mean test gain = 1.00 ± 0.14 SD), but percentage of meadow (mean test gain = 0.44 ± 0.08 SD) and precipitation (mean test gain = 0.50 ± 0.01 SD) had an equal contribution. Interestingly, the next most important variables were NDVI (mean test gain = 0.23 ± 0.06 SD) and temperature (mean test gain = 0.19 ± 0.06 SD), while the contribution of hedgerow density (mean test gain = 0.06 ± 0.02 SD) was clearly lower than for passerines. Again, TWI had a very weak importance in the model (mean test gain = 0.08 ± 0.04 SD).

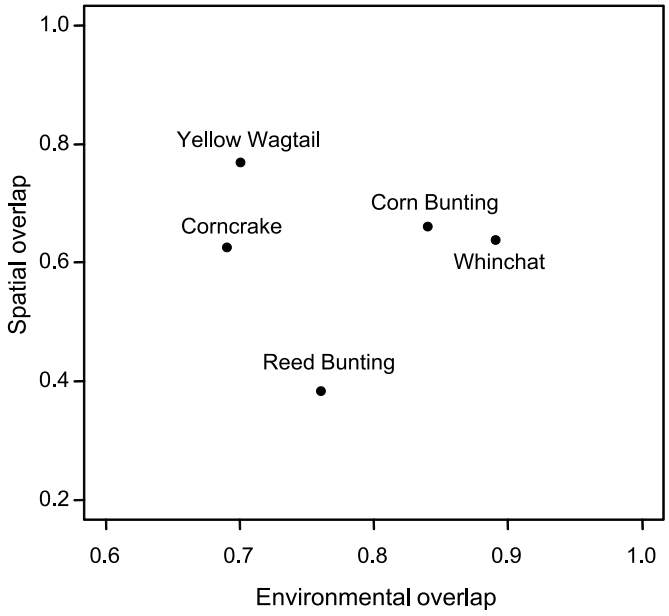


Figure 30: Overlap between each focal species and the merged four other species in the environmental space and the geographical space. Raw data are available in Table S1 and S3.

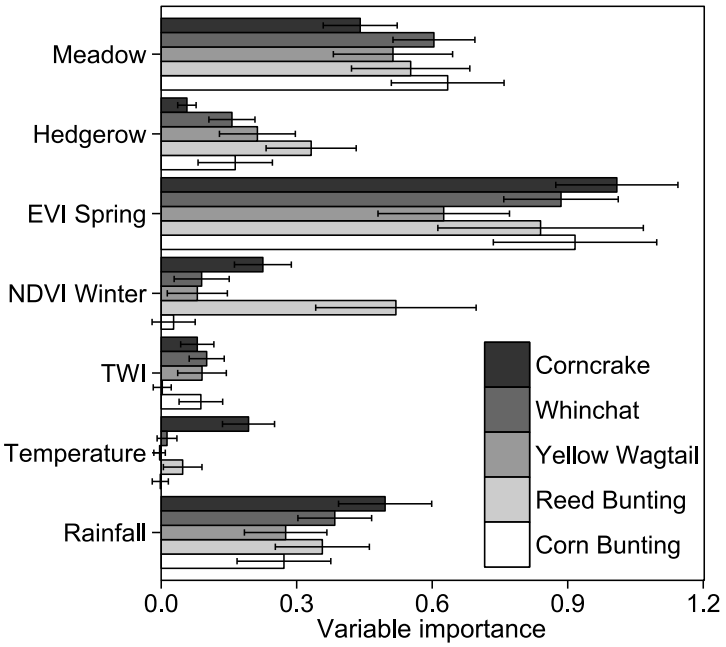


Figure 31: Importance of environmental predictors in MAXENT models based on the test gain with the variable only ± standard-deviation (100 replicates) for the five species distribution models.

5.1.4.3. Spatial overlap

Analysing the overlap between single-species binary maps and the merged distribution of the four other species (Figure 30, Table S3 in Supporting information) highlighted differences in the

potential of species to protect the community of grassland birds. Thus, we showed that the modelled range of the Yellow Wagtail had the highest spatial overlap with the 4 other species considered (77.0%) (Figure 30). On the contrary, the Reed Bunting distribution overlapped only 38.5% of the area potentially occupied by the other species. The two other passerine species, the Corn Bunting and the Whinchat, had intermediate values (66.2 and 63.9% respectively). The Corncrake covered 62.7% of the range of all passerines.

Pairwise spatial overlaps (Table S3) clearly showed that the Reed Bunting had the most divergent distribution compared to the other species since its modelled range covered only 48.9% ± 2.4 SD of all other bird's distribution. However, its own range was covered by the other modelled ranges at 87.3% ± 3.9 SD in average, revealing its limited area compared to the others. On the contrary, the Yellow Wagtail range covered 92.7% ± 1.4 SD of the range of other passerines but only 76.1% of Corncrake SDM was included in its range.

5.1.4.4. Landscape analysis

Percentage of meadows around bird locations was roughly similar to the percentage of meadow available at random locations whatever the spatial scale considered (Figure 32a). At 50 m, mean values ranged from 0.92 (Yellow Wagtail) to 0.98 (Corn Bunting) for bird locations and was 0.98 for random points. At 1000 m, mean value was 0.81 for random points, which was close to Corncrake (0.83), Yellow Wagtail (0.83), Whinchat (0.80) and Corn Bunting (0.82). Only Reed Bunting appeared to preferentially select areas with a higher proportion of meadow at that scale (0.88).

However, birds tended to prefer areas of low hedgerow density (Figure 32b). Buffers around random locations included 84 to 93 m/ha of hedgerow depending on buffer size. For passerines, hedgerow density was lower in buffers around bird locations (50 m: 5.4 – 23.6 m/ha; 1000 m: 51.8 – 62.5 m/ha). Corncrake was less sensitive to hedgerow density than passerines at small scale (50 m: 48.6 m/ha), but still occurred in areas of lower hedgerow density than random points. At larger scales, hedgerow density was similar to values observed for passerines (1000 m: 65.2 m/ha).

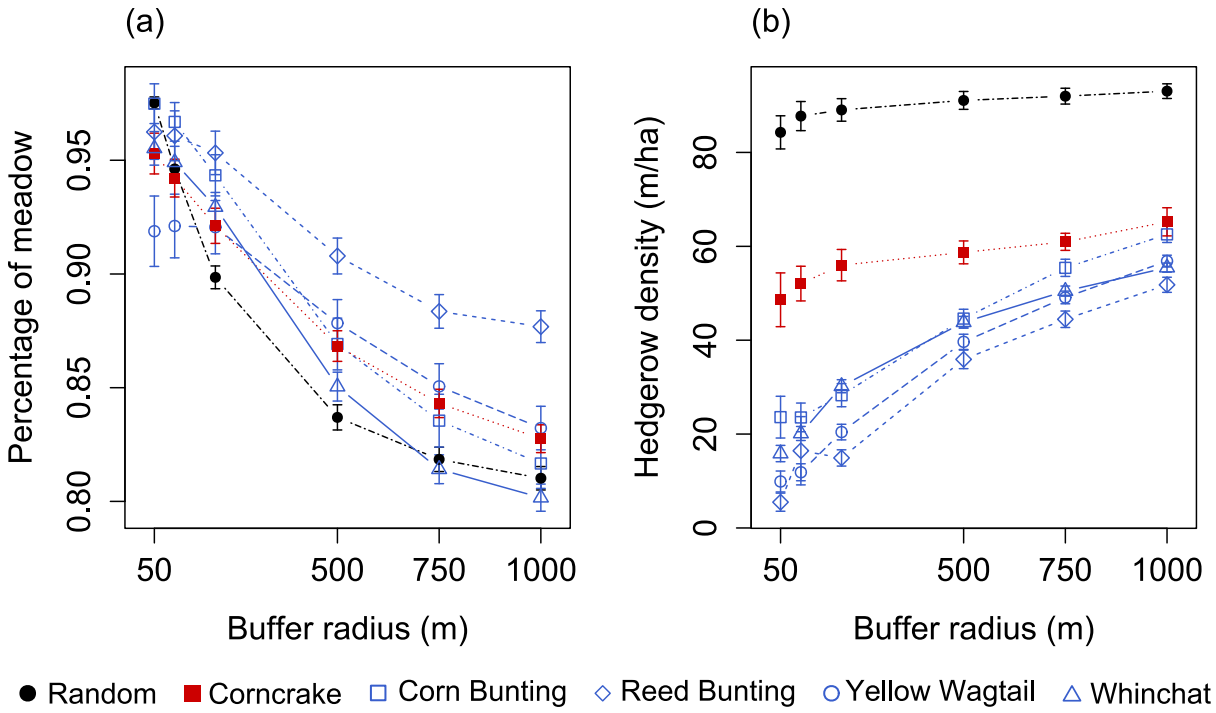


Figure 32: Percentage of meadow (a) and hedgerow density (b) around birds' occurrences and random points (± standard-error), for six buffer sizes ranging from 50 to 1000 m.

All species had a lower proportion of occurrences close to hedgerows than expected by chance (Figure 33). At 20 m, the proportion of Reed Bunting and Yellow Wagtail was 25.0% lower than expected by chance. A similar value was found for the Whinchat (-23.3%). This decrease of occurrence proportion was less pronounced for the Corncrake (-11.0%) and the Corn Bunting (-15.6%). The distance at which the proportion of birds crossed the random proportion of presence varied between 59 m (Corncrake) and 104 m (Reed Bunting). The three other species had intermediate values (Whinchat: 76 m, Corn Bunting: 85 m, Yellow Wagtail: 88 m).

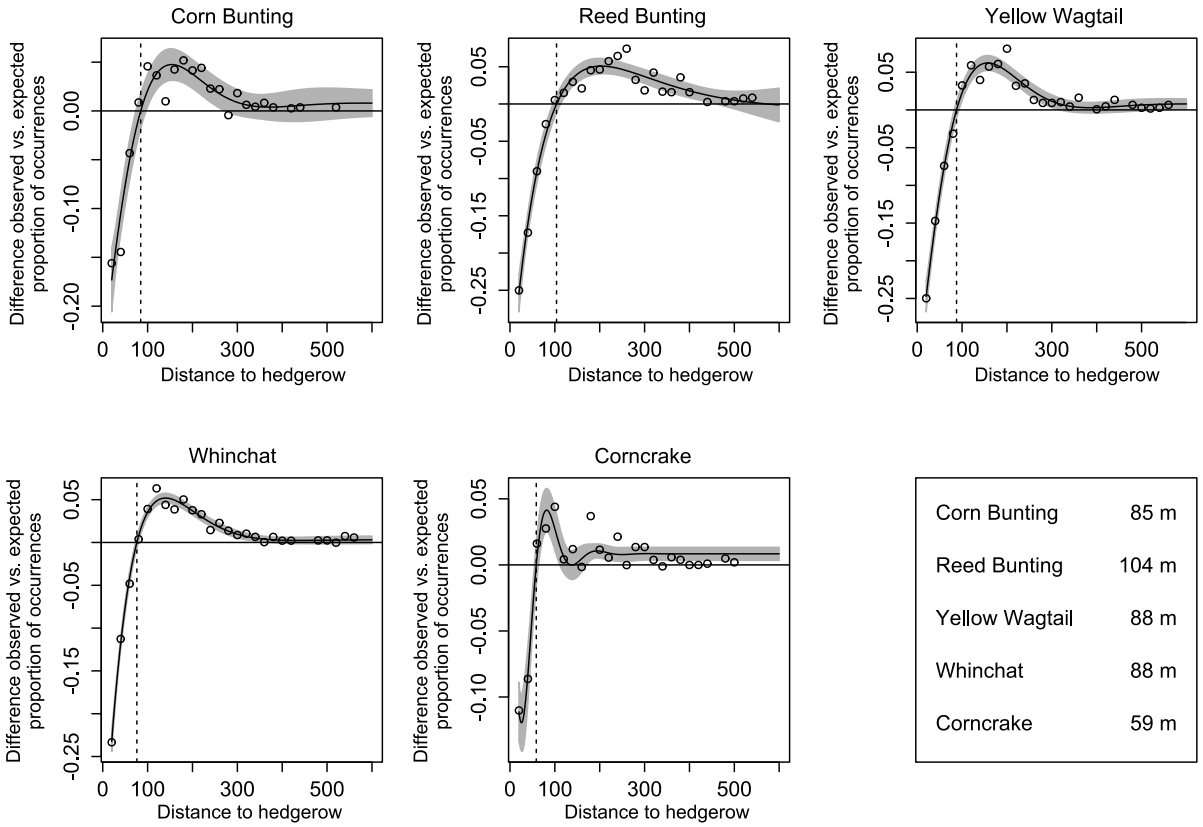


Figure 33: Difference between the observed proportion and the expected proportion (based on 1000 random points laid in grasslands) of bird occurrences plotted against the distance to the nearest hedgerow, for the five bird species. The shaded grey intervals show the fitted non-linear regression and its 95 % confidence intervals. The dotted vertical line shows the distance at which bird occurrences proportion crosses the expected (horizontal line). The bottom-right box indicates this distance for each species.

5.1.5. Discussion

We tested here the potential of the Corncrake to act as an umbrella species for the conservation of the whole community of grassland birds. Such an assessment highly depends on the environmental variables considered, from which will be estimated the ecological niches occupied by each species. We specifically developed a set of local environmental predictors that we selected in accordance with our knowledge of the floodplain ecosystem and the species' ecology in order to depict as accurately as possible local niches.

We detected clear evidence of habitat selection for all five species using a PCA and a marginality analysis. We also evaluated interspecific differences in habitat selection using the method developed by Broennimann (2012). We found that the mean overlap in the environmental space between passerines was 76%. Interestingly, Corncrake's ecological niche was relatively distant from most passerines' niche, especially the Reed Bunting's niche

(overlap = 54%). Overlap with the aggregated niche of all passerines was only 69%. Therefore, the Corncrake may not act as a good umbrella species able to protect the whole range of habitat variation used by grassland passerines. On the contrary, Corn Bunting and Whinchat occupied the largest environmental space and displayed a high overlap with the aggregated niche of the other species (84% and 89% respectively).

Although the analysis of environmental overlaps provides indications on the potential of species to perform as an umbrella, additional insights can be gained by projecting niches onto geographical space (Guisan *et al.* 2013). All SDMs differed in the contribution of environmental variables. Despite the fact that they all highlighted the importance of vegetation productivity (EVI) and percentage of meadow, they also revealed the lower importance of hedgerow density for Corncrake than for passerines, and the affinity of Reed Bunting for wet habitats such as marshes or fens (Gregory & Baillie 1998). The geographical overlaps of binary maps derived from SDMs highlighted again the limited capacity for the Corncrake to be an umbrella species. Indeed, its distribution covered only 63% of the aggregated distribution of all passerines. Based on these data only, the Yellow Wagtail would be a better candidate for the conservation of all grassland birds (77% overlap with aggregated ranges of the other species). Still, overlap remains rather low since it excludes almost one quarter of the distribution of the other target species. It is noteworthy that each species rank differently in terms of overlaps in the environmental and geographical spaces. Regarding spatial overlap, the potential of Corncrake to protect the other species is only outweighed by the Yellow Wagtail. However, the measures of environmental overlap reveal that its niche poorly covers the combined niche of all passerines. Therefore, basing conservation measures on this species only would exclude a large part of the ecological gradient used by other species of interest. Our double approach of overlaps – ecological niche and binary maps derived from SDM – allows considering simultaneously both the area and the ecological features that should be protected.

The landscape analysis provided supplementary evidence of interspecific ecological difference, especially between the Corncrake and the four passerines. First, we observed no clear evidence of habitat selection regarding the proportion of meadow around bird occurrences, except for the Reed Bunting which was preferentially found in areas with a higher percentage of meadows than occurs by chance. Second, all results pointed at an avoidance of hedgerows, a response already suggested in these grassland bird species (Besnard & Secondi 2014). Hedge avoidance may result from elevated predation risk close to these linear structures (Whittingham & Evans 2004). Yet, at small buffer scales, Corncrakes were observed in areas with higher hedgerow

density than passerines. Likewise, the distance to the nearest hedgerow below which fewer birds than expected were found was lower for Corncrake (59 m) than for passerines (from 76 m for Whinchat to 104 m for Reed Bunting). Such findings are in accordance with the contribution of environmental variables in SDMs which revealed a lower importance of hedgerow density for the Corncrake. Therefore, some habitats may be suitable for all birds but avoided by passerines only because of the presence of hedgerow. As a consequence, the area in which passerines really benefit from the protection of Corncrake may be smaller than expected, and therefore the cost-efficiency of measures may be lower than previously thought.

The three levels of analysis – environmental overlaps, geographical overlaps and fine-scale landscape features – suggested that the Corncrake represents a poor umbrella species for the conservation of grassland birds in Western France. Several studies promoted the use of the Corncrake for the conservation of the whole grassland ecosystems (Wettstein & Szép 2003; Boldogh *et al.* 2009). However, even if its high sensitivity to changes in agriculture practices makes it a potential good indicator species for Biodiversity, protecting solely the habitat occupied by this single species would run the risk of omitting areas used by grassland passerines only. For example in the present study area, the wettest areas like marshes are predicted to be occupied mainly by Reed Bunting, but not by the Corncrake. We also showed that basing actions on any of the four passerine species would exclude at least 20% of other species ranges. It is obvious that many levels and components of the ecosystem, like arthropods or plants, were not taken into account here, but this was not our purpose. We rather highlighted the necessity to consider objectively how much of the environmental and geographical space occupied by other species, groups of species or communities of interest could be effectively protected by focusing actions on a single species.

The concept of single-species umbrella has been largely criticized (Lambeck 1997; Simberloff 1998). More complex alternative approaches guild- or community-based have been developed (Lambeck 1997; Possingham *et al.* 2000) that are expected to encompass more ecological processes, offer refined assessment, and improve management efficiency. Despite the added value of including additional requirements in the design of conservation plans, the simplicity of the umbrella species concept allows the use of a limited number of protocols to monitor population trends, making this approach easier to fund and more feasible in many cases. However, one has to consider its limits and try to improve its relevance and efficiency. We proposed here a multi-approach assessment that quantifies the overlap between species or groups of interest in the environmental space but also in the geographical space. Generating robust

distribution maps is crucial to design conservation plans and assess their economic costs. We combined these outputs with a fine scale study of habitat selection, which could identify key ecological constraints at the site scale. So far, distribution models have been used in prioritization of conservation areas (Raxworthy *et al.* 2003; Ortega-Huerta & Peterson 2004) but rarely as a tool to explore a priori the potential of a species to be used as an umbrella species. However, it is important to note that the robustness of small-scale SDMs is dependent on the careful selection of the environmental predictors based on a good knowledge of the ecology of focal species . It may be necessary to build ad-hoc GIS layers, like in the present study, which may be highly time-consuming. Nevertheless such a step is needed to build reliable models and feed the multi-approach proposed here. Such a combination of analyses has a great potential to improve the umbrella species concept by helping to identifying a limited set of species that would optimize protection both in the environmental and the ecological space. The performance of this approach relative to more complex and economically challenging multilevel approaches is to be assessed for each study case though. It may depend on public or authority awareness, opportunity for funding and knowledge of the ecosystem. However, our increasing understanding of ecosystem processes may help to identify this limited set of species that could be selected and sustainably monitored as umbrella species for ecosystem management.

