Modelling species distribution when habitat occupancy departs from suitability

Application to birds in a landscape context

Nicolas TITEUX

Thèse présentée en vue de l'obtention du grade de docteur en sciences

Université catholique de Louvain

École doctorale en Biodiversité

Juin 2006

“I began my career basing my analyses of habitat relationships
on niche theory […] and then went through a multivariate phase […].
Each time I thought I had uncovered the ‘true’ habitat relationships
only to realize that my ‘other things being equal’ assumption contained
too much interesting ecology to ignore. Now I’m in a spatially explicit
landscape phase; we’ll see where that leads.”

John A. Wiens (2002)

Summary

Species Distribution Models (SDM) relate species presence or abundance to environmental predictor variables. They proved to be valuable tools in ecology and biogeography and are increasingly used to underpin the implementation of effective protection or management strategies for species conservation. However, existing SDM approaches largely overlook a widespread dual issue: a variety of factors may cause individuals to be present in unsuitable environmental conditions and/or absent from suitable ones. This thesis addresses this ecological question by enhancing the conceptualization of SDM.

Easy to survey and widely recognized indicators of environmental quality, birds offer ideal conditions for modelling. The habitat requirements of the Red-backed Shrike (*Lanius collurio* L.) in Southern Belgian rural landscapes were specifically investigated.

A niche-based modelling approach relying on presence-only information was designed to deal with the absence from suitable habitat. A statistical procedure screened out species presences not conveying reliable information about habitat suitability before building SDM. Breeding success data corroborated the ecological foundation of this screening approach. Spatial analyses revealed that the presence of individuals in locally unsuitable conditions was positively related to the presence of other individuals in the surroundings. These findings highlighted the fact that the decoupling between habitat suitability and occupancy was structured in space. Accordingly, we carried out conceptual and technical improvements to existing niche-based SDM approaches, so as to disconnect habitat suitability modelling from likelihood of occupancy modelling.

The ensuing SDM allowed (1) ranking the importance of different habitat components for the species conservation and (2) delineating the distribution of suitable areas and the potential species distribution. Similarities and discrepancies between both distributions allowed prioritizing areas in the landscape where protection and/or restoration measures are required.

Table of Contents

Remerciements

En commençant une thèse, on est relativement conscient de la direction globale que prendra cette aventure de plusieurs années, mais on est bien loin d'en imaginer le cours exact… Une thèse n'a en général rien de linéaire ! S'en sortir dans les embranchements que l'on emprunte et en voir le bout n'est pas un pari gagné d'avance… La concrétisation d'un tel travail repose sur le soutien, rapproché ou plus lointain, de nombreuses personnes que je tiens à remercier au travers de ces quelques lignes. Que celles et ceux qui arrivent à la fin de ce mot introductif en cherchant désespérément à y voir figurer leur nom m'en excusent, mais ma mémoire me lâche quelque peu ce soir…

Mes premiers remerciements s'adressent tout naturellement à mon promoteur, Pierre Defourny. Son incontestable ouverture d'esprit et l'immense confiance qu'il m'a accordée tout au long du parcours m'ont permis de m'épanouir dans mes recherches en empruntant les chemins parfois tortueux que je choisissais d'explorer. Malgré cette liberté d'action, Pierre est resté présent pour baliser la route aux bifurcations importantes… Mon espoir de lui faire découvrir les pies-grièches sur le terrain en Famenne, même s'il s'amenuise au fil des ans, ne s'est pas encore complètement volatilisé…

Marc Dufrêne m'a également été d'une aide inestimable aux moments importants de l'aventure, par ses conseils avisés et sa vision critique. Marc a la fâcheuse tendance à répondre à vos questions en vous en posant de nouvelles, ces dernières vous faisant momentanément oublier les premières… Mais le découragement qui s'ensuit fait toujours place à la constatation que, soudainement, la situation s'éclaircit. Merci Marc pour tes coups de mains efficaces, tes continuelles mises en doute, ton éclairage, ta confiance et ton assistance indispensables aux moments décisifs !

Les échanges d'idées à distance avec Alexandre Hirzel sont arrivés en cours de route, lors d'un grand virage de la thèse. Dès les premiers contacts, Alexandre s'est montré intéressé par mes recherches qui se basent sur les méthodes et utilisent les outils qu'il a développés et mis gracieusement à disposition de la communauté scientifique. Sans aucun doute, cela m'a permis d'aborder les choses sous un angle différent. Il s'est rapidement retrouvé impliqué dans diverses parties du travail et lui a apporté une vision essentielle et très appréciée. J'espère que sa présence parmi les membres du jury de cette thèse n'est que le début d'une collaboration scientifique plus rapprochée dans l'avenir.

Michel Baguette a gentiment accepté de jouer le rôle de répondant facultaire de ce doctorat réalisé à cheval sur deux facultés voisines. Mais j'aimerais surtout le remercier chaleureusement pour avoir amplement contribué, lors de mes années d'études, à éveiller et à aiguiser ma curiosité écologique. Quand je suis arrivé dans son bureau il y a 10 ans, alors qu'il était mon tuteur en première candi, j'ai directement su vers où je voulais me diriger !

Au coin d'un couloir, d'une table ou d'un auditoire, Éric Le Boulengé a toujours répondu gentiment et avec la grande patience qui le caractérise à mes questions d'ordre statistique. Par ailleurs, je garde un souvenir vraiment excellent de ma collaboration pédagogique avec Éric. En sa compagnie, l'enseignement des probabilités, des statistiques et de la biométrie aux futurs biologistes fut extrêmement enrichissant et bénéfique à de nombreux points de vue. Merci Éric pour ces années !

Je voudrais aussi exprimer ma reconnaissance envers Hans Van Dyck qui a immédiatement accepté de lire et de juger de la qualité de mon travail. Je raad en wetenschappelijke grondigheid hebben zeker bijgedragen tot het verbeteren van meerdere hoofdstukken. Dank ook voor je interesse voor mijn gedane, huidige en toekomstige onderzoek...

Merci également à Thierry Hance qui préside le jury de cette thèse.

Il est, au dernier étage du bâtiment Mendel, une équipe pluridisciplinaire de chercheurs ENGéliques au sein de laquelle règne une ambiance céleste. Évidemment, pour un biologiste, passer 6 ans de sa vie parmi eux relève du défi, de l'inconscience, voire de la folie… je ne sais pas vraiment en fait, sans doute un peu des trois. Même si j'ai toujours clamé, par pure provocation souvent, ne pas appartenir à leur « lignée », il n'en reste pas moins qu'ils m'ont tous, par leurs propres spécificités et à divers degrés, apporté des perspectives captivantes sur des horizons parfois insoupçonnés. Je ne voudrais évidemment vexer personne en mentionnant certains noms plutôt que d'autres, mais j'aimerais tout particulièrement remercier mes collègues de bureau actuels, Valérie et Xavier, qui m'ont supporté pendant de longues années. Leur soutien, leur complicité et leur volubilité (enfin, pour une des deux seulement) ont toujours compensé à merveille mes sautes d'humeur et mon sale caractère au sortir des salles de TP. Je tiens aussi à faire un clin d'œil à Vincent et Julien pour leur aide plus technique, mais précieuse. Allez, tant que j'y suis, je vais aussi remercier J…

ô non, il aurait trop bon ! Et puis, aux détours des couloirs de l'unité, il m'a été donné de rencontrer une ENGE qui m'a placé sur une orbite inaccessible aux satellites les plus sophistiqués…

De nombreuses autres personnes ont contribué à la réalisation de cette thèse. Jean-Baptiste Schneider, lors de son mémoire de fin d'étude, a effectué un travail de cartographie très utile. Nicolas Schtickzelle m'a accueilli lors de mes premiers pas à l'unité avant de choisir un autre chemin. Il m'a été d'une aide inestimable pendant quelques temps. Les nombreuses discussions avec Marc Paquay et Jean-Paul Jacob, les coordinateurs de l'Atlas de Lesse et Lomme, furent enrichissantes grâce à leur immense expérience de terrain. Par son enthousiasme et sa fascination incessante, Marc a également grandement contribué à me rendre accro de la Famenne… Daniel Borcard et Pierre Legendre m'ont ponctuellement conseillé sur l'utilisation des méthodes de partitionnement de la variance. Les discussions prolifiques que j'ai échangées avec Laurence Delahaye et Jean-Yves Paquet à propos de la sélection de l'habitat chez les oiseaux et de certaines méthodes d'analyses ont fait germer des idées nouvelles pour faire avancer mon travail. Norbert Lefranc nous a fait l'honneur, à Marc Paquay et à moi, de venir nous rendre visite en Famenne pour y observer cette fameuse pie-grièche hybride. Les discussions et les impressions que nous avons partagées alors m'apprirent énormément.

Plusieurs lecteurs ont contribué à l'amélioration de certaines parties de ce travail, chacun en y posant un regard particulier. Il s'agit de Dirk Maes, Mark Rounsevell, Hans Van Dyck, Wouter Vanreusel et Marie-Aline Wibrin. Merci à eux pour le temps qu'ils ont passé à essayer de décrypter des versions préliminaires parfois obscures de certains articles…

Ce travail n'aurait pas vu le jour sans l'aide précieuse de nombreux copains ou connaissances qui m'ont accompagné sur le terrain ou transmis leurs observations. Il m'est difficile de me souvenir de chaque sortie, mais je vais me risquer à citer Baptiste Bataille, Gaëtan Bottin, Vincent Bulteau, Marie Bylyna, Christophe Dehem, Frédéric Dermien, Jérémie Guyon, Xavier Janssens, Marc Paquay, Jean-Yves Paquet, Christophe Pontégnie, Michaël Pontégnie, Gilles San Martin y Gomez, Hugues Titeux, Christian Xhardez, ainsi que les membres de la Centrale Ornithologique Aves qui ont participé à la collecte des données de l'Atlas de Lesse et Lomme…

Je tiens à témoigner toute ma reconnaissance à l'Institut Géographique National qui nous a fourni gracieusement les bases de données cartographiques vectorielles au 1/50.000 et à prix démocratique celles au 1/10.000, indispensables à la réalisation de ce travail.

La recherche ne fût pas ma seule préoccupation professionnelle pendant ces 6 années, puisque la moitié de mon temps était consacrée aux charges de travaux pratiques en biologie animale et en statistiques. Je n'ai évidemment pas rigolé tous les jours, mais au bout du compte, je ne regrette absolument pas de m'être engagé dans cette voie pédagogique qui m'a énormément apporté à différents niveaux, grâce à une foule de collègues avec qui j'ai vraiment passé de bons moments…

Pour terminer, j'aimerais remercier tout spécialement mes proches, ma famille… J'ai l'immense chance d'être entouré par des personnes qui comprennent et acceptent mes passions et mes valeurs, sans jamais les juger. Je ne saurais jamais assez les remercier pour leur soutien de chaque instant et pour m'avoir inconditionnellement permis depuis toujours de m'engager dans les voies que je souhaitais !

… et merci à tous ceux que j'ai omis de mentionner bien malgré moi !

Nicolas

List of acronyms

AIC	Akaike's Information Criterion
AIC_c	Small sample version of AIC
A_g	Proportion of predicted suitable cells across the study area
Autcov	Autocovariate term
AVI	Absolute Validation Index
CCA	Canonical Correspondence Analysis
CVI	Contrast Validation Index
EFD	Environmental Functional Descriptor
ENFA	Ecological Niche Factor Analysis
Global-ENFA	ENFA performed on all species occurrences
Success-ENFA	ENFA performed on successful pairs only
Spatial-ENFA	ENFA performed with an autocovariate term
EO	Expected Occupancy
ENFA-EO	EO provided by ENFA
GLM-EO	EO provided by GLM
GIS	Geographic Information System
GLM	Generalized Linear Model
Spatial-GLM	GLM performed with an autocovariate term
HS	Habitat Suitability
ENFA-HS	HS provided by ENFA
GLM-HS	HS provided by GLM
PCA	Principal Component Analysis
PCNM	Principal Coordinates of Neighbour Matrices
RDA	Redundancy Analysis
ROC	Receiver Operating Characteristic
RS	Remote Sensing
RSF	Resource Selection Function

Introduction

Destruction and fragmentation of natural biotopes due to human activities are viewed as the main and ongoing threat to biodiversity worldwide (e.g. Wilcox and Murphy 1985, Andren 1994, Meffe and Carroll 1997, Henle *et al.* 2004). The establishment of protected areas such as nature reserves, though useful in many cases (e.g. Fabricius *et al.* 2003, Canova 2006, O'Dea *et al.* 2006, Lovejoy 2006), often proves to be insufficient to preserve some ecological functions (e.g. Dennis *et al.* 2003, De Klerk *et al.* 2004, Oldfield *et al.* 2004, Rodrigues *et al.* 2004a, 2004b, Van Teeffelen *et al.* 2006). Therefore, more attention needs to be paid to the wider environment (e.g. Holdgate 1994, Meffe and Carroll 1997, Tucker and Evans 1997, Pedrini and Sergio 2002, Watson and Whitfield 2002), especially to species living in the present-day man-shaped landscapes. This should enable to ensure their persistence outside protected areas (Pino *et al.* 2000, Donald and Evans 2006). The rapid decline of species in and outside protected areas has led many researches to get involved in the crisis disciplines of conservation biology and restoration ecology in the 1970s (Soulé 1986, Young 2000, Meffe 2001).

When embarking into those applied fields of conservation biology, one of our major needs is to improve our comprehension of the factors that determine the observed ecological patterns. Among others, understanding the foundation of the relationships between the species and their environment is a basic requirement for the assessment of impacts of anthropogenic activities, the maintenance of ecosystems in an acceptable conservation status, the restoration of disturbed systems or the development of specific recovery strategies (e.g. Van Horne 2002, Rushton *et al.* 2004, Guisan and Thuiller 2005).

Many research activities rely on the assumption that a relation exists between populations or communities and their ‘physical’ surrounding environment (e.g. Morrison *et al.* 1998, Heglund 2002, Mitchell 2005). It is further presumed that some quantification of population or community attributes on the one hand and of environmental conditions on the other hand will allow determining these relationships, with which adapted decisions can be made or efficient conservation strategies implemented (Wiens 2002). Apart from political willingness, the effectiveness of such actions is closely dependent on the way these relations are brought out and communicated (McCracken and Bignal 1998).

1. Species-Environment relations

This section intends to set the scene by giving a conceptual overview of the ecological foundations of the observed species-environment relations and by clarifying the terminology adopted throughout this thesis.

1.1. Patterns created by individual decision processes

Observed species distribution patterns in relation to environmental conditions are a consequence of decisions made by individuals[1] in selecting a place to live and reproduce (e.g. Wiens 1989b, Marzluff and Ewing 2001). Individuals possess an internal template of the environmental conditions that are suitable for accomplishing (parts of) their life cycle. This template is more or less precisely defined and either genetically determined, or learned, or both, depending on the species (see Wiens 1989b for more details).

In order to make optimal decisions about their living space, individuals have to rely on cues predicting local environmental quality (e.g. Williams and Nichols 1984, Wiens 1989c, Stamps 2001, Doligez *et al.* 2004). The settlement of individuals in a given area primarily depends on the intrinsic quality of the environment in terms of food abundance and accessibility, breeding sites availability, structures providing shelter from

---

[1] Species using mechanisms of passive dispersal only are not in the scope of this thesis and were therefore ignored in this section.

predators or other *direct* cues (*ultimate* factors) that enhance growth, survival or offspring production, that is, individual fitness (e.g. Wiens 1989b, Muller *et al.* 1997, Stamps 2001, Kristan 2003).

Individuals may take short-cuts and use *indirect* cues (*proximate* factors) that integrate the effects of many environmental factors on somatic condition or reproductive performance, therefore revealing the effect of environmental conditions on fitness. This should be more powerful and parsimonious than using each factor separately (e.g. Stamps 2001, Heglund 2002, Stamps and Krishnan 2005). Strong time constraints on breeding in migratory species, for instance, constitutes a selective pressure that may have evolutionarily shaped such mechanisms because more adaptative than relying on direct cues (Stamps 2001, Doligez *et al.* 2004). Examples of indirect cues are physiognomy of the environment that are not directly used by individuals but that reflect the existence of suitable resources (e.g. Heglund 2002), or conspecific presence, density or breeding success acting as a mirror of suitable environmental conditions (conspecific attraction and public information, see e.g. Stamps 1988, Doligez *et al.* 2004, Stamps and Krishnan 2005).

1.2. Ecological niche and related concepts

Several concepts could be used to describe those environmental conditions that are required by a given species for survival, reproduction or both. Paradoxically, two of the most important terms in ecology are among the most confusing ones: ‘niche’ and ‘habitat’ (e.g. Morrison and Hall 2002, Mitchell 2005). This confusion comes from the use of niche for different concepts, leading to a partial redundancy with habitat (Whittaker *et al.* 1973). A detailed review of these fundamental concepts is obviously out of the scope of the present study (see Morrison and Hall 2002 for a recent clarification). Here, we only address the different terms and definitions as they are applied throughout the present thesis.

1.2.1. Ecological niche

Grinnell (1917) defined the species niche as the ‘environmental requirements of the species’ and considered it as an ‘ultimate distributional unit of the species’. Later, Elton (1927) defined the niche as ‘the role of the species in the community’, which integrates its interactions (mainly competition) with other species (behaviour-based concept). Both definitions (geographical and functional) are conceptually vague (Whittaker *et al.* 1973, Heglund 2002) and were rigorously combined by Hutchinson (1957) who described the niche by ‘the coordinates of the species with *n*-dimensional resource axes’. The Hutchinsonian paradigm modelled the niche as a ‘hypervolume’ situated in a *n*-dimensional ‘hyperspace’; this hypervolume encloses ‘conditions that allow the species to exist indefinitely’ (‘fundamental’ niche). This Hutchinsonian niche amalgamates the notions of Grinnell and Elton (Whittaker *et al.* 1973). Because of interspecific interactions, the species may be excluded from some parts of its fundamental niche, reducing the hypervolume (‘realized’ niche). In the real world, unless manipulative experiments are designed, the fundamental niche is unlikely to be observed and we often *de facto* focus on describing the realized niche (Heglund 2002, Guisan and Thuiller 2005). These formulations are highly conceptual and rely on assumptions that could be largely violated, which explains the multiple contestations, denigrations and revisions formulated since the 1970s up to now (see Blondel 1995 for a clear historical review, but also Chase and Leibold 2003). However, Hutchinson provided a formalization of the niche concept that has since become the foundation of much ecological theory and reasoning (Morrison *et al.* 1998, Pulliam 2000, Heglund 2002).

1.2.2. Habitat and resources

Habitat is one of the most basic concept of theoretical and applied ecology (e.g. Whittaker *et al.* 1973, Southwood 1977, Smallwood 2002, Wiens 2002, Baguette and Mennechez 2004). Like the niche concept, it suffers from inconsistencies of definition (see e.g. Dennis *et al.* 2003 for a mini review or Mitchell 2005 for a provocative critique of the habitat concept applicability).

The habitat is viewed in this thesis as the set of environmental conditions meeting the individuals' ecological requirements for their survival and reproduction (according to Block and Brennan 1993). Apart from individual intrinsic quality and correcting for interspecific interactions, individuals in higher-quality habitats will have greater survival and reproduction, hence fitness, than those in lower-quality habitats. The environmental conditions required by the individuals are not stationary throughout the whole species distribution range: habitat is population-specific. Moreover, these relations between populations and environment are in perpetual evolution (Blondel 1995). Habitat selection is a behavioural process involving a suite of innate or learned decisions – based on direct or indirect cues – made by individuals about what site they will use for accomplishing (part of) their life cycle (see Morrison and Hall 2002).

The environmental conditions required by the individuals can be defined in terms of resources that are either consumables or utilities (Morrison and Hall 2002, Dennis *et al.* 2003). Consumables are directly used by the individuals (e.g. food for the different life stages). Utilities are all other required conditions for existence and persistence, like breeding sites, shelter from predators, foraging/mating areas, or microclimate (Dennis *et al.* 2003). Habitat therefore encapsulates several resources that are (1) all required by the individuals (non-substitutable resources) and (2) located in the man-shaped landscapes (overlapping, adjoining or disjoint resources). The different resources are connected to each other by individual movements (landscape complementation, e.g. Dunning *et al.* 1992, Hinsley 2000, Pope *et al.* 2000).

Accordingly, precisely defining habitat patch boundaries in the geographical space is not an easy task because of possible spatial discontinuity in resources distribution (Dennis *et al.* 2003, Baguette and

Mennechez 2004). Such delimitation of habitat patches as functional spatial entities can be achieved by integrating the movement behaviour of individuals between resources within habitats (Baguette and Mennechez 2004).

A resource-based definition of habitat is intimately linked to the Hutchinsonian concept of the multidimensional niche (Dennis *et al.* 2003, Mitchell 2005). For a species, habitat could be viewed as a geographical projection of the population hypervolume (hereafter called habitat hypervolume) defined in a multidimensional hyperspace representing the available environmental conditions within a given area (hereafter called environmental hyperspace). The interactions between species and the role of the species within the community are not explicitly integrated in this definition of the habitat. As defined here, the habitat is hence more related to the Grinnellian than the Eltonian niche (see Whittaker *et al.* 1973, James *et al.* 1984). Furthermore, when relying on field observations (and not manipulative experiments), the delineated habitat components are more linked to the realized than the fundamental niche.

In this thesis, we investigate habitats and not niche *sensu stricto* (Hutchinson 1957), but we still refer to the niche concept to point out that the formalism of Hutchinson was followed. We frequently mention species-habitat relations, but the reader should keep in mind that these relations are dynamic and population-specific. This shortcut is deliberate and underlines the fact that we do not formally explore population-environment systems *sensu stricto* (Blondel 1995). Actually, populations are here deemed homogeneous entities and we do not account for variations in the intrinsic quality of individuals. We focus on the spatial variability of environment and its influence on species distribution.

1.2.3. Landscape

A landscape can be defined as a spatially heterogeneous area encompassing several ecosystems (Forman and Godron 1986, Turner *et al.* 2001, Morrison and Hall 2002) but characterized by a geo-morphological and bio-climatic constancy (Blondel 1995). This mosaic of ecosystems encloses interacting species (e.g. Urban *et al.* 1987, Forman 1995, McGarigal and Marks 1995).

The concept of landscape refers to a level of organization, not a level of observation. This organization level is positioned above ecosystems and below biomes (Allen 1998). Its scale (spatial extent and grain) depends on the organisms and the ecological processes under study (McGarigal and Marks 1995, Allen 1998), which in turn determines the scale of the investigation. Whereas the landscape could be viewed in terms of geo-morphological and bio-climatic variables (geographical approach, e.g. Blondel 1995), the landscape level does not refer to an absolute spatial scale of organization but rather to a relative one that depends on the organisms' perception of their environment (functional approach, e.g. Wiens and Milne 1989, Allen 1998). This duality in the definitions is still under debate but these are not mutually exclusive.

The landscape level could be viewed as an organization level revealing a new set of constraints for individuals (e.g. regarding dispersal, migration or predation for instance), as compared with the local constraints involved at the habitat level. While the habitat definition deals with the spatial distribution of resources (see above), the landscape definition integrates the spatial arrangement of ecosystems and *a fortiori* of habitat patches. The influence of this landscape pattern (composition and configuration) on ecological processes at various spatial scales (e.g. Turner 1989, With *et al.* 1997) is the central theme of landscape ecology (e.g. Wiens 1992, Wiens *et al.* 1993, Turner *et al.* 2001), which combines the spatial approach of the geographer with the functional approach of the ecologist (Forman and Godron 1986).

1.3. The relationship between niche and distribution

Under ideal conditions in a given landscape, a population is expected to occupy a geographic area that strictly corresponds to the projection of its Grinnellian niche, i.e. to occur everywhere environmental conditions are suitable and nowhere else (Pulliam 2000). But in the real world, this congruence is very rarely observed (see Figure 1) due to a variety of factors, of which some are detailed below.

1. Interspecific competition may exclude individuals from some portions of their species fundamental niche (Wiens 1989b, 1989c). In this case and other factors apart, the realized niche (Hutchinson 1957) ends up narrower than the fundamental one (Figure 1). Such interspecific interactions are variable in time and space and vary with spatial scales (Wiens 1989a,c, Levin 1992).

2. Metapopulation dynamics (e.g. Hanski and Gilpin 1997, Baguette 2004, Hanski and Gaggiotti 2004) and source-sink dynamics (Pulliam 1988, Pulliam and Danielson 1991) further complicate the relationship between niche concept and species distribution. In particular, the metapopulation dynamics consists of a turnover of extinction-(re)colonisation of suitable habitat patches in the landscape (e.g. Baguette 2004). This dynamics is governed by a variety of factors (not detailed here) and explains the existence of unoccupied suitable habitat patches. Moreover, individuals may occur in sink habitats (Pulliam 1988, Pulliam and Danielson 1991, Dunning *et al.* 1992) or sink landscapes (With *et al.* 2006) that are not suitable for population persistence but that are *rescued* by the immigration from nearby source ones (Figure 1 and see e.g. Robinson *et al.* 1995, Donovan *et al.* 1995).

3. Social influence like conspecific attraction may incite individuals to establish in sub-optimal or even unsuitable environmental conditions near conspecifics (Stamps 2001, Stamps and Krishnan 2005).

4. When dispersal across the landscape is limited by a hostile matrix (e.g. Clobert *et al.* 2001, Wiens 2001, Stevens 2006), habitats may remain unoccupied whatever their local suitability (Figure 1, Pulliam 2000).
5. Populations fluctuate in abundance between years in response to various factors at several spatial (from global to local) and temporal scales (e.g. Wiens 1989c, Scott *et al.* 2002), affecting breeding output and survival of individuals. Such factors may hold populations densities below carrying capacities (Van Horne 1983, Wiens 1989b).

The species distribution pattern that results from summing up all these individual, populational, landscape, regional and global processes is therefore not an exact copy of the distribution of suitable habitat patches. The range of environmental conditions actually experienced by individuals may be greater[1] than the range of suitable conditions and/or suitable areas may remain unoccupied[2] (Figure 1). A considerable decoupling between habitat suitability and species distribution may therefore be observed in the real world and this distortion complicates any habitat study (Pulliam 2000). However, these studies should not deny this intrinsic property of natural systems (Wiens 2002, Guisan and Thuiller 2005) but instead deal with it for adequately identifying the resources that define habitat suitability for any species.

---

[1] Individuals occurring outside the environmental boundaries of species niche are called ‘outsiders’ throughout this thesis.

[2] Unoccupied suitable habitat patches lead to ‘meaningless absences’ from a strict habitat point of view.

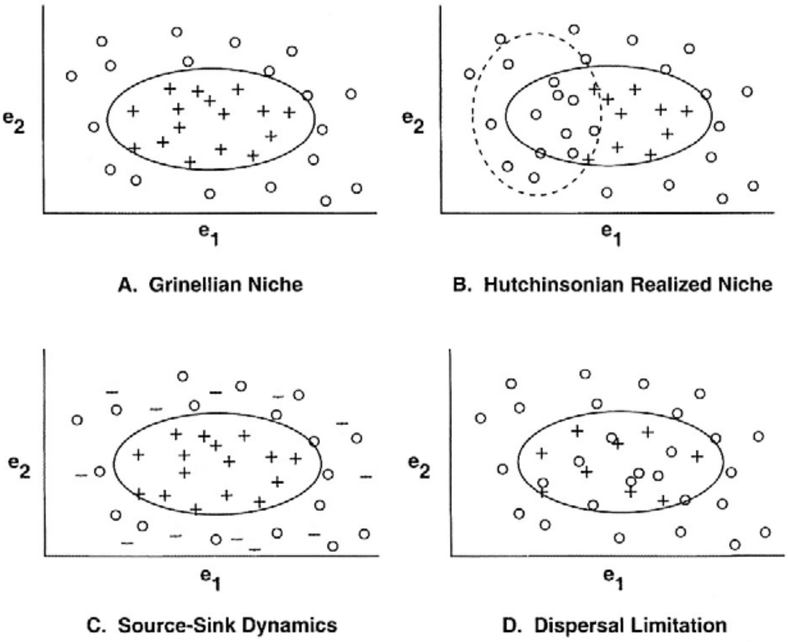


Figure 1. Four representations of the relationships between niche and distribution (in two-dimensions, copied from Pulliam 2000). The ellipses represented in each situation delineate the species fundamental niche, i.e. the combination of environmental factors e1 and e2 within which the population is expected to exist indefinitely. (A) Individuals occur everywhere conditions are suitable (+) and nowhere else (o). (B) Individuals do not occur in a part of the fundamental niche that is itself occupied by individuals of a dominant competitive species (dotted ellipse). (C) Individuals of the species frequently occur in unsuitable environmental conditions (sink habitats), leading to ‘misleading’ presences (–) outside the bounds of the fundamental niche (‘outsiders’). (D) The fundamental niche is not saturated due to dispersal limitations, leading to ‘meaningless’ absences within the bounds of the fundamental niche.