5.1.6. Acknowledgements

The study was funded by *Plan Loire Grandeur Nature*, European Regional Development Fund (ERDF), *Région des Pays de la Loire*, *Agence de l'eau Loire-Bretagne*, *Angers Loire Métropole*, *Direction Régionale de l'Environnement, de l'Aménagement et du Logement* (DREAL) and *Département Maine-et-Loire*. We are grateful to the *Conservatoire Régional des Rives de la Loire et de ses Affluents* (CORELA) for providing us part of environmental data layers. We thank *Ligue de Protection des Oiseaux* (LPO Anjou) and all volunteer ornithologists for Corncrake data, as well as Guillaume Berdin for his preliminary analyses of the landscape data. This project is part of Zone Atelier Loire.

5.1.7. Supporting information

Appendix S1: Details of the selection and generation of the environmental predictors used in the modelling of species distribution

We selected the mean precipitation and temperature calculated from March to June over the past 30 years as **climate** variables during the breeding season of birds. Although climatic variables are known to shape species distribution ranges (Hodkinson 1999; Beaumont *et al.* 2005), these may be important at a more local scale even if the range of precipitations and temperature is here rather limited across the study area. These layers were generated by spatial interpolation across the study area of data collected from 18 (precipitations) and 14 (temperature) weather stations of the national network (Météo France) during the 1980–2009 period. Stations were scattered across the area and seldom in the valleys. Temperature maps were corrected for altitude using a digital elevation model (DEM) with 250m pixel size and 1m altitude resolution (BD Alti® IGN).

The effect of **landscape** features on birds' presence, such as hedgerow density or proximity, is well documented (Hinsley & Bellamy 2000; Siriwardena *et al.* 2011). Therefore, we used two variables describing landscape features. GIS layers of land use (polygons) and hedgerows (lines) were mapped by photo-interpretation based on aerial pictures (BD Ortho® IGN). We derived two predictors: hedgerow density and proportion of meadow. Hedgerow density is the total length of hedgerow divided by pixel area whereas the proportion of meadow is the fraction of a pixel area covered by meadows. Constraints due to frequent flooding make land use very simple in the study area: meadow, crop fields and woodland.

As flood regime and soil moisture are major drivers of grassland habitat in the study area, wetness is expected to be relevant environmental descriptors. **Wetness** was assessed by two independent variables: TWI and NDVI. Topographic Wetness Index (TWI, Beven & Kirkby (1979)) represents the expected accumulation of water on each pixel of a drainage and is derived from a digital elevation model. It has been previously shown to be a predictor of abundance for some grassland passerines (Besnard *et al.* 2013). Second, we computed a layer of Normalized Difference Vegetation Index (NDVI, Goward *et al.* (1991)) using data provided by the MODIS project (http://modis.gsfc.nasa.gov). We computed mean NDVI values across five dates corresponding to contrasting water levels during the largest flood event available for the study area (i.e. winter 2002–2003). NDVI was originally designed as a remote-sensing tool to assess vegetation greenness and productivity (Reed *et al.* 1994). It has been efficiently used since then to detect open water and monitor floods (Xiao *et al.* 2002; Wang *et al.* 2003). Therefore, merging five NDVI images (see Table 12 for dates) at different times of the winter flood provided an estimate of flooding susceptibility of each 250 m pixel in the study area.

Finally, to characterise **vegetation productivity**, we used the Enhanced Vegetation Index (EVI), an improved version of NDVI which benefits from non saturation and lower sensitivity to

atmosphere noise (Huete *et al.* 2002). We computed, using again MODIS data, the mean EVI of May (the peak of the breeding period). We considered the last 10 years but we had to discard images with a too extended cloud cover. Thus, the final layer represented the average EVI values for the years 2004, 2005, 2008, 2009, 2010 and 2011. It has been shown that vegetation productivity, e.g. assessed by EVI, may predict arthropod abundance and species-richness in floodplains (Lafage *et al.* 2013). Thus, vegetation productivity may be a potential indicator of the availability of food resources for bird species. Moreover vegetation productivity can be a proxy of vegetation height that is a major criterion of habitat selection for grassland birds (Berg & Hiron 2012).

Table S1: Overlap in environmental space following the PCA-ENV approach of Broennimann (2012) between the five study species using birds occurrences and seven environmental variables. The last row shows the overlap between one species and the four other species datasets merged.

	Corn Bunting	Reed Bunting	Yellow Wagtail	Whinchat	Corncrake	4 other species
Corn Bunting	--					
Reed Bunting	0.75	--				
Yellow Wagtail	0.72	0.74	--			
Whinchat	0.82	0.79	0.71	--		
Corncrake	0.73	0.54	0.66	0.78	--	
4 other species	0.84	0.76	0.70	0.89	0.69	--

Table S2: Evaluation of species distribution models computed for each species, using the Area Under the receiver operating Curve (AUC), sensitivity, specificity and the True Skill Statistics (TSS). Evaluations are reported as the mean and standard-deviation of the 100 bootstrap replicates. Evaluation of the Corncrake model was also calculated using old occurrences (1991) as test points.

Species	AUC		Sensitivity		Specificity		TSS	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Corn Bunting	0.91	0.03	0.88	0.06	0.78	0.04	0.65	0.08
Reed Bunting	0.94	0.02	0.88	0.06	0.87	0.03	0.75	0.06
Yellow Wagtail	0.89	0.03	0.87	0.06	0.73	0.04	0.60	0.08
Whinchat	0.92	0.02	0.89	0.04	0.79	0.02	0.68	0.04
Corncrake								
Replicates	0.91	0.02	0.89	0.04	0.77	0.04	0.66	0.05
Old data	0.88	--	0.82	--	0.78	--	0.59	--

Table S3: Spatial overlap between single-species models and between single-species models and the 4 other species merged. The values represent the percentage (in modelled range surface) of species in columns covered by the species in rows. Therefore, the potential of a species to behave as an umbrella species corresponds to the last column, *i.e.* the percentage of the range of other species covered by a single one. The underlying models were converted to binary maps following the 10th percentile training presence threshold.

Overlapping distribution	Covered distribution					
	Corn Bunting	Reed Bunting	Yellow Wagtail	Whinchat	Corncrake	**4 other species**
Corn Bunting	--	83.3	78.5	93.9	70.6	**66.2**
Reed Bunting	49.4	--	45.5	51.3	49.3	**38.5**
Yellow Wagtail	93.0	91.1	--	93.9	76.1	**77.0**
Whinchat	89.9	82.9	75.8	--	72.5	**63.9**
Corncrake	70.7	83.4	64.3	75.8	--	**62.7**
4 other species	97.2	95.6	87.8	99.6	83.2	--

5.1.8. Conclusion Article 6

Les résultats obtenus indiquent que le Râle des genêts semble constituer une espèce parapluie de faible valeur pour la conservation des passereaux prairiaux. En effet, la distribution de l'espèce en Pays-de-Loire ne recouvre qu'à peine plus de 60% de celle des passereaux. De même, des différences majeures avec la niche écologique de ces derniers ont été observées, notamment avec le Bruant des roseaux qui occupe des habitats plus humides. Globalement, l'étude n'identifie pas d'espèce remplissant très efficacement le rôle d'espèce parapluie. Enfin, à l'échelle locale les résultats mettent en évidence une sélection d'habitat sensiblement différente entre Râle des genêts et passereaux vis-à-vis des variables paysagères. Ces derniers tendent en effet à éviter la proximité des haies de façon plus marquée que le Râle. En conséquence, une partie des surfaces occupées par le Râle des genêts risque de ne pas être adéquate pour l'installation des passereaux prairiaux, diminuant l'aire de protection réellement favorable à l'ensemble de la communauté.

La Bergeronnette printanière, plus généraliste, recouvre la plus large part de la distribution des autres espèces testées, mais occupe une niche relativement distante. On n'observe finalement pas d'espèce parmi les passereaux susceptible de fournir une protection suffisante vis-à-vis des autres espèces d'oiseaux prairiaux. Le concept d'espèce parapluie ne semble donc pas pouvoir s'appliquer dans cette situation. D'une manière générale, la mise en place de mesures de conservation centrées sur une seule espèce a été largement critiquée dans la littérature et l'utilisation de plusieurs espèces cibles a été conseillée (Lambeck 1997). Les résultats présentés ici tendent à confirmer qu'une approche globale de la communauté permettrait de conserver plus efficacement l'écosystème prairial.

Toutefois, dans un contexte budgétaire tendu, la mise en œuvre d'actions de conservation ciblant chaque espèce potentiellement menacée à moyen terme relève de l'utopie. Dès lors, il est souvent indispensable de cibler des espèces prioritaires, soit parce qu'il est urgent de stopper leur déclin, soit parce qu'on estime que leur préservation est encore possible. Dans cette optique, évaluer comment la protection d'une espèce peut bénéficier à un groupe d'autres espèces d'intérêt peut permettre de maximiser l'efficacité des actions. La méthode proposée ici, combinant analyse des niches écologiques, des distributions, et de l'utilisation du micro-habitat, offre une opportunité d'identifier ces espèces parapluies potentielles. Elle présente également une application des méthodes de modélisation de distribution dans un contexte où ces dernières ont été peu utilisées.

Discussion générale

6. DISCUSSION GENERALE

6.1. APPORTS A LA STRATEGIE DE CONSERVATION DU RALE DES GENETS

Les différentes approches développées au cours de ce travail ont permis de renforcer les connaissances sur l'écologie du Râle des genêts. La variété des approches et des échelles d'analyses ont mis en évidence les patrons de diversité et des processus agissant à une échelle continentale. Les connaissances acquises jusqu'alors dérivaient majoritairement de suivis très locaux et ne permettaient pas de dégager une vision large de la diversité des populations de Râles des genêts. Pour preuve, les tendances démographiques observées dans la plupart des populations d'Europe de l'ouest (Hudson *et al.* 1990; Broyer 1991) avaient initialement entrainé le classement de l'espèce comme globalement « menacée » (Birdlife International 2013). L'ouverture de collaborations internationales plus poussées entre experts occidentaux et de l'ancien bloc soviétique depuis les années 1990 (par exemple (Broyer *et al.* 2007)) a révélé la distribution extrêmement large de l'espèce en Europe de l'est et en Asie (Sibérie, Kazakhstan) et la situation bien plus favorable de ces populations qu'initialement évalué (Schäffer & Koffijberg 2004; Mischenko 2008). De la même façon, cette thèse a apporté un certain nombre de connaissances nouvelles autorisant la mise en perspective des actions et statuts de conservations locaux relativement aux processus globaux affectant les populations de Râles des genêts. Ces apports ont l'opportunité de bénéficier à la stratégie de conservation globale comme locale de l'espèce.

6.1.1. Distribution et déséquilibre des connaissances et des actions

L'un des premiers résultats de ce travail a été de modéliser la distribution globale du Râle des genêts. De façon quelque peu inattendue, ce chapitre a permis de renforcer les connaissances actuelles, acquises par la compilation d'avis d'experts et par extrapolation d'un nombre limité de suivis de population (Sukhanova & Mischenko 2003; Birdlife International 2012). La procédure de modélisation, tout comme la distribution estimée à dire d'experts, montre ainsi que le Râle des genêts est une espèce présente dans une grande part du Paléarctique. En conséquence, il s'avère que la très grande majorité des effectifs sont présents dans les pays de l'ex-Union Soviétique, principalement la Russie (de l'ouest jusqu'au centre de la Sibérie) mais également le Kazakhstan

(excepté sa partie la plus désertique), l'Ukraine ou encore les pays baltes. Pourtant, ces zones ont été très largement négligées dans la problématique de conservation du Râle.

Les mesures de conservations mises en place pour la protection de l'espèce le sont essentiellement par l'intermédiaire de programmes européens : plan d'action Européen (Crockford *et al.* 1997) décliné en versions nationales (Hennique *et al.* 2013) et locales, mesures agri-environnementales financées par la politique agricole commune européenne. La part des effectifs de l'espèce présents dans l'Union Européenne (UE) n'est pourtant que de l'ordre de 7-8% (Schäffer & Koffijberg 2004). A l'inverse, du fait de tendances démographiques plutôt positives (Keišs 2003), l'espèce est largement ignorée d'un point de vue conservation hors-UE. Toutefois, les connaissances sont encore largement limitées à la partie européenne de la distribution du Râle des genêts. L'état des populations en Asie est ainsi quasiment inconnu. Dès lors, il existe un **déséquilibre majeur dans la distribution des connaissances et des budgets** attribués à sa conservation par rapport à la distribution réelle de l'espèce. Or, même si l'on considère que les enjeux de conservation du Râle des genêts sont concentrés en Europe de l'ouest, ce manque d'attention sur toute la partie centrale et orientale de sa distribution pose au moins trois problèmes majeurs :

(i) **Les habitats occupés par le Râle en Europe de l'est sont en grande partie temporaires**. En effet, des augmentations de population ont été identifiées dans des anciennes parcelles agricoles abandonnées du fait de la déprise agricole qui a suivi la chute de l'URSS (Keišs 2003, 2005). A terme, ces friches agricoles favorables à la reproduction du Râle sont vouées à s'enfricher de plus en plus jusqu'à devenir inexploitables par les individus, ou au contraire à retourner à leur fonction agricole, avec les même effets qu'en Europe de l'ouest. Il semble donc que les tendances positives observées dans ces pays ne soient que temporaires et soient destinées à s'inverser à moyen terme.

(ii) **L'activité agricole n'est pas figée et risque de s'intensifier dans les zones autrefois épargnées**. Les variations dans la situation de conservation (tendances démographiques et tailles de populations) sont essentiellement dues aux variations d'intensité des pratiques agricoles. En effet, le développement économique des pays d'Europe de l'ouest a contribué, tout particulièrement depuis la fin de la 2ème guerre mondiale, à une augmentation des rendements agricoles en général et à une exploitation plus intensive des prairies de fauche. Or, l'augmentation de la population humaine et de la consommation alimentaire entraine des besoins toujours croissants

en production agricole (Foley *et al.* 2011). Il parait donc hautement probable que l'intensification de l'agriculture se poursuive, tout particulièrement dans les pays qui ont connu un développement limité jusqu'à maintenant. Par exemple, l'entrée dans l'union européenne des anciens pays communistes d'Europe de l'est s'accompagne d'un développement de leur agriculture, et par là même d'un déclin des populations d'oiseaux (Sanderson *et al.* 2013). De même, de nombreuses zones prairiales de Sibérie ou du Kazakhstan sont en cours de conversion à l'agriculture intensive, entrainant le déclin des oiseaux inféodés à ces milieux (Kamp *et al.* 2011). Ce basculement des régions steppiques vers l'agriculture intensive est d'autant plus susceptible de se produire que le réchauffement climatique tend à rendre ces zones de plus en plus propices à l'activité agricole. Le gradient d'intensité agricole de l'ouest vers l'est du Paléarctique encore actuellement observé pourrait ainsi s'atténuer voire s'inverser à terme. L'avenir des populations de Râles actuellement considérées non prioritaires pourrait ainsi être compromis dans l'avenir. Alors que les évaluations IUCN s'intéressent aux tendances prévues pour une fenêtre de temps relativement réduite (3 générations) (IUCN 2005), l'évolution de l'occupation du sol et de l'activité agricole est encore à trop long terme pour attirer l'attention.

(iii) Enfin, quelles que soient l'évolution de la situation de conservation de l'espèce dans l'avenir, il est probable que le caractère limité **des suivis de populations dans les zones les plus concernées** par les changements projetés empêche de s'en rendre compte. En effet, alors que les Râles des genêts sont recensés individuellement (du moins en ce qui concerne les mâles) dans plusieurs pays d'Europe occidentale, les populations est-européennes et surtout asiatiques restent très peu suivies. Cela n'est pas illogique compte tenu de la surface considérable de la région considérée et du fait qu'*a priori*, ces populations sont pour le moment non menacées, mais la faible résolution des informations peut s'avérer préjudiciable à l'aune des changements globaux en cours. Par exemple, la présence de l'espèce dans des régions telles que le Caucase, la Sibérie centrale ou la Chine occidentale n'est réellement attestée que par des rapports naturalistes ou des observations ponctuelles (exemple Ming & Qishan 2002). En Sibérie occidentale, quelques points de comptages réguliers ont été mis en place (Sukhanova & Mischenko 2003; Mischenko 2008), mais leur représentativité à l'égard de l'ensemble du Paléarctique oriental parait bien faible. De ce fait, les changements à l'œuvre dans ces populations, qui pourtant représente une grande

proportion de l'espèce, peuvent fort bien passer inaperçus pendant un moment significatif, et de là retarder l'application d'éventuelles mesures conservatoires.

6.1.2. Conséquences du flux de gènes

Au-delà des notions de distributions, et des questions qu'elles soulèvent vis-à-vis des changements à venir, cette thèse a permis d'apporter des précisions majeures en ce qui concerne la **structuration des populations** de Râles des genêts et les relations qu'elles entretiennent. On observe en effet, *via* la génétique, des différenciations très limitées entre sites pourtant éloignés de plusieurs milliers de kilomètres. En règle générale, la bibliographie faisait état d'une fidélité importante des mâles aux sites de reproduction d'une année sur l'autre (Green 1999). Toutefois, des déplacements à relativement longue distance ont également été observés au cours d'une saison, suggérant que des migrations entre sites de reproduction pouvaient avoir lieu (Keišs *et al.* 2004; Mikkelsen *et al.* 2013). Le taux de retour particulièrement faible des individus bagués, notamment dû à une forte mortalité, offre peu d'opportunités d'étudier précisément la **dispersion** du Râle des genêts (Green 2004). La génétique permet ici de s'affranchir de ce problème, chaque individu échantillonné fournissant une information, bien qu'en contrepartie elle ne permette pas de décrire précisément les processus de dispersion à l'œuvre. Dans notre étude à l'échelle européenne, la faiblesse de la structure génétique confirme donc les capacités de dispersion fortes entre sites de reproduction. Il est particulièrement notable que toutes les populations situées de l'Allemagne à la Russie présentent un profil génétique semblable. Il semble donc que l'on soit sans ambiguïté possible dans la cadre d'une seule population génétique où la dispersion des individus maintient un groupe génétique uniforme. Cela n'est d'ailleurs pas incompatible avec l'hypothèse d'une certaine fidélité au site, la dispersion pouvant ne concerner au final chaque année que quelques individus par site, mais sur une gamme de distances suffisamment large pour maintenir l'homogénéité génétique de la population. Cela ouvre donc la perspective à des mouvements à longue distance entre sites, par exemple dans le cadre de reports d'individus sur de nouveaux sites lorsque le premier subit des aléas climatiques (inondation) ou après sa destruction par l'activité agricole. Cela met en lumière l'importance considérable de prendre en compte, pour toute action de conservation, le fait que les individus observés à un endroit n'y sont pas nécessairement inféodés. Ces derniers peuvent en effet aller et venir entre sites de reproduction, y compris entre des pays relativement éloignés. Toute mesure de conservation, ou au contraire dégradation de l'habitat, intervenant sur un site donné peut alors avoir des répercutions plus larges.