2. Species-habitat modelling

In this section, the modelling approaches typically used to describe species-habitat relations are conceptually presented, as well as their usefulness for conservation applications. We then emphasize inherent limitations and dilemmas the modeller has to cope with. We finally stress some punctual issues that are of particular concern for selecting an appropriate modelling method.

2.1. Conceptual framework and underlying assumptions of modelling approaches

According to the paradigm of Hutchinson, locations in a landscape may be characterized by *n* environmental variables that may themselves be viewed as axes of a *n*-dimensional coordinate system representing a environmental hyperspace (Whittaker *et al.* 1973, Heglund 2002). Each population in the landscape occurs over a more or less wide range of these environmental variables. The range of the coordinates within the limiting values defines a *n*-dimensional habitat hypervolume within the environmental hyperspace. This hypervolume circumscribes all environmental conditions within which the individuals could exist or persist. The hypervolumes associated to the populations of different species are scattered in the hyperspace. Accordingly, populations of different species form a ‘complex population continuum’ within the hyperspace (Whittaker *et al.* 1973).

Individuals regularly meet environmental conditions situated beyond their limits of tolerance. Consequently, density or occurrence could be misleading indicators of habitat suitability (Van Horne 1983, Pulliam 2000) and one should therefore look for more reliable measurements (hereafter called species response) for adequately delineating the hypervolume (Guisan and Thuiller 2005). Moreover, within the environmental hyperspace, the species occurrences generally do not form a sharply bounded distribution and precise quantification of the species response distribution is required (e.g. Whittaker *et al.* 1973). This is the very aim of what is commonly called species-habitat modelling (e.g. Scott *et al.* 2002) that boomed from the 1970's (Stauffer 2002, Wiens 2002, Guisan and Thuiller 2005).

Species-habitat models may address single species populations or multi-species assemblages and try to quantify the relations between species responses and environmental conditions at a variety of spatial scales. The applied models in this field encompass a huge array of (non-)statistical techniques for relating species to environment, including e.g. expert opinion (e.g. Pearce *et al.* 2001, Seoane *et al.* 2005), (non-)linear regression (e.g. Chamberlain and Gregory 1999, Bonn and Schroder 2001, Austin 2002), regression tree (e.g. De'Ath 2002), discriminant analysis (e.g. Austin *et al.* 1996, Manel *et al.* 1999a), envelope approach (Walker and Cocks 1991, Busby 1991, Hirzel and Arlettaz 2003a, Pearce and Boyce 2005), ordination (e.g. Pasinelli *et al.* 2001) or direct gradient analysis (Borcard and Legendre 1994, Bio *et al.* 1998, De'Ath 1999). For a thorough review on the existing techniques we refer to Jongman *et al.* (1995), Legendre and Legendre (1998), Morrison *et al.* (1998), Guisan and Zimmermann (2000), Scott *et al.* (2002), Segurado and Araujo (2004) or Elith *et al.* (2006), among others. Modelling techniques are closely tied to modelling objectives (Van Horne 2002) and there is consequently no single best modelling approach. The adopted strategy should therefore result from a trade-off between ultimate objectives, available data and technique capabilities.

Most of these modelling techniques are underpinned by the Hutchinsonian niche theoretical framework (niche-based models) (e.g. Guisan and Zimmermann 2000, Guisan and Thuiller 2005). They try to disentangle the complex population continuum and to understand the species-habitat relations in the environmental hyperspace (e.g. Heglund 2002). While technically highly diversified, this is generally achieved by reducing the *n*-dimensional hyperspace to a coordinate system with a reduced number of major axes. Such axes are either *a priori* recognised as important resources for the species of interest or compositional axes derived from ordination techniques. Eventually, the representation of the species response distribution in this reduced space enables to interpret the relation between species and habitat, but also among species.

2.2. Species-habitat models as conservation planning tools

Identifying the key ecological requirements of individuals and delineating the distribution of the habitats for threatened species are increasingly needed in nature conservation (e.g. Huston 2002, Rushton *et al.* 2004, Guisan and Thuiller 2005, Seoane *et al.* 2005, Bayliss *et al.* 2005, Olivier and Wotherspoon 2006). Direct observations of species occurrence or abundance provide part of this qualitative information but are not sufficient (Scott *et al.* 2002, Elith *et al.* 2006), especially for detailed management decisions (e.g. Seoane *et al.* 2005). As far as conservation is concerned, a quantitative approach is recommended as a rigorous basis for decision-making.

In his reference book on the ecology of bird communities, Wiens (1989b) stated: ‘At a qualitative level, the general habitat associations of many species are known to any good birdwatcher, and the work of ecologists might be regarded as expressing (or sometimes obscuring) that knowledge in detailed quantitative analyses’. The transition from qualitative natural history research to quantitative approaches happened in the mid 1950s – when Hutchinson presented his niche formalization – and has evolved and diversified until now (see Stauffer 2002 for an interesting retro- and prospective overview).

Species-habitat models are currently recognised as helpful tools for providing valuable and quantitative information by displaying the most important resources required by the individuals in a spatial context (Stauffer 2002, Guisan and Thuiller 2005). For this reason, they allow to efficiently guide protection, management or restoration planning (e.g. Loiselle *et al.* 2003, Johnson *et al.* 2004, Swihart and Moore 2004, Rhodes *et al.* 2006) or to implement environmental impact assessment programs (e.g. Gates *et al.* 1993, Chamberlain and Fuller 2001). Furthermore, coupled with geographic information systems (GIS) technology, species-habitat models can be extended into a spatial dimension for producing maps that display the spatial configuration of the suitable habitats (e.g. Anderson and Martinez-Meyer 2004, Johnson *et al.* 2004, Gibson *et al.* 2004), which enables protection, management and restoration strategies to be implemented in a spatial context (e.g. Araujo and Williams 2000, Loiselle *et al.* 2003, Cabeza *et al.* 2004, Tole 2006). Besides illuminating the habitat selection pattern of the focal species, the application of such models to areas where environmental conditions are known but where species distributions are unknown provides habitat suitability maps (predictive models, e.g. Scott *et al.* 2002).

2.3. Limitations of modelling techniques

Most of the existing modelling techniques are increasingly accessible thanks to software development. As a result, they are commonly used and abused. Indeed, they are marred by several limitations regarding their ecological reliability and their usefulness. Such limitations are regularly overlooked and are therefore briefly presented here. Any modeller or model user should always keep them in mind to assess model's boundaries of applications before (1) choosing the most adequate method to match the objectives (e.g. Segurado and Araujo 2004) and (2) correctly interpreting the models' outcomes (e.g. Heglund 2002, Van Horne 2002).

2.3.1. Model, tell me the truth…

The goal of a modelling approach is not to reflect the full reality but to construct models which make biological sense, approximate this reality and constitute useful tools (Burnham and Anderson 2002). It is thus unsound and utopian to believe that a 'true' model will perfectly explain the biological data we observed (Hastie *et al.* 2001). Instead, biological systems are the results of many small effects, individual heterogeneity or interactions at multiple spatial and temporal scales. In other words, models are low-dimensional abstractions of infinite-dimensional forces acting on individuals. The way this abstraction is achieved mainly determines the usefulness of the ensuing models (see next sections for more considerations). Box (1976) wrote that ‘all models are wrong, but some are useful’ and Burnham and Anderson (2002) stated that ‘increased sample size allows to chase full reality, but never to catch it’.

2.3.2. Correlation and causality – Importance of resource-based definition of the habitats

A common criticism of species-habitat models is that most of them are based on correlations between species response and measured environmental variables, therefore providing little insight pertaining to the proximate mechanisms underlying such relations (e.g. Capen 1981) because correlation does not automatically mean causality or process (Van Horne 2002). However, understanding the processes creating the observed patterns is critical and needed if we are to devise or implement efficient conservation strategies (Heglund 2002). Ultimately, an experimental approach is often required to thoroughly understand the cause-and-effect relationship (but see Wiens *et al.* 1986 and Van Horne 2002 for limitations of manipulative experiments).

Furthermore, most of the models relate species response to coarse-scaled environmental variables (e.g. Gates *et al.* 1993, Natuhara and Imai 1996, Fleishman *et al.* 2001, Pasinelli *et al.* 2001, Atkinson *et al.* 2002, Storch *et al.* 2003, Maes *et al.* 2003), often including topography, climate or surrogate variables without any direct ecological foundation. Such approaches are successful for describing large-scale distribution pattern or for some purely predictive purpose but suffer from multiple criticism (Van Horne 2002, Guisan and Thuiller 2005). First, such variables are not necessarily within the scope of management or restoration actions (Guisan and Thuiller 2005, Vanreusel *et al.* 2006), leading the models to be of poor direct applicability for managers. Second, these surrogates have no straight ecological relation to the operative factors perceived by the individuals (direct cues) and hence give poor information about the functional processes (e.g. Austin 2002, Heglund 2002, Guisan and Thuiller 2005). Third, correlations between surrogate and causal factors may be area- or time-specific and scale-dependent. The causal factors may thus change without simultaneous changes in the surrogate measured variables (Van Horne 2002). Correlations between proximate and surrogate factors may then break down in space or time, leading the models to be spuriously overfitted to the local or current conditions (Luoto *et al.* 2002).

Adopting a functional resource-based concept of habitat (see e.g. Dennis *et al.* 2003) enable focusing on the fundamental requirements of the organisms and thus on the proximate operative factors driving species habitat selection. This requires a good qualitative knowledge of the species life history (Heglund 2002, Dennis *et al.* 2003). While not strictly proving the causality, one could be more confident in the models' outcomes, simply because the ecologically-scaled input is less likely to provide spurious outputs (Guisan and Thuiller 2005); this leads to ecologically-founded 'Resource Selection Functions' (RSF, Boyce and McDonald 1999).

2.3.3. Generality, specificity and overfitting – Matching techniques with objectives

Albert Einstein has said that 'everything should be made as simple as possible, but not simpler'. '*What is the best model to use?*' is the critical question when making inference from data (Burnham and Anderson 2002).

As caricaturised by Van Horne (2002), 'a model that works well for a given location, such as one that predicts the number of burrowing owls based on the number of prairie dog burrows available, may not be useful in another location, such as an area where prairie dogs are absent and burrowing owls nest in other sites than prairie dog burrows'. Obviously, while based on resources directly used by organisms (here, nest sites), such *specific* models are of no general interest and use, but it does not mean that they are not interesting at all.

Furthermore, the more environmental variables incorporated in the modelling process, the better the fit to the data at hand. However, Burnham and Anderson (2002) advocated following the ‘principle of parsimony’ and strongly called against this unsound and hazardous strategy because leading to *overfitted* models that are likely to be limited in their application, because of the inherent variability of species-habitat relations both in time and space (Wiens 1989b).

On the other hand, using fewer and quite coarse-scaled environmental variables when describing species-habitat relations enhances the clarity of the models and their applicability (*generality*) to a broad range of systems (see respectively Burnham and Anderson 2002 and Van Horne 2002, for number and nature of variables). Nevertheless, they could become too simple and even trivial, hence useless (Burnham and Anderson 2002). Moreover, owing to the use of surrogate coarse-scaled variables, they lose the ability to identify specific relations between the individuals and their environment, because they are less likely to be directly related to the underlying processes (Heglund 2002).

In face of this apparent dilemma, the researcher has to choose the adequate modelling framework, but models are neither universally valuable nor totally useless. Either they apply within restricted conditions, or they can be generalized but fail to identify processes accurately, or they might reveal spurious correlations. It follows that such a discussion remains vague and of poor interest without clear reference to the addressed question or issue. Recently, Van Horne (2002) called for de-emphasizing one of the fundamental components of habitat modelling, that is, validation of models using independent data for ensuring their generality (e.g. Fielding and Haworth 1995, Fielding and Bell 1997, Whittingham *et al.* 2003). She rather stimulated modellers to balance the relative gains of generality and precision to finally reach a trade-off tuned to the objectives of their study (see also Hastie *et al.* 2001). Mitchell (2005) pointed out that if modellers are not able to construct general models of species-habitat relationships, it may simply be due to the fact that such relations do not exist. Accordingly, adopting the adequate modelling strategy entails matching the objectives with an assessment of model abilities and boundaries (Van Horne 2002).

As seen above, a functional resource-based approach is crucial for understanding the operative factors governing species distribution. The information about the distribution of functional resources is though rarely readily available (Laymon and Barrett 1986) and its acquisition may be time-consuming (Seoane *et al.* 2005), hence limiting the spatial extent of the study area and potentially the applicability of the outcomes to wider areas. However, Vanreusel *et al.* (2006) showed that such a functional approach allowed building accurate predictive models that are transferable to other areas in the same eco-region, where resource availability and use are similar. Therefore, while limited in their spatial extent, such models may have wide conservation implications.

2.4. The importance of spatial scales in modelling

One may study ecological systems at a wide array of spatial scales and emerging patterns differ at each chosen scale (e.g. Wiens 1989c, Orians and Wittenberger 1991, Levin 1992, Blondel 1995, Dungan *et al.* 2002, Wu 2004). Virtually all ecological patterns or processes exhibit some degree of scale-dependency (e.g. Wiens 1989c, Bellehumeur and Legendre 1998, Mitchell *et al.* 2001, Scott *et al.* 2002). Accordingly, the insights gained at one scale could not be directly transferable to another scale (e.g. King 1991, Bissonette 1997, Scott *et al.* 2002), hence questioning the applicability of fine-scale models to broader scales and the reverse (Araujo *et al.* 2005).

This implies that there is no ‘best’ single scale for studying species-environment relationships (Turner *et al.* 2001, Wiens 2002). The arguments for the choice of a particular scale – regarding both grain and extent – result from a long-standing debate over the relative merits of holism and reductionism in ecology (Wiens 1989c, Blondel 1995). This debate is closely linked to the ‘generality versus specificity’ dilemma. Wiens (1989c) emphasized the fact that the selection of a scale of investigation depends primarily on the nature of the questions that are asked and of the phenomenon of interest. Investigations at the local scale of individuals may be subject to individual heterogeneity (‘idiosyncrasy’), sampling error, chance effects or other ‘noise-generating’ variations that could be lightened by averaging observations over a broader scale (Wiens 1989c). On the other hand, studies at a broad scale are prone to overlook significant elements that account for the population or community dynamics, such as competition, predation, resource use and limitation, which can be apprehended at finer scale (e.g. Vanreusel *et al.* 2006). Variation among individuals within local populations may reveal interesting mechanisms and provide crucial information to test causal hypotheses (Wiens 1989c), but this variation disappears at broader scales merely because individual observations are averaged.

In the case of species-habitat relations, local extent and fine resolutions are more appropriate for investigating the mechanistic response of individuals or populations to resource variation in space; broader extent and coarser resolution are more likely to reveal the distribution pattern of species along environmental gradients (Wiens 1989c, 2002, Guisan and Thuiller 2005). Nevertheless, the dichotomy is quite fuzzy and a single scale will not answer all the questions asked simultaneously.

Perspectives from multiple scales are obviously enlightening, because events occurring at one scale are not independent of other events that occur at different scales (e.g. Wiens and Rotenberry 1981, Ricklefs 1987, Naugle *et al.* 1999). Multi-scale approaches therefore provide the greatest insights (Urban *et al.* 1987, Wiens 1989c, Balcom and Yahner 1996, Böhning-Gaese 1997, Grand and Cushman 2003, Pearson *et al.* 2004, Driscoll *et al.* 2005) but are generally time- or budget-consuming. In any case, studies should be designed so that no discrepancy exists between the scale of the question asked, the processes thought to be important and the interpretation of the models' outcomes (Scott *et al.* 2002).

2.5. Understanding the processes within a landscape framework

Since the 1970s, there has been an increasing and urgent call for scientific underpinning of the management of large areas (Turner *et al.* 2001). Such needs notably catalyzed the emergence of landscape ecology (Forman and Godron 1986) that focuses on understanding ecological processes and proposing management solutions in a landscape context (Wiens 1992, Wiens *et al.* 1993). Nowadays, the incorporation of landscape context and heterogeneity into management decisions can not be overlooked, because its importance is widely recognised (e.g. Hansson and Angelstam 1991, Angelstam 1997, Burke 2000, Hirzel 2001).

Landscape ecology provides concepts, methods and tools that complement those developed simultaneously in other fields of ecology (Turner *et al.* 2001). This does not entail that broad-scale studies should replace fine-scale investigations, but that the latter need to be integrated in a broader context. Unfortunately, the temptation for the former solution – at the expense of mechanistic (process-based) studies – is often exacerbated by the explosive development and advances in numerous appealing technologies such as remote sensing or geographical information systems (Van Horne 2002). Instead, it is largely recognised and emphasized that the precise understanding of the ecological processes underlying the observed patterns within a spatial context (e.g. Dunning *et al.* 1992) guarantees the success of the ensuing management strategies undertaken at a broad scale (e.g. Fahrig 2001, Lee *et al.* 2002, Heglund 2002, Van Horne 2002).

2.6. Spatial autocorrelation is not a noise-generating factor

Besides environmental constraints, the spatial distribution of species may originate from intrinsic processes related to population dynamics or biotic interactions within the community (e.g. Legendre 1993, Koenig 1999). Such processes operate at various spatial scales (Wiens 1989c) and may result in spatial autocorrelation in the species response. Spatial autocorrelation arises when the intrinsic processes are such that the values of samples that are close together have a tendency to be more similar (for positive autocorrelation) or less similar (for negative autocorrelation) than those randomly located in the study area (Legendre 1993, Dale *et al.* 2002).

Moreover, the spatial arrangement of the environmental variables can play a significant part in structuring species distributions. This should not be left out of consideration when building species-habitat models (Augustin *et al.* 1996, Keitt *et al.* 2002).

Therefore, explicitly introducing the spatial structures of species distribution and of environmental variables into the models is necessary for assessing and quantifying the relative and synergetic roles of intrinsic processes and environment in creating the observed distribution patterns (e.g. Borcard *et al.* 1992, Legendre 1993, Legendre and Legendre 1998, Lichstein *et al.* 2002).

Failure to account for the deterministic – and not noisy – spatial autocorrelation and the spatially-structured environment can lead to incorrect conclusions regarding the importance of environmental variables as determinants of spatial distribution (e.g. Keitt *et al.* 2002). Presenting erroneous conclusions about the real factors governing species distribution could have far-reaching consequences as far as management and conservation of species are concerned.

Numerous methods for the explicit analysis of spatial structures have been developed in a wide range of scientific fields (reviews by Perry *et al.* 2002, Dale *et al.* 2002 and see e.g. Augustin *et al.* 1996, Keitt *et al.* 2002, Borcard *et al.* 2004 for more specific issues) therefore allowing the explicit incorporation of space into the species-habitat modelling framework. While conceptually and technically highly diversified, the methods generally represent the spatial structure of the data by creating a set of explanatory spatial variables from the geographic coordinates of the sampling sites. Those spatial variables describe the spatial relationships among sampling sites and are subsequently used to characterize the spatial structure of environmental variables and species response. This spatial information is then explicitly introduced within the modelling framework (e.g. Keitt *et al.* 2002). The adoption of a given method is obviously linked to the kind of available data (Legendre *et al.* 2002) since some apply to very sparse samples and others require a complete map of all the points in the area (Dale *et al.* 2002).

3. Bird habitats in Europe – Accounting for the wider environment

In this section, we justify the selection of birds as model-organisms and environmental quality indicators in the present species-environment study and we explain why we mostly focused on farmland species from a conservation perspective.

3.1. A call for bird-habitat studies

Birds are worldwide recognized as useful indicators of the state of the environment (e.g. Bibby 1999, Donald *et al.* 2002b, de Heer *et al.* 2005, Fleishman *et al.* 2005). More than 40% of bird species in Europe have currently an unfavourable conservation status (BirdLife International 2004), due to the small size, the decline or the limited extent of their populations (Tucker and Evans 1997). The greatest threats to birds in Europe lie in the increasing habitats degradation and fragmentation, principally due to the high environmental pressure following from human land use (e.g. Tucker and Evans 1997, Donald *et al.* 2002b). The ongoing decline of so many species shows a clear signal about the condition of a huge part of the biodiversity and the health of the environment (e.g. Pienkowski 1991). Birds are relatively easy to census as they are well known, easily recognisable and simpler to locate than many other taxonomic groups (Bibby *et al.* 1992). For all these reasons, birds were regarded as interesting and useful model-organisms for this species-environment study (see Pearman *et al.* 2006, Pereira and Cooper 2006).

Habitats and bird species conservation may be viewed as a strategy that focuses on statutory protected areas (Jackson *et al.* 2004, Marsden *et al.* 2005, O'Dea *et al.* 2006). Nevertheless, numerous studies on bird habitat requirements suggest that conservation action should be carried out by a mixture of site-protection measures and conservation of the wider environment (e.g. Pedrini and Sergio 2002, Seoane *et al.* 2006). This strategy should ensure preserving the ecological functions required for the persistence of many species that have become evolutionarily adapted to man-shaped landscapes (Pino *et al.* 2000, Donald and Evans 2006). As a consequence, even if necessary in some cases, the confinement of conservation strategies within protected areas only is still not sufficient for many bird species (e.g. Benayas and de la Montana 2003, De Klerk *et al.* 2004).

Accordingly, in a recent global assessment and updating of European birds' conservation status and their habitats, BirdLife International (2004) notably called for more scientific understanding of the species-habitat relationships, particularly in a landscape context and in the wider environment. An increased and detailed knowledge of such relations is required for underpinning the implementation of effective conservation strategies (see Tucker and Evans 1997).

3.2. Severe decline in farmland birds

Agricultural and grassland landscapes currently dominate Europe and hold a rich but declining avifauna (e.g. Pain and Pienkowski 1997, BirdLife International 2004). Whilst many farmland bird species have historically benefited from the conversion of woodlands to agricultural landscapes, the same group of species has declined both in distribution and abundance for several decades, especially since the early 1970s (e.g. Fuller *et al.* 1995, Pain *et al.* 1997, Siriwardena *et al.* 1998, Chamberlain *et al.* 2000, Donald *et al.* 2002b). This trend is far less dramatic in birds of other biotope types (e.g. Fuller *et al.* 1995, Pain and Pienkowski 1997, Gregory *et al.* 2005).

Simultaneously, the deterioration of many farming landscapes has been widespread throughout European Union, principally due to the increasing intensification of agricultural practices – like crop improvement, fertilizer and pesticide use, intensive grazing or mowing, removal of marginal landscape features or establishment of larger parcels – following the establishment of the Common Agricultural Policy (CAP, e.g. Stoate *et al.* 2001, Donald *et al.* 2002b, de Sadeleer and Born 2004, Petit and Firbank 2006).

The strong temporal and spatial correlations between such decline and the agricultural intensification suggest that changes in farming practices are at least partly responsible (e.g. Gates *et al.* 1993, Pain and Pienkowski 1997, Kyrkos *et al.* 1998, Peach *et al.* 1999, Chamberlain and Fuller 2000, 2001, Chamberlain *et al.* 2000, Fuller *et al.* 2001, Donald *et al.* 2001, 2002a, Donald and Evans 2006). The cause-and-effect relationship was clearly proved for some species such as the Corn Crake *Crex crex* (Stowe *et al.* 1993) or some granivorous passerines (Siriwardena *et al.* 2000a, 2001).

Currently, many conservation-interest farmland birds require resources that are only found in traditionally or non-intensively managed areas (e.g. Tucker and Evans 1997, Pain and Pienkowski 1997, Kleijn and Sutherland 2003, Donald and Evans 2006), stressing the need to precisely identify and quantify such features with bird-environment modelling.

Objectives

The present thesis focuses on the enhancement of species-environment modelling for understanding bird species spatial distribution patterns. We fundamentally follow a dichotomous modelling framework.

We first aim to describe the co-variation in space of (1) the main environmental conditions and (2) the bird assemblages on a broad spatial extent. A descriptive and synecological approach (Chapter 1) is therefore developed to identify the major factors governing the spatial arrangement of assemblages in a heterogeneous landscape. Secondly, we undertake a specific analysis of the fine-scale breeding habitat requirement for a depleted migratory bird species of extensive farming areas, the Red-backed Shrike *Lanius collurio L.* (Chapter 2 to 5).

These two contrasted approaches are conducted in a landscape context in Southern Belgium (Central Famenne *sensu lato*), but are situated on both sides of the holism – reductionism gradient. For this reason, we judged it very interesting to evaluate and compare their respective ecological and conservation interest, usefulness and limitations.

Identifying the environmental gradients related to spatial variation in bird assemblages

According to the Hutchinsonian formalism, we conceptually seek to untangle a complex mixture of population hypervolumes scattered in an infinite-dimensional environmental hyperspace by reducing the latter to some composite dimensions with the help of spatially-explicit direct gradient analyses (see Figure 2). The ultimate objective is to identify the main environmental and spatial gradients related to variations in bird species assemblages on a large extent (about 300 km²) (Chapter 1).

Such a broad and multi-species description requires relying on available information regarding both environmental variables and species assemblages. The environmental conditions are quantified on the basis of available cartographic data. The species assemblages are described using a semi-quantitative breeding bird atlas survey. Grid-based atlases are one of the most frequently – and often the only – used form of biological survey at various spatial scales in many European countries (e.g. Hagemeijer and Blair 1997, Donald and Fuller 1998, Schmid *et al.* 1998, Tobalske and Tobalske 1999, Pasinelli *et al.* 2001). A subordinated objective is therefore to assess the usefulness, the limitations and the ecological significance of fine-grained atlas data for describing bird-environment relations at the landscape scale within a conservation context.

We aim at explicitly accounting for the importance of spatially-structured intrinsic processes and environmental conditions in creating the observed species distribution patterns. Methods of variation partitioning are selected for incorporating these spatial structures into the modelling process. This approach allows to partition the assemblage variation between non-spatial environmentally-based variation, spatially-structured environmentally-based variation, spatial variation that is not shared by environmental variables, and unexplained variation.

Identifying the fine-scale habitat requirements for a farmland species

This auto-ecological approach aims at designing a functional and spatially-explicit species-habitat modelling framework in order to identify its fine-scale habitat requirements in farming landscapes, for ultimately proposing effective specific conservation measures. This species is retained as an indicator of extensive farming areas (see Lefranc and Worfolk 1997, Van Nieuwenhuyse 1999 and beginning of Chapter 2 for more details).

Such a single-species and functional approach allows to pay attention to the potentially distorted relation underlined above between the species niche and its observed spatial distribution (see Figure 2). This is very rarely and, to our knowledge, never thoroughly achieved in the vast majority of published species-habitat models (Wiens 2002, Guisan and Thuiller 2005). These largely assume only situations A or B as referred to the Figure 1 (see Pulliam 2000). Accounting for this decoupling (situations C and D in Figure 1) requires an ad-hoc field data collection on a consequently rather limited spatial extent (about 40 km²). Habitat is described on the basis of the available specific literature in order to reflect the ecological resources that determine the functional interactions between the organisms and their environment. The spatial scale (resolution) of investigation matches the key specific ecological processes thought to drive the species distribution in the landscape.

More specifically, the different objectives of this second section of the thesis are the following:

- The first specific objective (Chapter 2) is to design a statistical procedure that (1) accounts for the meaningless species absences (situation D in Figure 1) and (2) achieves a screening of the species occurrences not conveying reliable information about the suitability of environmental conditions (situation C in Figure 1), before delineating the habitat hypervolume (ellipse in Figure 1 an see Figure 2).
- The previous statistical approach is blind to any ecological foundation. Therefore, on the basis of breeding success data, we secondly aim at investigating the possible discrepancies between such a statistical procedure and the ecological reality (Chapter 3, see Figure 2). In particular, such breeding success information should enable (1) to identify the individuals that are occurring in unsuitable environmental conditions and therefore (2) to evaluate habitat suitability in a way that is more closely related to population productivity.
- The third specific objective (Chapter 4) is to relate the decoupling between habitat suitability and occupancy to conspecific proximity in

this migratory species (Figure 2). A niche-based modelling framework integrating (1) pairing and breeding success data and (2) spatial analyses is foreseen to assess the importance of conspecific proximity in a landscape context and from a conservation perspective.