Alors que l'on observe une homogénéité importante des populations de Râles des genêts à l'échelle européenne, on note toutefois une certaine signature de différenciation des populations françaises et européennes. Cette **différenciation génétique, même si extrêmement faible**, était attendue au vu de d'un certain nombre d'éléments qui suggéraient que les populations occidentales pouvaient présenter des caractéristiques propres : (i) les individus retrouvés en France ou en Ecosse sont en moyenne plus gros que ceux d'Europe de l'est (Keišs *et al.* 2004 et article 4), (ii) on a pu observer la migration de quelques individus écossais en Afrique de l'ouest (Green 2013), alors qu'on considérait jusque-là que l'hivernage du Râle se déroulait dans le sud-est de l'Afrique, (iii) les dates d'arrivée sur les sites de reproduction sont en moyenne plus précoces en Europe de l'ouest par rapport aux sites de l'est (communication personnelle de K. Koffijberg et O. Keiss). Bien qu'on ne puisse pas totalement exclure l'hypothèse que la structure génétique observée ait été très limitée depuis toujours, nos résultats tendent à montrer que le déséquilibre démographique qui s'accentue entre Europe de l'ouest et de l'est est responsable d'un flux de gènes asymétrique qui masquerait progressivement les caractéristiques génétiques des populations occidentales. Dans ce cas, se pose la question des conséquences de cette homogénéisation en cours. On peut poser deux hypothèses, non exclusives mais aux conséquences opposées en terme de conservation :

(i) Tout d'abord, ce flux de gènes important a permis de **maintenir une diversité génétique élevée** dans l'ensemble des populations, y compris celles qui ont connues un déclin important durant les dernières décennies. La théorie prédit que le goulet d'étranglement causé par la réduction rapide des effectifs entraîne une réduction de la diversité génétique, en éliminant des allèles lors de la réduction du pool génique, puis progressivement du fait de la dérive génétique plus rapide dans les petites populations (Frankham 2005). Cette perte de diversité serait susceptible d'avoir des conséquences délétères en favorisant la dépression de consanguinité qui peut réduire la survie ou la reproduction des individus (Keller 2002; Wright *et al.* 2007). Ici, le flux de gènes venant des populations abondantes d'Europe de l'est, voire plus loin, a probablement permis de compenser les chutes de populations à l'ouest en maintenant la diversité génétique par l'apport constant de nouveaux allèles. Les plus petites populations qu'on aurait pu croire menacées par des facteurs génétiques en plus du déclin des effectifs, bénéficient ainsi des capacités de dispersions importantes de cette espèce.

(ii) A l'inverse, le flux de gènes venant de l'est pourrait contribuer à la **perte d'adaptations locales** indispensables à la persistance des populations (Allendorf *et*

al. 2001; Stockwell *et al.* 2003). Le maintien d'adaptations locales malgré un flux de gènes important pose en effet question (Lenormand 2002). S'il se confirme que les populations occidentales utilisent une voie migratoire et une aire d'hivernage différentes, il est possible que ce comportement soit déterminé génétiquement. Si c'est le cas, les individus hybrides ou les migrants venant d'une population où la migration s'effectue différemment pourraient être incapables de compléter leur cycle migratoire (Veen *et al.* 2007). Outre la migration, on peut imaginer que les populations d'Europe de l'ouest possèdent des adaptations aux conditions climatiques locales, étant données qu'elles sont situées en limite d'aire de distribution. Quoi qu'il en soit, cette hypothèse reste purement spéculative et nécessiterait une investigation plus poussée afin de déterminer s'il existe encore des adaptations locales dont la perte pourrait affecter encore plus la *fitness* des populations déjà les plus menacées.

6.1.3. Zones d'ombre

Bien que l'objectif de ce travail fût d'apporter une vision plus globale à la problématique de conservation du Râle des genêts, il reste des parties de sa distribution dans lesquelles on ignore toujours largement le statut de l'espèce. Les analyses génétique et parasitaire menées au cours de cette thèse se sont en effet concentrées sur la partie européenne de la distribution du Râle, notamment du fait de la difficulté de mettre en place des campagnes d'échantillonnages dans les zones sibérienne et d'Asie centrale. Pourtant, les patrons de diversité observés à l'échelle européenne pourraient apparaître sensiblement différents lorsque l'on considère **l'ensemble de l'aire de reproduction de l'espèce**. On ignore notamment si l'absence de structure génétique constatée dans la population ouest-européenne est révélatrice d'une population panmictique couvrant l'ensemble du Paléarctique, ou s'il existe une structure génétique à plus large échelle. Parmi les espèces possédant sensiblement la même distribution, il a été observé une grande variété de structures génétiques, allant d'une structure longitudinale en deux groupes est et ouest (Gobemouche nain *Ficedula parva* ou Alouette des champs *Alauda arvensis*), à la panmixie la plus parfaite (Chevalier guignette *Actitis hypoleucos*), en passant par des structures beaucoup plus complexes (Bruant des roseaux *Emberiza schoeniclus*) (Zink *et al.* 2008). Malgré les capacités de dispersion apparemment importantes du Râle des genêts, la structure des populations à une si grande échelle reste inconnue. De même, l'analyse des pathogènes sanguins qui montrait un gradient d'infection est-ouest avec des prévalences relativement faibles pourrait donner des

résultats différents dans la partie orientale de la distribution du Râle, caractérisée par un climat, une occupation du sol et potentiellement des souches de pathogènes différents de l'Europe.

D'autre part, l'ensemble du travail présenté ne concerne que l'aire de reproduction du Râle des genêts. Or, au moins la moitié de son cycle de vie se déroule soit en **migration**, soit dans son **aire d'hivernage** africaine. Jusqu'alors, la distribution exacte de l'aire d'hivernage de l'espèce reste en partie inconnue. Les connaissances acquises se fondent essentiellement sur des observations ponctuelles d'individus et de très rares contrôles en Afrique d'individus bagués (Walther 2008). A l'aide de ces données, une modélisation de la distribution de Râle en période d'hivernage a été publiée (Walther *et al.* 2012), confirmant la prédominance des prairies d'Afrique du sud et du Zimbabwe dans l'habitat africain de l'espèce. On notera toutefois que cette distribution modélisée ne prédit pas la présence du Râle en Afrique occidentale, pourtant confirmée par des données de géolocalisateurs (Green 2013), au moins pour des individus écossais. D'une manière générale, il apparait que l'habitat d'hivernage et les couloirs de migration du Râle des genêts présentent encore de nombreuses incertitudes. De la même façon, alors que l'on considère le déclin de certaines populations de Râles comme résultant largement de facteurs agissant sur son aire de reproduction, l'existence de menaces également en aire d'hivernage reste largement inconnue et peu évaluée. Bien qu'aucune menace affectant le Râle en Afrique n'ait été identifiée pour le moment (Stowe & Green 1997b), les zones prairiales d'Afrique du sud sont soumises à des changement d'occupation du sol, par l'extension des zones urbaines et agricoles (Neke & Plessis 2004), qui pourraient à terme détériorer l'habitat du Râle en période d'hivernage. De plus, même si encore une fois cela n'est pas décrit à l'heure actuelle comme une menace majeure, on sait qu'il existe une mortalité d'origine anthropique relativement importante lors de la migration, par une pression de chasse importante sur la côte égyptienne (Baha El Din *et al.* 1996) et par l'observation de collisions avec les lignes électriques au Moyen-Orient (Shobrak 2012). A cela s'ajoute les effets potentiels du changement climatique qui perturbe la migration de nombreuses espèces d'oiseaux (Jenni & Kéry 2003; Both *et al.* 2006).

6.1.4. Conservation du Râle des genêts en France

L'approche globale développée au cours de ce travail permet également de dégager des enseignements pour la conservation du Râle des genêts à un niveau plus local, et notamment en France. Dans un premier temps, les précisions apportées quant à la distribution globale du Râle permettent de remettre en perspective la situation de la population française par rapport à

l'espèce prise dans sa globalité. En effet, ainsi que le prévoyaient les estimations à dire d'experts de l'IUCN (BirdLife International 2004), la modélisation met en lumière que les populations focalisant l'attention et les actions de conservation, et notamment la France, ne représentent que quelques *patchs* d'habitats périphériques, relativement négligeables par rapport à la distribution de l'espèce dans son ensemble. En conséquence, on s'attend à ce que les tendances démographiques toujours négatives observés en France n'aient aucune incidence sur la dynamique globale de l'espèce. De même, du fait de sa **situation ultrapériphérique** (très au sud et à l'ouest), il est probable que la population française soit naturellement extrêmement vulnérable aux **changements climatiques** (Parmesan & Yohe 2003; Jenouvrier 2013). Dans ce contexte, il est même envisageable que les conditions climatiques ouest-européennes deviennent inadéquates pour le maintien de l'espèce dans les années à venir, quelle que soit par ailleurs l'évolution des mesures de conservation (Figure 34).

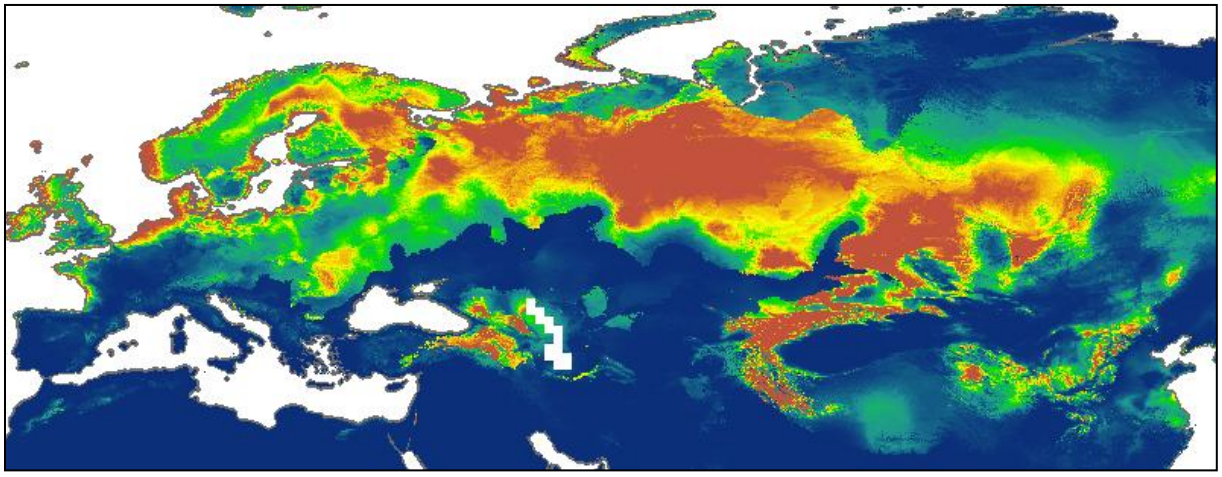


Figure 34 : Modélisation MAXENT de la distribution du Râle des genêts en 2080 sous l'hypothèse du scénario d'émission A2a (prédiction la plus pessimiste) avec le modèle de circulation HadCM3.

Par ailleurs, la réduction de ¾ des effectifs français de Râles des genêts en quelques décennies pouvait faire craindre une diminution sévère de la diversité génétique locale, avec des conséquences potentielles sur la *fitness* des individus. L'une de ces conséquences attendues était une susceptibilité accrue aux pathogènes (Hawley *et al.* 2005). Les données récoltées lors de la thèse, au niveau génétique et parasitaire, montrent que ce scénario n'est pas survenu. En effet, on a montré que les différences démographiques associées à une bonne capacité de dispersion entrainent un flux de gènes asymétrique est-ouest qui maintient la **diversité génétique** des populations les plus déclinantes. C'est le cas de la population française qui présente une

hétérozygotie moyenne de 66%, la mettant à l'abri des problèmes liés à la dépression de consanguinité. Dans le même temps, on observe une prévalence très faible en malaria aviaire, de l'ordre de 3%. Les dynamiques de populations à l'échelle de l'Europe permettent donc d'éviter à la population française de subir, en plus de l'effet des pratiques agricoles, des conséquences délétères d'un point de vue génétique. Toutefois, ce **flux de gènes** met en lumière une dynamique proche d'un système source-puits où une part importante des caractéristiques génétiques de la population observée en période de reproduction est sans doute due à l'apport régulier de migrants venant de l'est. Dans ce contexte, même si le maintien de la diversité génétique est assuré, on ignore la productivité locale réelle de la population française. D'autre part, ainsi qu'expliqué plus haut, ce flux de gènes est susceptible de contribuer à la perte de caractéristiques locales potentiellement essentielles à la survie des individus (Allendorf *et al.* 2001).

Finalement, alors que les effectifs de Râles des genêts en France ne cessent de décroitre et de se contracter, la justification majeure du maintien de mesures agro-environnementales (MAE), y compris dans les zones où l'espèce a disparu, tient à leur rôle de protection de l'ensemble de l'écosystème prairial. Or, malgré l'intérêt indéniable de ces mesures dans le maintien des zones en prairies, on démontre que focaliser celles-ci sur le Râle des genêts uniquement ne permet pas de protéger une part très importante des autres oiseaux prairiaux. Clairement, le Râle est incapable de représenter une **espèce parapluie** satisfaisante. Cela est d'autant plus criant que les tendances démographiques observées chez le Tarier des prés (Britschgi *et al.* 2006), qui occupe sensiblement le même habitat, sont également à la baisse. A l'heure actuelle, les mesures agro-environnementales censées assurer la protection du Râle des genêts apparaissent ainsi assez inefficaces et leur implémentation pose trois questions majeures :

(i) Leur **utilité** : Les actions de conservations menées en faveur du Râle des genêts semblent relativement vaines, puisqu'elles ne permettent pas le rétablissement des populations et ne fournissent pas une protection adéquate aux autres espèces de l'écosystème.

(ii) Leur **coût** : Au vu du budget considérable qu'elles mobilisent et de leur utilité toute relative, la pertinence des mesures agro-environnementales telles qu'elles sont actuellement implémentées peut être largement mise en question, à plus forte raison lorsque l'argent dépensé aurait potentiellement pu être redirigé vers des actions efficaces.

(iii) Leur **précarité** : En basant les mesures conservatoires sur le Râle des genêts, on s'expose à leur suppression à court terme après la disparition de l'espèce dans les sites

concernés, avec pour conséquence potentielle la conversion des sites prairiaux à la culture intensive ou la populiculture qui impacterait l'ensemble de la biodiversité prairiale. D'autre part, en lien avec le 2ème point, le coût considérable des MAE les contraint à des espaces réduits et peut remettre en cause leur pérennité à long terme dans un contexte de crise économique majeure.

Les tendances démographiques constatées ces dernières années en France pour la population de Râle met en lumière la nécessité de **refonder les MAE**, actuellement incapables d'enrayer le déclin de l'espèce. Parmi les pistes à envisager figurent l'obligation de laisser des « espaces » refuges non fauchés, une meilleure répartition des compensations financières vers les contrats les plus efficaces, le mosaïquage spatiotemporel des fauches ou encore une meilleure flexibilité des mesures en fonction des conditions annuelles. Ces contrats étant mis en place par les exploitants sur la base du volontariat, toute modification des mesures existantes doit prendre en compte le rapport entre efficacité environnementale et maintien d'une agriculture rentable. Si le Râle des genêts échoue dans son rôle d'espèce parapluie, il constitue à coup sûr une bonne **espèce indicatrice** (Noss 1990) de la détérioration des écosystèmes prairiaux, notamment en France. Il a été montré que l'intensification des pratiques agricole en prairies de fauche affecte l'ensemble de la communauté prairiale, allant des plantes (Klimek *et al.* 2006) aux oiseaux (Vickery *et al.* 2000) en passant par les insectes (Morris 2000; Giulio *et al.* 2001). Dès lors, outre la conservation du Râle des genêts en lui-même, le suivi des évolutions démographiques de l'espèce et la compréhension fine des déterminants de celles-ci pourrait permettre de surveiller l'état de santé de son habitat sans recourir à de coûteux suivis de l'ensemble des compartiments de l'écosystème.

6.2. IMPLICATIONS GENERALES

Au-delà de l'aspect appliqué relatif à la problématique de conservation du Râle des genêts, le travail réalisé lors de cette thèse montre l'intérêt de ce type d'approche intégrée, mêlant différentes méthodes d'analyse et différentes échelles. Les résultats obtenus permettent également de mettre en évidence des processus susceptibles d'affecter l'ensemble des communautés inféodées à l'habitat prairial.

6.2.1. Impact de l'activité anthropique sur l'écosystème prairial européen

Cette étude menée à la base dans le seul cadre de la conservation du Râle des genêts a permis de révéler des phénomènes d'intérêt plus large. On a en effet noté que les variations d'activité agricole en Europe affectent ici l'ensemble des phénomènes étudiés. Du fait de la réduction de la survie des individus dans les populations les plus touchées par l'activité agricole, un gradient de taille de populations se met en place proportionnellement à l'intensité agricole. Ces variations démographiques vont entrainer une asymétrie à l'échelle européenne influençant un grand nombre de processus. Dans un premier temps, ce déséquilibre démographique semble responsable d'un flux de gènes asymétrique provoquant une homogénéisation des caractéristiques génétiques des populations et un maintien de la diversité. Or, il y a tout lieu de penser que les variations d'activité anthropique affectent de la même façon une grande part de la communauté d'oiseaux, mais également les populations d'insectes, proie pour le nourrissage des jeunes et vecteurs de certains pathogènes, du fait de l'assèchement des zones humides (Brinson & Malvárez 2002) et de l'utilisation massive de pesticides (Geiger *et al.* 2010) qui caractérisent l'agriculture intensive. Cette hypothèse permet d'expliquer le gradient de prévalence en malaria aviaire constaté chez le Râle des genêts. Les variations de densité des hôtes et des vecteurs de parasites corrélées à l'intensité agricole entraîneraient alors un gradient d'infection longitudinal similaire.

Ce phénomène observé chez le Râle des genêts met en lumière la perturbation des interactions hôtes-pathogènes au sein de l'écosystème prairial. On met dans le même temps en évidence comment l'activité agricole façonne la diversité génétique chez cette espèce, paramètre pouvant être essentiel dans la susceptibilité aux pathogènes. Au-delà de leur intérêt relatif à la conservation du Râle des genêts, ces résultats suggèrent une **perturbation des interactions fonctionnelles** due à l'intensification agricole. Bien que des études complémentaires explorant en détail les relations entre agriculture, communautés d'insectes et interactions hôtes-parasites soient nécessaires pour valider ces conclusions (voir Perspectives), on peut supposer que les patrons constatés chez le Râle reflètent une réalité présente chez de nombreuses espèces inféodés aux mêmes milieux. Alors que les évaluations des statuts de conservation sont le plus souvent espèce-centrés, on tend à montrer ici que des processus affectant en même temps des communautés de groupes taxonomiques variés peuvent toucher le fonctionnement même de l'écosystème. Ces résultats sont rendus possibles grâce à une approche combinée prenant en compte à la fois les

variations démographiques, génétiques et parasitaires en lien avec l'activité agricole, le tout à l'échelle continentale.

6.2.2. Intérêt d'une approche intégrée globale pour la conservation des espèces paléarctiques

La multiplicité des acteurs impliqués dans les actions de conservation, tout spécialement en ce qui concerne les espèces à distribution continentale, rend particulièrement nécessaire la prise en compte de l'ensemble des processus régissant la dynamique de l'espèce d'intérêt. A cet égard, le couplage des approches de modélisation de distribution, de génétique des populations, de parasitologie et d'analyse de l'habitat local présenté ici représente un cadre d'analyse pertinent pour la conservation des espèces à large distribution, et notamment les espèces paléarctiques. L'analyse des processus affectant une espèce à l'échelle de l'ensemble de sa distribution, ou au moins d'une part importante de celle-ci, présente l'intérêt majeur de mettre en lumière des patrons ou dynamiques qui auraient pu rester invisibles à petite échelle.