- The fourth specific objective (Chapter 5) is to design an integrative spatial modelling approach that is able to deal with the following issues: meaningless absences (Chapters 2 and 3), existence of misleading presences (Chapters 2 and 3), non-exhaustive presence data (leading to meaningless absences) and influence of landscape context and of spatially-structured intrinsic population-based processes (Chapter 4). This information should enable to quantify (1) the habitat suitability and (2) the relative likelihood of species occupancy for any site within a given landscape, hence providing useful tools for management. The discrepancies between (1) and (2) are finally analyzed within a spatial context (Figure 2) and their conservation significance is evaluated.

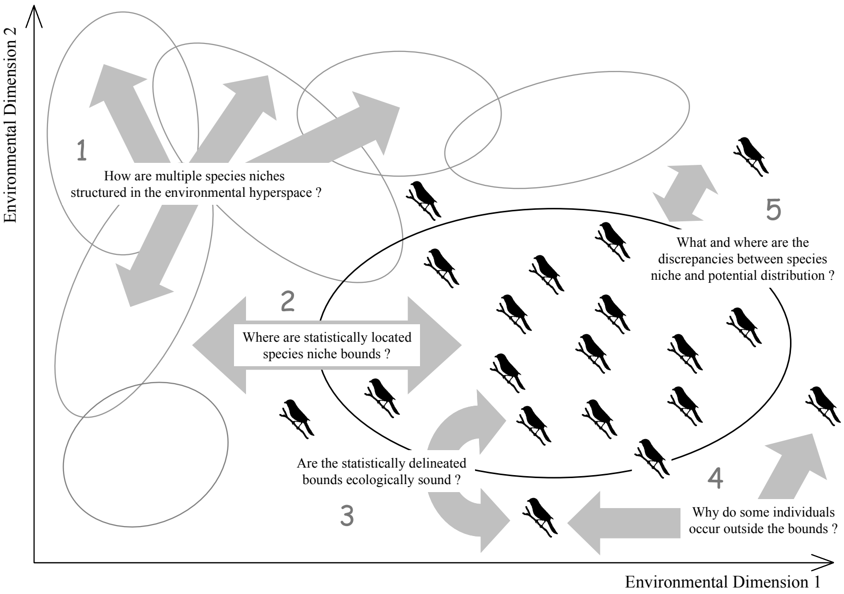


Figure 2. Research organization chart represented in a simplified 2-dimensions projection of the Hutchinsonian hyperspace and showing relations and connections between the different chapters of this thesis. Each chapter is summarized by the main question it raises. Ellipses delineate the environmental boundaries of species niche. Chapter 1 focuses on bird assemblages (grey ellipses) and Chapters 2 to 5 look into a single species niche (black ellipse). Bird silhouettes represent occurrence records of this latter species (Red-backed Shrike).

Chapter 1 – Untangling the Spatial Structuration of Bird Assemblages using a Coarse-scale Descriptive Approach

Note – This study has been published in Journal of Biogeography (Titeux, N., Dufrêne, M., Jacob, J. P., Paquay, M. and Defourny, P. 2004. Multivariate Analysis of a fine-scale Breeding Bird Atlas using Geographical Information System and Partial Canonical Correspondence Analysis – Environmental and Spatial Effects. Journal of Biogeography, 31: 1841-1856). The text presented here is slightly modified from the original publication for layout and terminology harmonization, as well as for references updating and methodological clarification.

Foreword [1]

According to the definitions given in the introduction of the thesis, this Chapter focuses on describing the coarse relations between bird species assemblages and environmental conditions, not on identifying the precise resources defining the habitats for the many species populations. For this purpose, the environmental conditions of the sampling sites were characterized by biotope-based and not habitat-based variables. These environmental variables were then used to build the composite dimensions of an environmental hyperspace that were related to bird assemblages thanks to a multi-species modelling framework.

Abstract

The aim of this study was to assess the relative roles of environment and space in coarsely structuring bird assemblage composition at the landscape level, on the basis of a fine-scale bird atlas data set. The study was carried out in southern Belgium using grid cells of 1x1 km, based on the distribution maps of the 'Oiseaux nicheurs de Famenne: Atlas de Lesse et Lomme' which

---

[1] This Foreword does not belong to the paper published in Journal of Biogeography.

contained abundance for 103 bird species. Species found in less than 10% or in more than 90% of the atlas cells were omitted from the bird data set for the analysis. Cells were characterised by 59 metrics, quantifying their environmental composition and spatial pattern, using a Geographic Information System (GIS). Partial Canonical Correspondence Analyses (CCA) were used to partition the variance of bird species matrix into independent components: (a) 'pure' environmental variation, (b) spatially-structured environmental variation, (c) 'pure' spatial variation and (d) unexplained, non-spatial variation. The variance partitioning method showed that the selected metrics explained 27.5% of the variation, whilst 'pure' spatial and spatially-structured environmental variables explained only a weak percentage of the variation in the bird species matrix (2.5% and 4%, respectively). Avian assemblages composition was primarily related to the degree of urbanisation and the amount and composition of forested and, to a lesser extent, of open areas. These variables explained more than half of the variation for three species and over one third of the variation for 12 species. The results indicated that the majority of explained variation in species assemblages was attributable to local environmental factors. However, the method did not appear to be appropriated for detecting and thoroughly extracting the spatial variation of assemblages. Consequently, the large amount of unexplained variation was probably due to missing spatial structures and to biased species abundance data. Furthermore, other relevant environmental factors – that were not taken into account in this study and which may operate at different spatial scales – could drive bird assemblage structure. Since a large proportion of ecological variation can be shared by environment and space, the partitioning procedure was found to be useful when analysing multi-specific atlas data, but it needed improvement to factor out all-scale spatial components of this variation (the source of 'false correlation') and to bring out the 'pure' environmental variation for ecological interpretation.

1. Introduction

The mapping of bird distributions using grid-based atlases at different spatial scales is one of the most frequently used form of ornithological survey (e.g. Jacob and Paquay 1992, Hagemeijer and Blair 1997, Schmid *et al.* 1998). Atlases have numerous potential uses including bird – environment relationship assessment (Donald and Fuller 1998) and the prediction of species distribution (Bibby *et al.* 1992, Tobalske and Tobalske 1999, Pearson and Dawson 2003). A constraint to such analysis is the availability of environmental data at the same spatial scale (grain and extent) as the atlas data (Donald and Fuller 1998). The development of Geographical Information Systems (GIS) and the increasing availability of digital maps allow the manipulation of spatially-distributed data (Burrough and McDonnell 1998, Schmit *et al.* 2006) to extract environmental data, i.e. various metrics, at selected scales (Berry and McGarigal 1998), and thus to avoid this technical problem.

The use of atlas data and GIS to develop species-habitat models (Gates *et al.* 1993, Tobalske and Tobalske 1999, Siriwardena *et al.* 2000b, Vanhinsbergh and Chamberlain 2001), to predict bird species distributions (e.g. Osborne and Tigar 1992), and to assess and predict relative distribution changes (Böhning-Gaese and Bauer 1996, Chamberlain and Fuller 2000, Chamberlain *et al.* 2001, Telfer *et al.* 2002) has already been explored in a species-specific way, often by means of Generalized Linear Models (GLM). Other taxonomic groups have also been investigated in the same way (e.g. Guisan *et al.* 1999, Dennis and Hardy 1999, Luoto *et al.* 2002). In the context of multi-species analysis, atlases and GIS have already been used to provide environmental interpretation of avifaunal zonation (Pasinelli *et al.* 2001), to model species richness distribution (Lobo and Martin-Piera 2002, Lobo *et al.* 2002, Maes *et al.* 2003) and to relate animal (Natuhara and Imai 1996, Storch *et al.* 2003) or floristic (Guisan *et al.* 1999) assemblages to environmental conditions. In order to explain relationships between species assemblages and environmental variables, ordination methods are often used, especially direct gradient analyses in which species occurrence or abundance are directly related to environmental variables (ter Braak 1986).

Spatial structuring is a fundamental and functional component of ecosystems: the observed distribution of organisms (response variables) result from the spatial dependence in these response variables (spatial autocorrelation) – due to biotic processes such as growth, mortality, dispersal or predation – or are caused by the dependence of response variables on several explanatory variables which are themselves spatially-structured, or both (Legendre and Legendre 1998). Spatial heterogeneity is therefore not the result of some random, noise-generating processes, and it is necessary to explicitly account for it. However, incorporating the spatial structure of the data (response and explanatory variables) into the modelling process is rarely done in the case of multi-specific atlas data analysis.

Appropriate approaches dealing with this very general property of ecological variables exist (Legendre 1993). Borcard *et al.* (1992) have proposed a method based on partial Canonical Correspondence Analysis (CCA) to factor out the spatial component of the ecological variation of species assemblages and to partition this variation as follows: (a) non-spatial environmental variation, (b) spatially-structured environmental variation, (c) spatial variation that is not shared by environmental variables and (d) unexplained, non-spatial variation. This method has been used for several purposes, notably in the assessment of the relative roles of environment and space in driving fauna or vegetation distributions (i.e. Borcard and Legendre 1994, Pinelalloul *et al.* 1995, Monti *et al.* 1996, Hobson *et al.* 2000). To our knowledge it has never been applied to atlas data sets in spite of these data being a special case of spatially-structured ecological data. Therefore, such an assessment should be performed. The study objective is to evaluate the applicability and usefulness of this variance partitioning technique for a fine-scale breeding bird atlas for a region of 279 km² encompassing various land use types, in order to quantify the relative roles of environment and space in coarsely structuring bird species assemblages, to display substantial environmental or spatial gradients and to highlight relevant environmental variables for bird assemblage composition.

2. Materials and methods

2.1. Study area

The region is located in the Southern part of Belgium and forms a vast depression (altitude: 140-300 m) characterized by schistose and calcareous rocks. The study area corresponds to the Central Famenne and covers 279 km². It is centred on the confluence of two rivers: the Lesse and the Lomme (Figure 3). This area has poorly drained clay soils derived from schist, which are unfavourable to agriculture (Thill 1964). Consequently, land use is dominated by grasslands with scattered hedgerows (39% of the total surface) and forests of varying extents and types (37%), while cultivated fields (16.5%) and urban areas (4%) are less frequent.

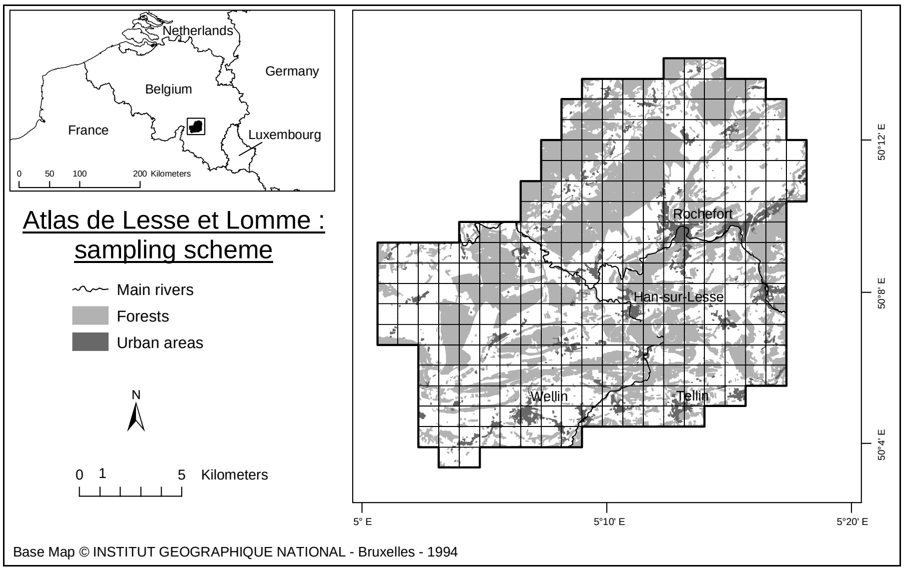


Figure 3. Study area covered by the Atlas de Lesse et Lomme (Jacob and Paquay 1992) showing the 279 atlas squares.

2.2. Avifauna and environmental data

2.2.1. Avifauna

The distribution maps presented in the 'Oiseaux nicheurs de Famenne: Atlas de Lesse et Lomme' (Jacob and Paquay 1992) are composed of 279 1 km² grid cells, hereafter called 'squares' (Figure 3). Bird censuses were conducted from 1985 to 1989 during spring and summer (breeding season) by 70 members of the COA (Centrale Ornithologique Aves, Belgium).

Each year between 1985 and 1988, a different set of squares was surveyed three times during the breeding season (March-April, May and June-July). Surveys began 15-20min before sunrise and ended within 4h after sunrise. Each square was totally surveyed during one season. The last year (1989) was devoted to verification to ensure that each square was surveyed with the same intensity and to limit the observer bias that could arise from *a priori* knowledge of the avifauna of certain squares or from differential census intensity. The survey time by square ranged from 5 to 25h depending on the environment complexity within the square.

For each square, the observer recorded the presence-absence and counted the number of breeding pairs (abundance) for each species. Abundance was aggregated into six classes as described in Table 1. For this study, the median of each interval was retained except for the last one (see Table 1). Only records of probable or confirmed breeding – as defined by the international EOAC code (Timothy and Sharrock 1974) – were considered, while possible breeding was excluded from the species set (for more details, see Jacob and Paquay 1992). Species abundance data were gathered in a species matrix Y.

Table 1. Classes of abundance as counted at the square level and used by the Atlas de Lesse et Lomme (Jacob and Paquay 1992).

Class	Number of breeding pairs	Median number of breeding pairs used for analysis
0	0	0
1	1 to 5	2.5
2	6 to 10	8
3	11 to 20	15
4	21 to 40	30
5	More than 40	50

2.2.2. GIS database

The GIS Software ArcInfo 7.0.4. (ESRI, Redlands, CA, USA) was used to combine several environmental data layers from different sources.

A 30m resolution digital elevation model (Institut Géographique National, edited in 1995) was used to represent the topography of the region. Digital 1:50,000 vector topographic maps (Institut Géographique National, edited in 1994) were used as a planimetric reference and land cover information. In addition, a set of elements that could potentially have a role in structuring the bird assemblages was retained for analysis. In particular, several types of linear and surface features were digitised from aerial 1:10,000 colour orthophotographs (acquired between 1985 and 1990) and digital 1:10,000 raster topographic maps (Institut Géographique National, edited in 1981 or 1988-89). Map digitizing based on visual interpretation also allowed to 'downdate' the environmental description to the bird atlas period (1985-1989).

Linear elements were hedgerows, railways, roads and the hydrographical network. Hedges were allocated to four classes by visual interpretation of the aerial photographs: (1) broken hedges formed by scattered and isolated shrubs, (2) continuous, but narrow hedges, (3) continuous and wide hedges and (4) tree rows. Roads and railways were allocated to four classes: (1) motorways, major and secondary roads, and railways, (2) smaller roads, (3) pathways and (4) footpaths. The hydrographical network comprised rivers (1) narrower than 3m, (2) from 3 to 15m wide, and (3) from 15 to 30m wide. No rivers exceeded 30m wide.

All surface features were classified into one of 10 land cover types: pond, marsh, urban area, hardwood forest, coniferous forest, mixed forest, orchard, grassland, crop field and heathland.

2.2.3. Environmental metrics computation

This study focussed on the relationships between bird assemblages on the one hand, and composition and spatial configuration of the environment on the other hand. Consequently it was necessary to characterise this composition and the spatial configuration for each of the 279 squares of the atlas.

For each atlas square, the total length of the 11 linear elements and the surface occupied by each of the 10 land cover types were calculated. These 21 metrics documented the environment composition in the squares (Table 2).

Spatial configuration metrics were computed using the standalone ArcInfo application FRAGSTATS*ARC version 2 (Berry and McGarigal 1998). Numerous metrics of spatial pattern were available, but many are highly correlated or redundant (Coppedge *et al.* 2001) and no consensus exists as to which metric best describes any particular ecological phenomenon (Gustafson 1998). These metrics quantified different aspects of the environment pattern, and 34 were selected to quantify the spatial configuration of the 279 atlas squares (Table 2). Two groups of indices were retained: class level metrics (for each land cover or linear element type) and square level metrics (for the whole atlas square). Several core area metrics were computed. Core area was defined as the area within a patch beyond some specified buffer width. Core area metrics were arbitrarily based, for the purpose of this study, on a 50m-wide buffer along the perimeter of each patch. Core area metrics reflected both environment composition and configuration, and were related to the 'edge effect' (e.g. Woodward *et al.* 2001). To avoid the effect of patch truncation at atlas square boundaries on the calculation of core areas, the full size of patches that overlapped two squares was used, that is, the feature area extending outside of the squares was included for the calculation of core areas (see Trzcinski *et al.* 1999). In each atlas square, several statistics representing the total edge length of a particular patch type (class level) were computed. Several interface metrics were also calculated, representing the total edge length between two particular patch types or between a patch type and a linear element.

Square level metrics that quantified the square mosaic as a whole were also computed. The area-weighted mean shape index measured the average perimeter-to-area ratio for all patches in the square, by weighting patches according to their size, so that the larger the patch the more strongly it contributed to the value of the metric. Shannon's diversity and evenness indices referred to the number of patch types and the distribution of area among these different types.

Four metrics were computed to describe the topographic characteristics of each square: minimum elevation, maximum elevation, mean elevation and standard deviation of elevation. The latter metric was a measure of the altitude variability in a square.

The nature and number of explanatory variables (59) resulted from a trade-off between concerns to maintain a high ratio of sites-to-variables for subsequent multivariate analysis (CCA) and a relevant and coherent description of the environmental conditions of the squares, within the framework of this study. All the environmental metrics (hereafter called 'environmental variables') are detailed in Table 2 and were gathered in an environmental matrix X.

Table 2. The 59 metrics used to describe the environmental conditions of the 1-km atlas squares.

Acronym	Units	Description
Hedge1	meters	Total length of scattered and isolated shrubs
Hedge2	meters	Total length of continuous but narrow hedges
Hedge3	meters	Total length of continuous and wide hedges
Trees	meters	Total length of tree rows
Riv1	meters	Total length of rivers from 0 to 3m wide
Riv2	meters	Total length of rivers from 3 to 15m wide
Riv3	meters	Total length of rivers from 15 to 30m wide
RoA	meters	Total length of motorways, major and secondary roads and railways
RoB	meters	Total length of smaller roads
Path	meters	Total length of pathways
Foot	meters	Total length of footpaths
Pond	ha	Total area of ponds
Marsh	ha	Total area of marshes
Urban	ha	Total area of urban zones
HFor	ha	Total area of hardwood forests
CFor	ha	Total area of coniferous forests
MFor	ha	Total area of mixed forests
Orch	ha	Total area of orchards
Grass	ha	Total area of grasslands
Field	ha	Total area of fields
Heath	ha	Total area of heathlands
HForCore	ha	Total amount of hardwood forests core area
CForCore	ha	Total amount of coniferous forests core area
MForCore	ha	Total amount of mixed forests core area
GrassCore	ha	Total amount of grasslands core area
FieldCore	ha	Total amount of fields core area
RivFor	meters	Total length of edges between rivers and forested areas
RivOpen	meters	Total length of edges between rivers and open areas
PondFor	meters	Total length of edges between ponds and forested areas
PondOpen	meters	Total length of edges between ponds and open areas
RoAFor	meters	Total length of edges between roads A and forested areas
RoAOpen	meters	Total length of edges between roads A and open areas
RoBFor	meters	Total length of edges between roads B and forested areas
RoBOpen	meters	Total length of edges between roads B and open areas
HForCFor	meters	Total length of edges between hardwood and coniferous forests
HForMFor	meters	Total length of edges between hardwood and mixed forests
CForMFor	meters	Total length of edges between coniferous and mixed forests
HForField	meters	Total length of edges between hardwood forests and fields
HForGrass	meters	Total length of edges between hardwood forests and grasslands

(Table 2 continued)

CForField	meters	Total length of edges between coniferous forests and fields
CForGrass	meters	Total length of edges between coniferous forests and grasslands
MForField	meters	Total length of edges between mixed forests and fields
MForGrass	meters	Total length of edges between mixed forests and grasslands
FieldGrass	meters	Total length of edges between fields and grasslands
HForTE	meters	Total length of edges involving hardwood forests
CForTE	meters	Total length of edges involving coniferous forests
MForTE	meters	Total length of edges involving mixed forests
GrassTE	meters	Total length of edges involving grasslands
FieldTE	meters	Total length of edges involving fields
NP	none	Number of patches (all patch types together)
PR	none	Patch richness (number of patch types)
Awmsi	none	Area-weighted mean shape index $= \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left(\frac{p_{ij}}{2\sqrt{\pi a_{ij}}}\right)\left(\frac{a_{ij}}{A}\right)\right]$
TE	meters	Total length of edges
Shdi	none	Shannon's diversity index $= -\sum_{i=1}^{m}\left(P_i \ln P_i\right)$
Shei	none	Shannon's evenness index $= \frac{-\sum_{i=1}^{m}\left(P_i \ln P_i\right)}{\ln m}$
Altmin	meters	Minimal elevation
Altmax	meters	Maximal elevation
Altmean	meters	Mean elevation
Altstd	meters	Elevation standard deviation

p_{ij} = perimeter of patch *j* for type *i*
a_{ij} = surface of patch *j* for type *i*
A = total surface of the square (i.e. 1 km²)
P_i = proportion of the total surface of the square occupied by patch type *i*
m = number of patch types in the square

2.3. Data analysis

2.3.1. Canonical Correspondence Analysis

The avian assemblages (matrix Y) were related to environmental variables (matrix X) using Canonical Correspondence Analysis (CCA) with the program CANOCO 4.0 for Windows (ter Braak and Smilauer 1998). CCA is an eigenvector ordination technique for multivariate direct gradient analysis (ter Braak 1986) and constitutes an extension of multiple regression analysis for multivariate responses, i.e. matrix Y (Palmer 1993). The ordination axes are constrained to be linear combinations of the explanatory variables (ter Braak 1986, Natuhara and Imai 1996, Guisan *et al.* 1999). As a consequence, CCA produces an ordination of Y that is constrained to be related to variables in matrix X (ter Braak 1986).

This technique is appropriate when the responses of the dependent variables are expected to be unimodal along environmental gradients (ter Braak and Smilauer 1998). Redundancy Analysis (RDA) is another ordination technique where the underlying response model is a monotonic distribution of species along environmental gradients, which limits its use when large gradients are analysed (Guisan and Zimmermann 2000). CCA is also appropriate when dealing with presence-absence data (ter Braak 1986, Legendre and Legendre 1998).

Ideally, rare species would also be examined with gradient analysis. Unfortunately these species tend to obscure assemblage patterns because their weak prevalence introduces a large number of zeros in the bird data set. These species may have an unduly large influence on the analysis (ter Braak and Smilauer 1998), by creating an increase in the total inertia of the species data set or a distortion of the ordination. Furthermore, rare species are generally associated with specific environmental conditions not summarised by the 59 environmental variables. For these reasons, species found in less than 10% of the atlas squares were discarded from the bird data set for the subsequent analysis. Likewise, ubiquitous species did not exhibit a relationship with the computed environmental variables and might skew or obscure community patterns. Consequently, species found in more than 90% of the atlas squares were also omitted.

2.3.2. Distinction between environmental and spatial effects

Spatial autocorrelation can be defined as the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations (Legendre 1993). In the case of atlas squares, the observed bird species assemblage in one square might be influenced by the species assemblages in the surrounding squares because of contagious biotic processes. Similarly, environmental variables used to describe the squares are also neither randomly nor uniformly spatially-distributed, but structured by physical processes causing gradients and/or patchy structures. One consequence of this general property of ecological variables is that the assumption of independence of the observations is not respected (Legendre 1993).

Therefore, it is necessary to explicitly incorporate the spatial structure of the data within the modelling process of species-environment relations. In the present case of multivariate dependent data set, partial (constrained) ordination analyses are advisable (Legendre 1993, Jongman *et al.* 1995). The ordination axes are computed after the effects of known or undesirable variables (hereafter called covariables, gathered in a third matrix W) have been removed by multiple linear regression from variables present in Y and X. A partial CCA therefore relates residuals in Y to residuals in X. Conversely, a partial CCA can relate residuals in Y to residuals in W after the effects of X have been removed.

Borcard *et al.* (1992) proposed to use a third-order function of the geographical coordinates of the sampling locations for removing the spatial structure in species (Y) and environmental (X) matrices by polynomial regression. A matrix W of x and y geographical coordinates of the centres of the atlas squares was compiled by adding all terms for a cubic trend surface:

$$x \quad y \quad x^2 \quad xy \quad y^2 \quad x^3 \quad x^2y \quad xy^2 \quad y^3$$

This ensures the detection of more complex spatial features in the species data set than linear gradient patterns only (Borcard *et al.* 1992), because third order polynomial regressions simulate either skewed or bimodal responses, or even combinations of both of these (Guisan and Zimmermann 2000).

In total, three matrices were constructed: species assemblages Y, environmental variables X and spatial matrix of geographical coordinates W. By making two canonical ordinations of Y, each of them constrained by one set of explanatory variables (X or W) one obtained a measure of the effects of the environmental conditions (X) and the spatial structure (W) that are not independent because of spatially-structured environmental variables. Consequently, the two runs were partially redundant (Borcard *et al.* 1992) and we partitioned the variation of the bird species matrix Y as follows (Legendre 1993): (a) non-spatial environmental variation, (b) spatially-structured environmental variation, (c) spatial variation that is not shared by environmental variables and (d) unexplained, non-spatial variation. This partition was possible by using the sum of all canonical eigenvalues of two canonical ordinations, each of them constrained by one set of explanatory variables (X and W), and of two partial canonical ordinations, each of them constrained by one set of explanatory variables while controlling for the effect of the others (deemed as covariables) (Borcard *et al.* 1992).

2.3.3. Environmental and spatial variables selection

In the canonical ordination constrained by the polynomial function of geographical coordinates of the squares (hereafter called 'spatial' CCA), the geographical terms were submitted to forward selection and Monte Carlo permutation tests (999 permutations; $P < 0.05$) in CANOCO to retain only the most important ones and to avoid the artificial increase of explained variation by chance. These terms only will be incorporated into subsequent partial canonical ordination constrained by spatial effects while controlling for the environmental effects (hereafter called 'spatial' partial CCA). In the partial canonical ordination constrained by environmental variables while controlling for the selected geographical terms (hereafter called 'environmental' partial CCA), variables were submitted to forward selection. Their significance was tested by Monte Carlo permutations tests (999 permutations) to retain the 10 ones that best explained the variation in the breeding bird abundance. Furthermore, the interactions of these 10 environmental variables were computed in CANOCO and added as explanatory variables. An additional 'environmental' partial CCA was run using the 10 selected environmental variables and their interactions, and the best 10 variables or interactions were again selected manually during a forward selection procedure. These selection procedures were not species specific: selections were made of the variables which explained successively the highest proportion of variation in the species data set as a whole (Guisan *et al.* 1999). Forward selection included collinear variables if they made important contributions to the variation in species composition, but eliminated completely redundant variables (Palmer 1993). Finally, only the 10 retained environmental (interactions included) and the selected geographical variables were used in all 'environmental' or 'spatial' (partial) canonical ordinations.

2.3.4. CCA and partial CCA fit and interpretation measures

Canonical eigenvalues measured the amount of variation in the species data set that was explained by the explanatory variables. The total variation in the species matrix (called 'Total Inertia') was measured by the chi-square statistic of the sample-by-species table divided by the table's total (ter Braak and Smilauer 1998). This was equal to the sum of all eigenvalues of a CA on species assemblages. Dividing the sum of all canonical eigenvalues by the total inertia gave the percentage of total variance in the species data set that was explained by the explanatory variables, i.e. an overall measure of the CCA fit (ter Braak and Smilauer 1998). Similarly, the proportion of the total inertia in the species data that was explained by each canonical axis was computed. Unrestricted Monte Carlo permutations were used to test the statistical significance of the first canonical axis alone and of all canonical axes together. Tests of significance in CCA did not depend on parametric distributional assumptions (Palmer 1993). Therefore, species data were not transformed and environmental and spatial variables were simply standardized.

Canonical coefficients and intraset correlations were examined to evaluate the relative contributions of environmental variables to the structuring of the assemblage composition (Boren *et al.* 1999). Canonical coefficients were the coefficients of a weighted multiple regression of the sample scores on the environmental variables and thus, defined the ordination axes as linear combinations of the environmental variables (ter Braak 1986). Intraset correlations provided the correlation between an environmental variable and a canonical axis, adjusted for the covariables (partial correlations) (ter Braak 1986). In order to interpret the ecological signification of the canonical axes, this latter coefficient should be used instead of the canonical coefficient in the case of multicollinearity of environmental variables.

The marginal effect of an environmental variable was equal to the eigenvalue of a partial CCA if the corresponding variable was the only environmental variable (additionally to the variance explained by covariables). The conditional effect of an environmental variable was equal to the additional amount of variance in species assemblages explained by the corresponding variable at the time it was included into the model during a selection procedure (additionally to the variance explained by covariables) (ter Braak and Smilauer 1998). Such effects were also examined to assess the relative contributions of environmental variables for structuring the assemblage composition.

2.3.5. Species–Environment biplots

CCA allowed a visual interpretation of multiple environmental dimensions by plotting species and environmental variables in the ordination diagram formed by two canonical axes obtained by the 'environmental' partial CCA. The type-2 scaling of CANOCO was used, where the species were represented at the centroids of the sampling sites where they occur. On such a biplot, environmental variables were depicted by lines. The length of each line represented the relative importance of the environmental variable to the assemblage structuring (in the two-dimensional plane under consideration) (ter Braak 1986). The further the species points were from the origin, the more informative the ordination: a cluster of species near the origin is difficult to interpret and indicates a weak species-environment relationship (MacFaden and Capen 2002). A perpendicular line could virtually be drawn from each species position to each environmental variable. Endpoints of the perpendicular lines indicated relative positions of species distribution centres along this environmental variable (ter Braak and Smilauer 1998).

3. Results

The bird census program detected 103 breeding bird species during 1985 – 1989. The species richness of the squares was 45.7 ± 8.9 (range: 9 – 68 breeding species by km²). The number of atlas squares occupied by each species was highly variable. Only 80 species were found consistently enough (more than 10% of atlas squares) to allow environmental analysis. Among them, 17 species were detected in more than 90% of atlas squares and were therefore discarded (Table 3), resulting in 63 species.

3.1. Distinction between environmental and spatial effects

The following geographical terms were retained by the forward selection procedure ($P < 0.05$) during the 'spatial' CCA: $x \quad y \quad x^2 \quad xy^2$

The 10 environmental variables or interactions selected during 'environmental' partial CCA were summarised in Table 4. The number of explanatory variables in matrix X was arbitrarily limited to 10. The use of additional variables in the analysis did not substantially improve the results because these explained only a negligible part of the variation. Species abundance transformation, taking Napierian logarithms $[z' = \ln(z+1)]$, did not considerably improve the results either.

The four (partial) CCA runs gave the following results:

1. 'environmental' CCA of the species assemblages: sum of all canonical eigenvalues = 0.532 = E;
2. 'spatial' CCA of the species assemblages: sum of all canonical eigenvalues = 0.109 = S;
3. 'environmental' partial CCA of the species assemblages: sum of all canonical eigenvalues = 0.464 = E(S);
4. 'spatial' partial CCA of the species assemblages: sum of all canonical eigenvalues = 0.042 = S(E).

The total variation in the species assemblages (total inertia = G) was 1.69. The proportion of variance explained by the environmental variables and by the polynomial function of geographical coordinates was equal to the sum of all canonical eigenvalues of a CCA [E or S] plus the sum of all canonical eigenvalues of the complementary partial CCA [S(E) or E(S), respectively], divided by total inertia:

$$[E + S(E)] / G = [0.532 + 0.042] / 1.69 = 0.34$$

$$[S + E(S)] / G = [0.109 + 0.464] / 1.69 = 0.34$$

The variation of species assemblages could be partitioned as follows (Borcard *et al.* 1992):

(a) non-spatial environmental variation
$0.464 / 1.69 = 0.2746$

(b) spatially-structured environmental variation
$[0.532 - 0.464]\cdot/ 1.69 = [0.109 - 0.042] / 1.69 = 0.0402$

(c) spatial variation of the species assemblages that were not shared by environmental variables
$0.042 / 1.69 = 0.0249$

(d) unexplained, non-spatial variation
$1 - 0.3396 = 0.6604$

The partitions of species assemblages variation are shown on Figure 4.

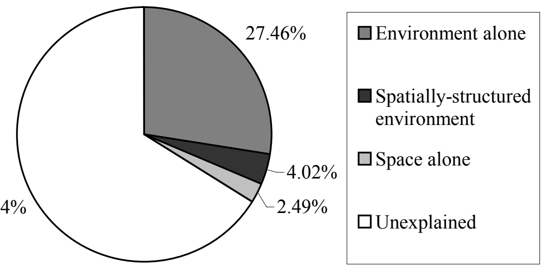


Figure 4. Variation partitioning of the bird species assemblages.

Table 3. List of all recorded species with English and scientific names (taxonomic order), code (based on scientific names), total number of occupied atlas squares (*N*) and proportion of variance explained by environmental variables (%E) and by space (%S) as a result of the respective partial CCA (only for species present in > 10% and < 90% of atlas squares).

Species					
English name	Scientific name	Code	*N*	%E	%S
Little Grebe	*Tachybaptus ruficollis**	Taru	13	–	–
Grey Heron	*Ardea cinerea**	Arci	2	–	–
Common Mallard	*Anas platyrhynchos*	Anpl	91	28	< 1
European Honey–buzzard	*Pernis apivorus*	Peap	34	7	3
Red Kite	*Milvus milvus**	Mimi	1	–	–
Eurasian Sparrowhawk	*Accipiter nisus*	Acni	86	19	2
Common Buzzard	*Buteo buteo*	Bubu	181	37	5
Common Kestrel	*Falco tinnunculus*	Fati	50	10	4
Hazel Grouse	*Bonasia bonasia**	Bobo	4	–	–
Common Quail	*Coturnix coturnix**	Coct	24	–	–
Common Pheasant	*Phasianus colchicus*	Phco	138	18	2
Water Rail	*Rallus aquaticus**	Raaq	3	–	–
Corn Crake	*Crex crex**	Crcr	5	–	–
Common Moorhen	*Gallinula chloropus**	Gach	22	–	–
Eurasian Coot	*Fulicula atra**	Fuat	3	–	–
Northern Lapwing	*Vanellus vanellus*	Vava	67	11	2
Little Ringed Plover	*Charadrius dubius**	Chdu	3	–	–
Eurasian Woodcock	*Scolopax rusticola*	Scru	44	8	5
Stock Pigeon	*Columba oenas*	Cooe	29	3	2
Common Wood–pigeon	*Columba palumbus***	Copa	266	–	–
Eurasian Collared–dove	*Streptopelia decaocto*	Stde	102	6	3
European Turtle–dove	*Streptopelia turtur*	Sttu	177	8	2
Eurasian Cuckoo	*Cuculus canorus*	Cuca	177	13	2
Barn Owl	*Tyto alba*	Tyal	27	19	1
Eurasian Eagle–owl	*Bubo bubo**	Bubb	3	–	–
Little Owl	*Athene noctua*	Atno	62	14	2
Tawny Owl	*Strix aluco*	Stal	80	16	3
Long–eared Owl	*Asio otus*	Asot	71	12	4
Eurasian Nightjar	*Caprimulgus europaeus**	Caeu	2	–	–
Common Swift	*Apus apus*	Apap	49	37	3
Common Kingfisher	*Alcedo atthis**	Alat	10	–	–
Eurasian Green Woodpecker	*Picus viridis*	Pivi	104	9	6
Black Woodpecker	*Dryocopus martius*	Drma	44	14	3
Great Spotted Woodpecker	*Dendrocopos major*	Dema	227	44	< 1

(Table 3 continued)

Middle Spotted Woodpecker	*Dendrocopos medius**	Deme	24	–	–
Lesser Spotted Woodpecker	*Dendrocopos minor*	Demi	68	16	< 1
Wood Lark	*Lullula arborea**	Luar	12	–	–
Skylark	*Alauda arvensis*	Alar	216	45	4
Sand Martin	*Riparia riparia**	Riri	11	–	–
Barn Swallow	*Hirundo rustica*	Hiru	130	40	1
Northern House–martin	*Delichon urbica*	Deur	78	31	3
Tree Pipit	*Anthus trivialis***	Antr	261	–	–
Meadow Pipit	*Anthus pratensis*	Anpr	122	20	2
Grey Wagtail	*Motacilla cinerea*	Moci	59	25	2
White Wagtail	*Motacilla alba*	Moal	223	20	1
White-throated Dipper	*Cinclus cinclus*	Cici	36	17	4
Winter Wren	*Troglodytes troglodytes***	Trtr	271	–	–
Hedge Accentor	*Prunella modularis***	Prmo	268	–	–
European Robin	*Erithacus rubecula***	Erru	271	–	–
Common Nightingal	*Luscinia megarhynchos*	Lume	116	15	2
Black Redstart	*Phoenicurus ochruros*	Phoc	137	25	1
Common Redstart	*Phoenicurus phoenicurus*	Phph	73	21	1
Common Stonechat	*Saxicola torquata*	Sato	34	8	2
Common blackbird	*Turdus merula***	Tume	277	–	–
Fieldfare	*Turdus pilaris*	Tupi	80	13	5
Song Thrush	*Turdus philomelos***	Tuph	270	–	–
Mistle Thrush	*Turdus viscivorus*	Tuvi	183	22	1
Common Grasshopper–Warbler	*Locustella naevia*	Lona	95	11	1
Marsh Warbler	*Acrocephalus palustris*	Acpa	91	14	3
Eurasian Reed–warbler	*Acrocephalus scirpaceus**	Acsc	21	–	–
Melodious Warbler	*Hippolais polyglotta*	Hipo	59	7	1
Lesser Whitethroat	*Sylvia curruca*	Sycu	221	18	3
Greater Whitethroat	*Sylvia communis*	Syco	227	28	1
Garden Warbler	*Sylvia borin***	Sybo	274	–	–
Blackcap	*Sylvia atricapilla***	Syat	276	–	–
Wood Warbler	*Phylloscopus sibilatrix*	Phsi	204	57	< 1
Common Chiffchaff	*Phylloscopus collybita***	Phcl	276	–	–
Willow Warbler	*Phylloscopus trochilus*	Phtr	269	–	–
Goldcrest	*Regulus regulus*	Rere	247	36	4
Firecrest	*Regulus ignicapillus*	Reig	214	27	4
Spotted Flycatcher	*Muscicapa striata*	Must	107	10	2
Long–tailed Tit	*Aegithalos caudatus*	Aeca	215	21	3
Willow Tit	*Parus montanus***	Pamo	265	–	–
Marsh Tit	*Parus palustris*	Papa	202	32	1
Crested tit	*Parus cristatus*	Pacr	162	37	5
Coal Tit	*Parus ater*	Paat	154	26	10
Blue Tit	*Parus caeruleus***	Paca	268	–	–
Great Tit	*Parus major***	Pama	277	–	–
Eurasian Nuthatch	*Sitta europaea*	Sieu	199	52	2

(Table 3 continued)

Eurasian Tree–creeper	*Certhia familiaris**	Cefa	17	–	–
Short–toe Tree–creeper	*Certhia brachydactyla*	Cebr	243	38	1
Eurasian Golden–oriole	*Oriolus oriolus**	Oror	22	–	–
Red–backed Shrike	*Lanius collurio*	Laco	55	8	3
Great Grey Shrike	*Lanius excubitor*	Laex	45	22	1
Eurasian Jay	*Garrulus glandarius*	Gagl	244	35	1
Black–billed Magpie	*Pica pica*	Pipi	158	18	2
Eurasian Jackdaw	*Corvus monedula*	Como	40	22	3
Rook	*Corvus frugilegus**	Cofr	13	–	–
Carrion Crow	*Corvus corone***	Coco	258	–	–
Common Starling	*Sturnus vulgaris*	Stvu	187	27	3
House Sparrow	*Passer domesticus*	Pado	146	54	1
Eurasian Tree Sparrow	*Passer montanus*	Pamn	105	17	5
Chaffinch	*Fringilla coelebs***	Frco	274	–	–
European Serin	*Serinus serinus**	Sese	24	–	–
European Greenfinch	*Carduelis chloris*	Cach	178	25	1
European Goldfinch	*Carduelis carduelis*	Caca	69	6	6
Eurasian Linnet	*Carduelis cannabina*	Cacn	222	27	< 1
Red Crossbill	*Loxia curvirostra**	Locu	26	–	–
Eurasian Bullfinch	*Pyrrhula pyrrhula*	Pypy	165	11	1
Hawfinch	*Coccothraustes coccothraustes*	Cocc	220	29	3
Yellowhammer	*Emberiza citrinella***	Emci	263	–	–
Reed Bunting	*Emberiza schoeniclus*	Emsc	47	12	2
Corn Bunting	*Miliaria calandra**	Mica	5	–	–

* species present in < 10% of atlas squares / ** species present in > 10% of atlas squares

Table 4. Environmental variables selected by the manual forward procedure, in order of their inclusion in the model during the 'environmental' partial CCA of the species assemblages. The canonical coefficients and intraset correlations, the additional variance each variable explained at the time of inclusion (i.e. conditional effect, λ-A), the marginal effect of each variable (λ-1), the statistics of the Monte Carlo significance test for the forward procedure (F) and the associated probability (P) are reported for each variable.

Variable	Canonical coefficients				Intraset correlations				λ-A	λ-1	F	P
	Axis1	Axis2	Axis3	Axis4	Axis1	Axis2	Axis3	Axis4				
HFor	0.63	-0.65	-0.60	-0.04	0.75	-0.34	-0.44	0.03	0.16	0.16	31.31	0.001
Urban*GrassTE	0.11	-0.10	-0.43	1.18	-0.67	-0.56	-0.08	0.34	0.12	0.15	24.27	0.001
MFor	0.32	-0.33	0.16	0.09	0.39	-0.21	0.29	0.07	0.05	0.05	10.35	0.001
CFor	0.26	-0.22	0.57	-0.12	0.46	-0.15	0.61	-0.08	0.03	0.07	7.79	0.001
FieldCore	0.08	0.29	-0.24	0.37	0.02	0.70	-0.15	0.38	0.03	0.06	6.85	0.001
RivOpen	-0.02	-0.23	0.43	-0.41	-0.42	-0.18	0.35	-0.16	0.02	0.06	3.61	0.001
Urban^2	0.52	0.66	0.61	0.73	-0.40	-0.38	0.04	0.54	0.01	0.07	3.67	0.001
Urban	-0.95	-1.21	-0.37	-1.14	-0.63	-0.55	-0.05	0.39	0.02	0.14	2.94	0.001
Altmin*GrassTE	-0.03	-0.11	-0.22	-0.98	-0.38	0.01	-0.21	-0.58	0.01	0.05	3.40	0.001
Pond^2	-0.06	-0.11	0.29	0.26	-0.12	-0.12	0.34	0.24	0.01	0.02	3.15	0.006

For the 'environmental' partial CCA of the species assemblages, the Monte Carlo procedures found that the first partial CCA axis (F = 50.465; P < 0.001) and the overall analysis (F = 10.98; P < 0.001) were highly significant, indicating that axes of partial CCA significantly explained species assemblages after accounting for 'pure' spatial variations. For the 'spatial' partial CCA of the species assemblages, the Monte Carlo procedures showed that the first partial CCA axis (F = 3.619; P < 0.001) and the overall analysis (F = 2.468; P < 0.001) were highly significant, indicating that axes of partial CCA significantly explained species assemblages after accounting for 'pure' environmental variations.

3.2. Environmental drivers of bird species assemblages

All subsequent results are relative to 'environmental' partial CCA.

The first two partial CCA axes accounted respectively for 54.6% (eigenvalue = 0.254) and 25.4% (eigenvalue = 0.118) of the extracted variance in the species – environment relationship, and for 16% and 7.5% of the inertia of species data, after having removed the effects due to 'pure' spatial variation of the species assemblages. Therefore, the first two canonical axes explained about 80% of the bird species – non-spatial environment relationships and 23.5% of the non-spatially-structured variation of species assemblages. The third (eigenvalue = 0.03) and fourth axes (eigenvalue = 0.02) taken together explained about 10% of the bird species – non-spatial environment relationships and 3.2% of the non-spatially-structured variation of species assemblages. Eigenvalues calculated for each ordination axis indicated the degree of separation of species along this axis and served as a measure of the importance of the axis (ter Braak and Smilauer 1998). Consequently, the first axis is about twice as important as the second, and so on.

The percentage of the variance explained varied markedly from species to species (Table 3) and ranged from 3% for Stock Pigeon (*Columba oenas*) to about 57% for Wood Warbler (*Phylloscopus sibilatrix*). About half of the species had a percentage of the variance explained higher than 20% while more than half of the variation was explained for three species.

No multicollinearity was detected in the 10 selected environmental variables since Variance Inflation Factors were all less than 20 (see ter Braak and Smilauer 1998). Therefore, both intraset correlations and canonical coefficients might be used to assess the relative contributions of environmental variables to the structuring of assemblages composition, as well as marginal and conditional effects of selected variables, respectively λ-1 and λ-A (Table 4).

Several species – environment biplots were reported (Figure 5 and Figure 6), illustrating species assemblages in the three-dimensional ordination space of the first three canonical axes.

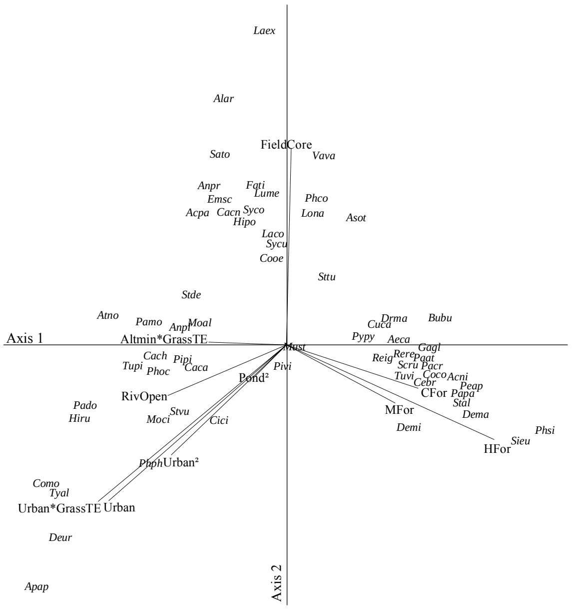


Figure 5. Ordination biplot depicting the first two axes of the 'environmental' partial CCA of the species assemblages. Environmental variables are represented by lines and their acronyms (see Table 2), and species locations by their code (see Table 3).

These ordination diagrams showed well-spread distributions of species and environmental variables, indicating continuous changes in avifauna along ecological gradients. Therefore, the relationships between avian assemblages and environmental variables in terms of the main axes of variation were easy to interpret. The positions of samples on the diagrams were not shown here, but did not demonstrate substantial discontinuities. In this unimodal ordination analysis, the species score, depicted on the biplots, was a weighted average of the sample scores, i.e. the centre of the species bell-shaped distribution along the ordination axis, namely an approximation of the species optimum within the studied range of variation.

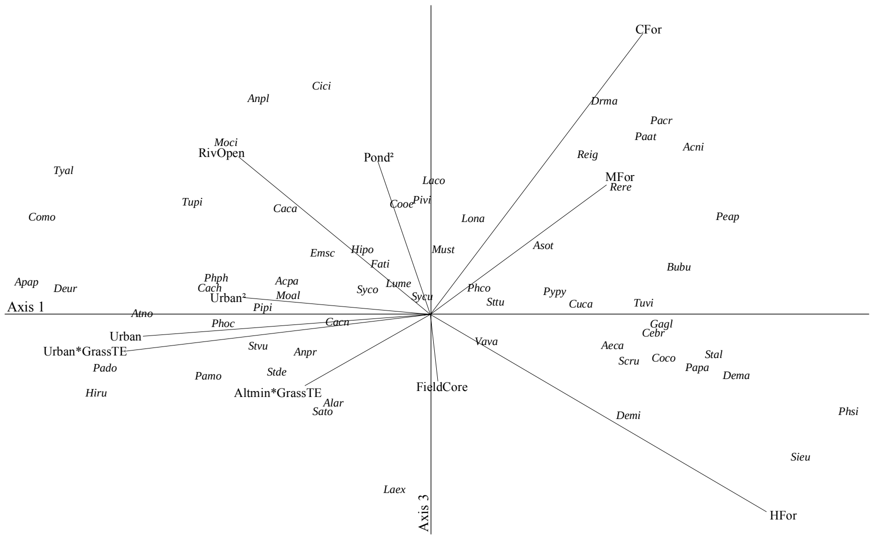


Figure 6. Ordination biplot depicting the first and third axes of the 'environmental' partial CCA of the species assemblages. Environmental variables are represented by lines and their acronyms (see Table 2), and species locations by their code (see Table 3).

The examination of intraset correlations in Table 4 and of biplots on Figure 5 and Figure 6 indicated that the first canonical axis was positively correlated with forested areas and negatively correlated with the amount of urban – grassland areas. Therefore, the first axis separated urban and grassland species, such as House and Eurasian Tree Sparrows (*Passer domesticus* and *P. montanus*), Little Owl (*Athene noctua*) or Eurasian Jackdaw (*Corvus monedula*), from forest species such as Wood Warbler (*Phylloscopus sibilatrix*), Goldcrest (*Regulus regulus*), Great Spotted Woodpecker (*Dendrocopos major*) or Eurasian Nuthatch (*Sitta europaea*). The second axis was negatively correlated with urban and forested areas but positively correlated with the variable *FieldCore*, while the variable *Altmin*GrassTE* obtained an intermediate canonical coefficient. Thus, within species occupying open areas, the second axis separated those which were more associated to urban areas, such as Common Swift (*Apus apus*) or Barn Owl (*Tyto alba*), from those that were found in very open lowlands or fields, such as Great Grey Shrike (*Lanius excubitor*) or Skylark (*Alauda arvensis*), including species of grassland and hedges, such as the European Greenfinch and Goldfinch (*Carduelis chloris* and *C. carduelis*). The third axis notably separated two forest types and species: hardwood forests on one side, with Wood Warbler (*Phylloscopus sibilatrix*) or Eurasian Nuthatch (*Sitta europaea*) for instance, and coniferous and mixed forests on the other side, with Eurasian Sparrowhawk (*Accipiter nisus*), Goldcrest and Firecrest (*Regulus regulus* and *R. ignicapillus*) or Crested and Coal Tits (*Parus cristatus* and *P. ater*). The third axis was also responsible for the separation of aquatic species: as seen on Figure 5 and Figure 6, White-throated Dipper (*Cinclus cinclus*) and Grey Wagtail (*Motacilla cinerea*) were clustered in the three-dimensional ordination space formed by the first three axes, and closely associated to rivers in open areas and ponds.

4. Discussion

4.1. Environmental and spatial effects

The spatial pattern of species assemblages was assessed by means of a cubic trend surface function of the geographical coordinates of atlas squares which ensured the extraction of not only linear gradients in the species assemblages, but also more local structures such as patches or gaps (Borcard *et al.* 1992, Guisan and Zimmermann 2000). About 2.5% of the species assemblage variation (i.e. 7.4% of explained variation) was attributable to these spatial variables without being related to the 10 selected environmental variables (‘spatial’ partial CCA). This fraction varied from species to species, ranging from 0.1% for Wood Warbler (*Phylloscopus sibilatrix*) to 10% for Coal Tit (*Parus ater*) (Table 3). In the 'spatial' CCA (step 2), it can be observed that the percentage of variance explained by species was substantially higher than the percentage following the ‘spatial’ partial CCA (step 4), especially for forest species (for instance, 14.17% for Great Spotted Woodpecker (*Dendrocopos major*) and 17.12% for Eurasian Nuthatch (*Sitta europaea*)). This is due to the spatial arrangement of forests in large patches within the study area. This broad-scale pattern was well captured by the selected terms of geographical coordinates while more local patches were probably not detected (Meot *et al.* 1998) and would have required many additional terms. After factoring out the effect of environment ('spatial' partial CCA), however, space appeared to be a minor descriptor of species assemblage patterns undetected by the 10 environmental variables. The majority (about two thirds) of the spatial distributions of species detected by the geographical coordinates was due to the spatial arrangement of the environment (see below for the interpretation of this fraction). The remaining 'pure' spatial variation reflected some contagious biological processes without relationship to the environmental variables included in the analysis, such as predation or social aggregation for instance (Borcard *et al.* 1992), that could have created spatial autocorrelation in the multivariate dependent data set (bird species matrix). At a broader scale, this fraction may reflect the role of spatial dynamics and/or historical patterns of dispersal in producing contemporary distributional patterns. This effect could be potentially strong and well captured by a polynomial trend surface when larger areas are investigated, i.e. at national or continental extent, including several biogeographic regions and where some species reach their distributional limits (Storch *et al.* 2003). In such cases, geographical coordinates could affect the shape and structure of distributional ranges. The low influence of 'pure' space in our ordinations indicated that bird assemblages are weakly spatially-structured. However, most of the spatial structure in bird assemblages at this landscape level could have been missed by the polynomial trend surface. These results hence suggest that the use of a polynomial function of geographical coordinates of the sampling sites is appropriate for modelling broad-scale variation but may not be adequate for quantifying fine spatial structures at the landscape level. Accordingly, to detect spatial patterns over a wider range of scales within a given landscape, an improvement in the method is required.

Environment explained about 31.5% (fractions (a) and (b)) of the variation in the species assemblages, from which about 4% was shared by the polynomial function of geographical coordinates of the atlas squares. In similar studies (e.g. Borcard *et al.* 1992, Borcard and Legendre 1994, Hobson *et al.* 2000), spatially-structured environmental variation can reach highly variable proportions, sometimes surpassing by far the 'pure' spatial and 'pure' environmental variations. The fairly low fraction observed here suggests that species assemblages and environmental variables did not have a common spatial pattern. Again, a large part of the spatially-structured environmental conditions could have been missed by the polynomial trend surface. The remaining 27.5% of the environmental variation reflected local effects of the environmental variables after partialling out the spatial component. This partition is rarely higher than 40-50% in other studies and is sometimes close to the value obtained here or smaller (e.g. Borcard *et al.* 1992, Borcard and Legendre 1994, Pinelalloul *et al.* 1995, Hobson *et al.* 2000). As shown by the biplots, by the order of variable selection during manual forward procedure, by canonical coefficients and intraset correlations, and by the relative marginal and conditional effects of these variables (Table 4), avian assemblage composition was primarily related to the degree of urbanisation and the amount of forested and open areas. To a lesser extent, the composition of forests (coniferous / mixed forests *versus* hardwood forests) and of open areas (grassland *versus* fields) were also important, as well as aquatic biotopes. Other environmental variables appeared to be less related to bird assemblages structuring. For instance, hedges did not appear during the manual forward selection, maybe because they were not described with sufficient accuracy in terms of floristic composition, vertical structure or spatial arrangement. Likewise, square level variables were less important, probably because they were highly correlated with other composition and configuration metrics and thus, were not included during the forward selection procedure. Biologically, they also correspond to the environmental requirements for a minority of species. Nevertheless, the number of species per square was positively and significantly explained by square level metrics such as *TE*, *Shdi* and *PR* (Titeux 2002), indicating a strong relationship between environmental heterogeneity and species richness.

4.2. Unexplained variation interpretation

The unexplained variation of species assemblages was quite high (about 66%) and could be ascribed to the stochastic space-time fluctuations of the communities, to the fact that some species did not occupy all suitable environmental conditions, to unmeasured environmental variables or to spatially-structuring processes that have been missed by the selected geographical terms. Incorporation of the following best explanatory variables in the analysis did not substantially increase the sum of all canonical eigenvalues in steps 1 or 3. Moreover, a CCA run including ubiquitous species, i.e. detected in more than 90% of atlas squares, and with the same explanatory variables showed that these species are clustered near the origin of the ordination space, indicating a weak species – environment relationships, and strengthening their omission from the species matrix. Furthermore, inclusion of rare species (less than 10% of atlas squares) did not increase the total explained variation.

This fairly low percentage of explained variation is not uncommon in ecological studies because species abundance or occurrence data are often very 'noisy' (ter Braak 1986, ter Braak and Smilauer 1998, Guisan *et al.* 1999). Other studies using partial CCA have obtained similar levels of unexplained variation, ranging from about 40 to 85% (see e.g. Borcard *et al.* 1992, Borcard and Legendre 1994, Pinelalloul *et al.* 1995, Hobson *et al.* 2000).