Les approches à large échelle permettent ainsi de prendre en compte les **variations spatiales** existantes au sein de l'aire de répartition d'une espèce qui vont conditionner la diversité des réponses aux menaces et les actions de conservation (Whittingham *et al.* 2007). En premier lieu, de nombreux paramètres **intrinsèques** à l'espèce, importants dans une optique de conservation, vont varier spatialement lorsque les populations sont structurées génétiquement, ou même lorsqu'elles présentent des variations génétiques continues. On peut observer ainsi de la diversité dans les exigences écologiques des populations à la faveur d'adaptations à des conditions locales. De même, l'histoire phylogéographique des populations peut conditionner les patrons de diversité génétique observés actuellement. La diversité génétique des populations influencera alors grandement les possibilités d'adaptation aux conditions changeantes de l'environnement et par là-même peut déterminer le risque d'extinction de la population dans un contexte de perturbation anthropique. Dans le même temps, des paramètres **extrinsèques** sont susceptibles de présenter des variations considérables à travers la distribution d'une espèce. Ainsi, le niveau des menaces s'exerçant sur l'espèce va varier au gré de l'activité anthropique, elle-même largement liée à des considérations économiques ou politiques. De même, les actions menées en faveur de la conservation d'une espèce dépendront des mêmes facteurs, et peuvent être extrêmement variables lorsque l'aire de répartition de l'espèce couvre de nombreux pays aux réalités socio-économiques différentes. Cela est particulièrement criant dans le cadre du Râle des

genêts, dont l'ensemble des distributions en hivernage, en reproduction et en migration va comprendre pratiquement l'ensemble de l'Afrique, de l'Europe, et une partie importante de l'Asie. La prise en compte de ces variations à grande échelle est alors indispensable pour évaluer les menaces et les opportunités de conservation applicables à l'espèce.

Dans cette optique, l'analyse globale des processus affectant une espèce s'avère nécessaire afin de comprendre le degré de **connectivité** ou de **discontinuité** des populations. En fonction du niveau d'interrelations entre populations, la portée des menaces ou des actions de conservations qui s'exercent en un lieu donné va varier considérablement. En effet, dans un contexte de populations isolées, les mesures appliquées en vue de la conservation d'une population auront peu d'impact sur les autres populations de l'espèce. A l'inverse, lorsque les caractéristiques de l'espèce et de son habitat permettent le déplacement des individus entre populations, les actions menées dans un site, si tant est qu'elles influencent réellement la dynamique de la population, auront des répercussions dans l'ensemble des populations connectées. Dès lors, la coordination des évaluations et des plans de conservation nécessite une bonne connaissance des relations entre sites, et donc de disposer d'une vision large du fonctionnement des populations. Cet aspect est ici exploré *via* la génétique mais la question du déplacement des individus peut également être traitée via d'autres méthodes. En effet, des approches de suivis conventionnels par capture-marquage-recapture (Lebreton *et al.* 2003) permettent entre autres d'obtenir des renseignements sur les mouvements entre sites. D'autre part, les déplacements à échelle fine peuvent par exemple être étudiés par radiotélémétrie (Pride & Swift 1992) tandis que des technologies telles que la télémétrie satellite permettent de suivre des migrations à longue-distance (Perras & Nebel 2012).

Dans le même temps, la **multiplicité des approches** peut s'avérer indispensable à la compréhension des processus déterminants l'état de conservation des populations. Dans le cas présent, les approches retenues se focalisent sur la modélisation de distribution, la génétique des populations, le chant, la parasitologie ou encore l'analyse de l'habitat local. Toutefois, l'utilisation d'autres approches complémentaires peut s'avérer précieuse pour évaluer les processus à l'œuvre. On peut ainsi s'intéresser plus précisément aux réponses physiologiques des individus à certains paramètres environnementaux d'intérêt. De même, des approches comportementales peuvent apporter des connaissances relatives à l'utilisation des ressources, aux mouvements des individus ou aux interactions sociales. C'est finalement l'ensemble des caractéristiques biologiques et écologiques de l'espèce qui peuvent bénéficier à la conservation d'une espèce. Dans ce cas,

l'évaluation du compromis coût-efficacité doit précéder la définition des méthodes d'analyse à privilégier.

6.3. PERSPECTIVES

Les résultats obtenus au cours de cette thèse apportent un nombre important de connaissances concernant le fonctionnement des populations de Râle des genêts. De nombreuses questions restent malgré tout en suspens, pouvant justifier des analyses complémentaires dans l'avenir. Bien que leur mise œuvre dans l'avenir soit purement hypothétique à l'heure actuelle, on peut lister différentes questions dont l'exploration permettrait d'aller plus loin dans la compréhension des processus décrits lors de la thèse :

1- En premier lieu, plusieurs **questions relatives au modèle Râle des genêts** restent à explorer à l'issue de ce travail. Ainsi, l'approche de génétique des populations menée lors de la thèse se concentrait sur la partie européenne de la distribution du Râle des genêts. Une extension de ces analyses aux populations plus orientales pourrait permettre de préciser la structure génétique globale de l'espèce. De même, l'étude des patrons d'infection parasitaire dans l'ensemble de l'aire de reproduction du Râle, mais également en zone d'hivernage, apporterait des éléments plus globaux à la compréhension des interactions hôtes-parasites chez cette espèce.

Par ailleurs, on a posé l'hypothèse que l'homogénéisation génétique des populations, et notamment le remplacement des caractéristiques génétiques des populations occidentales, pouvait avoir des conséquences négatives si ce phénomène conduisait à la perte d'adaptations locales (Allendorf *et al.* 2001). En effet, la suspicion d'une voie migratoire différente et le caractère extrêmement périphérique de ces populations pourraient être en faveur de l'existence d'adaptations dans ces populations. A l'inverse, on a mis en évidence que des effets écologiques conduisaient à une réduction de prévalence parasitaire, relâchant vraisemblablement la pression de sélection sur les locus impliqués dans la réponse immunitaire. De plus, le maintien éventuel d'adaptations locales dans un contexte de flux de gènes élevé entre populations est incertain (Lenormand 2002). La présence ou non d'adaptations locales au sein des populations de Râles de genêts ne peut ainsi être précisée que par des analyses complémentaires.

Il est également possible d'apporter des éléments plus appliqués à la mise en place de mesures de conservations efficaces. Tout d'abord, le rôle du Râle des genêts comme espèce parapluie peut être évalué *a posteriori* par une évaluation des bénéfices des mesures actuellement en place sur la biodiversité prairiale, en comparant par exemple les tendances démographiques de

nombreuses espèces sympatriques dans et en dehors des zones couvertes par les MAE basées sur le Râle. Enfin, la modélisation des réponses des populations aux ajustements des actions de conservation, incluant les caractéristiques de l'espèce, le jeu des acteurs impliqués, les implications économiques des mesures, etc., permettrait sans aucun doute d'améliorer l'aménagement des mesures agro-environnementales.

Enfin, alors que l'analyse de la prévalence en pathogène menée ici se concentre sur les malarias aviaires, il serait intéressant d'étendre cette étude à d'autres types de parasites afin de déterminer si les variations observés sont généralisables à l'ensemble de la communauté de parasites présents chez le Râle des genêts. Dans ce but, des échantillons fécaux avaient été collectés lors de l'échantillonnage des populations. Les premières analyses de ces échantillons montrent la présence de **coccidies** (Figure 35), parasite intestinal déjà identifié chez le Râle (Jeanes *et al.* 2013). Le faible nombre d'échantillons disponibles a toutefois fait que l'analyse n'a pas été poursuivie.

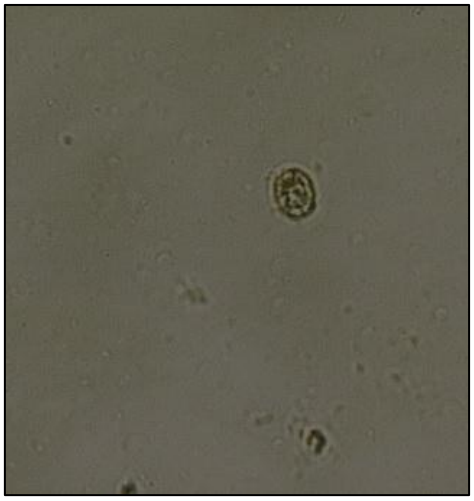


Figure 35 : Spore de coccidie (*Eimeria sp.*) identifié à la loupe binoculaire dans un échantillon fécal provenant d'un individu français capturé en 2011.

2- Les approches développées lors de cette thèse posent également un certain nombre de questionnements plus théoriques. Les espèces distribuées au sein d'une large aire de répartition continentale présentent la particularité d'avoir un habitat recouvrant une gamme de conditions écologiques souvent très importante. De plus, les espèces migratrices telle que celle étudiée au cours de ce travail possèdent en plus deux aires de distribution distinctes selon la saison, tout en étant capable d'utiliser l'espace entre les deux lors de leur migration. Ainsi, les migrateurs transsahariens effectuant leur reproduction dans le Paléarctique tandis que leur aire d'hivernage se trouve en Afrique vont se rencontrer, selon la période de l'année, dans une zone pouvant couvrir

une très grande partie de l'Afrique, de l'Eurasie et du Moyen-Orient. Dès lors, chez ces espèces, va se poser la question de la **stabilité spatio-temporelle de la niche** et le problème afférent à la sélection des **descripteurs** de celle-ci.

D'un point de vue spatial, on constate que les conditions climatiques s'exerçant à l'échelle d'un continent peuvent varier considérablement, en fonction de la latitude ou de l'éloignement de des océans notamment. Pourtant, on observe par exemple de nombreuses espèces d'oiseaux distribuées au sein de l'ensemble du Paléarctique (Bruant de roseaux, Busard des roseaux, Vanneau huppé, etc.) ou même sur tous les continents (Faucon pèlerin, Balbuzard pêcheur). Ces espèces occupent-elles une seule et même niche dans l'ensemble de leur large aire de distribution ou bien la niche occupée varie-t-elle spatialement ? La question qui se pose alors est celle des prédicteurs utilisées pour décrire la niche écologique. A l'échelle de la distribution d'une espèce, on considère généralement que les variables climatiques représentent les meilleurs descripteurs de la niche (Soberón & Nakamura 2009). Pourtant, on a pu observer au cours du travail de modélisation de la distribution du Râle des genêts que seul le partitionnement de la niche permettait de prédire l'ensemble de la distribution de l'espèce, les deux sous-modèles étant influencés par des variables différentes. S'agit-il réellement d'une hétérogénéité spatiale de la niche ou bien est-ce que les prédicteurs climatiques utilisés ne permettaient pas de décrire les exigences écologiques de l'espèce ? Dans le cas présent, on peut aisément imaginer que la variable expliquant la majeure partie de la distribution du Râle des genêts est la présence de prairies naturelles ou semi-naturelles. L'existence de ces prairies pourrait être liée à des espèces végétales différentes selon la zone géographique concernée, chacune avec une niche écologique distincte. Le cas des espèces invasives nous indiquent également que des changements de niche peuvent survenir lors d'un processus d'invasion (Hill *et al.* 2012). Toutefois, ces changements hypothétiques de niche écologique peuvent parfaitement refléter simplement l'hétérogénéité des conditions au sein de la niche de l'espèce (Soberón & Peterson 2011), sans aucun processus évolutif sous-jacent, ou bien encore un mauvais choix des variables utilisées pour décrire cette niche (Rödder *et al.* 2009a). Si l'hétérogénéité de la niche est constatée, la nature de la réponse de l'espèce aux conditions écologiques variables est également un processus méritant de plus amples investigations. Il peut en effet s'agir d'une capacité d'acclamation physiologique de l'espèce à une large gamme de conditions ou bien on peut être en présence de différentes populations avec des adaptations aux caractéristiques locales de la niche.

Chez les espèces migratrices, un autre phénomène s'ajoute à l'hétérogénéité spatiale des conditions écologiques. En effet, ces espèces occupent séquentiellement deux aires de

distribution parfois situées sur des continents différents. Or, on ignore le plus souvent le degré de conservatisme temporel de la niche, entre aires de reproduction et d'hivernage. La chronologie et le chemin de migration peuvent ainsi être dictés par le suivi de conditions écologiques particulières au cours de la saison. Ce suivi temporel de la niche écologique permettrait ainsi à l'espèce de se trouver en permanence au sein des conditions environnementales qui lui sont favorables. L'hypothèse d'une stabilité de la niche est favorisée par la comparaison empirique des niches d'hivernage et de reproduction (Nakazawa *et al.* 2004) ainsi que par le fait que la distribution des espèces, notamment durant la migration, tend à se déplacer en suivant finement le réchauffement climatiques (Monahan & Hijmans 2008; Tingley *et al.* 2009). Toutefois, il n'est pas exclu que des espèces puissent alternativement exploiter deux niches écologiques différentes selon la période. Une analyse de la cohérence des deux niches pourrait ainsi être menée sur un large spectre d'espèces migratrices afin de définir plus précisément la généralité des phénomènes de suivi temporel de niche ou au contraire de changement saisonnier de la niche. Cela pourrait être réalisé en testant les capacités de projection d'un modèle de distribution en période de reproduction européen dans l'habitat d'hivernage et vice-versa. L'étude des variations spatiales et temporelles de la niche chez des espèces migratrices à large distribution présente ainsi un cadre d'analyse théorique intéressant mais peut également s'avérer utile dans un objectif de biologie de la conservation en permettant d'anticiper la stabilité spatiotemporelle des paramètres écologiques contraignants.

3- Au-delà des variations de la niche écologique, il existe des variations considérables de l'activité anthropique à l'échelle d'un continent, pouvant impacter de différentes façons les populations naturelles. Dans le travail réalisé ici sur le Râle des genêts, on a constaté un gradient de prévalence en malaria aviaire corrélé à l'activité agricole. On a posé l'hypothèse que les variations d'intensité agricole avaient contribué à des effets variables sur les populations de vecteurs et d'hôtes et ainsi conduit aux variations observées dans l'abondance des pathogènes. Toutefois, bien que cette hypothèse ouvre des perspectives intéressantes en ce qui concerne l'influence de l'activité agricole sur la **perturbation des interactions hôtes-pathogènes**, nous ne disposons pas à l'heure actuelle de preuve formelle de ce phénomène. Dès lors, une analyse plus poussée des relations entre intensification agricole, abondance des insectes vecteurs de parasites et prévalence de ces parasites dans les populations hôtes pourrait apporter des éléments de réponse.

L'effet de l'intensification des pratiques dans les agrosystèmes sur les populations d'insectes commence à être de mieux en mieux documenté, même si la situation de conservation globale des insectes demeure largement inconnue (Dunn 2005). Toutefois, la question de la conservation des insectes a très souvent été biaisée en faveur des espèces porteuses de services écosystémiques tels que les pollinisateurs (Biesmeijer *et al.* 2006; Aizen & Harder 2009; Potts *et al.* 2010) ou des espèces plus emblématiques telles que les papillons (Maes & Van Dyck 2001; Thomas *et al.* 2004). De la même façon, la question des dynamiques de populations de pathogènes et de leurs vecteurs a plus souvent été prise en compte du point de vue du contrôle des maladies humaines (Meyrowitsch *et al.* 2011). De ce fait, l'effet des pratiques agricoles sur la prévalence en pathogènes dans les populations animales naturelles a été peu étudiée. Il a toutefois été montré que l'activité agricole pouvait favoriser l'infection en parasites sanguins dans des populations d'oiseaux lorsque la présence d'élevage de volaille fournit à la fois un réservoir de vecteurs (points d'eau artificiels) et de pathogènes (volailles porteurs de parasites) (Gonzalez-Quevedo *et al.* 2014). Cette corrélation positive entre activité humaine et abondance des vecteurs de pathogènes est toutefois peu probable à une échelle plus large, ou l'impact anthropique est plus certainement responsable d'un déclin de ces populations de fait de l'effet combiné du drainage des zones humides (Brinson & Malvárez 2002) et de l'utilisation massive de pesticides (Geiger *et al.* 2010). Il est alors parfaitement vraisemblable que les variations d'intensité agricole qu'on observe à l'échelle continentale puisse entrainer des variations dans l'abondance des insectes et par conséquent dans la prévalence des pathogènes. Cette relation, si elle se confirme, révélerait un effet de l'activité humaine, non pas seulement sur les dynamiques de populations mais également sur les interactions fonctionnelles hôtes-parasites.

Afin d'explorer en détail cette hypothèse, une simple analyse des relations entre intensité agricole et prévalence en pathogène, comme menée dans le cas du Râle des genêts, s'avère insuffisante. En effet, des données relativement précises des variations d'abondance en insectes sont nécessaires pour confirmer le lien avec les prévalences observées dans les populations hôtes. D'autre part, la généralité de l'hypothèse proposée ici doit être testée à travers une large gamme d'espèces hôtes. Dans cette optique, les oiseaux représentent un taxon dans lequel l'infection en malaria aviaire est très étudiée (Valkiūnas 2005). De même, les organismes gouvernementaux ou internationaux tels que la FAO (Organisation des Nations-Unis pour l'agriculture et l'alimentation) fournissent des données précises des variations d'intensité agricole entre pays, même si une description infranationale des pratiques agricoles permettrait d'atteindre une meilleure description du phénomène. Une méta-analyse de résultats déjà publiés peut d'ores et déjà permettre de déterminer si le patron prédit se retrouve chez une majorité d'espèces. Enfin,

ce type d'analyse n'a pas à être confiné au modèle oiseau/malaria. D'autres parasites transmis par des vecteurs peuvent faire l'objet du même effet de l'activité agricole sur leur prévalence. Par exemple, les bactéries du genre *Borrelia*, responsables de la maladie de Lyme infectent une grande variété d'hôtes parmi les mammifères (Steere *et al.* 2004). Il serait intéressant d'étendre l'analyse des relations entre agriculture, abondance des vecteurs et prévalence à ce type de parasites transmis par l'intermédiaire de tiques, des vecteurs appartenant cette fois-ci au groupe des arachnides. La confirmation d'un effet généralisé de l'intensification agricole sur les dynamiques d'interactions hôtes-pathogènes pourrait finalement fournir un moyen de diagnostic de l'état de santé d'un écosystème. En effet, une prévalence parasitaire exceptionnellement faible chez un certain nombre d'espèces hôtes pourrait refléter dans le site concerné la détérioration des interactions fonctionnelles de l'écosystème.

7. BIBLIOGRAPHIE

Acevedo P, Jimenez-Valverde A, Lobo JM, Real R (2012) Delimiting the geographical background in species distribution modelling. *Journal of Biogeography*, **39**, 1383–1390.

Adams WM, Aveling R, Brockington D *et al.* (2004) Biodiversity conservation and the eradication of poverty. *Science*, **306**, 1146–1149.

Aizen M a., Harder LD (2009) The Global Stock of Domesticated Honey Bees Is Growing Slower Than Agricultural Demand for Pollination. *Current Biology*, **19**, 915–918.

Alho JS, Välimäki K, Merilä J (2010) Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. *Molecular Ecology Resources*, **10**, 720–2.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids : setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Allouche O, Steinitz O, Rotem D, Rosenfeld A, Kadmon R (2008) Incorporating distance constraints into species distribution models. *Journal of Applied Ecology*, **45**, 599–609.

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology & Evolution*, **18**, 589–596.

Amos W, Wilmer JW, Fullard K *et al.* (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 2021–2027.

Aparicio JM, Ortego J, Cordero PJ (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology*, **15**, 4659–65.

Appleby B, Redpath S (1997) Variation in the male territorial hoot of the Tawny Owl Strix aluco in three English populations. *Ibis*, **139**, 152–158.

Araújo MB, Cabeza M, Thuiller W, Hannah L, Williams PH (2004) Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, **10**, 1618–1626.

Araújo MB, Pearson R, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

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

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in ecology & evolution*, **22**, 42–7.

Arino O, Bicheron P, Achard F *et al.* (2008) GLOBCOVER The most detailed portrait of Earth. *ESA Bulletin - European Space Agency*, **136**, 24–31.

Asghar M, Hasselquist D, Bensch S (2011) Are chronic avian haemosporidian infections costly in wild birds? *Journal of Avian Biology*, **42**, 530–537.