The unexplained variation could also be due to the species abundance data being recorded at a spatial scale that did probably not match the fundamental ecological processes for many species: some ones could react more strongly to microhabitat characteristics at a finer spatial resolution than 1 km² (MacFaden and Capen 2002), e.g. tits or warblers, whilst others, e.g. the Black Woodpecker (*Dryocopus martius*) or the Great Grey Shrike (*Lanius excubitor*) react to environmental factors at a broader spatial scale. Therefore, the abundance or occurrence of all species within inflexible 1 km squares may not be appropriate for landscape-scale species-environment analyses even if it represented an acceptable compromise from an inventory perspective (Hustings *et al.* 1985, Jacob and Paquay 1992). Considering bird habitat selection as a multi-scale phenomenon would certainly increase the relevance of the results but is technically challenging within a multi-species framework.

Furthermore, at this spatial scale of investigation wide ecological gradients or spatial structures could not be displayed within the study area because it is quite homogeneous in terms of climate, altitudinal extent and land cover, unlike the areas chosen for other studies (Pasinelli *et al.* 2001, Storch *et al.* 2003).

Finally, the fairly low percentage of explained inertia was certainly due to the poorly explained distribution of some species requiring very specific and local environmental conditions, not summarised by the 59 environmental metrics, e.g. the Red-backed Shrike (*Lanius collurio*), the

European Turtle-dove (*Streptopelia turtur*) or the Eurasian Cuckoo (*Cuculus canorus*). Species – environment relationships could not have been identified for several species, which tended to obscure the ordination. This suggests that an effort should be made to compute more relevant or additional environmental variables, but this is difficult to achieve on the basis of existing spatial land cover data (Schmit *et al.* 2006). For instance, the 59 metrics characterise environment in terms of quantity but none of them describe quality of the mapped patches, such as forest succession stages or degree of intensification of crop- or grasslands. An increase in the amount of explained variation of species assemblages would therefore be expected by computing additional metrics if ecologically-sound and accurate spatial land cover data were readily available at this spatial scale.

Nevertheless, it is impossible to quantify the fraction of currently unexplained variation that could potentially be explained by additional or more pertinent environmental variables (Borcard *et al.* 1992), because a fair amount of unexplained variation was probably due to spatiotemporal stochasticity or any deterministic processes. Keeping this in mind, an ordination diagram that explained only a low percentage of species – environment relationships may still be quite informative (ter Braak and Smilauer 1998, Guisan *et al.* 1999). Therefore the canonical ordination technique provided an interesting overview of avian assemblages (see Guisan *et al.* 1999), but requires improvement especially regarding the spatial dissection.

5. Conclusion

This study presents a holistic approach to assess the relative roles of environment and space in structuring bird assemblages and to identify substantial spatial and environmental variables related to bird assemblage variations, using a fine-scale bird atlas data set. Although a traditional canonical ordination without partialling out the spatial effects would have provided quite similar results, the factoring out procedure used here was essential because ecological phenomena arise from processes that are not exclusive and which can overlap considerably (Legendre 1993). This is clearly the case in the context of atlas data where individual samples (squares) are not independent, as sites that are close to each other share both environmental conditions (spatially-structured explanatory variables) and species (due to spatial autocorrelation). Such an approach should be adopted in order to provide a general overview of the multivariate data set, prior to specific analyses. Nevertheless, the polynomial function of geographical coordinates of the sampling sites was most probably not appropriate for detecting spatial patterns of assemblages resulting from fine-scale processes operating at the landscape level. Consequently, the method should be improved to detect hidden spatial variations that are unexplained by present environmental gradients so as to partial out this spatial component of the variation. This way, it would be possible to ensure the observed species – environment relationships are not due to underlying spatially-structured causes.

Acknowledgements

We thank all the members of the COA (Centrale Ornithologique Aves, Belgium) that have participated in the mapping schemes. We are grateful to Dirk Maes and Mark Rounsevell for useful discussions and for commenting on a draft of the manuscript. We thank the Institut Géographique National for providing us with 1:50,000 maps.

Postscript [1]

One of the major conclusion of this article was that the broad-scale pattern of environmental variables and bird assemblages was well captured by the trend surface while more local patches were probably not detected (see also Borcard *et al.* 2004). Significant spatial structure of the bird assemblages

---

[1] This part does not belong to the published paper in Journal of Biogeography

resulting from fine-scale intrinsic processes could therefore have been missed by the polynomial function of geographical coordinates. Hence, we called for an improved method able to decorticate the spatial structure of the data at multiple spatial scales.

Such an improved technique (called ‘Principal Coordinates of Neighbour Matrices’, PCNM) was designed by the team of Pierre Legendre and Daniel Borcard in Montréal (see Borcard and Legendre 2002, Borcard *et al.* 2004, Legendre and Borcard 2006) simultaneously to our study. This method achieves a spectral decomposition of the spatial relationships among the sampling sites (Borcard *et al.* 2004) and enables to detect spatial patterns over a wide array of spatial scales (Borcard and Legendre 2002). PCNM may be integrated into a canonical (CCA or RDA) and variance partitioning framework to model the spatial variation of species assemblages at all spatial scales (Dray *et al.* 2005). Borcard and Legendre (2002) stated that the efficiency of such modelling will be greatly improved in terms of total explained inertia by replacing the polynomial function by the spatial variables issued from PCNM, but notified that one would expect an increase of the spatial fractions (b) and (c), sometimes at the expense of the pure environmental fraction (a) and not necessarily (d).

The PCNM approach has since been used in some studies that tried to dissect the spatial variation of ecological data (see e.g. Avois-Jacquet *et al.* 2005, Brind'Amour *et al.* 2005, Paquet *et al.* 2006), but was unfortunately not used in the present study that was conducted in 2001-2002.

Finally, we report here that Daniel Borcard recently advised us (pers. com.) to use redundancy analysis (RDA) instead of CCA in any variance partitioning, because RDA proved to be more reliable than the initially proposed CCA framework for quantifying the different fractions (see also Peres-Neto *et al.* 2006).

Chapter 2 – A Statistical Method to Discard Misleading Occurrences within a Functional Habitat Modelling Framework

Note – This study has been submitted to Ecography (Titeux, N., Dufrêne, M., Hirzel, A.H. and Defourny, P. Modelling habitat suitability with presence-only data: we need to account for unreliable species occurrences). The text presented here is slightly modified from the submitted paper for layout and terminology harmonization, as well as for references updating.

Foreword [1]

As mentioned in the general introduction of this thesis, a functional approach entails gathering relevant ecological information regarding the specific resources and tuning the spatial resolution to the ecological processes thought to underlie the specific distribution patterns. Accordingly, in face of limited funding and time (see Roberge and Angelstam 2004), such investigation can generally not be simultaneously conducted for many species within a given area and shortcuts have been proposed whereby the attention is focussed on a single (or few) species (e.g. Landres *et al.* 1988, Simberloff 1998, Bani *et al.* 2006).

In front of the pandemic and ongoing collapse of farmland birds throughout Europe, we chose to target this functional approach at farming areas. We decided to further focus on hedgerow network because hedgerows are endangered farming areas features across Europe (e.g. Baudry *et al.* 2000) and have an important role in maintaining biodiversity (see Burel 1996 for general overview and Sparks *et al.* 1996, Hinsley and Bellamy 2000 for birds specifically). We did not formally use the fashionable paradigm of ‘umbrella’ or ‘focal’ species (see e.g. Landres *et al.* 1988,

---

[1] This Foreword does not belong to the submitted paper to Ecography.

Simberloff 1998, Lindenmayer and Fischer 2003, Roberge and Angelstam 2004, Padoa-Schioppa *et al.* 2006) for selecting an optimal model species of farming and hedgerow areas. 'Umbrella' species are those species whose conservation is expected to confer protection to some other co-occurring species (Lambeck 1997, Roberge and Angelstam 2004). The 'focal' species concept is rather similar but identifies several species that are most sensitive to a particular threatened ecological process (Lambeck 1997). Both concepts are obviously non exclusive. While widely used and championed, these concepts are much debated merely because different co-occurring species are inevitably limited by different ecological factors or processes, so that single-species management can conflict with the management of other species (e.g. Simberloff 1998, Lindenmayer and Fischer 2003). This is an inherent property of natural systems wherein species naturally evolved so as to scatter their respective hypervolumes all over the environmental hyperspace. Accordingly, within a limited area, enhancing the geographical projection of a given compartment of the hyperspace would unavoidably impinge upon the geographical extent of the other ones.

In this study, the Red-backed Shrike (*Lanius collurio L.*) was retained as an indicator species of non-intensively managed farming areas (Lefranc and Worfolk 1997) and medium-width hedgerow networks (Padoa-Schioppa *et al.* 2006). The habitat requirements of this species encapsulate the needs of many other species of semi-open areas (Van Nieuwenhuyse *et al.* 1999, Lefranc 2004).

In the subsequent paragraphs we provide a brief overview of the qualitative knowledge about the habitat requirements of the species. Most of the Western European Red-backed Shrike populations breed in a rather wide range of semi-open areas created and maintained by extensive farming activities (e.g. Van Nieuwenhuyse 1998a, Van Nieuwenhuyse *et al.* 1999, Lefranc 2004). Highly attracted by the combination of hay production and livestock rearing and by scattered thorny bushes or hedges (used as nest, larder or perch sites), the Red-backed Shrike avoids large and homogeneous extents of cultures or intensive grasslands (e.g. Lefranc 1993, Lefranc and Worfolk 1997, Fuisz *et al.* 1998, Van Nieuwenhuyse *et al.* 1999, Van Nieuwenhuyse 1999, Lefranc 2004).

In the farming areas, nesting sites of the Red-backed Shrike are mainly composed of thorny isolated bushes or short hedges fragments that are thick from base to top and not overgrazed by the cattle (e.g. Van Nieuwenhuyse and Vandekerkhove 1992, Fuisz *et al.* 1998, Tryjanowski *et al.* 2000, Lefranc 2004).

The species is mainly insectivorous (Coleoptera, Hymenoptera and Orthoptera, see Tryjanowski *et al.* 2003, Karlsson 2004). Intensive grazing pressure as well as nitrogen fertilizer alter the composition and the structure of the herbaceous vegetation, inducing a dramatic decline in invertebrate richness and abundance (e.g. Andrews and Rebane 1994, Morris 2000, Kruess and Tscharntke 2002a, O' Neill *et al.* 2003). Soil moisture (e.g. Couvreur and Godeau 2000, Holland 2002), proximity of forest (e.g. Magura *et al.* 2001), semi-natural patches abundance and their spatial cohesion (e.g. Stoner and Joern 2004, Pywell *et al.* 2005) and field margins length and management (e.g. Meek *et al.* 2002, Holland 2002, Backman and Tiainen 2002) are other factors enhancing arthropod’s diversity, density or biomass in the farming areas.

The majority of preys are caught on the ground but detected from exposed perches (mainly bushes, hedges and fences – 1 to 4 meters high), most often located at 20-40 meters from the nest (e.g. Van Nieuwenhuyse 1998a, Lefranc 2004, Karlsson 2004). Catches are mainly executed within the first 10-20 meters from the perches (‘usable area’, see Van Nieuwenhuyse 1998a) on low or gappy sward (mainly pastures, mowed hay meadows or xeric grasslands), bare soils (e.g. agricultural lanes or small country roads) or along interfaces between contrasted herbaceous vegetation heights (Van Nieuwenhuyse and Vandekerkhove 1992, Van Nieuwenhuyse *et al.* 1999, Lefranc 2004).

Predation on eggs or nestlings is mainly performed by corvids (Söderström *et al.* 1998, Horvath *et al.* 1998, Roos and Part 2004). Forests edges, cultivated and urban areas are frequently inhabited or foraged by different predatory species, like Eurasian Jay (*Garrulus glandarius*), Carrion Crow (*Corvus corone*), Eurasian Jackdaw (*Corvus monedula*) or Black-billed Magpie (*Pica pica*) (Söderström *et al.* 1998, Söderström 2001).

Unlike females, males frequently keep their territories from one year to another (Van Nieuwenhuyse *et al.* 1999, Van Nieuwenhuyse 2000b, Simek 2001, Lefranc 2004). This site fidelity strongly depends on the success of the reproduction in the preceding years (Lefranc and Worfolk 1997, Van Nieuwenhuyse *et al.* 1999, Van Nieuwenhuyse 2000b). Individuals do not show any marked philopatry in the choice of their first breeding territory (e.g. Lefranc 2004). The mean territory area of the Red-backed Shrike is about 1.5 ha (most often between 0.5-1 and 3 ha), but varies according to environmental conditions and population density (Lefranc 1993, Söderström 2001, Lefranc 2004, Karlsson 2004). Males usually sing and call in groups, leading breeding pairs to be aggregated in the landscape (e.g. Van Nieuwenhuyse *et al.* 1999, Van Nieuwenhuyse 2000b, Lefranc 2004). Several authors suggested that this clustering behaviour increases the chances for males to find a mate or for males and females to achieve extra-pair copulations (e.g. Jakober and Stauber 1994, Van Nieuwenhuyse 2000b, Lefranc 2004).

The Red-backed Shrike has shown a marked decline in Western and Northern Europe between 1970 and 1990, and then more slightly in several Western countries during 1990-2000 (Yosef 1994, Lefranc and Worfolk 1997, Lefranc 2004, BirdLife International 2004). Although causal factors remain largely unclear (Yosef 1994), it was suggested that reduction in several types of resources are among the main reasons (Lefranc and Worfolk

1997). Human development, agricultural intensification and cessation of agricultural activities in some areas are most probably among the main causes of these changes (Lefranc and Worfolk 1997). Therefore, the conservation of the species strongly depends on directions that will be taken in the European agricultural policy (Van Nieuwenhuyse 1999).

While the species clearly requires the persistence of extensive farming techniques, its habitat selection flexibility makes the detection of its main requirements challenging. Yet, it is crucial to gather accurate and fine-grained information about how the various habitat components drive its spatial distribution and breeding performances, if we are to implement efficient management strategies. The Natura 2000 network currently implemented throughout European countries aims to preserve or restore – into a favourable conservation status – the species populations and biotopes that are threatened on the European scale (European Directives 79/409/CEE and 92/43/CEE). The Red-backed Shrike is in the Annex 1 of the Birds Directive 79/409/CEE and for this reason conservation professionals urgently need more precise information about its habitat requirements in farming areas.

Abstract

Niche-based models of species spatial distribution are of paramount importance in ecology and conservation because they enhance our understanding of individuals requirements and allow the identification of potentially suitable areas. Most methods rely on the environmental patterns of species presence and absence to produce Habitat Suitability (HS) maps. However, the frequent lack of valid and reliable absences has led to the development of ‘presence-only’ techniques. These methods assume that species presence is a reliable indicator of habitat suitability. Nevertheless, for historical, behavioural, dynamical or social reasons, some individuals are often found in unsuitable conditions. Such ‘outsiders’ can have a huge and spurious effect on habitat suitability models. We propose here a simple approach for identifying and discarding such unreliable occurrences. First, we build an Ecological Niche Factor Analysis (ENFA) model. Second, we use this model to delineate unsuitable areas, which include a chosen proportion of species presences (outsiders). Third we draw pseudo-absences from this area and use them to compute a more accurate Generalized Linear Model (GLM). We illustrate our method with data on the Red-backed Shrike (*Lanius collurio*) in Southern Belgium. We find that discarding unreliable presences considerably improves the quality of the ensuing models and HS maps. Furthermore, we give general advice for estimating the optimal proportion of presences to discard before selecting pseudo-absences and calibrating the presence/absence predictive model. Unreliable species occurrences have often been neglected. However, both ecologists and managers should account for them to pertinently highlight species habitat requirements and to compute credible HS Maps. The simple screening procedure proposed here ensures a relevant delineation of suitable areas for species whose occurrences do not all reflect suitable conditions.

1. Introduction

Mapping species spatial distribution is an important issue in conservation and population management (e.g. Guisan *et al.* 2002, Rushton *et al.* 2004, Guisan and Thuiller 2005). Indeed, efficiently delineating the habitats of a threatened species allows species or resource management to focus on important areas (Johnson *et al.* 2004). Habitat modelling techniques, linked to Geographic Information Systems (GIS), can produce maps of habitat suitability (Guisan and Zimmermann 2000). Quantitative models predicting the spatial distribution of a species are highly diverse (see Guisan and Zimmermann 2000, Segurado and Araujo 2004, Guisan and Thuiller 2005, Pearce and Boyce 2005 for reviews). Most rely on the relationship between the species and its environment (Heglund 2002, Guisan and Thuiller 2005) to generate ‘resource selection functions’ (RSF), which allow the prediction of potentially suitable areas for the species, given the environmental conditions (e.g. Boyce and McDonald 1999, Guisan and Zimmermann 2000, Zaniewski *et al.* 2002, Johnson *et al.* 2004). These functions highlight the habitat selection pattern of the focal species, hence enhancing our understanding of the main ecological drivers of species distributions. Their application to areas where environmental conditions are known but where species distribution is unknown provides Habitat Suitability (HS) maps (e.g. Brotons *et al.* 2004, Rushton *et al.* 2004).

The vast majority of these models (including Generalized Linear Models, Generalized Additive Models, Artificial Neural Networks, Discriminant and Classification Analyses, etc…) relies on data about presence and absence of the species to generate the statistical RSF, assuming that they respectively indicate suitable and unsuitable environmental conditions (Wiens 2002). Among others, Generalized Linear Models (GLM) have proved to be robust in many situations and are widely used (e.g. Manel *et al.* 1999a, Pearce and Ferrier 2000, Guisan *et al.* 2002, Rushton *et al.* 2004). Nevertheless, presence/absence surveys are often characterized by false absences (i.e. failure to detect a species when in fact it is present) when detection probability is less than 1 for inconspicuous species (Tyre *et al.* 2003, Wintle *et al.* 2004). Methods for incorporating estimated detection probabilities in the analysis of site occupancy data from presence/absence surveys usually require detection/non-detection data acquired through repeated visits (see Wintle *et al.* 2004, MacKenzie *et al.* 2005 for reviews), which is often unachievable due to diverse constraints. Moreover, large biological databases are often collected for inventory purpose and thus characterized by a relatively scarce set of occurrences gathered without ad-hoc sampling design and without reliable information on species absence, especially for uncommon or rare species (Hirzel *et al.* 2002a, Anderson 2003, Ottaviani *et al.* 2004, Engler *et al.* 2004). Non exhaustive presence-only data are therefore frequently the only available ones (Zaniewski *et al.* 2002, Pearce and Boyce 2005, Elith *et al.* 2006). Finally, the assumption that absences indicate unsuitable environmental conditions may be violated due to a variety of factors like population dynamics, fragmentation effects, interspecific interactions, dispersal or history (e.g. Araujo and Williams 2000, Pulliam 2000). All these reasons result in severe limitations and biases for fitting ‘discrimination models’ that distinguish between presence and absence (e.g. Tyre *et al.* 2003, Gu and Swihart 2004, Engler *et al.* 2004, Brotons *et al.* 2004).

Alternative methods adapted to presence-only data are particularly suitable when absences are fallacious or meaningless (see Pearce and Boyce 2005, Elith *et al.* 2006 for reviews). Among others, ‘profile methods’ are based on Hutchinson’s concept of the ecological niche (Hutchinson 1957) and rely on the delineation of environmental envelopes around recorded species occurrences in the multidimensional space (hereafter called hyperspace) of environmental conditions (e.g. Walker and Cocks 1991, Busby 1991, Carpenter *et al.* 1993, Robertson *et al.* 2001, Hirzel and Arlettaz 2003a, Robertson *et al.* 2004). A recent example is the Ecological Niche Factor Analysis (ENFA), a factor analysis comparing, in a environmental hyperspace, the distribution of the sites where the species is present to a reference set (called ‘global set’) describing the whole study area, and giving a Habitat Suitability (HS) value to any site (Hirzel *et al.* 2002a). ENFA considers the density of occurrences within the hyperspace (see below for details) and is therefore an improvement on other profile techniques (Pearce and Boyce 2005).

However, ENFA often tends to provide over-optimistic HS maps and less accurate predictions than GLM (Olivier and Wotherspoon 2005, 2006, Pearce and Boyce 2005). This is because there is no discriminating absence helping to ‘fix the floor’ of what is really unsuitable environmental conditions, and thus to limit the extent of the envelopes (Hirzel *et al.* 2001, Zaniewski *et al.* 2002, Engler *et al.* 2004, Brotons *et al.* 2004). Overestimating areas of suitable habitats may obviously have highly undesirable consequences when these HS maps are used for conservation planning (Loiselle *et al.* 2003). The discrepancy between ENFA and GLM predictions depends on the reliability of absence data and the species prevalence or tolerance (Hirzel *et al.* 2001, Brotons *et al.* 2004).

When absence data are unavailable or unreliable, the ‘discrimination approach’ can still be used by conditioning it to the results of a preliminary ‘profile approach’. Recently, Engler *et al.* (2004) proposed a modelling strategy that combines ENFA and GLM, using ENFA to weight the random generation of ‘pseudo-absences’ used subsequently in GLM and deemed as ‘true’ absences (see also Lobo *et al.* 2006 for a similar approach).

This elegant method reduces the ‘false’ absences issue and achieves promising results on species for which each occurrence is supposed to be correlated to truly suitable environmental conditions. Nevertheless, in many specific cases, some occurrences (hereafter called ‘outsiders’) may not convey reliable information about habitat suitability for various reasons like historical, behavioural, population or social processes (see Pulliam 2000). Such occurrences in unsuitable conditions are thus unreliable from a purely habitat viewpoint and could have a huge and spurious effect on the subsequent modelling output. This key aspect has been too often overlooked although accounting for its existence may be necessary to pertinently describe the habitat requirements for the species and to elaborate relevant HS maps (see Wiens 2002, Guisan and Thuiller 2005).

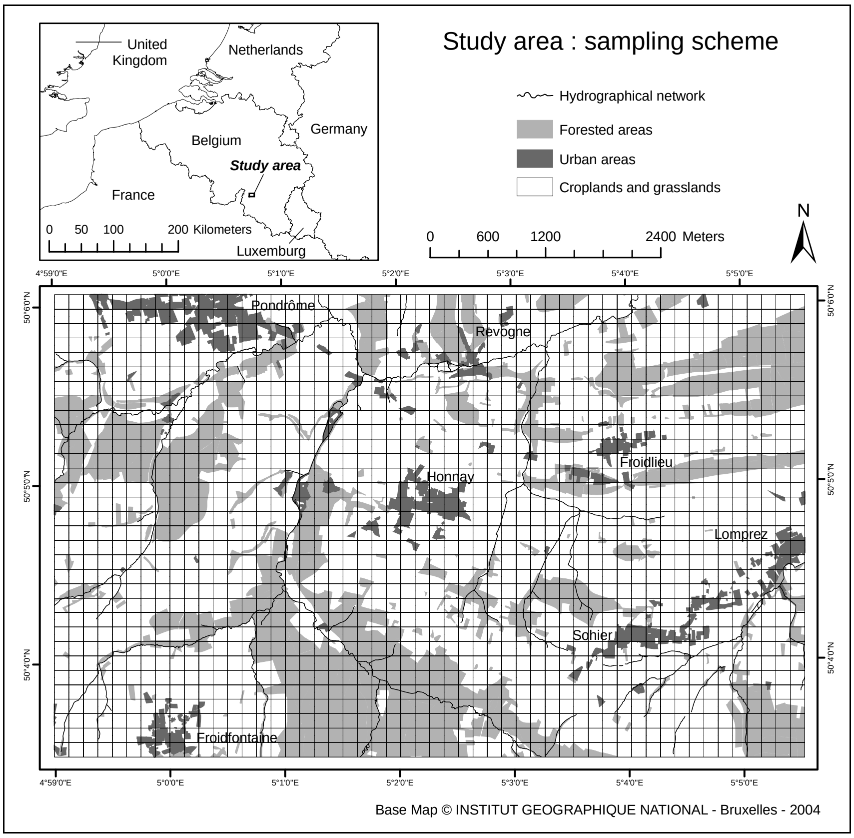
In this paper, we focus on the potentially drawbacks of not explicitly accounting for putative outsiders during the modelling process, questioning the usefulness of the models and the ecological reliability of the ensuing HS maps. We then propose an easy-to-implement approach within the most advanced ENFA-GLM modelling framework to identify and to discard the less representative and informative occurrences from a habitat perspective. We subsequently examine and discuss the effect of increasing proportions of discarded putative outsiders on the quality of the models and we propose some recommendations for choosing the most adequate proportion on statistical grounds. Throughout this paper, we emphasize the necessity of considering the existence of unreliable occurrences rather than ignoring them, to ensure ecologists and managers of delineating relevant suitable areas for species whose occurrences do not all reflect suitable conditions.

We used the Red-backed Shrike (*Lanius collurio L.*), an umbrella and flagship bird species that has been declining in Western and Northern Europe during the last 30-40 years (Yosef 1994, Lefranc 2004, BirdLife International 2004), to illustrate our method.

2. Materials and methods

2.1. Study area

The study area (Figure 7) covers about 38 km² and is situated in Calestienne, Belgium. Calestienne is punctuated by numerous calcareous rocks forming a narrow west-east string of hills (altitude range: 170-300 meters) interspersed with schistose areas (Thill 1964). Hills are typically forested and slopes are pastured or cultivated depending on the soil. Depressions are characterized by poorly drained clay soils derived from schist and hence unfavourable to cultivation. Instead, they are devoted to cattle breeding, thus covered by extensive to high-intensive mesic grasslands (pastures, hay meadows or aftermath grazed meadows) and often scattered with bushes and hedgerow networks. Finally, human settlements (mostly small villages) are scattered along the hydrographical network or inserted between woodlands (Thill 1964). Calestienne is characterized by a by a geo-morphological and bio-climatic constancy.



Base Map © INSTITUT GEOGRAPHIQUE NATIONAL - Bruxelles - 2004

Figure 7. Localisation of the study area, showing the 1664 150m-by-150m cells.

2.2. The species

The reader is referred to the Foreword of this Chapter for general informations about the biology of the Red-backed Shrike.

Populations may fluctuate in abundance from one year to another in response to a variety of factors independently of local environmental conditions. These factors may thus hold the breeding densities below carrying capacity. The pattern of settlement of a population within a given area can change considerably between consecutive years, with suitable sites not occupied each year (Van Nieuwenhuyse 2000b, pers. obs.). Absences are thus unreliable from a habitat standpoint.

2.3. Species data

The study area was sampled according to a regular 150m-by-150m resolution grid formed by 1664 (32x52) square cells (reference cells). Cell surface (2.25 ha) is close to the mean territory area of the Red-backed Shrike (see Foreword of this Chapter). This sample unit size is a good trade-off between the desired spatial accuracy of the predictions, the scale at which key processes driving the local distribution pattern operate (resource selection) and the scale at which environmental data were collected[1]. This matching of scales is of primary importance for understanding and quantifying the relationship between pattern and processes accurately (Huston 2002, Van Horne 2002, Wiens 2002).

All territories of Red-backed Shrikes in the study area were surveyed during a comprehensive field survey in May-July 2004. All open areas were visited at least three times during the breeding season. Each site was declared unoccupied if species remained undetected after the third visit. Territory boundaries were delineated by connecting the outermost observations of territorial disputes or foraging during at least 20 min with sunshine and no wind (15-20 per male). Territories were subsequently aggregated at the reference cell level (150m-by-150m grid). A value of 0 or 1 was assigned to each cell for absence or presence of the species respectively. Depending on their size and geographical location, territories could either straddle up to three cells or be completely included within a single cell. In the first case, the value of 1 was assigned to cells when more

---

[1] We did not intend to delineate the precise boundaries of habitats across the study area but rather to quantify the relative suitability of environmental conditions within the designed sample units (reference cells). In the Red-backed Shrike, all resources are required within the bounds of the territory (about 2 ha) and are connected to each other by individual movements (e.g. Van Nieuwenhuyse and Vandekerkhove 1992). Therefore, the size of the reference cells appeared to be the most appropriate.

than half of their surface was intersected by a territory. In the second case, a value of 1 was assigned to the cell whatever the territory size.