Ashton K (2002) Patterns of within-species body size variation of birds: strong evidence for Bergmann's rule. *Global Ecology and Biogeography*, **11**, 505–523.

Atkinson CT (2008) Avian malaria. In: *Parasitic diseases of wild birds* (eds Atkinson CT, Thomas NJ, Hunter DB), pp. 35–53. Wiley-Blackwell.

Austin M, Belbin L, Meyers J, Doherty M, Luoto M (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling*, **199**, 197–216.

Baha El Din SM, Salama W, Grieve A, Green RE (1996) Trapping and shooting of Corncrakes Crex crex on the Mediterranean coast of Egypt. *Bird Conservation International*, **6**, 213–227.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

Barbet-Massin M, Thuiller W, Jiguet F (2010) How much do we overestimate future local extinction rates when restricting the range of occurrence data in climate suitability models? *Ecography*, **33**, 878–886.

Barnosky AD, Matzke N, Tomiya S *et al.* (2011) Has the Earth' s sixth mass extinction already arrived ? *Nature*, **470**, 51–57.

Barton K (2013) MuMIn: Multi-model inference.

Barve N, Barve V, Jimenez-Valverde A *et al.* (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4.

Beadell JS, Ishtiaq F, Covas R *et al.* (2006) Global phylogeographic limits of Hawaii's avian malaria. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 2935–44.

Beardmore J (1983) Extinction, survival, and genetic variation. In: *Genetics and conservation: a reference for managing wild animal and plant populations* (eds Schonewald-Cox C, Chambers S, MacBryde F, Thomas L), pp. 125–151. Benjamin/Cummings, Menlo Park, California.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beaumont MA (2005) Adaptation and speciation: what can F(st) tell us? *Trends in ecology & evolution (Personal edition)*, **20**, 435–40.

Beaumont LJ, Hughes L, Poulsen M (2005) Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecological Modelling*, **186**, 251–270.

Beaumont MA (2010) Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 379–406.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Beck J, Böller M, Erhardt A, Schwanghart W (2013) Spatial bias in the GBIF database and its effect on modelling species' geographic distributions. *Ecological Informatics*.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–5.

Beier J, Strandberg J, Stoskopf MK, Craft C (1981) Mortality in robins (Turdus migratorius) due to avian malaria. *Journal of Wildlife Diseases*, **17**, 247–250.

Bensch S, Stjernman M, Hasselquist D *et al.* (2000) Host specificity in avian blood parasites: a study of Plasmodium and Haemoproteus mitochondrial DNA amplified from birds. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1583–9.

Bensch S, Hellgren O, Pérez-Tris J (2009) MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Molecular Ecology Resources*, **9**, 1353–8.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. *Nucleic Acids Research*, **33**, D34–8.

Berg A, Gustafson T (2007) Meadow management and occurrence of corncrake Crex crex. *Agriculture, Ecosystems and Environment*, **120**, 139–144.

Berg A, Hiron M (2012) Occurrence of Corncrakes Crex crex in mosaic farmland landscapes in south-central Sweden – effects of habitat and landscape structure. *Bird Conservation International*, **22**, 234–245.

Berkes F (2004) Rethinking community-based conservation. *Conservation biology*, **18**, 621–630.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–25.

Besnard AG, La Jeunesse I, Pays O, Secondi J (2013) Topographic wetness index predicts the occurrence of bird species in floodplains. *Diversity and Distributions*, **19**, 955–963.

Besnard AG, Secondi J (2014) Hedgerows diminish the value of meadows for grassland birds: potential conflicts for agri-environment schemes. *Agriculture Ecosystem & Environment*, in press.

Beven KJ, Kirkby MJ (1979) A physically based , variable contributing area model of basin hydrology. *Hydrological Sciences*, **24**, 43–69.

Biesmeijer JC, Roberts SPM, Reemer M *et al.* (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, **313**, 351–354.

Bijlsma R (2000) Does inbreeding affect the extinction risk of small populations?: predictions from Drosophila. *Journal of Evolutionary Biology*, **13**, 502–514.

BirdLife International (2004) *Birds in Europe: Population Estimates, Trends and Conservation Status*. BirdLife International, Wageningen.

Birdlife International (2012) Crex crex. In: *IUCN Red List of Threatened Species Version 2012.1*. <www.iucnredlist.org>.

Birdlife International (2013) *Species factsheet: Crex crex*. Downloaded from http://www.birdlife.org on 17/12/2013.

Bivand R, Rundel C (2013) rgeos: Interface to Geometry Engine - Open Source (GEOS).

Blaustein A, Kiesecker J (2002) Complexity in conservation: lessons from the global decline of amphibian populations. *Ecology letters*, **5**, 597–608.

Blum MGB, François O (2010) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, **20**, 63–73.

Boldogh S, Szegedi Z, Szentgyörgyi P, Petrovics Z (2009) Distribution, population size and conservation status of Corncrake Crex crex in North-east Hungary, 1997 – 2006. *Vogelwelt*, **130**, 153–158.

Boria R a., Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, **275**, 73–77.

Both C, Bouwhuis S, Lessells C, Visser M (2006) Climate change and population declines in a long-distance migratory bird. *Nature*, **441**, 81–83.

Boubou A, Migeon A, Roderick GK *et al.* (2012) Test of colonisation scenarios reveals complex invasion history of the red tomato spider mite Tetranychus evansi. *PloS one*, **7**, e35601.

Boyce M, McDonald L (1999) Relating populations to habitats using resource selection functions. *Trends in Ecology & Evolution*, **14**, 268 – 272.

Bradbury JW, Vehrencamp SL (2011) *Principles of animal communication, 2nd ed*. Sinauer Associates.

Brambilla M, Ficetola GF (2012) Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species. *The Journal of Animal Ecology*, **81**, 781–787.

Brambilla M, Pedrini P (2011) Intra-seasonal changes in local pattern of Corncrake Crex crex occurrence require adaptive conservation strategies in Alpine meadows. *Bird Conservation International*, 1–6.

Branton M, Richardson JS (2011) Assessing the value of the umbrella-species concept for conservation planning with meta-analysis. *Conservation Biology*, **25**, 9–20.

Braunisch V, Coppes J, Arlettaz R *et al.* (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**.

Bray J, Curtis J (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **7**, 325–349.

Breiman L, Friedman J, Stone CJ, Olshen RA (1984) *Classification and Regression Trees*. Chapman & Hall.

Breiman L (2001a) Random forests. *Machine learning*, 5–32.

Breiman L (2001b) Statistical Modeling : The Two Cultures. *Statistical Science*, **16**, 199–231.

Brenowitz EA (1991) Evolution of the vocal control system in the avian brain. *Seminars in Neuroscience*, **3**, 399–407.

Bretagnolle V (1996) Acoustic communication in a group of nonpasserine birds, the petrels. In: *Ecology and evolution of acoustic communication in birds* (eds Kroodsma DE, Miller EH), pp. 160–177. Cornell Univ. Press.

Bretagnolle V, Genevois F (1997) Geographic variation in the call of the Blue Petrel: Effects of sex and geographical scale. *Condor*, **99**, 985–989.

Brinson MM, Malvárez AI (2002) Temperate freshwater wetlands: types, status, and threats. *Environmental Conservation*, **29**, 115–133.

Britschgi A, Spaar R, Arlettaz R (2006) Impact of grassland farming intensification on the breeding ecology of an indicator insectivorous passerine, the Whinchat Saxicola rubetra: Lessons for overall Alpine meadowland management. *Biological Conservation*, **130**, 193–205.

Broennimann O, Fitzpatrick MC, Pearman PB *et al.* (2012) Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecology and Biogeography*, **21**, 481–497.

Brown JH (1984) On the relationship between abundance and distribution of species. *American naturalist*, **124**, 255–279.

Brown JH, Stevens GC, Kaufman DM (1996) The Geographic Range: Size, Shape, Boundaries, and Internal Structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.

Broyer J (1991) The situation of the Corncrake in France. *Vogelwelt*, **112**, 71–77.

Broyer J (1994) The decline of the corncrake Crex crex in France and the management of meadow habitats. *Alauda*, **62**, 1–7.

Broyer J (1995) Project of a hay-making calendar tolerable for the successful breeding of corncrakes Crex crex in France. *Alauda*, **63**, 209–212.

Broyer J (1996) Les « fenaisons centrifuges », une méthode pour réduire la mortalité des jeunes râles de genêts crex crex et cailles des blés coturnix coturnix. *Revue d'Ecologie (La Terre et la Vie)*, **51**, 269–276.

Broyer J (2002) Contribution to the development of a methodology to monitor Corncrake Crex crex population during the breeding period. *Alauda*, **70**, 195–202.

Broyer J (2003) Unmown refuge areas and their influence on the survival of grassland birds in the Saône valley (France). *Biodiversity and Conservation*, **12**, 1219–1237.

Broyer J, Curtet L, Mischenko A, Sukhanova O (2007) Les prairies de la Russie européenne: mise en place du suivi d'un écosystème primordial pour l'avifaune migratrice hivernant en France. *Faune sauvage*, **276**, 23–26.

Brussard PF (1984) Geographic patterns and environmental gradients: the central-marginal model in Drosophila revisited. *Annual Review of Ecology and Systematics*, **15**, 25–64.

Budka M, Mikkelsen G, Turcokova L *et al.* (2014) Macrogeographic variations in the call of the corncrake Crex crex. *Journal of Avian Biology*, **45**, 65–74.

Budka M, Osiejuk TS (2013a) Neighbour–stranger call discrimination in a nocturnal rail species, the Corncrake Crex crex. *Journal of Ornithology*.

Budka M, Osiejuk TS (2013b) Habitat preferences of Corncrake (Crex crex) males in agricultural meadows. *Agriculture, Ecosystems & Environment*, **171**, 33–38.

Budka M, Osiejuk TS (2013c) Formant Frequencies are Acoustic Cues to Caller Discrimination and are a Weak Indicator of the Body Size of Corncrake Males (M Manser, Ed,). *Ethology*, **119**, n/a–n/a.

Burdon JJ, Chilvers G a (1982) Host Density as a Factor in Plant Disease Ecology. *Annual Review of Phytopathology*, **20**, 143–166.

Bürger P, Pykal J, Hora J (1998) Distribution, numbers and ringing results of Corncrake (Crex crex) in the Czech republic in te period 1993-1997. *Sylvia*, **34**, 73–84.

Burgman M, Carr A, Godden L *et al.* (2011) Redefining expertise and improving ecological judgment. *Conservation Letters*, **4**, 81–87.

Burgman MA, Fox JC (2003) Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation*, **6**, 19–28.

Burnham KP, Anderson DR (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Busby J (1991) BIOCLIM: a bioclimate analysis and prediction system. *Plant Protection Quarterly*, 64 – 68.

Busche G (1994) The decline of wet-meadow birds in Schleswig-Holstein/Germany from 1950 to 1992. *Journal fur Ornithologie*, **135**, 167–177.

Bustamante J, Seoane J (2004) Predicting the distribution of four species of raptors (Aves: Accipitridae) in southern Spain: statistical

models work better than existing maps. *Journal of Biogeography*, **31**, 295–306.

Butlin R (2010) Population genomics and speciation. *Genetica*, **138**, 409–18.

Bystriakova N, Peregrym M, Erkens RHJ, Bezsmertna O, Schneider H (2012) Sampling bias in geographic and environmental space and its effect on the predictive power of species distribution models. *Systematics and Biodiversity*, **10**, 1–11.

Caizergues A, Rätti O, Helle P *et al.* (2003) Population genetic structure of male black grouse (Tetrao tetrix L.) in fragmented vs. continuous landscapes. *Molecular ecology*, **12**, 2297–2305.

Calenge C (2006) The package adehabitat for the R software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 516–519.

Campana MG, Hunt H V, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, **11**, 349–52.

Carlsson J (2008) Effects of microsatellite null alleles on assignment testing. *The Journal of heredity*, **99**, 616–23.

Caro T (2003) Umbrella species: critique and lessons from East Africa. *Animal Conservation*, **6**, 171–181.

Caro TM, O'Doherty G (1999) On the Use of Surrogate Species in Conservation Biology. *Conservation Biology*, **13**, 805–814.

Carpenter G, Gillison A, Winter J (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and conservation*, **2**, 667–680.

Cassini MH (2011) Ranking threats using species distribution models in the IUCN Red List assessment process. *Biodiversity and Conservation*, **20**, 3689–3692.

De Castro F, Bolker B (2004) Mechanisms of disease-induced extinction. *Ecology Letters*, **8**, 117–126.

Catchpole CK, Slater PJB (2008) *Bird Song: Biological Themes and Variations, 2nd ed.* Cambridge Univ Press.

Channell R, Lomolino M (2000) Dynamic biogeography and conservation of endangered species. *Nature*, **403**, 84–86.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology*, **18**, 2746–65.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution*, **24**, 621–31.

Charlesworth D (2003) Effects of inbreeding on the genetic diversity of populations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **358**, 1051–70.

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, **18**, 237–268.

Chauvenet A, Ewen J, Armstrong D, Blackburn T, Pettorelli N (2013) Maximizing the success of assisted colonizations. *Animal Conservation*, **16**, 161–169.

Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, **210**, 478–486.

Chen C, Durand E, Forbes F, Fran??ois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747–756.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**, 1024–1026.

Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution*, **53**, 1259–1267.

Conrad KF, Warren MS, Fox R, Parsons MS, Woiwod IP (2006) Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis. *Biological Conservation*, **132**, 279–291.

Corander J, Sirén J, Arjas E (2007) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Cosens SE, Falls JB (1984) A comparison of sound propagation and song frequency in temperate marsh and grassland habitats. *Behavioral Ecology and Sociobiology*, **15**, 161–170.

Costa GC, Nogueira C, Machado RB, Colli GR (2009) Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, **19**, 883–899.

Courchamp F, Clutton-Brock T, Grenfell B (1999) Inverse density dependence and the Allee effect. *Trends in Ecology and Evolution*, **14**, 405–410.

Cramp S, Simmons K (1980) *The birds of the Western Palearctic, vol 2*. Oxford University Press.

Crandall KAK, Bininda-Emonds OORP, Mace GGM, Wayne RRK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**, 290–295.

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. *Heredity*, **83**, 260–70.

Crockford N, Green RE, Rocamora G *et al.* (1997) A summary of the European Action plan for the Corncrake Crex crex. *Vogelwelt*, **118**, 169–173.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–8.

Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

Cumming GS, Shepard E, Okanga S *et al.* (2013) Host associations, biogeography, and phylogenetics of avian malaria in southern African waterfowl. *Parasitology*, **140**, 193–201.

Czech B (2000) Economic growth as the limiting factor for wildlife conservation. *Wildlife Society Bulletin*, **28**, 4–15.

Czech B, Krausman PR (1997) Distribution and Causation of Species Endangerment in the United States. *Science*, **277**, 1116–1117.

Daszak P, Cunningham A, Hyatt A (2000) Emerging Infectious Diseases of Wildlife-- Threats to Biodiversity and Human Health. *Science*, **287**, 443–449.

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity*, **80**, 531–7.

Dawson DA, Horsburgh GJ, Küpper C *et al.* (2010) New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources*, **10**, 475–494.

Dawson DA, Ball AD, Spurgin LG *et al.* (2013) High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. *BMC Genomics*, **14**, 176.

Deceuninck B, Fantin P, Jolivet C *et al.* (1997) Chronologie de la reproduction du râle de genêts Crex crex en France: Particularités régionales et évaluation des mesures de conservation. *Alauda*, **65**, 91–104.

Deceuninck B, Mourgaud G, Beslot E (2011) *Plan National d'Action Râle des genêts (2005-2009). Bilan final.* Rapport LPO/MEEDDTL-DREAL PL.

Dennis R, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73–77.

Dietz K (1988) Density-dependence in parasite transmission dynamics. *Parasitology Today*, **4**, 91–97.

Dizon AE, Locker C, Perrin WF, Demaster DP, Sisson J (1992) Rethinking the Stock Concept : A Phylogeographic Approach. *Conservation Biology*, **6**, 24–36.

Dobzhansky T, Wright S (1941) Genetics of Natural Populations. V. Relations between Mutation Rate and Accumulation of Lethals in Populations of Drosophila Pseudoobscura. *Genetics*, **26**, 23–51.

Donald PF, Green RE, Heath MF (2001) Agricultural intensification and the collapse of Europe's farmland bird populations. *Proceedings of the Royal Society of London B: Biological Sciences*, **268**, 25–9.

Van Dongen WFD, Mulder RA (2006) Habitat density, song structure and dialects in the Madagascar paradise flycatcher Terpsiphone mutata. *Journal of Avian Biology*, **37**, 349–356.

Dormann CF, McPherson JM, Araújo MB *et al.* (2007) Methods to account for spatial autocorrelation in the analysis of species

distributional data: a review. *Ecography*, **30**, 609–628.

Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Drew CA, Perera AH (2011) Expert Knowledge as a Basis for Landscape Ecological Predictive Models. In: *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications* (eds Drew CA, Wiersma YF, Huettmann F), pp. 229–248. Springer, New York.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S (2013) Demographic inference reveals African and European admixture in the North American Drosophila melanogaster population. *Genetics*, **193**, 291–301.

Dudík M, Schapire RE, Phillips SJ (2005) Correcting sample selection bias in maximum entropy density estimation. In: *Appearing in Advances in Neural Information Processing Systems*

Dunn R (2005) Modern insect extinctions, the neglected majority. *Conservation Biology*, 1030–1036.

Durrant KL, Marra PP, Fallon SM *et al.* (2008) Parasite assemblages distinguish populations of a migratory passerine on its breeding grounds. *Journal of Zoology*, **274**, 318–326.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Ebert D, Zschokke-Rohringer C, Carius H (2000) Dose effects and density-dependent regulation of two microparasites of Daphnia magna. *Oecologia*, **122**, 200–209.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–88.

Edwards TC, Cutler DR, Zimmermann NE, Geiser L, Moisen GG (2006) Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling*, **199**, 132–141.

Elith J, Graham CH, Anderson RP *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith J, Kearney M, Phillips SJ (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Elith J, Phillips SJ, Hastie T *et al.* (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–47.

Elith J, Burgman M (2002) Predictions and their validation: Rare plants in the Central Highlands, Victoria, Australia. In: *Predicting Species Occurrences: Issues of Accuracy and Scale* , pp. 303–313. Island Press, Washington, USA.

Elith J, Leathwick JR (2006) Conservation prioritisation using species distribution modelling. In: *Spatial conservation prioritization: quantitative methods and computational tools.* (eds Moilanen A, Wilson KA, Possingham HP). Oxford University Press, Oxford, UK.

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.

Elton C (1927) *Animal Ecology*. Sedgwick and Jackson, London, UK.

Engler J, Rödder D, Elle O, Hochkirch A, Secondi J (2013) Species distribution models contribute to determine the effect of climate and interspecific interactions in moving hybrid zones. *Journal of Evolutionary Biology*, **26**, 2487–2496.

Esselman PC, Allan JD (2011) Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. *Freshwater Biology*, **56**, 71–88.

Estrada A, Real R, Vargas JM (2011) Assessing coincidence between priority conservation areas for vertebrate groups in a Mediterranean hotspot. *Biological Conservation*, **144**, 1120–1129.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Evans ML, Neff BD (2009) Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations

of Chinook salmon (Oncorhynchus tshawytscha). *Molecular Ecology*, **18**, 4716–29.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–98.

Excoffier L, Foll M (2011) Fastsimcoal: a Continuous-Time Coalescent Simulator of Genomic Diversity Under Arbitrarily Complex Evolutionary Scenarios. *Bioinformatics*, **27**, 1332–4.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–7.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

FAO (2007) *FAOSTAT - Machinery and Equipment statistics*. Food and Agriculture Organization of the United Nations, Rome, Italy.

Féon V Le, Schermann-Legionnet A, Delettre Y *et al.* (2010) Intensification of agriculture, landscape composition and wild bee communities: a large scale study in four European countries. *Agriculture, Ecosystems & Environment*, **137**, 143–150.

Ferrier S (1984) *The status of the Rufous Scrub-bird Atrichornis rufescens : habitat, geographical variation and abundance*. University of New England, Armidale, Australia.

Ficetola GF, Thuiller W, Miaud C (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species - the American bullfrog. *Diversity and Distributions*, **13**, 476–485.

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Filz KJ, Schmitt T, Engler JO (2013) How fine is fine-scale? – Questioning the use of fine-scale bioclimatic data in species distribution models to forecast abundance patterns in butterflies. *European Journal of Entomology*, **110**, 311–317.

Fischer J, Lindenmayer D (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography*, **16**, 265–280.

Fisher RA (1930) *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fitch WT (1999) Acoustic exaggeration of size in birds via tracheal elongation: comparative and theoretical analyses. *Journal of zoology*, **248**, 31–48.

Fletcher NH (2004) A simple frequency-scaling rule for animal communication. *The Journal of the Acoustical Society of America*, **115**, 2334–2338.

Foley J a., Ramankutty N, Brauman K a. *et al.* (2011) Solutions for a cultivated planet. *Nature*.

Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–93.

Fonderflick J, Besnard A, Martin J-L (2013) Species traits and the response of open-habitat species to forest edge in landscape mosaics. *Oikos*, **122**, 42–51.

Fontaine MC, Austerlitz F, Giraud T *et al.* (2013) Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen Plasmopara viticola. *Molecular Ecology*, **22**, 2771–2786.

Food and Agriculture Organization of the United Nations FAO Statistical Database (FAOSTAT).

Forester BR, DeChaine EG, Bunn AG (2013) Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions (B Wintle, Ed,). *Diversity and Distributions*.

Forstmeier W, Schielzeth H (2012) Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology*, **21**, 3237–3249.

Fourcade Y, Richardson DS, Keišs O *et al.* Range dynamics at the continental scale: a test of the

central-marginal hypothesis in the Corncrake, a species with a complex conservation status.

Fourcade Y, Engler JO, Besnard AG, Rödder D, Secondi J (2013) Confronting expert-based and modelled distributions for species with uncertain conservation status: a case study from the Corncrake (Crex crex). *Biological Conservation*, **167**, 161–171.

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

François O, Durand E (2010) Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **5**, no–no.

Frankel OH, Soulé ME (1981) *Conservation and Evolution*. Cambridge University Press, Cambridge, UK.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.

Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, **10**, 1500–1508.

Frankham R, Briscoe DA, Ballou JD (2002) *Introduction to Conservation Genetics*. Cambridge University Press.

Frankham R (2005) Genetics and extinction. *Biological Conservation*, **126**, 131–140.

Frankham R (2009) Inbreeding in the wild really does matter. *Heredity*, **104**, 124.

Franklin J (2009) *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge, UK.

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular ecology*, **10**, 2741–52.

Freeman TG (1991) Calculating catchment area with divergent flow based on a regular grid. *Computers & Geosciences*, **17**, 413–422.

Freeman EA, Moisen G (2008) {PresenceAbsence}: An R Package for Presence Absence Analysis. *Journal of Statistical Software*, **23**, 1–31.

Friedman J (1991) Multivariate adaptive regression splines. *The annals of statistics*, **19**, 1–67.

Fuller R, Gregory R, Gibbons DW *et al.* (1995) Population declines and range contractions among lowland farmland birds in Britain. *Conservation Biology*, **9**, 1425–1441.

Galeotti PR, Appleby BM, Redpath SM (1996) Macro and microgeographical variations in the "hoot" of Italian and English tawny owls (Strix aluco). *Italian Journal of Zoology*, **63**, 57–64.

García-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, **51**, 21–28.

Gaston KJ, Blackburn T, Greenwood J *et al.* (2000) Abundance–occupancy relationships. *Journal of Applied Ecology*, **37**, 39–59.

Gautschi B, Klug Arter M, Husi R, Wettstein W, Schmid B (2002) Isolation and characterization of microsatellite loci in the globally endangered Corncrake, Crex crex Linné. *Conservation Genetics*, **3**, 451–453.

Geiger F, Bengtsson J, Berendse F *et al.* (2010) Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. *Basic and Applied Ecology*, **11**, 97–105.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P (2010) Calculations of population differentiation based on GST and D: forget GST but not all of statistics! *Molecular Ecology*, **19**, 3845–52.

Gilbert M, Xiao X, Domenech J (2006) Anatidae migration in the western Palearctic and spread of highly pathogenic avian influenza H5NI virus. *Emerging Infectious Diseases*, **12**, 1650–6.

Giulio MDI, Edwards PJ, Meister E (2001) Enhancing insect diversity in agricultural grasslands: the roles of management and landscape structure. *Journal of Applied Ecology*, **38**, 310–319.

Godsoe W, Harmon LJ (2012) How do species interactions affect species distribution models? *Ecography*, no–no.

Gonzalez S, Soto-Centeno J, Reed D (2011) Population distribution models: species distributions are better modeled using biologically relevant data partitions. *BMC Ecology*, **11**, doi:10.1186/1472–6785–11–20.

Gonzalez-Quevedo C, Davies RG, Richardson DS (2014) Predictors of malaria infection in a wild bird population: landscape-level analyses reveal climatic and anthropogenic factors. *The Journal of animal ecology*.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F -statistics. *Molecular Ecology*, **2**, 184–186.

Goward SN, Markham B, Dye DG, Dulaney W, Yang JL (1991) Normalized Difference Vegetation Index Measurements From The Advanced Very High-Resolution Radiometer. *Remote Sensing Of Environment*, **35**, 257–277.

Grand J, Cummings MP, Rebelo TG, Ricketts TH, Neel MC (2007) Biased data reduce efficiency and effectiveness of conservation reserve networks. *Ecology Letters*, **10**, 364–74.

Green RE (1995) The decline of the corncrake Crex crex in Britain continues. *Bird Study*, **42**, 66–75.

Green RE (1996) Factors affecting the population density of the corncrake Crex crex in Britain and Ireland. *Journal of Applied Ecology*, **33**, 237–248.

Green RE, Rocamora G, Schäffer N (1997a) Populations, ecology and threats to the Corncrake Crex crex in Europe. *Vogelwelt*, **118**, 117–134.

Green RE, Tyler GA, Stowe TJ, Newton A V (1997b) A simulation model of the effect of mowing of agricultural grassland on the breeding success of the corncrake (Crex crex). *Journal of Zoology*, **243**, 81–115.

Green RE (1999) Survival and dispersal of male Corncrakes Crex crex in a threatened population. *Bird Study*, **46**, 218–229.

Green RE (2004) A new method for estimating the adult survival rate of the Corncrake Crex crex and comparison with estimates from ring-recovery and ring-recapture data. *Ibis*, **146**, 501–508.

Green RE (2013) Tracking Scotland's Corncrakes. *Birdwatch*, **april**, 26–28.

Green RE, Gibbons DW (2000) The status of the Corncrake Crex crex in Britain in 1998. *Bird Study*, **47**, 129–137.

Green RE, Rayment MD (1996) Geographical variation in the abundance of the Corncrake Crex crex in Europe in relation to the intensity of agriculture. *Bird Conservation International*, **6**, 201–211.

Green RE, Stowe TJ (1993) The decline of the corncrake Crex crex in Britain and Ireland in relation to habitat change. *Journal of Applied Ecology*, **30**, 689–695.

Green RE, Tyler GA (2005) Estimating the age of Corncrake Crex crex chicks from body weight and the development of primary feathers. *Ringing and Migration*, **22**, 139–144.

Greenwell BM (2013) investr: Inverse Estimation with Linear and Nonlinear Regression Models in R.

Gregory RD, Baillie SR (1998) Large-scale habitat use of some declining British birds. *Journal of Applied Ecology*, **35**, 785–799.

Grinnell J (1917) The Niche-Relationships of the California Thrasher. *The Auk*, **34**, 427 – 433.

Guillot G, Mortier F, Estoup A (2005) GENELAND: A computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Guisan A, Tingley R, Baumgartner JB *et al.* (2013) Predicting species distributions for conservation decisions. *Ecology letters*, **16**, 1424–1435.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Guthery F, Brennan L, Peterson M, Lusk J (2005) Information Theory in Wildlife Science: Critique and Viewpoint. *Journal of Wildlife Management*, **69**, 457–465.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanley A, Mcneil J (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Hansen P (1979) Vocal learning: its role in adapting sound structures to long-distance propagation, and a hypothesis on its evolution. *Animal Behaviour*, **27**, 1270–1271.

Hansson B, Bensch S, Hasselquist D, Akesson M (2001) Microsatellite diversity predicts recruitment of sibling great reed warblers. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 1287–1291.

Hansson B, Westerdahl H, Hasselquist D, Akesson M, Bensch S (2004) Does linkage disequilibrium generate heterozygosity-fitness correlations in great reed warblers? *Evolution*, **58**, 870–879.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, **11**, 2467–74.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Harris G, Pimm SL (2008) Range size and extinction risk in forest birds. *Conservation Biology*, **22**, 163–171.

Harrison S, Hastings A (1996) Genetic and evolutionary consequences of metapopulation structure. *Trends in Ecology and Evolution*, **11**, 180–183.

Hawley DM, Sydenstricker K V, Kollias G V, Dhondt A a (2005) Genetic diversity predicts pathogen resistance and cell-mediated immunocompetence in house finches. *Biology Letters*, **1**, 326–9.

Hayward MW (2009) The need to rationalize and prioritize threatening processes used to determine threat status in the IUCN Red List. *Conservation Biology*, **23**, 1568–76.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629–636.

Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution*, **56**, 1902–1908.

Hedrick PW (2004) Recent developments in conservation genetics. *Forest Ecology and Management*, **197**, 3–19.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–8.

Hedrick PW (2011) *Genetics of Populations*.

Hedrick PW, Kalinowski ST (2000) Inbreeding Depression in Conservation Biology. *Annual Review of Ecology and Systematics*, **31**, 139–162.

Heikkinen RK, Luoto M (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **6**, 751–777.

Hellgren O, Waldenström J, Bensch S (2004) A new PCR assay for simultaneous studies of Leucocytozoon, Plasmodium, and Haemoproteus from avian blood. *The Journal of Parasitology*, **90**, 797–802.

Helyar SJ, Hemmer-Hansen J, Bekkevold D *et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources*, 123–36.

Hendrickx F, Maelfait J-P, Van Wingerden W *et al.* (2007) How landscape structure, land-use intensity and habitat diversity affect components of total arthropod diversity in agricultural landscapes. *Journal of Applied Ecology*, **44**, 340–351.

Hengeveld R, Haeck J (1982) The distribution of abundance. I. Measurements. *Journal of Biogeography*, **9**, 303–316.

Hennique S, Mourgaud G, Deceuninck B, Chanson C (2013) *Deuxième plan national d'actions en faveur du Râle des genêts (Crex crex) 2013-2018*. LPO, LPO Anjou, Ministère de l'Écologie, du Développement Durable et de l'Énergie, DREAL des Pays de la Loire.

Herzog SK, Maillard Z. O, Embert D, Caballero P, Quiroga D (2012) Range size estimates of Bolivian endemic bird species revisited: the importance of environmental data and national expert knowledge. *Journal of Ornithology*, **153**, 1189–1202.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans RJ (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, **93**, 679–688.

Hijmans RJ, Elith J (2012) *Species distribution modeling with R*. http://cran.r-

project.org/web/packages/dismo/vignettes/dm.pdf.

Hill MP, Hoffmann AA, Macfadyen S, Umina PA, Elith J (2012) Understanding niche shifts: using current and historical data to model the invasive redlegged earth mite, Halotydeus destructor. *Diversity and Distributions*, **18**, 191–203.

Hinsley S., Bellamy P. (2000) The influence of hedge structure, management and landscape context on the value of hedgerows to birds: A review. *Journal of Environmental Management*, **60**, 33–49.

Hirzel A, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027 – 2036.

Hochachka WM, Dhondt AA (2000) Density-dependent decline of host abundance resulting from a new infectious disease. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 5303–6.

Hodkinson ID (1999) Species response to global environmental change or why ecophysiological models are important: a reply to Davis et al. *Journal of Animal Ecology*, **68**, 1259–1262.

Hoekstra J, Boucher T, Ricketts T, Roberts C (2005) Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, **8**, 23–29.

Hoffmann M, Hilton-Taylor C, Angulo A *et al.* (2010) The impact of conservation on the status of the world's vertebrates. *Science*, **330**, 1503–1509.

Hoffmann AA, Blows MW (1994) Species borders: Ecological and evolutionary perspectives. *Trends in Ecology and Evolution*, **9**, 223–227.

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Horan RD, Shogren JF, Gramig BM (2008) Wildlife conservation payments to address habitat fragmentation and disease risks. *Environment and Development Economics*, **13**, 415–439.

Hoshen MB, Morse AP (2004) A weather-driven model of malaria transmission. *Malaria journal*, **3**, 32.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology*, **9**, 1322–1332.

Hudson A V, Stowe TJ, Aspinall SJ (1990) Status and distribution of corncrakes in Britain in 1988. *British Birds*, **83**, 173–187.

Hudson R (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Huete A, Didan K, Miura T *et al.* (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195–213.

Hutchinson G (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–27.

Iglecia M, Collazo J, McKerrow A (2012) Use of Occupancy Models to Evaluate Expert Knowledge-based Species-Habitat Relationships. *Avian Conservation and Ecology*, **7**.

Ihlow F, Dambach J, Engler JO *et al.* (2012) On the brink of extinction? How climate change may affect global chelonian species richness and distribution. *Global Change Biology*, **18**, 1520–1530.

Ilgūnas M, Palinauskas V, Iezhova TA, Valkiūnas G (2013) Molecular and morphological characterization of two avian malaria parasites (Haemosporida: Plasmodiidae), with description of Plasmodium homonucleophilum n. sp. *Zootaxa*, **3666**, 49–61.

Ippi S, Vasquez RA, van Dongen WFD, Lazzoni I (2011) Geographical variation in the vocalizations of the suboscine Thorn-tailed Rayadito Aphrastura spinicauda. *Ibis*, **153**, 789–805.

Isaksson C, Sepil I, Baramidze V, Sheldon BC (2013) Explaining variance of avian malaria infection in the wild: the importance of host density, habitat, individual life-history and oxidative stress. *BMC Ecology*, **13**.

Isler ML, Isler PR, T. RB (2005) Clinal variation in vocalizations of an antbird (thamnophilidae) and implications for defining species limits. *The Auk*, **122**, 433–444.

IUCN (2001) *IUCN Red List Categories and Criteria, Version 3.1*. IUCN – The World Conservation Union, Gland, Switzerland.

IUCN (2005) Resolution 3.013: The Uses of the IUCN Red List of Threatened Species. In: *Resolutions and Recommendations: World Conservation Congress, Bangkok, Thailand, 17–25 November 2004* , pp. 14–16. IUCN ed.

IUCN Standards and Petitions Subcommittee (2010) *Guidelines for Using the IUCN Red List Categories and Criteria. Version 8.1*. Standards and Petitions Subcommittee, <http://intranet.iucn.org/webfiles/doc/SSC/RedList/RedListGuidelines.pdf>.

Jadwiszczak P (2009) Rundom Pro 3.14. Software for classical and computer-intensive statistics.

Jaffee B, Phillips R, Muldoon A, Mangel M (1992) Density-dependent host-pathogen dynamics in soil microcosms. *Ecology*, **73**, 495–506.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

James FC, Johnston RF, Wamer NO, Niemi GJ, Boecklen WJ (1984) The Grinnellian Niche of the Wood Thrush. *The American Naturalist*, **124**, 17–47.

Jeanes C, Vaughan-Higgins R, Green RE *et al.* (2013) Two new Eimeria species parasitic in Corncrakes (Crex crex) (Gruiformes: Rallidae) in the United-Kingdom. *Journal of Parasitology*, **99**, 634–638.

Jenni L, Kéry M (2003) Timing of autumn bird migration under climate change: advances in long-distance migrants, delays in short-distance migrants. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 1467–71.

Jenouvrier S (2013) Impacts of climate change on avian populations. *Global Change Biology*, **19**, 2036–2057.

Jeschke J, Strayer D (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, **1134**, 1–24.

Jetz W, Sekercioglu CH, Watson JEM (2008) Ecological Correlates and Conservation Implications of Overestimating Species Geographic Ranges. *Conservation Biology*, **22**, 110–119.

Jiménez-Alfaro B, Draper D, Nogués-Bravo D (2012) Modeling the potential area of occupancy at fine resolution may reduce uncertainty in species range estimates. *Biological Conservation*, **147**, 190–196.

Jimenez-Valverde A, Peterson AT, Soberón J *et al.* (2011) Use of niche models in invasive species risk assessments. *Biological Invasions*, **13**, 2785–2797.

Jimenez-Valverde A, Acevedo P, Barbosa a. M, Lobo JM, Real R (2012) Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*, **22**, 508–516.

Jimenez-Valverde A (2014) Threshold-dependence as a desirable attribute for discrimination assessment: implications for the evaluation of species distribution models. *Biodiversity and Conservation*, **23**, 369–385.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–5.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, **11**, 94.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, **14**, 401–413.

Kamp J, Urazaliev R, Donald PF, Hölzel N (2011) Post-Soviet agricultural change predicts future declines after recent recovery in Eurasian steppe bird populations. *Biological Conservation*, **144**, 2607–2614.

Kark S, Hadany L, Safriek U *et al.* (2008) How does genetic diversity change towards the range periphery? An empirical and theoretical test. *Evolutionary Ecology Research*, **10**, 391–414.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Keišs O (2003) Recent increases in numbers and the future of Corncrake Crex crex in Latvia. *Ornis Hungarica*, **12-13**, 151–156.

Keišs O (2004) Results of a survey of Corncrake Crex crex in Latvia, 1989-1995. In: *Proceedings of the International Conference and 13th Meeting of the European Bird Census Council* , pp. 73–76. Bird Census News, Pärnu.

Keišs O, Granāts J, Mednis A (2004) Use of biometrical data to study Corncrake Crex crex population in Latvia. *Acta Universitatis Latviensis*, **676**, 119–126.

Keišs O (2005) Impact of changes in agricultural land use on the Corncrake Crex crex population in Latvia. *Acta Universitatis Latviensis*, **691**, 93–109.

Keller LF (2002) Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, **17**, 230–241.

Kenta T, Gratten J, Haigh NS *et al.* (2008) Multiplex SNP-SCALE: a cost-effective medium-throughput single nucleotide polymorphism genotyping method. *Molecular Ecology Resources*, **8**, 1230–8.

Keyel AC, Strong AM, Perlut NG, Reed JM (2013) Evaluating the roles of visual openness and edge effects on nest-site selection and reproductive success in grassland birds. *The Auk*, **130**, 161–170.

Kilpatrick A, LaPointe D, Atkinson CT *et al.* (2006a) Effects of chronic avian malaria (Plasmodium relictum) infection on reproductive success of Hawaii Amakihi (Hemignathus virens). *The Auk*, **123**, 764–774.

Kilpatrick A, Chmura AA, Gibbons DW *et al.* (2006b) Predicting the global spread of H5N1 avian influenza. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19368–73.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications*, **13**, 235–248.

Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *American Naturalist*, **150**, 1–23.