2.4. Raw environmental data

The GIS Software ArcGIS 8.3 (ESRI, Redlands, CA, USA) was used to combine several environmental data layers from different sources. A 30m resolution digital elevation model (Institut Géographique National, 1995) was used to represent the topography of the region. A digital 1:20,000 vector soil map (I.R.S.I.A. 1966) allowed to derive pedologic descriptors. A digital 1:10,000 vector topographic map (Institut Géographique National, 2004) was used as a planimetric reference and basic environmental description. No point feature was retained from this map. Selected linear elements are railways, roads and hydrographical network. Roads were subdivided into (1) suitable for hunting by shrikes (secondary roads and pathways) and (2) unsuitable for hunting (main roads). Land use information was used as basic surface information.

In addition, a set of elements that could potentially have a role in explaining the species distribution was collected. In particular, several types of punctual, linear and surface features were surveyed on the field during May-August 2004 and digitized from 40 cm resolution aerial colour orthophotographs printed at 1:2,500 (Walloon Region, 1998). Field map digitizing also allowed to 'update' the landscape description to the species census period (2004).

Punctual and linear elements reported from the field were trees and bushes on the one hand, and tree lines, hedges and fences on the other hand. Bushes and hedges were allocated to (1) thorny and (2) others. Height of each point element, tree line and hedge segment was recorded.

Each parcel was digitized as unique surface feature and classified into one of 31 vegetation types. EUNIS typology (Davies and Moss 2002) was followed for describing each feature but some types were subdivided according to biological relevance in this specific context. All surface types are listed in Table 5. The intensification level of grazed grasslands was categorized in three classes according to the physiognomy and the floristic diversity of the vegetation. The vast majority of hay meadows in Famenne and Calestienne are in fact aftermath grazed meadows, but were classified as such only if they were grazed during the reproductive season of the Red-backed Shrike.

2.5. Environmental descriptors computation

Fitting a predictive distribution model relies on the existence of explanatory variables that define suitable environmental conditions for a species (Guisan and Zimmermann 2000, Austin 2002).

Several environmental descriptors (see Table 6) were computed and stored in GIS layers (ArcGIS 8.3, ESRI, Redlands, CA, USA). They were designed on the basis of the available ecological knowledge in order to reflect more or less directly the ecological resources that determine the functional interactions between the organisms and their environment among the variety of occupied conditions (see Foreword of this Chapter).

Most of the descriptors were calculated within the 150m reference cells (see Species Data). Some descriptors needed more spatial details and were computed using a 75m resolution grid. By contrast, to account for a wider neighbourhood, some descriptors were computed within 300m cells. However, the adopted modelling approach needed the same resolution for all descriptors. Each 150m reference cell was assigned the maximum value onto those computed within the four 75m cells it contained. For those computed within 300m cells, we used overlapping moving windows centred on the 150m reference cells. Finally, environmental conditions for each reference cell were described at three different spatial scales.

Table 5. List of land use types adapted from EUNIS typology.

Type	Land use
Aquatic biotopes	Permanent oligotrophic ponds
	Hydrographical network
Shrub-covered biotopes	Scrubs
	Densely shrub-covered area
	Sparsely shrub-covered area
Cultivated biotopes	Arable land (1)
	Recently abandoned arable land * (1.5)
Forested biotopes	Broadleaved deciduous woodland
	Mixed woodlands
	Coniferous woodland
	Low-stem tree orchards
Grazed/Mowed biotopes	Permanent extensive mesotrophic pastures * (2) (patches of ungrazed and diversified vegetation)
	Permanent intensive mesotrophic pastures * (1) (short but quite diversified and heterogeneous vegetation)
	Permanent very intensive mesotrophic pastures * (0.5) (overgrazed and homogeneous vegetation)
	Hay meadows * (3)
	Aftermath grazed meadows * (2.5)
Unmanaged biotopes	Wet grasslands *
	Dry grasslands *
	Grassy fallow *
	Wet fallow *
	Dry heath *
	Dry heath with scattered trees
	Dry heath with numerous bushes
	Wet heaths *
	Wet heaths with numerous bushes
Anthropogenic biotopes	Rail networks
	Road networks (* if secondary roads)
	Agricultural constructions
	Buildings of cities, towns and villages
	Domestic gardens of villages and urban peripheries
	Artificial and sterile biotopes

* Land use type suitable for hunting.
Unmanaged biotopes are patches (temporarily) not used for cultivation or rearing, classified according to herbaceous composition-structure and ligneous density and composition. Weights between brackets reflect relative prey density in several types of pastured or cultivated patches (see Kruess and Tscharntke 2002a, 2002b) and were used for environmental descriptors computation (see Table 6 and Table 7).

Table 6. List of environmental descriptors calculated for each cell, with their spatial scale of computation, their functional significance and a brief description.

Name	Resolution (m)			Function	Description
	75	150	300		
Nest		x		Nest	Abundance of bushes and/or hedges suitable for nest installation (1-6m high / thorny / not completely inside a very intensive pasture)
NestDist		x		Nest	= Nest where each bush and/or hedge is positively weighted by its distance to urban area
NestStd		x		Nest	Standard deviation of the heights of bushes and/or hedges suitable for nest installation
Arable		x	x	Prey density	Area of arable lands
Hay		x		Prey density	Area of hay meadows
Aftermath		x		Prey density	Area of aftermath grazed meadows
ExtPast		x		Prey density	Area of extensive pastures
Past		x	x	Prey density	Area of intensive pastures
IntPast		x	x	Prey density	Area of very intensive pastures
SemiNat			x	Prey density	Area of (temporarily) unmanaged biotopes and/or hay meadows (considered as 'semi-natural' biotopes)
SemiNatW		x	x	Prey density	= SemiNat, where area of each patch is weighted by the area of 'semi-natural' biotope patches within a radius of 250m
Extensive		x	x	Prey density	Area of pastured and cultivated biotopes where each type is weighted by a coefficient representing the relative Prey density (see Table 5)
RatioPC			x	Prey density	Ratio of pastures to arable lands areas
SoilDepth			x	Prey density	Mean phreatic table depth

(Table 6 continued)

Name	Resolution (m)			Function	Description
	75	150	300		
SoilDry				Prey density	Euclidean distance between the centre of the cell and the nearest point where phreatic table depth is less than 15 cm
DiForest				Prey density – Predation	Euclidean distance between the centre of the cell and the nearest forest
DiUrban				Predation – Disturbance	Euclidean distance between the centre of the cell and the nearest urban area (building)
DiNaRo				Disturbance	Euclidean distance between the centre of the cell and the nearest major road
Hunt	x	x		Prey detectability	‘Usable area’ for hunting defined by a 20m-buffer around each suitable perch for hunting (height 1-4m / bush, hedge or fence / distance to suitable nest site < 40 m) and intersecting suitable patches for hunting
HuntVg	x	x		Prey detectability	= Hunt, where the area of each intersected patch is negatively weighted by the mean vegetation height of this patch type
DenPerch	x	x		Prey detectability	Density of adequate perches for hunting (ratio number of perches / surface of open areas)
DenLign		x		Prey detectability	Density of suitable ligneous perches for hunting (ratio number of suitable ligneous perches / surface of open areas)
Boc		x	x	Prey detectability	Geometric mean of DenHedge and the mean of all ligneous perches height
DenHedge		x	x	Prey detectability	Density of hedges (ratio length of linear hedges / surface of open areas)
NbHunt		x		Prey density – detectability	Number of patches intersecting the ‘usable area’ for hunting
Interface		x	x	Prey density – detectability	Length of all interfaces between the different open patches
Contrast		x	x	Prey density – detectability	= Interface, where each interface type is positively weighted by the difference of vegetation height between both patch types
SuOpen			x	Prey density – detectability	Surface of open areas

(Table 6 continued)

Name	Resolution (m)			Function	Description
	75	150	300		
NbOpen			x	Prey density – detectability	Number of open area patches
NbType			x	Prey density – detectability	Number of open area patch types
MaxSlo		x		Topography	Maximal slope
AveSlo		x		Topography	Average slope

It is worthwhile to indicate that forests and urban areas outside the study area were taken into account for distances computations (external buffer of 1 km).

2.6. Statistical methodology

2.6.1. Exploratory analyses

First an *a priori* screening of the 1664 cells was realised to eliminate those where land-use is known to be incompatible with the species occurrence (interior of woodland or cells without any bush or hedge required for nest installation). Final model predictions were set to zero for these cells (n=969). The remaining cells (n=695) were used for the subsequent steps of the modelling process.

All descriptors were then normalized using napierian logarithms $x' = \ln(x+1)$ and standardized (mean=0 / variance=1) in order to be brought to some common scale and to allow independent parameters estimation. Departure from normality after transformation was detected for some descriptors but ENFA is considered not too sensitive to such violation of the assumptions (Hirzel *et al.* 2002a).

2.6.2. Ecological Niche Factor Analysis (ENFA)

ENFA (Hirzel *et al.* 2002a) is a multivariate approach which does not require absence data. This technique is like a PCA of sites, i.e. the cells, where principal components are ecologically meaningful. Two distributions are used for computing the components. The global distribution is the multidimensional distribution of all cells of the study area along the *p* descriptors. The species distribution[1] is that defined by the occupied cells only.

Using these two distributions, the first component, called 'marginality factor', maximizes the difference between the species mean and the global mean, describing how far the species optimum is from the mean environmental conditions in the study area. Next factors, called 'specialization factors', are then extracted orthogonally in order to maximise

---

[1] ‘Species distribution’ has here nothing to do with the spatial distribution of individuals.

the ratio between the remaining variance in the global distribution and the remaining variance in the species distribution.

The few first factors of the ENFA, gathering the majority of the information, are then used to compute Habitat Suitability (ENFA-HS) values between 0 and 1 in the ENFA hyperspace. Several algorithms are available for HS computation (Hirzel *et al.* 2002b) and the distance geometric-mean algorithm was used. This algorithm was shown to be a good trade-off between two contradictory constraints, i.e. precision and generality (Hirzel and Arlettaz 2003b). While making no assumption about the shape of the species distribution along the different factors, this algorithm takes into account the occurrence density in the ENFA hyperspace to increase the influence of occurrences that are close to each other (see Hirzel and Arlettaz 2003b). The distance geometric-mean algorithm relies thus on the assumption that the higher the density of occurrences in the hyperspace, the higher the suitability of the corresponding environmental conditions (Hirzel *et al.* 2002b, Hirzel and Arlettaz 2003b). For each point P in this hyperspace, the density is modelled by the geometric mean H_G of its distances δ to the N occurrences O_i (Hirzel and Arlettaz 2003b), creating a multidimensional H_G field (see equation 1).

$$H_G(P) = \sqrt[N]{\prod_{i=1}^{N} \delta(P, O_i)} \tag{1}$$

High values of H_G reflect low occurrence density in the hyperspace. Envelopes can then be delineated by circumscribing all points in the hyperspace that have a value lower than a certain threshold. Several envelopes can thus be defined enclosing different proportions of occurrences, from the innermost to the outermost part of the occurrence cloud. Finally, a Habitat Suitability (ENFA-HS) value is associated to each envelope by counting the proportion of occurrences they encompass (Hirzel and Arlettaz 2003a). This ENFA-HS field is simply inversely proportional to the H_G field.

ENFA was performed using all occurrences and ENFA-HS values were computed for all cells with the standalone software BIOMAPPER 3.1. (Hirzel *et al.* 2002b).

2.6.3. Generalized Linear Models (GLM)

GLM are a class of statistical models that is a generalization of classical linear models, allowing non-normal response variables to be modelled (McCullagh and Nelder 1989).

Their application for establishing Habitat Suitability maps or for modelling species richness distribution is widely recognised: they are frequently used for binary (e.g. Zimmermann and Kienast 1999, Manel *et al.* 1999b, Guisan *et al.* 2002), ordinal (Guisan and Harrell 2000, Guisan 2002) or continuous (e.g. Lobo *et al.* 2002, Maes *et al.* 2003) response variables. In our case of presence-absence of the species (see below for the problem of absences), a binomial distribution and a logit link-function were specified (logistic regression model, Hosmer and Lemeshow 2000), allowing to describe the relationship between the response variable and the environmental descriptors[1]. Applying the final model to each cell provides the conditional mean of the response, given the values of the descriptors, in other words, a probability of occurrence, ranging from 0 to 1. These values can be used to generate HS maps.

---

[1] Habitat models should explicitly incorporate the spatial structure of the data (Lichstein *et al.* 2002 and see Chapter 5) because spatial autocorrelation violates the assumption of independence among observations (Keitt *et al.* 2002). This could result in incorrect inference, resulting in overestimating the effects of explanatory variables in predictive models (Legendre 1993, Augustin *et al.* 1996, Klute *et al.* 2002). However, we did not intend here to compute the 'true' HS models, but simply to rank the influence of differential outsider elimination rates on a relative scale. Non-spatial GLM were therefore used as a common denominator for all models.

2.6.4. Pseudo-absences generation: current practice and improved methodology

When absence data are unavailable or unreliable, one can build ‘discrimination models’ by generating 'pseudo-absences' (Zaniewski *et al.* 2002, Engler *et al.* 2004, Pearce and Boyce 2005, Lobo *et al.* 2006, Olivier and Wotherspoon 2006). The methodology used to select 'pseudo-absences' is important and can affect the quality of the final model because of biased absence data (Boyce *et al.* 2002, Olivier and Wotherspoon 2005).

Engler *et al.* (2004) proposed a strategy of ENFA-weighted pseudo-absences generation that reduces the chance of selecting absences where the species does in fact occur (but not detected) or where the environmental conditions are suitable even if the species is absent. According to their methodology, an envelope enclosing all occurrences in the ENFA hyperspace is delineated. Pseudo-absences are then randomly chosen outside this envelope and combined with real occurrences for GLM calibration. The number of randomly-chosen pseudo-absences is the same as the total number of real occurrences because it is easier to find optimal threshold in this situation (Liu *et al.* 2005).

This method reduces the problem of ‘false’ absences and achieves good results on species for which each occurrence record reflects suitable environmental conditions. As developed in the introduction, this procedure is expected to be less adapted when some occurrences do not reflect suitable environmental conditions, these occurrences stretching spuriously the envelope delineated for selecting the pseudo-absences (see Figure 1 and Pulliam 2000).

To highlight this effect, envelopes were delineated at each 5th percentile of ENFA-HS value along the ENFA-HS field, from marginal (ENFA-HS value 0.01) to core area (ENFA-HS value 0.75) in the hyperspace created by the ENFA. These concentric envelopes encompassed progressively fewer occurrences towards the edge of the occurrence ‘cloud’, but always included occurrences at the innermost part of the cloud (highest ENFA-HS values). According to the distance geometric-mean algorithm and isopleths scaling (BIOMAPPER user's manual, Hirzel *et al.* 2002b), these envelopes excluded from 0 to 75% of putative outsiders (lowest ENFA-HS values). Excluded occurrences were deemed uninformative about species habitat and were therefore discarded from the occurrence dataset[1].

Pseudo-absences were then randomly chosen outside or on the delineated envelopes, i.e. among cells with ENFA-HS values lower than or equal to ENFA-HS values used for delineating the envelopes. They were combined with the 'informative' presences (as compared with putative outsiders that were discarded) for GLM calibration. In order to give the same importance to presences and pseudo-absences in GLM (see Liu *et al.* 2005), the number of randomly-chosen pseudo-absences was, when possible, the same as the number of 'informative' presences kept in the envelopes (see Results for why this was not always possible). As explained by Manly *et al.* (2002) and Pearce and Boyce (2005), when using presence-only and randomly selected pseudo-absence data, the prevalence of the species is not accounted for and it is therefore impossible to calculate the absolute probability of occurrence for any given site. It is, however, possible to predict the relative suitability or likelihood of occurrence.

---

[1] Note that the envelope enclosing all occurrences (0% of discarded putative outsiders – HS value = 0.01) is similar to the envelope used by Engler *et al.* (2004).

2.6.5. Model calibration

Models were calibrated for each pseudo-absences generation resulting from each envelope's delineation. For each GLM model, we used a stepwise procedure to select the combination of descriptors that best explained the dataset during the GLM fitting[1]. This procedure consists of a series of alternating step-up and step-down phases (SAS Institute 1990). The former adds variables to the model in their order of significance level, while the latter removes variables from the model. A significance level of 0.05 was specified to allow a variable to enter and to stay in the model. Parameter estimates of the model were then used to predict a final HS value (hereafter called GLM-HS value), ranging from 0 to 1, for each of the 695 cells not masked out a priori. For the 969 discarded cells, the model GLM-HS prediction was simply set to 0. The Receiver Operating Characteristic (ROC) curve was used to select a threshold of GLM-HS above which the cell was deemed suitable. The same cost was assigned to false-negative and false-positive classifications (because we considered pseudo-absences as 'true' absences for GLM calibration), so that the optimum probability threshold was found by reading the point on the curve at which the sum of sensitivity and specificity was maximized (Zweig and Campbell 1993, Fielding and Bell 1997, Manel *et al.* 2001, Liu *et al.* 2005).

Because chance could play a part in the ENFA-weighted selection of pseudo-absences (Engler *et al.* 2004), each modelling procedure described above was repeated 100 times. During each run, models were calibrated using stepwise procedure ensuring that the best set of descriptors was selected for each data set, since they included different pseudo-absences sets.

---

[1] Sequential null-hypothesis tests like stepwise procedures suffer from multiple criticisms. They are currently viewed as heretical by numerous wildlife biologists, notably because they are naive and create spurious effects (e.g. Burnham and Anderson 2002, Johnson and Omland 2004, Stephens *et al.* 2005, Welch and MacMahon 2005 and see Chapter 5). However, we did not intend here to compute the 'true' HS models. Stepwise procedures were therefore used as a common denominator for all GLM.

2.6.6. Model evaluation

The lack of absence data makes the assessment of the predictive power of the models difficult with classical methods related to confusion matrix, including Kappa coefficient (Pearce and Ferrier 2000, Boyce *et al.* 2002).

An Absolute Validation Index (AVI) could be the proportion of occurrences correctly predicted as suitable according to the ROC-based threshold. But following this simple index, a model predicting all cells throughout the study area to be suitable would be a perfect model, which is of course not relevant. A solution is to compare the performance of the model to the performance that could be expected by chance alone. Two proportions were compared (Hirzel and Arlettaz 2003b): the proportion of occupied cells correctly predicted as suitable (AVI) on the one hand and the proportion of cells predicted as suitable across the entire study area (A_g) on the other hand. A_g indicates what proportion of occupied cells correctly predicted as suitable could have been expected by chance alone in any model identifying the same number of sites as suitable. AVI and A_g were compared by calculating the difference between them, yielding the Contrast Validation Index (CVI). CVI is always positive (Hirzel and Arlettaz 2003b): a value near 0 means that the model's performance is not better than by chance alone and the higher the value the better the model, i.e. the more accurate and more 'contrasted' (see Hirzel and Arlettaz 2003b).

AVI, A_g and CVI were computed for each ENFA-GLM procedure (100 pseudo-absences generations for each proportion of discarded occurrences) using the same data set as used for calibration[1]. These indices

---

[1] We are conscious habitat models should be evaluated with an independent data set (e.g. Fielding and Bell 1997, see Chapter 5) or through cross-validation (e.g. Boyce *et al.* 2002, see Chapters 3 and 4) before being used to map species spatial distribution or habitats, so as to reduce the risk of overestimating model accuracy. However, we only intended here to rank the influence of differential outsider elimination rates on a relative scale and not to compute valid HS maps in absolute terms.

were also computed for the different concentric envelopes delineated by ENFA (ENFA-only model), using the same increasing thresholds of ENFA-HS value as those used for putative outsiders designation.

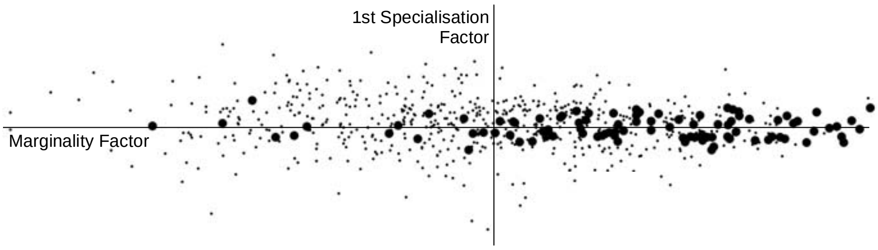
3. Results

Figure 8A illustrates the Red-backed Shrike occurrences (n=96) plotted in the two-dimensional space depicted by the marginality and the first specialization factors of the ENFA. By comparing the eigenvalues to Mac-Arthur's broken-stick distribution (Jackson 1993, Hirzel *et al.* 2002a), 7 significant factors were retained for ENFA-HS computation. Together, they explained about 83.5% of the information (100% of the marginality and 67% of the specialization). The marginality factor showed an association of the Red-backed Shrike with humid and extensive areas, with abundance of nest sites and of semi-natural patches. The distance to national roads and urban areas for nest site selection also appears along this factor. Niche breadth (reflected by specialization factors) of the Red-backed Shrike appeared to be restricted by nest sites abundance, distance to human settlements and usable areas for hunting.

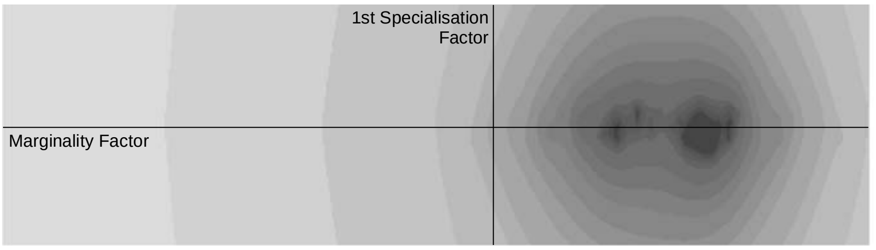
The density of occurrences in the hyperspace was modelled with the distance geometric-mean algorithm. This ENFA-HS field, represented in 2D on Figure 8B, was actually computed in 7D and used for assigning a ENFA-HS value to each cell of the study area (see Figure 8C). Only 32 cells of the global data (n=695) fell below or on the lowest ENFA-HS value assigned to an occurrence (i.e., ENFA-HS = 0.01), as illustrated on Figure 9. For a comparison, Figure 9 also shows the cells of the global data that could randomly be selected as pseudo-absences after the filtration of 10% of putative outsiders, i.e. 10% of occurrences with the lowest ENFA-HS values. This example illustrates the effect of leaving putative outsiders out of consideration on the delineation of the envelopes and the ensuing selection of pseudo-absences. These putative outsiders were occurrences scattered at the margins of the axes, reflecting occupied sites that could be less or not at all correlated to truly suitable environmental conditions. Envelopes were considerably wider when circumscribing all occurrences compared with envelopes delineated after some proportion of discarded occurrences (see also Figure 8B).

The evaluation measures (CVI, AVI and A_g) for the different ENFA-GLM and ENFA-only models are given in Figure 10, as well as their trend according to the proportions of discarded occurrences. ENFA globally tended to provide over-optimistic HS maps and less accurate predictions than ENFA-GLM procedures (see also Engler *et al.* 2004). As expected, the proportion of occurrences correctly predicted as suitable (AVI, Figure 10B) and the proportion of cells predicted as suitable across the entire study area (A_g, Figure 10C) both decreased as the percentage of occurrences discarded as outsiders during the ENFA-GLM procedure increased. The decrease of A_g was more abrupt for the lower proportions of discarded occurrences, indicating that a high number of cells were declared suitable when very few or no outsiders were discarded and that this number fell rapidly with a small increase in the number of occurrences discarded, tending to decrease less dramatically afterwards. The combination of AVI and A_g trends gave the plot of CVI illustrated on Figure 10A. CVI increased rapidly from 0% to 15-20% of discarded occurrences, reached a ‘plateau’ up to 40-50% and gently decreased afterwards. Note that there was no variation in the evaluation measures for 0% of discarded occurrences (i.e. the procedure proposed by Engler *et al.* 2004) because only 32 cells of the global data fell below or on the lowest ENFA-HS value assigned to an occurrence. Consequently, these 32 cells only were always selected as pseudo-absences for the 0% threshold, giving identical results for the 100 runs of the procedure. The low CVI for this threshold could be due to the imbalance between the numbers of occurrences (96) and pseudo-absences (32). Nevertheless, trends of AVI, A_g and CVI were similar when assigning a weight to pseudo-absences in the GLM so that the sum of these weights adds up to give the number of occurrences (for counterbalancing the lower proportion of pseudo-absences).

A.



B.



C.

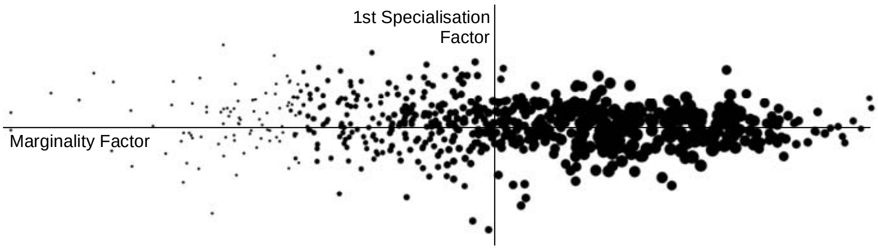


Figure 8. Representation of the first two dimensions of the environmental hyperspace created by the ENFA. (A) Red-backed Shrikes occurrences (n=96) are represented by black dots and all other cells of the study area by little black points. (B) Occurrences density is modelled by the distance geometric-mean algorithm (ENFA-HS field - the darker, the more suitable). Envelopes are delineated at each 5th percentile of ENFA-HS value. (C) ENFA-HS values assigned to each cell (n=695) of the study area following the modelled occurrence density in the 7D hyperspace. The size of the bubbles is proportional to the ENFA-HS values.

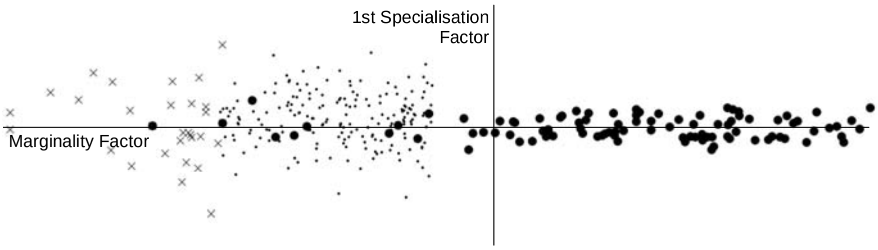
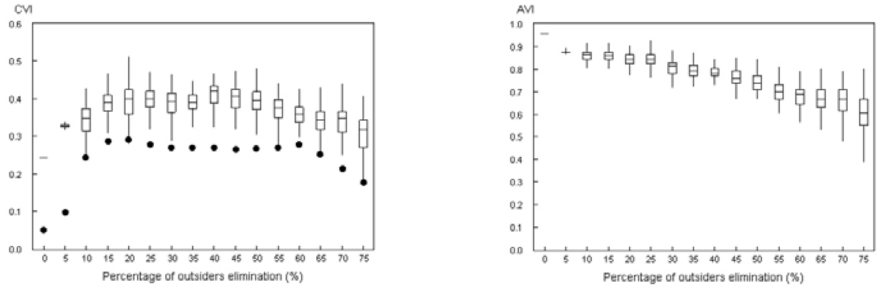


Figure 9. Distribution of the pool of potentially selectable pseudo-absences after no screening (black crosses, see Engler *et al.* 2004) and after discarding 10% of putative outsiders (black crosses + little black points), in the 2D environmental hyperspace represented by the first two factors of the ENFA. Occurrences (n=96) are represented by black dots.

A. B.



C.

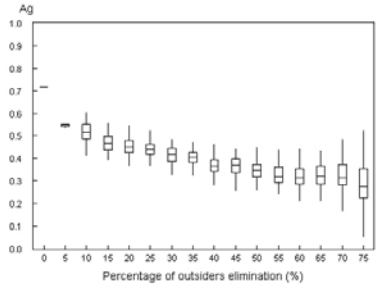


Figure 10. Trends of the evaluation measures of the ENFA-GLM models (n=1600, 100 repetitions for each of the 16 thresholds of HS value) against the percentage of discarded occurrences (from 0% to 75%) before ‘pseudo-absences’ generation. (A) Box-plot of Contrast Validation Index (CVI). Black points represents the same evaluation measure but for ENFA-only models. (B) Box-plot of Validation Index (AVI) (proportion of occurrences correctly predicted as suitable). (C) Box-plot of Ag (proportion of cells predicted as suitable among the 695 cells of the global data).