Klimek S, Richter A, Hofmann M, Isselstein J (2006) Plant species richness and composition in managed grasslands : The relative importance of field management and environmental factors. *Biological Conservation*, **134**, 559–570.

Knowles SCL, Palinauskas V, Sheldon BC (2010) Chronic malaria infections increase family inequalities and reduce parental fitness: experimental evidence from a wild bird population. *Journal of Evolutionary Biology*, **23**, 557–69.

Knowles SCL, Wood MJ, Alves R *et al.* (2011) Molecular epidemiology of malaria prevalence and parasitaemia in a wild bird population. *Molecular Ecology*, **20**, 1062–1076.

Koffijberg K, Schäffer N (2005) Species Action Plan Corncrake Crex crex. *BirdLife International*.

Koffijberg K, Schäffer N (2006) International Single Species Action Plan for the Conservation of the Corncrake Crex crex. *CMS/AEWA Technical Series*, **14-9**.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in ecology & evolution*, **23**, 141–8.

Kramer-Schadt S, Niedballa J, Pilgrim JD *et al.* (2013) The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*.

Krone O, Waldenström J, Valkiūnas G *et al.* (2008) Haemosporidian blood parasites in European birds of prey and owls. *The Journal of parasitology*, **94**, 709–15.

Lachish S, Knowles SCL, Alves R, Wood MJ, Sheldon BC (2011) Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *The Journal of Animal Ecology*, **80**, 1196–206.

Lachish S, Knowles SCL, Alves R *et al.* (2013) Spatial determinants of infection risk in a multi-species avian malaria system. *Ecography*, **36**, 587–598.

Lafage D, Secondi J, Georges A, Bouzillé J-B, Pétillon J (2013) Satellite-derived vegetation indices as surrogate of species richness and abundance of ground beetles in temperate floodplains (A Stewart, R Ewers, Eds,). *Insect Conservation and Diversity*.

Lambeck R (1997) Focal species: a multi-species umbrella for nature conservation. *Conservation biology*, **11**, 849–856.

Lambrechts M (1996) Organization of birdsong and constraints on performance. In: *Ecology and evolution of acoustic communication in birds* (eds Kroodsma DE, Miller EH), pp. 305–320. Cornell Univ. Press.

Lanciani C (1975) Parasite-induced alterations in host reproduction and survival. *Ecology*, **56**, 689–695.

Lande R (1988) Genetics and demography in biological conservation. *Science*, **241**, 1455.

Lande R (1993) Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *American Naturalist*, **142**, 911–927.

Lande R (1994) Risk of population extinction from fixation of new deleterious mutations. *Evolution*, **48**, 1460–1469.

Latta SC, Ricklefs RE (2010) Prevalence patterns of avian haemosporida on Hispaniola. *Journal of Avian Biology*, **41**, 25–33.

Lebreton JD, Hines JE, Pradel R, Nichols JD, Spendelow JA (2003) Estimation by capture-recapture of recruitment and dispersal over several sites. *Oikos*, **101**, 253–264.

Lein MR (2008) Song Variation In Buff-Breasted Flycatchers (Empidonax fulvifrons). *The Wilson Journal of Ornithology*, **120**, 256–267.

Leitão PJ, Moreira F, Osborne PE (2011) Effects of geographical data sampling bias on habitat models of species distributions: a case study with steppe birds in southern Portugal. *International Journal of Geographical Information Science*, **25**, 439–454.

Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.

Lesica P, Allendorf F (1995) When are peripheral populations valuable for conservation? *Conservation Biology*, **9**, 753–760.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–95.

Lima S (1986) Predation risk and unpredictable feeding conditions: determinants of body mass in birds. *Ecology*, **67**, 377–385.

Liu C, Berry P, Dawson T, Pearson R (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Loaiza JR, Miller MJ (2013) Seasonal pattern of avian Plasmodium-infected mosquitoes and implications for parasite transmission in central Panama. *Parasitology Research*, **112**, 3743–51.

Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

Loiseau C, HARRIGAN R, Robert A *et al.* (2012) Host and habitat specialization of avian malaria in Africa. *Molecular Ecology*, **21**, 431–441.

Loiselle B, Jørgensen P, Consiglio T *et al.* (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *Journal of Biogeography*, **35**, 105–116.

Lopes JS, Boessenkool S (2009) The use of approximate Bayesian computation in conservation genetics and its application in a case study on yellow-eyed penguins. *Conservation Genetics*, **11**, 421–433.

Luikart G, Sherwin W, Steele B, Allendorf F (1998) Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology*, **7**, 963–974.

Luikart G, England PR, Tallmon D a, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–94.

MacDougall-Shackleton EA, Derryberry EP, Foufopoulos J, Dobson AP, Hahn TP (2005) Parasite-mediated heterozygote advantage in an outbred songbird population. *Biology Letters*, **1**, 105–7.

Mace GM, Collar NJ, Gaston KJ *et al.* (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, **22**, 1424–42.

Maes D, Van Dyck H (2001) Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biological Conservation*, **99**, 263–276.

Mager JN, Walcott C, Evers D (2007) Macrogeographic Variation in the Body Size and Territorial Vocalizations of Male Common Loons (Gavia immer). *Waterbirds*, **30**, 64–72.

Maisongrande P, Duchemin B, Dedieu G (2004) VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation

of new standard products. *International Journal of Remote Sensing*, **25**, 9–14.

Manly BFJ, McDonald LL, Thomas D (1993) *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Chapman & Hall.

Martel A, Spitzen-van der Sluijs A, Blooi M *et al.* (2013) Batrachochytrium salamandrivorans sp. nov. causes lethal chytridiomycosis in amphibians. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15325–9.

Martin TG, Burgman M a., Fidler F *et al.* (2012) Eliciting Expert Knowledge in Conservation Science. *Conservation Biology*, **26**, 29–38.

Martinsen ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites (Plasmodium and closely related genera): evolution of life-history traits and host switches. *Molecular Phylogenetics and Evolution*, **47**, 261–73.

Marzal A, Albayrak T (2012) Geographical variation of haemosporidian parasites in Turkish populations of Krüper's Nuthatch Sitta krueperi. *Journal of Ornithology*, **153**, 1225–1231.

Matthews SN, Iverson LR, Prasad AM, Peters MP (2011) Changes in potential habitat of 147 North American breeding bird species in response to redistribution of trees and climate following predicted climate change. *Ecography*, **34**, 933–945.

May L (1998) Individually distinctive corncrake Crex crex calls : A further study. *Bioacoustics*, **9**, 135–148.

Mayaud N (1936) *Inventaire des oiseaux de France*. Société d'Etudes Ornithologiques, Paris.

McBride MF, Fidler F, Burgman M a. (2012) Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research. *Diversity and Distributions*, **18**, 782–794.

McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modelled? *Trends in ecology & evolution*, **16**, 295–300.

McCallum H, Dobson A (1995) Detecting disease and parasite threats to endangered species and ecosystems. *Trends in Ecology & Evolution*, 190–194.

McCallum H, Dobson A (2002) Disease, habitat fragmentation and conservation. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 2041–2049.

McCarthy K, Fletcher RJ (2012) Predicting Species Distributions from Samples Collected along Roadsides. *Conservation Biology*, **26**, 68–77.

McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman & Hall, London, UK.

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Meyrowitsch DW, Pedersen EM, Alifrangis M *et al.* (2011) Is the current decline in malaria burden in sub-Saharan Africa due to a decrease in vector population? *Malaria journal*, **10**, 188.

Mikkelsen G, Dale S, Holtskog T (2013) Can individually characteristic calls be used to identify long-distance movements of Corncrakes Crex crex? *Journal of Ornithology*, **154**, 751–760.

Ming M, Qishan W (2002) New records of Corncrake Crex crex in Xinjiang, China. *Forktail*, **18**, 158–158.

Mischenko A (2008) Corncrake Crex crex monitoring in European Russia in 2002-2003: a pilot study. *Revista Catalana d'Ornitologia*, 65–70.

Mittermeier R, Myers N, Thomsen J, Da Fonseca G, Olivieri S (1998) Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology*, **12**, 516–520.

Moilanen A (2005) Reserve selection using nonlinear species distribution models. *The American Naturalist*, **165**, 695–706.

Moilanen A, Franco AM a, Early RI *et al.* (2005) Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1885–91.

Monahan WB, Hijmans RJ (2008) Ecophysiological constraints shape autumn migratory response to climate change in the North American field sparrow. *Biology letters*, **4**, 595–598.

Morgenstern H (1982) Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health*, **72**, 1336–44.

Moritz C (1994) Defining "evolutionarily significant units" for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Morris M (2000) The effects of structure and its dynamics on the ecology and conservation of arthropods in British grasslands. *Biological Conservation*, **95**, 129–142.

Morton ES (1975) Ecological Sources of Selection on Avian Sounds. *The American Naturalist*, **109**, 17.

Mundinger P (1982) Microgeographic and macrogeographic variation in the acquired vocalizations of birds. In: *Ecology and evolution of acoustic communication in birds* (eds Kroodsma DE, Miller EH), pp. 147–208. Cornell Univ. Press.

Murray J V., Goldizen AW, O'Leary RA *et al.* (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies Petrogale penicillata. *Journal of Applied Ecology*, **46**, 842–851.

Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nakazawa Y, Peterson A, Martinez-Meyer E, Navarro-Sigüenza AG (2004) Seasonal niches of Nearctic-Neotropical migratory birds: implications for the evolution of migration. *The Auk*, **121**, 610–618.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3.

Neke K, Plessis M Du (2004) The threat of transformation: quantifying the vulnerability of grasslands in South Africa. *Conservation Biology*, **18**, 466–477.

Nelson A (2008) *Travel time to major cities: A global map of Accessibility*. Office for Official Publications of the European Communities, Luxembourg.

Nicholls JA, Goldizen AW (2006) Habitat type and density influence vocal signal design in satin bowerbirds. *The Journal of Animal Ecology*, **75**, 549–558.

Nichols RA, Bruford MW, Groombridge JJ (2001) Sustaining genetic variation in a small population: evidence from the Mauritius kestrel. *Molecular Ecology*, **10**, 593–602.

Njabo KY, Cornel a J, Bonneaud C *et al.* (2011) Nonspecific patterns of vector, host and avian malaria parasite associations in a central African rainforest. *Molecular Ecology*, **20**, 1049–61.

Norris C (1947) Report on the distribution and status of the Corncrake. *British Birds*, **40**, 226–244.

Noss RF (1990) Indicators for Monitoring Biodiversity: A Hierarchical Approach. *Conservation Biology*, **4**, 355–364.

Nottebohm F (1969) The song of the chingolo (Zonotrichia capensis) in Argentina: description and evaluation of a system. *The Condor*, **71**, 299–315.

NPWS (2005) *All-Ireland Species Action Plan: Corncrake Crex crex*. Environment & Heritage Service, Northern Ireland and National Parks & Wildlife Service, Ireland.

O'Brien M, Green RE, Wilson JD (2006) Partial recovery of the population of Corncrakes Crex crex in Britain, 1993-2004. *Bird Study*, **53**, 213–224.

O'Grady JJ, Brook BW, Reed DH *et al.* (2006) Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservation*, **133**, 42–51.

O'Leary RA, Choy SL, Murray J V. *et al.* (2009) Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. *Environmetrics*, **20**, 379–398.

Ockendon N, Hewson CM, Johnston A, Atkinson PW (2012) Declines in British-breeding populations of Afro-Palaearctic migrant birds are linked to bioclimatic wintering zone in Africa, possibly via constraints on arrival time advancement. *Bird Study*, **59**, 111–125.

Odom KJ, Mennill DJ (2012) Inconsistent geographic variation in the calls and duets of Barred Owls (Strix varia) across an area of genetic introgression. *The Auk*, **129**, 387–398.

Olesen JE, Bindi M (2002) Consequences of climate change for European agricultural productivity,

land use and policy. *European Journal of Agronomy*, **16**, 239–262.

Ortega-Huerta MA, Peterson AT (2004) Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions*, **10**, 39–54.

Ortego J, Cordero PJ, Aparicio JM, Calabuig G (2007) No relationship between individual genetic diversity and prevalence of avian malaria in a migratory kestrel. *Molecular Ecology*, **16**, 4858–4866.

Osborne PE, Suárez-Seoane S (2002) Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, **157**, 249–259.

Osiejuk TS, Olech B, Ratyńska K, Owsiński A, Gromadzka-Ostrowska J (2004) Effects of season, plasma testosterone and body size on corncrake (Crex crex) call rhythm. *Annales Zoologici Fennici*, **41**, 647–659.

Osiejuk TS, Olech B (2004) Amplitude spectra of corncrake calls: What do they signalise? *Animal Biology*, **54**, 207–220.

Ottvall R (1999) Female Corncrake (Crex crex) singing in the wild. *Journal fur Ornithologie*, **140**, 453–456.

Oyler-McCance SJS, Kahn NW, Burnham KPK *et al.* (1999) A population genetic comparison of large- and small- bodied sage grouse in Colorado using microsatellite and mitochondrial DNA markers. *Molecular Ecology*, **8**, 1457–1465.

Ozaki K, Isono M, Kawahara T, Iida S, Kudo T (2006) A Mechanistic Approach to Evaluation of Umbrella. *Conservation Biology*, **20**, 1507–1515.

Palsbøll PJ, Bérubé M, Allendorf FW (2007) Identification of management units using population genetic data. *Trends in ecology & evolution*, **22**, 11–6.

Paquette SR, Garant D, Pelletier F, Bélisle M (2013) Seasonal patterns in Tree Swallow prey (Diptera) abundance are affected by agricultural intensification. *Ecological Applications*, **23**, 122–133.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Peake TM, McGregor PK, Smith KW *et al.* (1998) Individuality in Corncrake Crex crex vocalizations. *Ibis*, **140**, 120–127.

Peake TM, McGregor PK (1999) Geographical variation in the vocalisation of the corncrake Crex crex. *Ethology Ecology and Evolution*, **11**, 123–137.

Peake TM, McGregor PK (2001) Corncrake Crex crex census estimates: A conservation application of vocal individuality. *Animal Biodiversity and Conservation*, **24**, 81–90.

Pearce JL, Cherry K, M. D, S. F, Whish G (2001) Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *Journal of Applied Ecology*, **38**, 412–424.

Perera AH, Drew CA, Johnson CJ (Eds.) (2012) *Expert Knowledge and Its Application in Landscape Ecology*. Springer, New York.

Pérez-Rodríguez A, Fernández-González S, de la Hera I, Pérez-Tris J (2013) Finding the appropriate variables to model the distribution of vector-borne parasites with different environmental preferences: climate is not enough. *Global change biology*, **19**, 3245–53.

Perras M, Nebel S (2012) Satellite Telemetry and its Impact on the Study of Animal Migration. *Nature Education Knowledge*, **3**, 4.

Peterson A (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, 59–72.

Peterson A, Soberón J, Pearson R *et al.* (2011) *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ, USA.

Phalan B, Green RE, Balmford A, B PTRS (2014) Closing yield gaps : perils and possibilities for biodiversity conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In: *Proceedings of the twenty-first international conference on Machine learning* , p. 83. ACM, New York, New York, USA.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species

geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips SJ (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography*, **31**, 272–278.

Phillips SJ, Dudík M, Elith J *et al.* (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Pinto J, Sousa CA, Gil V *et al.* (2000) Malaria in São Tomé and Príncipe: parasite prevalences and vector densities. *Acta Tropica*, **76**, 185–193.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502–503.

Plowright RK, Sokolow SH, Gorman ME, Daszak P, Foley JE (2008) Causal inference in disease ecology: investigating ecological drivers of disease emergence. *Frontiers in Ecology and the Environment*, **6**, 420–429.

Podos J, Warren PS (2007) The Evolution of Geographic Variation in Birdsong. *Advances in the Study of Behavior*, **37**, 403–458.

Possingham HP, Ball I, Andelman S (2000) Mathematical methods for identifying representative reserve netorks. In: *Quantitative methods for conservation biology* (eds Ferson S, Burgman M), pp. 291–305. Springer-Verlag, New York.

Potts SG, Biesmeijer JC, Kremen C *et al.* (2010) Global pollinator declines : trends, impacts and drivers. *Trends in Ecology & Evolution*, **25**, 345–353.

Potvin D a (2013) Larger body size on islands affects silvereye Zosterops lateralis song and call frequency. *Journal of Avian Biology*, **44**, 221–225.

Pounds JA, Bustamante MR, Coloma LA *et al.* (2006) Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature*, **439**, 161–7.

Pride G, Swift S (1992) *Wildlife telemetry : remote monitoring and tracking animals*. Ellis Horwood, New York.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–59.

La Puente JM, Merino S, Tomás G *et al.* (2010) The blood parasite Haemoproteus reduces survival in a wild bird: a medication experiment. *Biology Letters*, **6**, 663–5.

Pulliam R (1988) Source, sinks and population regulation. *The American Naturalist*, **132**, 652–661.

Pulliam HR (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1947–1952.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing.

Radwan J, Biedrzycka A, Babik W (2010) Does reduced MHC diversity decrease viability of vertebrate populations? *Biological Conservation*, **143**, 537–544.

Ralls K, Ballou JD, Templeton A (1988) Estimates of Lethal Equivalents and the Cost of Inbreeding in Mammals. *Conservation Biology*, **2**, 185–193.

Raxworthy CJ, Martinez-Meyer E, Horning N *et al.* (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837–41.

Read AF, Weary DM (1990) Sexual selection and the evolution of bird song: A test of the Hamilton-Zuk hypothesis. *Behavioral Ecology and Sociobiology*, **26**, 47–56.

Reddy S, Davalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.

Reed BC, Brown JF, VanderZee D *et al.* (1994) Measuring Phenological Variability from Satellite Imagery. *Journal of Vegetation Science*, **5**, 703–714.

Reed DH, Frankham R (2003) Correlation between Fitness and Genetic Diversity. *Conservation Biology*, **17**, 230–237.

Reed K, Meece J (2003) Birds, migration and emerging zoonoses: West Nile virus, Lyme

disease, influenza A and enteropathogens. *Clinical Medicine & Research*, **1**, 5–12.

Reid W (1998) Biodiversity hotspots. *Trends in Ecology & Evolution*, **13**, 275–280.

Ręk P (2013a) Corncrake males learn new signal meanings during aggressive interactions. *Animal Behaviour*, 1–7.

Ręk P (2013b) Acoustic location of conspecifics in a nocturnal bird: the corncrake Crex crex. *Acta Ethologica*.

Ręk P (2013c) Soft calls and broadcast calls in the corncrake as adaptations to short and long range communication. *Behavioural processes*, **99**, 121–129.

Rek P, Osiejuk TS (2011) Nonpasserine bird produces soft calls and pays retaliation cost. *Behavioral Ecology*, **22**, 657–662.

Rek P, Osiejuk TS (2013) Temporal patterns of broadcast calls in the corncrake encode information arbitrarily. *Behavioral Ecology*, **24**, 547–552.

Ręk P, Osiejuk TS (2010) Sophistication and simplicity: Conventional communication in a rudimentary system. *Behavioral Ecology*, **21**, 1203–1210.

Ręk P, Osiejuk TS (2011) No male identity information loss during call propagation through dense vegetation: The case of the corncrake. *Behavioural processes*.

Renner IW, Warton DI (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274–81.

Richardson DS, Jury FL, Blaakmeer K, Komdeur J, Burke T (2001) Parentage assignment and extra-group paternity in a cooperative breeder: the Seychelles warbler (Acrocephalus sechellensis). *Molecular Ecology*, **10**, 2263–73.

Ridgeway G (1999) The state of boosting. *Computing Science and Statistics*, **31**, 172–181.