4. Discussion

Habitat Suitability values were computed using a modelling strategy that combines a ‘profile method’ (ENFA) and a ‘discrimination approach’ (GLM). This procedure reduced the problem of misleading absences by increasing the chances to select pseudo-absences where environmental conditions were truly unsuitable (Engler *et al.* 2004), while addressing the issue of individuals occurring in unsuitable conditions (Pulliam 2000).

According to the procedure proposed by Engler *et al.* (2004), the pool of pseudo-absences, in our case study, was shown to be constituted by a majority of cells situated on the margins of the environmental hyperspace defined by the ENFA. In the subsequent GLM, the risk associated with this strategy is to compare occurrences to a set of completely ‘atypical’ or marginal cells (pseudo-absences) that are not representative of the unsuitable environmental conditions for the species. A possible consequence is to infer the obvious, i.e. the species is not present in those completely ‘atypical’ and marginal cells, which is of course not what is hoped from predictive models. In other words, the existence of a few marginal occurrences in the hyperspace can force pseudo-absences to be selected too far from the main and relevant cloud of occurrences, as illustrated on Figure 9 (black points). This could involve a biased generation of pseudo-absences resulting in a general overestimation of the GLM-HS values for many cells. More than 70% of the cells were predicted as suitable when the percentage of discarded occurrences was set to 0 (see Figure 10C) [1]. These predicted cells of course include the most marginal occurrences, explaining the high AVI as illustrated on Figure 10B. Overestimating areas of suitable habitats may have highly undesirable consequences when HS maps are used for resource management planning (Loiselle *et al.* 2003). Furthermore, when comparing two clouds of cells excessively separated in the hyperspace with logistic

---

[1] Procedure proposed by Engler *et al.* (2004).

models, a possible effect is that GLM find (by mere chance) a small collection of descriptors that completely explains change in the log-odds. This could lead to a complete separation of occurrences and pseudo-absences, i.e. an almost perfectly discriminating but ecologically unsound model. Such discrimination could cause numerical problems because the maximum likelihood estimates of parameters may not exist or may be infinite, with huge standard errors (Hosmer and Lemeshow 2000, Agresti 2002). Although not happening in our case study, this is a phenomenon to be aware of.

With regard to this marginal occurrences issue, we assumed a straight link between occurrence density in the hyperspace and the suitability of environmental conditions (see also Hirzel *et al.* 2002b, Hirzel and Arlettaz 2003b). As a consequence, isolated or loosely distributed occurrences in the hyperspace were deemed to be unrepresentative of suitable conditions merely because they were unrepresentative of commonly used conditions. By progressively discarding these occurrences, our methodology intended to eliminate potentially unreliable information (i.e. putative outsiders) and to select pseudo-absences that more likely are relevant. In fact, when screening out marginal occurrences, the generated pseudo-absences could be more representative of the range of unsuitable environmental conditions and 'fix the floor' more pertinently. On the one hand the validation index (AVI) decreased when the percentage of discarded occurrences increased, because marginal occurrences were logically less and less predicted as suitable. On the other hand the proportion of cells predicted as suitable among the global data (A_g) fell drastically, indicating that models were less optimistic. The consequence was a global increase of the Contrast Validation Index (CVI) (see Figure 10A). The 'plateau' of CVI reached between 15-20% and 40-50% of discarded occurrences (Figure 10A) is explained by the similar trend of AVI and A_g within this range of percentages. The envelopes were closer to each other starting from 15-20% because the occurrence cloud became more densely distributed towards its innermost part, as illuminated by the

Figure 8B. Consequently, the pseudo-absences generation did not differ outrageously from one threshold to the next within this range, explaining the gently negative slope of A_g. Since the trend of AVI was fairly similar up to 40-50% of discarded occurrences, the CVI was stable, A_g and AVI making up for each other. Of course, discarding too high a proportion of occurrences as putative outsiders could induce a loss of reliable occurrences on the one hand, and the selection of some pseudo-absences that are suitable cells on the other hand. Moreover, discarding too many putative outsiders caused pseudo-absences to be generated at both extremes of the 'marginality axis' (see ENFA-HS field on Figure 8B), which may be ecologically irrelevant in some cases. This could explain the negative slope of CVI and the larger variability (reflecting instability) for AVI and A_g, starting from 40-50%. Note that discarding more than 50% of occurrences as outsiders is clearly ecologically unsound but was achieved here to illustrate the effect of too drastic discarding rate.

Looking at these results, we suggest CVI trend based on distance geometric-mean modelling could be used to reflect the relative variation in occurrence cloud density and in global cloud density along the ENFA-HS field. In particular, this trend enables to detect the envelope in the ENFA hyperspace beyond which occurrences begin to be loosely scattered among unoccupied cells. This envelope corresponds to the beginning of the CVI plateau. Therefore, we would suggest considering here about 15% of the occurrences with the lowest ENFA-HS values as possible outsiders.

The trend of CVI based on ENFA-only and ENFA-GLM models appeared to be closely related (Figure 10A). Further investigations based on simulated distributions in varied ecological conditions are required to gauge the generality of these parallel trends. Should this happen, we would encourage modellers not to go through the entire ENFA-GLM procedure. They could rather calculate CVI for ENFA-only models at varying discard levels and identify where the plateau begins with their specific data before deciding on the appropriate cut-off for generating pseudo-absences.

Obviously, ecological soundness of ENFA-HS-based filtering of putative outsiders remains to be confirmed. Future studies should therefore focus on drawing a parallel between ENFA-HS values and measures that are closely linked to the quality of the habitat. In birds, the parental reproductive success could prove to be informative (Wiens 1989b, Penteriani *et al.* 2003, Muller *et al.* 2005). This way the assumption of a straight relation between occurrence density in the hyperspace on the one hand and habitat suitability on the other hand would be discussed carefully. The veracity of this assumption and the very existence of outsiders most probably depend on the type of species to be modelled (e.g. species using active *versus* passive dispersal or mobile *versus* sessile species). Moreover, such analyses would assess the foundation of the statistically-defined cut-off between reliable and unreliable occurrences. In some species, the distinction could not be clear-cut but quite fuzzier. This is of paramount importance for generating pseudo-absences relevantly. We stimulate modellers to further look into marginal occurrences whenever a similar CVI trend is detected along the ENFA-HS field, because CVI informs of the existence of unrepresentative occurrences, which could convey unreliable indication about habitat suitability. We regard our procedure as enabling to blow the whistle but the final decision (whether or not the identified putative outsiders are to be discarded for HS modelling) should be made following thorough analyses guaranteeing that the function and relevance of these individuals are not important from a habitat suitability perspective.

We would like to finally specify that we do not regard outsiders as insignificant individuals in absolute terms. Analyzing the factors that incite such individuals to settle down in such conditions could help to understand the environmental boundaries of the species niche. Moreover, outsiders could provide valuable information about the spatial distribution pattern of the species. They should therefore be useful for delineating the potential distribution of the species.

In conclusion, neglecting the existence of occurrences that may not convey reliable information on habitat suitability is to be avoided. Instead, explicitly considering them in the modelling framework is required to ensure that ecologists and managers gain a more pertinent understanding of observed patterns and are able to delineate suitable areas for such species.

Acknowledgements

We thank Jean-Baptiste Schneider and Julien Radoux for their technical help on the field and with GIS software, respectively. We are grateful to Éric Le Boulengé, Nicolas Schtickzelle, Dominique Fasbender and Wouter Vanreusel for useful discussions or for commenting on a draft of the manuscript. We thank the Institut Géographique National (Belgium) and the Walloon Region for providing us with the vector 1:10,000 maps and the aerial colour orthophotographs (Licence number 031205-1000), respectively.

Chapter 3 – Breeding Success Information to Better Circumscribe the Species Niche

Note – This study has been submitted to Journal of Applied Ecology (Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. and Defourny, P. Identifying the Habitat Requirements of the Red-backed Shrike (*Lanius collurio*): the Importance of Breeding Success). The text presented here is slightly modified from the submitted paper for layout and terminology harmonization, as well as for references updating. Some sections of 'Materials and method' and some Tables are partly – but not fully – redundant with Chapter 2 and we therefore apologize to the reader.

Foreword [1]

The statistical approach proposed in the Chapter 2 was blind to any ecological foundation. The main aim of the present Chapter is to assess, on the basis of breeding success data, the ecological relevance of the ENFA-based elimination of outsiders. This justifies the use of ENFA-only models in this Chapter, whereas ENFA-GLM combinations were used in the previous one.

Abstract

Many bird species of semi-open areas throughout Western Europe are strongly dependent on the directions of the agriculture policies. Hence, it is crucial to gather accurate and fine-grained information about how the various environmental components are driving species habitat selection and breeding performances, if we are to devise efficient management strategies. We use the case of the depleted Red-backed Shrike (*Lanius collurio*), a species of man-shaped farming landscapes, to illustrate how this can be achieved. Fine-scale 'Resource Selection Functions' (RSF) were built to identify the main environmental forces causing the spatial distribution of the Red-backed Shrike in Southern Belgium. Breeding success data measured on

---

[1] This Foreword does not belong to the submitted paper to Journal of Applied Ecology.

the field were integrated into the modelling framework in order to identify the key breeding habitat requirements. The computation of ecologically-founded descriptors reflecting diverse resources allowed focussing on the functional interactions between organisms and their environment, dealing with the flexibility of the habitat selection pattern of the species and offering direct applications regarding management and restoration planning. As absence data were presumed unreliable because the population could not saturate all suitable sites in the landscape, a presence-only method was used to generate the RSF. The incorporation of breeding performances into the functional modelling procedure revealed the existence of occurrences conveying unreliable information about breeding habitat suitability and strengthened the importance of discarding them for accurately delineating the suitable areas, instead of using all occurrences indiscriminately. Combined with breeding success information, RSF showed the importance of multiple and non-substitutable resources for the successful reproduction of individuals, regarding nesting sites availability, distance to human settlements, suitable perching sites, foraging areas and insect abundance. This stressed the importance of accounting for every aspect of habitat requirement when planning management or restoration.

Synthesis and Applications – Incorporating breeding success into a predictive modelling approach may prove to be required for discarding misleading occurrences. The ensuing models illuminated the key breeding habitat requirements for the Red-backed Shrike and called attention to the negative impacts of homogenization and intensification of Western European farming areas.

1. Introduction

The Red-backed Shrike (*Lanius collurio*) has shown a marked decline in Western and Northern Europe between 1970 and 1990. Decline continued more slightly in several Western countries during 1990-2000 (Yosef 1994, Lefranc and Worfolk 1997, Lefranc 2004, BirdLife International 2004). Although causal factors remain unclear (Yosef 1994), it was suggested that reduction in suitable habitats, decline in food resources, climatic change and nest predation by corvids are the main reasons (review by Lefranc and Worfolk 1997). Human development and agricultural intensification (e.g. intensive grazing or mowing, intensive monocultures, hedges cutting, larger parcels, agrochemicals use), as well as cessation of agricultural activities (leading to field and rural landscape encroachment, see Scozzafava and De Sanctis 2006), are among the main causes of these changes (Lefranc and Worfolk 1997, Van Nieuwenhuyse 1999). Therefore, its conservation is strongly dependent on directions that will be taken in the European agricultural policy (Van Nieuwenhuyse 1999).

Most of the Western European Red-backed Shrike populations breed in a rather wide range of semi-open areas created or maintained by extensive farming activities (combination of hay production and livestock rearing) and punctuated by an appropriate network of thorny hedges and bushes (e.g. Van Nieuwenhuyse 1998a, Van Nieuwenhuyse *et al.* 1999, Lefranc 2004). This species of man-shaped landscapes globally requires the persistence of extensive farming techniques but the flexibility of its habitat selection pattern makes the detection of its main requirements challenging. It is though crucial to gather accurate and fine-grained information about how the various environmental components are driving its habitat selection and breeding performances, if we are to implement efficient management strategies.

Fine-scale predictive habitat models can be valuable tools for providing useful and precise information about species distributions (see

Scott *et al.* 2002, Guisan and Thuiller 2005). Most commonly adopted modelling approaches investigate the relationships between the species and its environment (Heglund 2002, Van Horne 2002) and generate ‘Resource Selection Functions’ (e.g. Boyce and McDonald 1999, Guisan and Zimmermann 2000, Zaniewski *et al.* 2002, Johnson *et al.* 2004, Guisan and Thuiller 2005). Coupled with geographic information systems (GIS), RSF can produce maps displaying the suitability of the habitats (e.g. Johnson *et al.* 2004, Gibson *et al.* 2004).

Most of the predictive models relate species occurrences to coarse-scaled environmental variables (e.g. Gates *et al.* 1993, Engler *et al.* 2004, Brotons *et al.* 2004), most often including topography, climate or surrogate predictors with some indirect relations to the causal factors. Such approaches are useful for describing broad-scale distribution patterns but may suffer from multiple criticism, especially regarding direct conservation applications (e.g. Guisan and Thuiller 2005), transferability (e.g. Vanreusel *et al.* 2006) or ecological meaning (e.g. Austin 2002, Guisan and Thuiller 2005). At a fine spatial scale it is essential to focus on ecological resources determining the functional interactions between the organisms and their environment (Tischendorf 2001, Vos *et al.* 2001, Vaughan and Ormerod 2003, Dennis *et al.* 2003). Besides remedying the above cited drawbacks, investigating the causal species-environment relationships allow dealing with the flexibility of the species habitat selection pattern. It becomes therefore necessary to compute ecologically-founded functional predictors reflecting the fine-scale resources that are common to the variety of occupied habitats.

Furthermore, the regional pattern of settlement of a Red-backed Shrike population can change considerably between successive years, with suitable sites not occupied each year (Van Nieuwenhuyse 2000b, Söderström 2001). We assumed the population size could not reach the carrying capacity of the environment and absences were therefore deemed meaningless, leading to severe limitations and biases for fitting classical ‘discrimination models' relying on presence-absence data (like Generalized Linear models, GLM) because these techniques assume that they respectively indicate suitable and unsuitable environmental conditions (e.g. Guisan *et al.* 2002, Gu and Swihart 2004, Engler *et al.* 2004, Brotons *et al.* 2004, Guisan and Thuiller 2005). Models based on presence-only data are particularly advisable in such a case (Hirzel *et al.* 2001, Pearce and Boyce 2005, Elith *et al.* 2006).

Based on Hutchinson's (1957) niche paradigm, presence-only models generally delineate envelopes around species occurrences in an environmental hyperspace (see Pearce and Boyce 2005), assuming that species occurrences are more or less closely associated to suitable environmental conditions. Nevertheless, this assumption can be violated in many cases (e.g. Robinson *et al.* 1995, Pulliam 2000). Indeed, behavioural, social, historical or population processes could cause some individuals to occur outside the environmental bounds of the species fundamental niche (Pulliam 2000 and see introduction of this thesis), hence not conveying reliable information about habitat suitability (Guisan and Thuiller 2005). However, despite potential distortion between habitat suitability and occupancy, most published studies use indiscriminately all occurrence records to build predictive models, without actually establishing what does and what does not constitute suitable environmental conditions (Wiens 2002). Titeux *et al.* (2006b, Chapter 2) proposed a statistical approach to discard such misleading occurrences before calibrating relevant presence-only habitat models. Nevertheless, since habitat suitability is recognised as an important factor influencing breeding success in many birds (e.g. Wiens 1989b, Penteriani *et al.* 2003, Muller *et al.* 2005), a rigorous assessment of habitat suitability should relate this success to environmental conditions, which is rarely achieved in the context of habitat modelling (but see Railsback *et al.* 2003, Olson *et al.* 2004, Ozesmi *et al.* 2006).

In this study, we relate the Red-backed Shrike breeding success to environmental conditions within a niche-based modelling framework. We first describe these conditions by computing ecologically-founded functional descriptors reflecting the fine-scale resources common to all occupied habitats. Using breeding success, we assess the ecological meaning of considering and discarding misleading occurrences within a modelling context, by comparing breeding success-supported and classical occupancy-supported habitat models. We take advantage of breeding success data to (1) gather accurate and fine-grained information about how the various environmental components are driving species habitat selection and breeding performances and (2) identify the key resources. We finally emphasize the implications of these outcomes for the conservation of the species in Western European farming areas.

2. Materials and methods

2.1. Study area

The reader is referred to Chapter 2 and Figure 7 for the description of the study area.

2.2. Species data

The reader is referred to Chapter 2 for description of species data collection. For this study, species data were gathered in May-July 2005.

Various aspects of reproduction may serve as measures of breeding success, e.g. timing of breeding, clutch size, fledging success or fledgling weight (Leugger-Eggiman 1997). While the presence or abundance of offspring is an important component of breeding success, survival to their first breeding season is crucial for their contribution to future generations. Consequently, fledging success alone is not the best measure of parental breeding success (Leugger-Eggiman 1997) but was assumed here to be an informative indicator of reproductive performance. Nesting pairs were considered successful if they produced at least one fledgling (see Muller *et al.* 2005), because this information was relatively easy to collect on the field.

Each presence was allocated to (1) unpaired male, (2) unsuccessful pair (fledging failure) or (3) successful pair (fledging success). Replacement clutches following nesting failures and leading to fledging failure or success on the same sites were allocated to categories 2 or 3, respectively. Such pairs did not count double.

2.3. Environmental functional descriptors

The reader is referred to Chapter 2 for a description of environmental data acquirement (Table 5) and of multi-scale environmental descriptors computation (Table 6). We assumed that landscape modifications between 2004 (environmental data acquirement) and 2005 (species census period for this study) were negligible.

A subset of the most functional descriptors was retained from the Table 6. These resource-based descriptors (hereafter called ‘environmental functional descriptors’, EFDs) are listed in Table 7. Their underlying ecological rationale is detailed in the Foreword of Chapter 2.

Table 7. List of EFDs calculated for each cell, with their spatial scale of computation, their functional significance (type) and a brief description (EFDs are a functional subset of environmental descriptors used in Chapter 2, see Table 6).

EFD	Resolution (m)			Function	Description
	75	150	300		
Nest		x		Nest	Abundance of bushes and/or hedges (transformed to points separated by 5m intervals) suitable for carrying a nest (thorny / height 1-6m / length < 50m / not completely inside a very intensive pasture / distance to forest > 25m)
NestDist		x		Nest	= Nest, where each bush and/or hedge is positively weighted by its distance to urban area
NestStd		x		Nest	Standard deviation of the heights of bushes and/or hedges suitable for nest installation
Arable		x	x	Prey density	Area of arable lands
IntPast		x	x	Prey density	Area of very intensive pastures
SemiNat			x	Prey density	Area of (temporarily) unmanaged biotopes and/or hay meadows (considered as 'semi-natural' biotopes)
SemiNatW		x	x	Prey density	= SemiNat, where area of each patch is weighted by the area of 'semi-natural' biotopes patches within a radius of 250m
Extensive		x	x	Prey density	Area of pastured and cultivated biotopes where each type is weighted by a coefficient representing the relative Prey density (see Table 5)
SoilDepth			x	Prey density	Mean phreatic table depth
SoilDry				Prey density	Euclidean distance between the centre of the cell and the nearest point where phreatic table depth is less than 15 cm
DiForest				Prey density – Predation	Euclidean distance between the centre of the cell and the nearest forest

(Table 7 continued)

Name	Resolution (m)			Function	Description
	75	150	300		
DiUrban				Predation – Disturbance	Euclidean distance between the centre of the cell and the nearest urban area (building)
Hunt	x	x		Prey detectability	'Usable area' for hunting defined by a 20m-buffer around each suitable perch for hunting (height 1-4m / bush, hedge or fence / distance to suitable nest site < 40 m) and intersecting suitable patches for hunting (see * in Table 5)
HuntVg	x	x		Prey detectability	= Hunt, where the area of each intersected patch is negatively weighted by the mean vegetation height of this patch type
NbHunt		x		Prey density – detectability	Number of patches intersecting the 'usable area' for hunting
Interface		x	x	Prey density – detectability	Length of all interfaces between the different open patches
Contrast		x	x	Prey density – detectability	= Interface, where each interface type is positively weighted by the difference of vegetation height between both patch types

It is worthwhile to indicate that forests and urban areas outside the study area were taken into account for distances computations (external buffer of 1 km).

2.4. Statistical methodology

In order to quantify the ecological requirements for the Red-backed Shrike, we analysed the relationship between two GIS data sets, the spatial distribution of the species on the one hand and the set of EFDs on the other hand.

A first *a priori* screening discarded those cells that were completely wooded or with neither bush nor hedge because they were clearly unsuitable for Shrike settlement.

EFDs were standardized (mean=0 and variance=1) and normalized using the Box-Cox algorithm (Sokal and Rohlf 1998). Departure from normality after transformation was detected for some EFDs but the adopted ordination technique is not too sensitive to this assumption (Hirzel *et al.* 2002a). Spearman correlation coefficients were computed for each pair of EFDs.

2.4.1. Ecological Niche Factor Analysis

As Red-backed Shrike absences are meaningless, the species niche was described using a presence-only approach, the Ecological Niche Factor Analysis (ENFA, Hirzel *et al.* 2002a). ENFA uses the distribution of species occurrences to summarise EFDs into independent components (as Principal Component Analysis) that are related to the species ecological niche.

The first component explains the ‘marginality’ of the species regarding the EFDs, describing how far the species optimum is from the mean environmental conditions within the whole study area. Next factors are then extracted orthogonally in order to explain the ‘specialization’ of the species, describing the narrowness of its niche (see Hirzel *et al.* 2002a and Chapter 2 for a detailed description of the method).

ENFA was applied using (1) all occurrences indiscriminately (global-ENFA) and (2) successful pairs only (success-ENFA), and was performed with BIOMAPPER 3.1 (Hirzel *et al.* 2002b).

2.4.2. Habitat Suitability

The few first factors of the ENFA, gathering the majority of the information, were used to compute Habitat Suitability (HS[1]) values between 0 and 1 for all sites within the study area. The distance geometric-mean algorithm was used and the reader is referred to Chapter 2 for a thorough description of this distance-based algorithm.

2.4.3. Accounting for outsiders

Titeux *et al.* (2006b, Chapter 2) suggested that about 15% of Red-backed Shrike occurrences with the lowest HS values following the ENFA (hereafter called 'outsiders') could convey misleading information about habitat suitability. The ecological relevance of this suggestion is assessed and discussed below in light of breeding success data. For supporting this discussion, the species-environment relationship was approached in a different but complementary way. The Euclidian distance between each occurrence and the species average environmental conditions was computed in a p-dimensional environmental space (one dimension for each EFD, here p=24). As the EFDs were standardised, their relative contributions to the distance values are comparable (Legendre and Legendre 1998). For a given occurrence, this Euclidian distance reflects the departure of its environmental conditions from the average environmental conditions occupied by the species, i.e. its marginality compared with the species mean conditions. Moreover, this distance metric is sensitive to the extent of the departure from the mean in individual dimensions. High values reflect cells for which EFDs are marginal for one or few of them, while low values indicate cells characterized by EFDs that are all similar to or not excessively distant from the species mean conditions. A parallel will be drawn between these Euclidian distances and ENFA-based HS values of occurrences.

---

[1] This acronym is equivalent to ENFA-HS used in Chapter 2.

2.4.4. Model evaluation

The lack of reliable absence data makes the assessment of presence-only models difficult with classical methods related to confusion matrix, including Kappa coefficient and ROC curves (Pearce and Ferrier 2000, Boyce *et al.* 2002, Ottaviani *et al.* 2004).

Two evaluation indices were computed (Hirzel and Arlettaz 2003b). The Absolute Validation Index (AVI) is the proportion of occupied cells for which HS value is higher than 0.5. The Contrast Validation Index (CVI) is equal to the difference between the AVI and A_g (A_g is the proportion of cells for which HS values are higher than 0.5 across the entire study area). CVI ranges from 0 to $1 - A_g$ and reflects model accuracy (Hirzel *et al.* 2004), values near 0 indicating that the model does not outperform a random one (Hirzel and Arlettaz 2003b).

AVI and CVI were estimated through a cross-validation procedure (Manly 1997, Fielding and Bell 1997, Sokal and Rohlf 1998, Hirzel and Arlettaz 2003b). The data set was partitioned into 20 subsets. In turn, 19 of these were used for model calibration and the remaining one for model evaluation (AVI and CVI computation). This procedure provided mean and standard deviation of AVI and CVI for global- and success-ENFA.

3. Results

A priori screening of cells where land-use was thought to be incompatible with the species settlement and reproduction eliminated 480 cells (Figure 11). The remaining 1184 were used for the subsequent analyses. No strong correlation was detected between EFDs belonging to different functional types (all spearman $|\rho| < 0.7$), while few ones are quite correlated ($|\rho| > 0.7$) within some of these types (especially among ‘prey detectability’ EFDs) or between spatial scales for multi-scale EFDs (Extensive, Arable, Hunt).

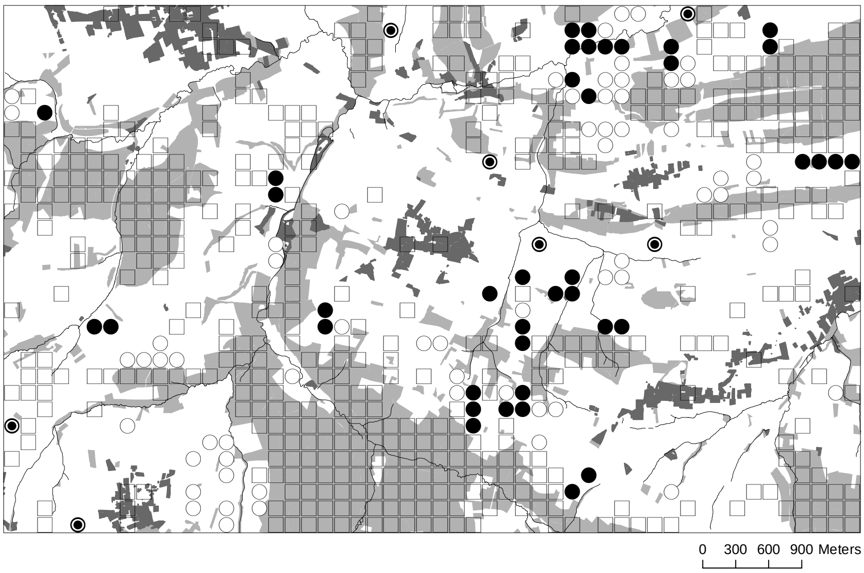


Figure 11. Localisation of occurrences within the study area. Successful and unsuccessful pairs are represented by white and black dots, respectively. Unpaired males (n=7) are depicted by encircled black points. Squares represent a priori masked-out cells (n=480), where land use is incompatible with the Red-backed Shrike settlement. See Figure 7 for background features.

3.1. Global-ENFA

In 2005, 74 males settled down in the study area, from which 67 were paired with females. Territories were aggregated at the cell level, reaching a total of 110 occupied cells because of straddling territories (Figure 11). ENFA computed global marginality and tolerance coefficients of 1.32 and 0.72 respectively, indicating that the Red-backed Shrike settled down in environmental conditions highly different from the global average, but with a rather large niche breadth. By comparing the eigenvalues to Mac-Arthur's broken-stick distribution (Jackson 1993, Hirzel *et al.* 2002a), 4 significant factors of the ENFA were retained for HS computation. They altogether explained about 72% of the information (100% of the marginality and 44% of the specialization). This means that 4 factors were sufficient to describe the niche of the species when using all occurrences. The projection of the

1184 cells on the first two factors is illustrated in Figure 12. Besides marginality, the first factor explained only 8% of specialization. This low fraction means that the combination of EFDs that explained the species marginality did not explain its specialization adequately.

The cross-validation procedure provided a mean AVI of 0.53 (SD = 0.20) and a mean CVI of 0.35 (SD = 0.20). These results indicate that the model prediction power was good but the difference between AVI and CVI shows that part of model performance was maybe attributable to randomness. On the 201 cells in the core area (HS > 0.5, as defined by Hirzel *et al.* (2004) for the bearded vulture), 148 (74%) were unoccupied in 2005. The Figure 13 shows the distribution of cells' frequencies in different envelopes.