Rintamäki P, Ojanen M (1998) Blood parasites of migrating willow warblers (Phylloscopus trochilus) at a stopover site. *Canadian journal of Zoology*, **988**, 984–988.

Van Riper III C, Van Riper SG, Lee Goff M, Laird M (1986) The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs*, **56**, 327–344.

Ripley BD (1996) *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK.

Roberge J, Angelstam P (2004) Usefulness of the umbrella species concept as a conservation tool. *Conservation Biology*, **18**, 76–85.

Robinson JD, Simmons JW, Williams AS, Moyer GR (2012) Population structure and genetic diversity in the endangered bluemask darter (Etheostoma akatulo). *Conservation Genetics*, **14**, 79–92.

Rödder D, Schmidtlein S, Veith M, Lötters S (2009a) Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? *PloS one*, **4**, e7843.

Rödder D, Kielgast J, Bielby J *et al.* (2009b) Global Amphibian Extinction Risk Assessment for the Panzootic Chytrid Fungus. *Diversity*, **1**, 52–66.

Rödder D, Engler JO (2011) Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography*, **20**, 915–927.

Rödder D, Lötters S (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (Hemidactylus turcicus). *Global Ecology and Biogeography*, **18**, 674–687.

Rodrigues ASL, Pilgrim JD, Lamoreux JF, Hoffmann M, Brooks TM (2006) The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution*, **21**, 71–6.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–6.

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution*, **1**, 9–10.

Saccheri I, Kuussaari M, Kankare M *et al.* (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491–494.

Sagarin RD, Gaines SD, Gaylord B (2006) Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecology & Evolution*, **21**, 524–30.

Sagarin RD, Gaines SD (2002) The "abundant centre" distribution: to what extent is it a biogeographical rule? *Ecology Letters*, **5**, 137–147.

Samis KE, Eckert CG (2007) Testing the abundant center model using range-wide demographic surveys of two coastal dune plants. *Ecology*, **88**, 1747–1758.

Sanderson FJ, Kucharz M, Jobda M, Donald PF (2013) Impacts of agricultural intensification and abandonment on farmland birds in Poland following EU accession. *Agriculture, Ecosystems & Environment*, **168**, 16–24.

Saupe EE, Barve V, Myers CE *et al.* (2012) Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, **237-238**, 11–22.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–20.

Schaffer N (1999) Habitatwahl und Partnerschaftssystem von Tüpfelralle Porzana porzana und Wachtelkönig Crex crex. *Ökologie der Vögel*, **21**, 1–267.

Schäffer N (1995) Calling behaviour and functions of calls in the Corncrake Crex crex. *Vogelwelt*, **116**, 141–151.

Schäffer N, Green RE (2001) The Global Status of the Corncrake. *RSPB Conservation Review*, **13**, 18–24.

Schäffer N, Koffijberg K (2004) Crex crex Corncrake. *Bwp Update*, **6**, 57–78.

Schäffer N, Weisser WW (1996) A strategy for the conservation of the Corncrake Crex crex. *Journal fur Ornithologie*, **137**, 53–75.

Scheuerlein A, Ricklefs RE (2004) Prevalence of blood parasites in European passeriform birds. *Proceedings of the Royal Society of London B: Biological Sciences*, **271**, 1363–70.

Schoener T (1968) The Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology*, **49**, 704–726.

Schrag S, Wiener P (1995) Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends in Ecology & Evolution*, **10**, 319–324.

Schulte-Hostedde A, Zinner B, Millar J, Hickling G (2005) Restitution of mass-size residuals: validating body condition indices. *Ecology*, **86**, 155–163.

Schwarz G (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461–464.

Seddon PJ, Leech T (2008) Conservation short cut, or long and winding road? A critique of umbrella species criteria. *Oryx*, **42**, 240–245.

Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–29.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 415–436.

Sharrock J (1976) *The Atlas of Breeding Birds in Britain and Ireland*. Poyser.

Shobrak M (2012) Electrocution and collision of birds with power lines in Saudi Arabia. *Zoology in the Middle East*, **56**, 45–52.

Simberloff D (1998) Flagships, umbrellas, and keystones: is single-species management passé in the landscape era? *Biological conservation*, **83**, 247–257.

Sinclair S, White M, Newell G (2010) How useful are species distribution models for managing biodiversity under future climates. *Ecology and Society*, **15**, 8.

Sing T, Sander O, Beerenwinkel N, Lengauer T (2012) ROCR: Visualizing the performance of scoring classifiers.

Siriwardena GM, Cooke IR, Sutherland WJ (2011) Landscape, cropping and field boundary influences on bird abundance. *Ecography*, **35**, 162–173.

Slabbekoorn H, Smith TB (2002) Bird song, ecology and speciation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **357**, 493–503.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–62.

Smith K, Acevedo-Whitehouse K, Pedersen a. B (2009) The role of infectious diseases in biological conservation. *Animal Conservation*, **12**, 1–12.

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–23.

Soberón J, Nakamura M (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences*, **106**, 19644–19650.

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Soberón J, Peterson AT (2011) Ecological niche shifts and environmental space anisotropy : a cautionary note. *Revisita Mexicana de Biodiversidad*, **82**, 1348–1355.

Sol D, Jovani R, Torres J (2000) Geographical variation in blood parasites in feral pigeons: the role of vectors. *Ecography*, **23**, 307–314.

Sosa-López JR, Mennill DJ, Navarro-Sigüenza AG (2013) Geographic variation and the evolution of song in Mesoamerican rufous-naped wrens Campylorhynchus rufinucha. *Journal of Avian Biology*, **44**, 27–38.

Specht R (2007) *Avisoft-SASLab Pro version 4.39*. Bioacoustics, Avisoft, Berlin, Germany.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004) Does Inbreeding and Loss of Genetic Diversity Decrease Disease Resistance? *Conservation Genetics*, **5**, 439–448.

Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 979–88.

Steere AC, Coburn J, Glickstein L (2004) The emergence of Lyme disease. *The Journal of clinical investigation*, **113**, 1093–101.

Stephens P a., Buskirk SW, Hayward GD, Martinez Del Rio C (2005) Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology*, **42**, 4–12.

Stiels D, Schidelko K, Engler JO, Elzen R, Rödder D (2011) Predicting the potential distribution of the invasive Common Waxbill Estrilda astrild (Passeriformes: Estrildidae). *Journal of Ornithology*, 769–780.

Stoate C, Báldi A, Beja P *et al.* (2009) Ecological impacts of early 21st century agricultural change in Europe--a review. *Journal of Environmental Management*, **91**, 22–46.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94–101.

Stowe TJ, Green RE (1997a) Response of Corncrake Crex crex populations in Britain to conservation action. *Vogelwelt*, **118**, 161–168.

Stowe TJ, Green RE (1997b) Threats to the Corncrake Crex crex on migration and in the winter quarters. *Vogelwelt*, **118**, 175–178.

Sukhanova O, Mischenko A (2003) Monitoring Corncrake Crex crex numbers in European Russia: the first stage. *Ornis Hungarica*, **12**, 135–141.

Sunnåker M, Busetto AG, Numminen E *et al.* (2013) Approximate Bayesian Computation (S Wodak, Ed,). *PLoS Computational Biology*, **9**, e1002803.

Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. *Trends in ecology & evolution*, **19**, 305–8.

Svoboda A, Marthinsen G, Turčoková L, Lifjeld J, Johnsen A (2009) Identification of blood parasites in old world warbler species from the Danube River Delta. *Avian Diseases*, **53**, 634–636.

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–93.

Syfert MM, Smith MJ, Coomes DA (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PloS one*, **8**, e55158.

Szöllősi E, Cichoń M, Eens M *et al.* (2011) Determinants of distribution and prevalence of

avian malaria in blue tit populations across Europe: separating host and parasite effects. *Journal of Evolutionary Biology*, **24**, 2014–24.

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: a time for reappraisal. *Evolution*, **64**, 1202–1217.

Templeton AR (1999) Using gene trees to infer species from testable null hypothesis: cohesion species in the Spalax ehrenbergi complex. In: *Evolutionary Theory and Processes: Modern Perspectives, Papers in Honour of Eviatar Nevo* (ed Wasser S), pp. 171–192. Kluwer Academic Publishers, Dordrech.

Terry AMR, Peake TM, McGregor PK (2005) The role of vocal individuality in conservation. *Frontiers in Zoology*, **2**, 10.

Thomas J, Telfer M, Roy D *et al.* (2004) Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science*, **303**, 1879–1881.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–80.

Thorn JS, Nijman V, Smith D, Nekaris KAI (2009) Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). *Diversity and Distributions*, **15**, 289–298.

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

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Tingley MW, Monahan WB, Beissinger SR, Moritz C (2009) Birds track their Grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **106 Suppl** , 19637–19643.

Tockner K, Stanford JA (2002) Riverine flood plains: present state and future trends. *Environmental Conservation*, **29**, 308–330.

Tôrres NM, De Marco P, Santos T *et al.* (2012) Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. *Diversity and Distributions*, **18**, 615–627.

Townsend AK, Clark AB, McGowan KJ *et al.* (2009) Disease-mediated inbreeding depression in a large, open population of cooperative crows. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 2057–64.

Trape JF, Lefebvre-Zante E, Legros F *et al.* (1992) Vector density gradients and the epidemiology of urban malaria in Dakar, Senegal. *The American Journal of Tropical Medicine and Hygiene*, **47**, 181–189.

Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.

Tucker M, Heath MF, Tomialojc L, Grimmett R (1994) *Birds in Europe: Their Conservation Status*. BirdLife International, Cambridge.

Tyler GA (1996) The Ecology of the Corncrake with special reference to the effect of mowing on breeding. University of Cork.

Tyler GA, Green RE, Casey C (1998) Survival and behaviour of Corncrake Crex crex chicks during the mowing of agricultural grassland. *Bird Study*, **45**, 35–50.

Tyler GA, Green RE (1996) The incidence of nocturnal song by male Corncrakes Crex crex is reduced during pairing. *Bird Study*, **43**, 214–219.

Tyler GA, Green RE (2004) Effects of weather on the survival and growth of Corncrake Crex crex chicks. *Ibis*, **146**, 69–76.

Väli U, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808–3817.

Valkiūnas G (2005) *Avian Malaria Parasites and other Haemosporidia*. CRC Press, Boca Raton, FL, USA.

VanDerWal JJ, Shoo L, Graham CH, Williams S (2009a) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

VanDerWal JJ, Shoo LP, Johnson CN, Williams SE (2009b) Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *The American Naturalist*, **174**, 282–91.

Varela S, Anderson RP, García-Valdés R, Fernández-González F (2014) Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*, 10.1111/j.1600–0587.2013.00441.x.

Veen T, Svedin N, Forsman JT *et al.* (2007) Does migration of hybrids contribute to post-zygotic isolation in flycatchers? *Proceedings of the Royal Society B: Biological Sciences*, **274**, 707–712.

Veloz SD (2009) Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography*, **36**, 2290–2299.

Ventim R, Morais J, Pardal S *et al.* (2012) Host-parasite associations and host-specificity in haemoparasites of reed bed passerines. *Parasitology*, **139**, 310–316.

Verissimo D, MacMillan DC, Smith RJ (2011) Toward a systematic approach for identifying conservation flagships. *Conservation Letters*, **4**, 1–8.

Vickery PD, Herkert JR, Knopf FL, Keller CE (2000) Grassland Birds : An Overview of Threats and Recommended Management Strategies Major Threats to Grassland. In: *Strategies for bird conservation: The Partners in Flight planning process; Proceedings of the 3rd Partners in Flight Workshop* (eds Bonney R, Pashley DN, Cooper RJ, Niles L), pp. 74–77. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Voegeli B, Saladin V, Wegmann M, Richner H (2012) Parasites as mediators of heterozygosity-fitness correlations in the Great Tit (Parus major). *Journal of Evolutionary Biology*, **25**, 584–90.

Vrijenhoek RC (1994) *Genetic diversity and fitness in small populations*. Birkhauser.

Waldenström J, Bensch S, Kiboi S, Hasselquist D, Ottosson U (2002) Cross-species infection of blood parasites between resident and migratory songbirds in Africa. *Molecular Ecology*, **11**, 1545–1554.

Waldenström J, Bensch S, Hasselquist D, Ostman O (2004) A new nested polymerase chain reaction method very efficient in detecting Plasmodium and Haemoproteus infections from avian blood. *The Journal of Parasitology*, **90**, 191–4.

Wallander J, Isaksson D, Lenberg T (2006) Wader nest distribution and predation in relation to man-made structures on coastal pastures. *Biological Conservation*, **132**, 343–350.

Walther G, Post E, Convey P *et al.* (2002) Ecological responses to recent climate change. *Nature*, 389–395.

Walther B (2008) Ringing recoveries ofthe Corncrake Crex crex in Africa and Sinai. *Vogelwelt*, **29**, 103–108.

Walther B a., Barry Taylor P, Schäffer N, Robinson S, Jiguet F (2012) The African wintering distribution and ecology of the Corncrake Crex crex. *Bird Conservation International*, 1–14.

Wang Q, Watanabe M, Hayashi S, Murakami S (2003) Using NOAA AVHRR data to assess flood damage in China. *Environmental monitoring and assessment*, **82**, 119–48.

Waples RS (1991) *Status review for Snake River sockeye salmon*. NOAA Technical mMemorandum NMFS F/NWC.

Waples R (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium*, **17**, 8–27.

Waples RS, Gaggiotti OE (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–39.

Warren PS, Katti M, Ermann M, Brazel A (2006) Urban bioacoustics: It's not just noise. *Animal Behaviour*, **71**, 491–502.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–83.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Westerdahl H, Waldenström J, Hansson B *et al.* (2005) Associations between malaria and MHC genes in a migratory songbird. *Proceedings of the Royal Society of London B: Biological Sciences*, **272**, 1511–8.

Westerdahl H (2007) Passerine MHC: genetic variation and disease resistance in the wild. *Journal of Ornithology*, **148**, 469–477.

Western D (1987) Africa's elephants and rhinos: Flagships in crisis. *Trends in Ecology & Evolution*, **2**, 343–346.

Wettstein W, Szép T, Kéry M (2001) Habitat selection of Corncrakes (Crex crex L.) in Szatmár-Bereg (Hungary) and implications for further monitoring. *Ornis Hungarica*, **11**, 9–18.

Wettstein W (2003) Conservation biology, population structure and large-scale spatial behaviour of corncrakes (Crex crex L.) in Europe (Doctoral dissertation). University of Zurich, Switzerland.

Wettstein W, Szép T (2003) Status of the Corncrake Crex crex as an indicator of biodiversity in eastern Hungary. *Ornis Hungarica*, **12**, 143–149.

Whittingham MJ, John R, Swetnam RD *et al.* (2007) Should conservation strategies consider spatial generality ? Farmland birds show regional not national patterns of habitat association. *Ecology Letters*, **10**, 25–35.

Whittingham MJ, Evans KL (2004) The effects of habitat structure on predation risk of birds in agricultural landscapes. *Ibis*, **146**, 210–220.

Wickramasinghe LP, Harris S, Jones G, Vaughan Jennings N (2004) Abundance and Species Richness of Nocturnal Insects on Organic and Conventional Farms: Effects of Agricultural Intensification on Bat Foraging. *Conservation Biology*, **18**, 1283–1292.

Wilcox BA (1984) In situ conservation of genetic resources: determinants of minimum area requirements. In: *National parks: conservation and development* , pp. 639–647. Smithsonian Institution Press, Washington DC, USA.

Wirth T, Meyer A, Achtman M (2005) Deciphering host migrations and origins by means of their microbes. *Molecular Ecology*, **14**, 3289–306.

Wisz MS, Hijmans RJ, Li J *et al.* (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

Wood MJ, Cosgrove CL, Wilkin TA *et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, Cyanistes caeruleus. *Molecular Ecology*, **16**, 3263–73.

Woolhouse MEJ, Haydon DT, Antia R (2005) Emerging pathogens : the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, **20**, 238–244.

Worley K, Collet J, Spurgin L *et al.* (2010) MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, **19**, 3064–3075.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323 – 354.

Wright LI, Tregenza T, Hosken DJ (2007) Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, **9**, 833–843.

Xiao X, Boles S, Frolking S *et al.* (2002) Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. *International Journal of Remote Sensing*, **23**, 3009–3022.

Zink RM, Pavlova A, Drovetski S, Rohwer S (2008) Mitochondrial phylogeographies of five widespread Eurasian bird species. *Journal of Ornithology*, **149**, 399–413.

Zink R, Barrowclough G (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 2107–2121.

Table des matières

Thèse de Doctorat

Yoan FOURCADE

Approche intégrative de la stratégie de conservation du Râle des genêts

An integrative approach to the conservation of the Corncrake

Résumé

Les espèces distribuées sur une large aire de répartition continentale rencontrent des situations socio-économiques et des conditions écologiques très variables qui impactent directement la répartition des menaces et déterminent les actions consacrées à leur conservation. A cette échelle, une approche intégrée de l'étude du fonctionnement des populations peut s'avérer nécessaire pour reconstituer l'ensemble des processus affectant la situation de conservation de l'espèce. Le Râle des genêts, oiseau prairial distribué au sein du Paléarctique en période de reproduction, présente ainsi un statut de conservation très variable lié à l'intensité de l'agriculture. En Europe de l'ouest, l'intensification agricole a conduit à un déclin considérable des populations tandis que l'espèce a été épargnée dans les parties centrales et orientales de sa distribution. Cette thèse vise à apporter des éléments de connaissances pour la conservation de l'espèce via une approche intégrée mettant en œuvre des approches de modélisation de distribution, de génétique des populations, de parasitologie et d'analyses écologiques de sélection d'habitat.
Les résultats obtenus ont permis de confirmer la distribution de l'espèce jusqu'alors estimée à dire d'experts. On a également montré que les variations d'intensité agricole au sein de l'Europe semblent favoriser une homogénéisation de la structure génétique ainsi qu'un gradient longitudinal de prévalence en malaria aviaire. Enfin, on a démontré que le Râle des genêts constituait une espèce parapluie d'intérêt modéré pour les passereaux prairiaux en Pays-de-Loire. Ces résultats permettent d'améliorer la compréhension du fonctionnement des populations à différentes échelles, et peuvent ainsi contribuer à la conservation du Râle des genêts.

Mots clés
Conservation, oiseaux prairiaux, Paléarctique, agriculture, distribution, modélisation, génétique, malaria aviaire, espèce parapluie, plaine alluviale, inondation, prairie

Abstract

Species distributed across a large continental range encounter different socio-economic and ecological situations, which directly impact the distribution of threats and determine the actions for their conservation. At this scale, an integrative approach to the study of populations' dynamics may be necessary to retrace all processes that affect the conservation status of the species. The Corncrake, a grassland bird distributed across the Palearctic during its reproduction season, exhibits a variable conservation status depending on the intensity of agriculture. In western Europe, agricultural intensification lead to a high decrease of populations while numbers remained stable in the central and eastern parts of its range. This work aims at improving knowledge for the conservation of this species using an integrative approach involving species distribution modelling, population genetics, parasitology and the ecological analysis of habitat selection.
Results confirmed the distribution of the species, which was previously estimated through expert knowledge. We also showed that the variations of agricultural intensity in Europe seem to contribute to a homogenisation of genetic structure as well as a longitudinal gradient of avian malaria prevalence. Finally, we demonstrated that the Corncrake is an umbrella species of poor value for the protection of grassland passerines in *Pays-de-Loire*. These results should provide new insights to the understanding of Corncrake populations and therefore favour the conservation strategy of the species.

Key Words
Conservation, grassland birds, Palearctic, agriculture, distribution, modelling, genetics, avian malaria, umbrella species, floodplain, flooding, grassland