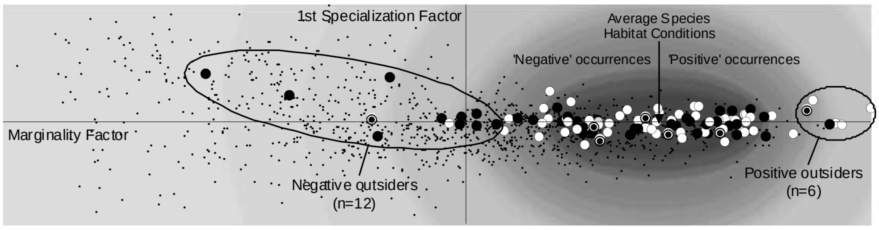


Figure 12. Red-backed Shrikes occurrences plotted in the 2D environmental space represented by the two first factors of the global-ENFA. Successful and unsuccessful pairs are represented by white and black dots, respectively. Unpaired males are depicted by encircled black points. Other cells of the study area are depicted by little black spots. The HS field modelled by the distance geometric-mean algorithm is represented here in 2D (the darker, the higher the HS values) although the occurrence density was modelled using 4D. Occurrences surrounded by ellipses are those with the lowest HS values, i.e. 15% of putative outsiders.

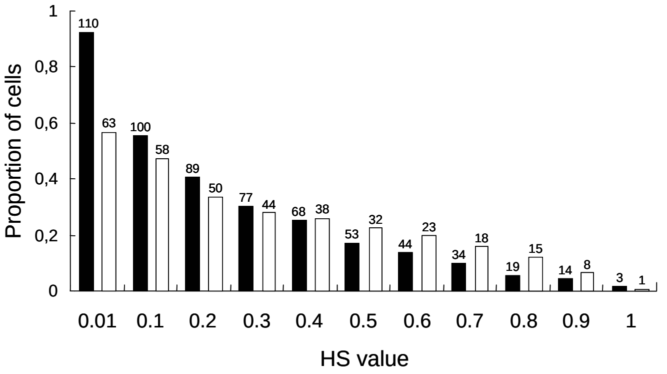


Figure 13. Proportions of cells (n=1184) enclosed by several envelopes, from marginal (low HS values) to core areas (high HS values), following the global- (black bars) and success-ENFA (white bars). The number of occurrence cells (total or successful only, for black and white bars respectively) enclosed in each envelope is indicated at the top of each bar. Envelopes circumscribing all occurrences (global-ENFA) or all successful pairs (success-ENFA) correspond to a HS value of 0.01, according to the distance geometric-mean algorithm and HS isopleths scaling.

Coefficients on the marginality factor were positive for most of the EFDs (Table 8), showing that the Red-backed Shrike was found in areas where these EFDs are higher than average conditions. By decreasing order of importance, the marginality factor was mainly correlated to nest carriers' abundance and their distance to urban areas, to suitability for hunting (prey detectability) and to prey density. Negative coefficients on this factor indicated that the Red-backed Shrike settled down in areas where the soil dryness and the amount of very intensive pastures were slightly lower than average conditions.

Niche breadth of the Red-backed Shrike was mainly restricted by nest carrier's abundance and by distance to human settlements. On a lesser extent, the species was less tolerant regarding distance to forest, usable areas for hunting and field margin abundance. Areas of arable lands were weakly correlated to the retained factors. For multi-scale EFDs, marginality and specialization coefficients were generally not very sensitive to the scale.

Table 8. Correlation between the retained global-ENFA factors and the EFDs. Percentages indicate the proportion of specialization explained by each factor (100% of the marginality is accounted for by the first factor).

	Factor 1 (8%)	Factor 2 (16%)	Factor 3 (11%)	Factor 4 (9%)
Nest-150	+++	*****	0	****
NestDist-150	++++	*******	**	*******
NestStd-150	+++	0	0	*
Arable-150	0	*	**	*
Arable-300	0	0	0	*
IntPast-300	–	*	0	0
SemiNat-300	++	0	*	*
SemiNatWe-150	++	0	0	0
SemiNatWe-300	+	0	*	**
Extensive-150	++	0	*	**
Extensive-300	++	*	*	*
SoilDry	–	*	*	0
SoilDepth-300	– –	**	*	*
DiForest	0	**	***	0
DiUrban	++	****	**	**
Hunt-75	+++	*	**	***
Hunt-150	+++	0	*****	*
HuntVg-75	+	0	**	0
HuntVg-150	++	0	*	*
NbHunt-150	++	0	*	0
Interface-150	++	*	***	*
Interface-300	++	**	****	0
Contrast-150	++	*	0	*
Contrast-300	++	*	***	*

For the first factor (marginality factor), the symbols + and – mean the Red-backed Shrike settled down in areas with higher and lower values than average, respectively (0 indicates a very weak or no correlation). The number of symbols is proportional to the strength of the correlation.

For next factors (specialization factors), the symbol * indicates that the Red-backed Shrike occupies a narrower range of values than available (0 indicates a very low specialization). The number of symbols is proportional to this narrowness.

3.2. Success-ENFA

On the 74 territories (67 pairs and 7 unpaired males) found in 2005, 44 pairs (60%) reproduced successfully at first (n=36) or second (replacement clutch, n=8) attempt (63 cells of the study area, Figure 11). For these successful pairs only, the global marginality coefficient increased compared to the whole data set (1.52 vs. 1.32), indicating that the Red-backed Shrike bred successfully in more marginal environmental conditions. Moreover, the global tolerance decreased (0.41 vs. 0.72) showing that the niche breadth for these successful pairs was much more restricted compared with all settled-down individuals. By comparing the eigenvalues with Mac-Arthur's broken-stick distribution, only 2 significant factors of the ENFA were retained for HS computation. They altogether explained about 84% of the information (100% of the marginality and 68% of the specialization). The marginality of successful pairs was explained by about the same EFDs as the marginality of all occurrences. On the other hand, unlike with all occurrences, the marginality factor explained nearly half (47%) the specialization, indicating that niche breadth was mainly defined by the same combination of EFDs, that is, nest carriers abundance, distance to human settlements and suitability for hunting (Table 9).

Table 9. Correlation between the retained success-ENFA factors and the EFDs. Percentages indicate the proportion of specialization explained by each factor (100% of the marginality is accounted for by the first factor).

	Factor 1 (47%)	Factor 2 (21%)
Nest-150	+++	*****
NestDist-150	++++	*******
NestStd-150	+++	0
Arable-150	0	0
Arable-300	0	0
IntPast-300	–	0
SemiNat-300	++	0
SemiNatWe-150	++	0
SemiNatWe-300	++	0
Extensive-150	++	0
Extensive-300	++	*
SoilDry	–	0
SoilDepth-300	– –	0
DiForest	0	0
DiUrban	++	****
Hunt-75	+++	*
Hunt-150	+++	*
HuntVg-75	+	*
HuntVg-150	++	*
NbHunt-150	++	0
Interface-150	++	0
Interface-300	+	0
Contrast-150	++	0
Contrast-300	++	0

For the first factor (marginality factor), the symbols + and – mean the Red-backed Shrike settled down in areas with higher and lower values than average, respectively (0 indicates a very weak or no correlation). The number of symbols is proportional to the strength of the correlation.
For next factors (specialization factors), the symbol * indicates that the Red-backed Shrike occupies a narrower range of values than available (0 indicates a very low specialization). The number of symbols is proportional to this narrowness.

The cross-validation procedure gave a mean AVI of 0.58 (SD = 0.33) and a mean CVI of 0.35 (SD = 0.32). These indices did not significantly differ from their counterpart for the global-ENFA (Student's tests, $df = 38$, $p = 0.59$ and 0.96, respectively[1]). While a bit more optimistic for the success-ENFA (leading to a larger mean difference between AVI and CVI, i.e. 0.23 compared to 0.18 for global-ENFA), core areas defined by both procedures were fairly similar. On the other hand, the distribution of cells' frequencies in the different envelopes was very different for both ENFA, especially for marginal areas, as illustrated on Figure 13. This effect was due to the fact that, for global-ENFA, distant and scattered occurrences in the factorial space were almost only unsuccessful pairs (see Figure 12 and Figure 14) that stretched spuriously the marginal envelopes. These unsuccessful pairs being eliminated from success-ENFA dataset, the remaining occurrences were more densely distributed in the factorial space.

The proximity of potential predators (corvids) nesting or foraging sites was quantified by DiForest, DiUrban and Arable (Table 7). These EFDs did not significantly explain the differences between successful (n=44) and unsuccessful pairs (n=23) (logistic regressions, logit link, log-likelihood ratio tests: $p = 0.75$, 0.71, 0.89 and 0.44 for DiForest, DiUrban, Arable-150 and Arable-300 respectively, $p = 0.55$ for the full model including these four EFDs). For territories straddling adjacent cells, the mean values of the EFDs were retained for logistic regressions.

---

[1] These *p*-values are given here for indication only because tests are not fully valid due to redundancy (hence non-independence) in success and global data sets.

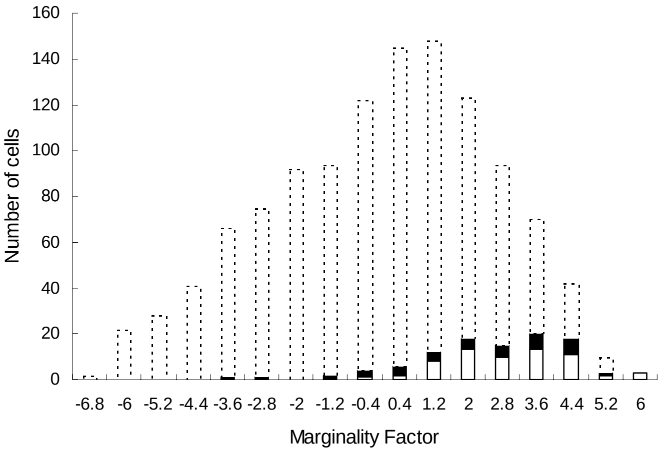


Figure 14. Distribution of occupied (white and black histograms, for successful and unsuccessful pairs, respectively) and unoccupied (dotted histogram) cells along the marginality factor of the global-ENFA (unpaired males were not considered here).

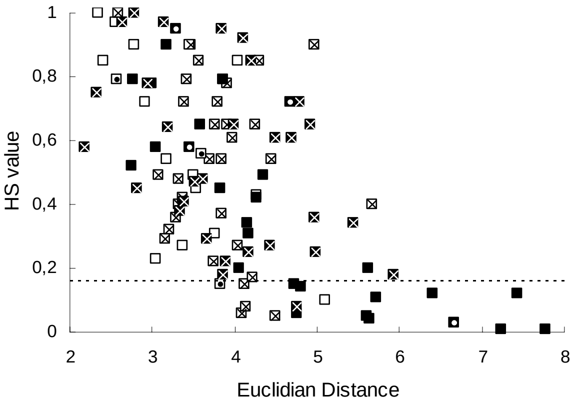


Figure 15. Representation of each occurrence depicted in a 2D scatter plot defined by Euclidian distance and HS values. The dotted line indicates the threshold of HS value for designating outsiders. 'In-' and 'Outsiders' are located above and below this line, respectively. White and black squares are occurrences situated on the 'positive' and 'negative' ends of the marginality factor, compared with the average environmental conditions occupied by the species within the study area (see Figure 12). Crossed and punctuated squares represent successful pairs (fledging success) and unpaired males, respectively.

3.3. Outsiders

Following global-ENFA, outsiders were designated as the most distant occurrences from the average environmental conditions occupied by the species in the study area, as illustrated in Figure 12. As a consequence, 15% of the occurrences with the lowest HS values (n=18) formed two groups at both ends of the marginality factor, hereafter called the ‘negative’ and ‘positive’ outsiders (see Figure 12). Despite their similar HS values, these two subsets of outsiders differed largely regarding four aspects: (1) Positive outsiders were found in rare environmental conditions, while negative ones were found in quite common ones (Figure 14). (2) The environmental conditions occupied by the positive outsiders were highly prized, while those on the other side were sporadically used by negative outsiders (Figure 14). (3) The Euclidian distance to average environmental conditions occupied by the species was significantly shorter (p=0.0016, one-sided Student test, df=16) for positive outsiders (mean = 4.29) than for negative ones (mean = 5.96), indicating that one or few EFDs for the latter were considerably more marginal, as compared with the average conditions used by the species (Figure 15). (4) The fledging success was much higher (Fisher’s exact test of independence, p=0.0128, n=18) for the positive outsiders (4 successful pairs, 1 unpaired male, n=6) than for the negative ones (1 successful pair, 1 unpaired male, n=12), indicating that environmental conditions with positive marginality were highly suitable for the reproduction of the Red-backed Shrike (Figure 14). On the negative end of the marginality factor, the dramatic decrease of fledging success appeared to match up to the proposed outsiders’ designation, but not on the positive end. All these elements converge to indicate that negative outsiders were pairs nesting in unsuitable environmental conditions, while positive outsiders were most probably an artefact of the habitat suitability model.

4. Discussion

4.1. Fledging success and outsiders designation

Nest predation – mainly performed by corvids (Söderström *et al.* 1998, Horvath *et al.* 1998, Roos and Part 2004) – accounts for a significant part of breeding failures in the Red-backed Shrike (Söderström 2001, Muller *et al.* 2005). Because of (1) the interspersion of the different land uses frequently inhabited or foraged by different predatory species (see Figure 7) and (2) the large flying capacities of corvids between their nest sites and foraging areas, the predation pressure by corvids was thought to exist everywhere throughout the study area, even if certainly not homogeneous. This quasi-ubiquitous predation, combined with other factors influencing fledging success (like parental quality, timing of breeding, conspecific density or weather conditions – see Muller *et al.* 2005), could explain why nest can be lost even in highly suitable habitats (Figure 14).

But our results show that differences between fledging successes and failures were not mainly related to the proximity of potential nesting sites or foraging areas of predatory species. This should nonetheless be confirmed with data acquired during multiple years. Such a claim does absolutely not deny the globally strong selective pressure of predation and its importance in shaping the habitat selection patterns of the species (Söderström 2001, Roos 2002, Roos and Part 2004 and see below). Nevertheless, looking at these results, it could reasonably be assumed that nest predation alone did not explain the dramatic decrease of fledging success for the negative outsiders (Figure 14). Instead, these occupied cells lacked one or few other essential components of habitat requirements for the reproduction of the species, as revealed by the Euclidian distance (Figure 15). These missing components were (from farthest to nearest negative outsiders in the factorial space) usable areas for hunting, nesting sites availability and prey density. As they are probably all required (non-substitutable) for the reproduction, shortage in one or several of these resources could explain the failure of the reproduction, possibly in synergy with predation. On the other hand, Euclidian distances of the positive outsiders were quite similar to those of the ‘insiders’ (85% of occurrences with the largest HS values), suggesting that all EFDs characterizing these cells did not excessively differ from the average species conditions. Simply, they were all ‘better’ than the average conditions used by the species and a fortiori than the global average conditions, explaining their marginalization by the ENFA.

Furthermore, our results indicated that, despite their similar HS values, the two subsets of outsiders differed largely regarding breeding success. On the one hand, the negative outsiders were pairs that attempted but failed to breed in unsuitable conditions. On the other hand, the positive outsiders were clearly pairs located in highly suitable conditions, rewarded by a high fledging success, which explains the quasi-saturation of these environmental conditions by the population (Figure 14). Because of their particularities (see just above – Euclidian distances) and relative scarcity (Figure 14), these latter occurrences were quite isolated at the positive end of the marginality factor in the environmental hyperspace, hence not representative of the average (although maybe less optimal) environmental conditions occupied by the species. According to the distance geometric-mean algorithm (Hirzel and Arlettaz 2003b), the absolute density of occurrences was assumed to reflect the suitability of the environmental conditions. Consequently, whatever their high suitability, the positive outsiders cells were not attractive enough when modelling the HS field. Hence, even if this algorithm was shown to be a good trade-off between precision and generality (Hirzel and Arlettaz 2003b), accounting for the availability of different environmental conditions would most probably increase the ecological relevance of the HS field under such specific circumstances. Even if unoccupied cells are meaningless in absolute terms, provided that detectability of the species is constant across occupied environmental conditions, saturation information (occupancy *versus* availability) is meaningful on a relative scale. Note that the other proposed algorithms did not provide more relevant HS field in this case (not shown here). Future works should look into this issue for improving existing algorithms. HS field was thus not modelled relevantly here and HS maps were not displayed.

Besides calling for the improvement of HS computation algorithms, fledging success data allowed defining habitat requirements for the reproduction of the Red-backed Shrike (see below) more accurately than with all occurrences indiscriminately. Indeed, despite the similarities of the EFDs coefficients between global- and success-ENFA, higher marginality, lower tolerance and higher percentage of information explained with fewer factors for the latter indicated that the niche for successful pairs was more accurately delineated and narrower than for all occurrences of the data set. Even if not strictly assessed here, this carefully delineated niche should enclose resources that lead to a population growth rate higher than 1. Similarities of AVI and CVI for both approaches were simply explained by the fact that main divergences did not lie in the core areas but rather in the marginal ones, as illustrated on Figure 13. This breeding success pattern strengthened the relevance and need of identifying and discarding misleading occurrences before describing the species niche. Overlooking these outsiders when drawing HS maps would induce a global overestimation of the areas of suitable habitat, which could obviously have highly undesirable consequences, as far as conservation is concerned (Loiselle *et al.* 2003).

A question remains unsolved about incentives of the settlement of these pairs in unsuitable environmental conditions. Rather than ignoring them, this issue should be further investigated and explicitly integrated in the process of spatial modelling in order to accurately predict the species spatial distribution.

4.2. Breeding habitat requirements for the Red-backed Shrike

Every species may respond to a combination of rather different sets of environmental forces in different parts of its distribution range (e.g. Gibson *et al.* 2004). This is undoubtedly the case with the Red-backed Shrike in the forested Ardenne region in Southern Belgium for instance, where the species is typically found in recently felled areas or early-stages plantations (Jacob 1999). However, ENFA was based on the comparison between the locations where the species has been observed and the available environmental conditions within the study area. Even if the HS model performed well within the study area, its efficiency was evaluated using a k-fold cross-validation procedure with occurrences in this same area and not in an independent one (Fielding and Bell 1997, Whittingham *et al.* 2003). Consequently, transferring the model to other areas must be achieved very cautiously (Fielding and Haworth 1995, Whittingham *et al.* 2003). While this issue is still under debate (e.g. Whittingham *et al.* 2003, Seoane *et al.* 2005, Ozesmi *et al.* 2006), focalizing here on ecological resources that determine the functional interactions between the organisms and their environment, instead of computing environmental surrogates, should allow more confidently transferring our main findings to populations established in similar areas (Vanreusel *et al.* 2006), at least at the Southern Belgian scale and most probably to other comparable Western European areas.

The Red-backed Shrike showed a mainly marked association with and selectivity for areas where suitable nest carriers are abundant (Nest EFD[1]) and diversified (NestStd EFD), especially when distant from human settlements (NestDist and DiUrban EFDs). Nest site selection and structural characteristics of the immediate nest environment have been shown having a

---

[1] We refer in the text to the important EFDs highlighted by the success-ENFA, which allow describing the habitat requirements of the Red-backed Shrike in the study area.

huge effect on the breeding performance of the Red-backed Shrike (Tryjanowski *et al.* 2000, Muller *et al.* 2005). Locally, the probability of finding an adequate nest carrier and of suitably concealing the nest obviously increases where the bushes or hedges density is high and their physiognomy is heterogeneous (Van Nieuwenhuyse 1998a, Muller *et al.* 2005). The Red-backed Shrike did not nest in the close proximity of urban areas (Kuzniak and Tryjanowski 2000), maybe because this EFD act as a surrogate for unmeasured resources, but also probably due to direct disturbance and to a potentially higher predation pressure by Black-billed Magpies (*Pica pica*) or cats (Muller *et al.* 2005). Note that this global assertion is absolutely not in contradiction with the interpretation proposed above about the minor role of predation in inducing fledging failure of the negative outsiders.

While less specialized regarding the subsequent aspects, the species was highly linked to sites where suitable perches for hunting were abundant, but particularly where their spatial arrangement provided a maximal foraging area on a minimal total surface (Hunt EFDs). This was already shown as a key factor for habitat occupancy in the case of the Great Grey Shrike (*Lanius excubitor*) by Rothhaupt and Klein (1998) in Southern Germany.

On the other hand, even if the Red-backed Shrike forages preferentially on low vegetation or bare soils (Van Nieuwenhuyse and Vandekerkhove 1992, Lefranc 2004), inversely weighting the ‘usable area’ by the mean vegetation height of each intersecting patch type (Huntvg EFDs) appeared not to be more informative. This is most probably due to the fact that (1) the relationship between hunting preference and vegetation height is not straightforward and (2) this hunting process happens at a finer spatial scale, involving unmapped landscape features, like paths side slopes or rocky outcrops. Nonetheless, the species was associated to high field margins density (Interface EFDs) and to alternation of high and low vegetation heights (Contrast EFDs), because this mosaic favours (1) the abundance of prey (e.g. Meek *et al.* 2002, Holland 2002, Backman and

Tiainen 2002, Pywell *et al.* 2005) and (2) their accessibility (Van Nieuwenhuyse 1998a, Van Nieuwenhuyse *et al.* 1999, Lefranc 2004). Boundaries between different herbaceous vegetation heights are particularly looked for when foraging (Van Nieuwenhuyse *et al.* 1999). The interspersion of tight and tiny patches of different vegetation heights (quantified here by Interface and Contrast EFDs) allows maximizing the heterogeneity on a minimal total surface (Van Nieuwenhuyse 1999).

Furthermore, the Red-backed Shrike did not settle down in high-intensity farming areas (Extensive EFDs) where ‘semi-natural’ patches were less abundant and scattered (SemiNat EFDs). This could be explained by the fact that invertebrate richness and abundance were considerably reduced in high-intensity farming areas (e.g. Kruess and Tscharntke 2002b), increasing parental-expenditure in terms of time devoted to hunting flight activity and energy allocation (Leugger-Eggiman 1997). Moreover, low prey density in high-intensity farming areas could induce the production of fewer or lighter nestlings, indicating a lower survival probability of offspring (Leugger-Eggiman 1997). Both consequences most probably considerably shape the habitat selection pattern of the species, but this reluctance was a bit less marked than other habitat requirements. Undoubtedly, more precise measures of reproductive performance would have revealed more clearly the impact of agricultural intensification on the Red-backed Shrike (see Leugger-Eggiman 1997).

Finally, soil moisture (SoilDry and SoilDepth EFDs) was on average slightly higher in occupied sites than in available ones, most probably because poorly drained soils enhance arthropod’s biomass or density, especially for large species of Orthoptera (e.g. Couvreur and Godeau 2000) and for Carabidae (e.g. Holland 2002). Soil moisture also influences agricultural practices and for this reason may act as a surrogate factor.

4.3. Implications for conservation

The required resources are related to the composition and the configuration of the farming landscape and are probably non-substitutable for the reproduction of the Red-backed Shrike. This latter statement is strengthened here by the fact that the significance of each functional type (nest, prey density and detectability, predation) was revealed by the ENFA, while no strong correlation was detected between these types. Hence, despite variations in their relative importance for explaining species marginality and specialization (Table 9), multiple resources appeared to be required for the successful reproduction of the species, hence for a population growth rate higher than 1. This stresses the importance of accounting for every aspect of the habitat when planning management or restoration.

The results presented above could be used to formulate conservation suggestions for improving the suitability of environmental conditions for the reproduction of the Red-backed Shrike in Western European farming areas. First, human disturbances and ensuing influences should be minimized by alleviating the extension of many villages and infrastructure installation in rural areas (DiUrban). The maintenance of any thorny feature (hedge fragments and bushes) along and inside grazed and mowed areas – even those that seem to be insignificant – is obviously of paramount importance (Nest and Hunt EFDs). Their spatial configuration (Hunt EFDs) and their structural diversity (NestStd) create a beneficial network of potential nesting and perching sites. This allows the Red-backed Shrike finding the most adequate site for concealing the nest. A distance of about 15-20 meters between suitable hunting perches was suggested by Van Nieuwenhuyse *et al.* (1999). This spatial arrangement creates an optimal system of perching sites that increases the accessible foraging area and the foraging efficiency by minimizing the energy expenditure. The spatio-temporal rotation in bushes and hedges pruning could ensure the maintenance of an adequate structural diversity (Van Nieuwenhuyse 1998a). The fragmentation of overgrown hedges could even prove useful in some cases (Van Nieuwenhuyse 1998a).

In the direct vicinity of such ligneous features, individuals require an alternation of tight and tiny open patches (Interface EFDs) of various vegetation heights (Contrast EFD), which simultaneously improves prey density and detectability. This particular requirement undoubtedly leads to the exclusion of the species from landscapes affected by re-allotment schemes because they induce a structural trivialization of the farming landscape. At more local scale, within mowed and grazed complexes, such spatial heterogeneity could be enhanced via phased and rotational mowing and grazing (Van Nieuwenhuyse 1998a).

Beside landscape spatial configuration, the composition of the open patches is crucial because the Red-backed Shrike is clearly linked to areas dedicated to a combination of hay production and/or livestock rearing using extensive farming techniques (Extensive and SemiNat EFDs). Such techniques favour the physiognomic heterogeneity and the floristic diversity of the herbaceous vegetation, therefore increasing prey densities. In areas where the conservation of the species is focused, unmanaged semi-open biotopes should be preserved and fertilizer application or grazing-mowing intensity should be limited in pastures and hay meadows. At this stage, precise threshold of fertilizer amount can not be proposed. Finally, soil wetness is less determining but still non negligible (SoilDry and SoilDepth EFDs), indicating that under-field drainage and filling of ditches should be avoided in order to maintain the abundance of particular prey taxa.

In conclusion, more attention needs to be paid to the wider environment than by creating protected reserves (Van Nieuwenhuyse 1999), in order to ensure the suitability of environmental conditions for the Red-backed Shrike in Western European man-shaped farming landscapes. With this end and within the Common Agricultural Policy framework, more initiatives (imposing rules or financial incentives) should be taken to promote less intensive agricultural techniques, hence encouraging farmers to take the advices of conservationists into account and to implement some of the proposed measures. In areas where management or restoration is focused, measures should be taken in order to simultaneously heighten all resource types that are required by the species.

Acknowledgements

We thank Jean-Baptiste Schneider for its technical help on the field. We are grateful to Frédéric Dermien, Norbert Lefranc, Marc Paquay and Jean-Yves Paquet for useful discussions. We thank the Institut Géographique National (Belgium) and the Walloon Region for providing us with the vector 1:10,000 maps and the aerial colour orthophotographs (Licence number 031205-1000), respectively.

Chapter 4 – Conspecific Proximity and Conservation of a Territorial Migratory Bird Species

Note – This study is to be submitted to Landscape Ecology (Titeux, N., Dufrêne, M. and Defourny, P. The Importance of Conspecific Proximity for the Conservation of Territorial Migratory Bird Species – The Case of the Red-backed Shrike). The text presented here is slightly modified from the paper to be submitted for layout and terminology harmonization. Some sections of 'Materials and method' are partly – but not fully – redundant with Chapter 2 and 3, and we therefore apologize to the reader.

Abstract

Localizing the sites where to target protection or restoration for a threatened species is a shared field of landscape ecology and conservation biology. Identifying the habitat requirements of individuals and delineating the boundaries of the habitats is needed but may be insufficient when habitat occupancy departs from habitat suitability. For planning relevant conservation measures, it becomes necessary to account for this decoupling. In this study, we first integrate breeding performance information within a niche-based species-habitat modelling framework to quantify this decoupling. Second, we achieve spatial analyses to test whether this decoupling was related to the neighbouring conspecific density. We used the Red-backed Shrike (*Lanius collurio*), a depleted migratory bird species breeding in Southern Belgium, as model species. In light of breeding performances, a considerable decoupling was observed between habitat suitability and occupancy. This decoupling was related to the neighbouring conspecific density, revealing the presumed use of direct (intrinsic suitability of the habitat) and indirect (conspecific attraction) cues during the settling and pairing phases. The implications of the observed pattern are discussed in a landscape context and from a conservation standpoint. Specific recommendations are suggested for enhancing the population management effectiveness and sustaining ecological processes at the landscape scale. The

maintenance of traditional agro-grazing practices is obviously of paramount importance, but management practices should be adjusted in light of the decoupling between habitat occupancy and suitability.

1. Introduction

Identifying the habitat requirements of the individuals of a species and the locations where to optimally target management measures is an ecological field situated at the convergence of landscape ecology and conservation biology (e.g. Hansson and Angelstam 1991, Burke 2000, Opdam *et al.* 2001, Swihart and Moore 2004, Guisan and Thuiller 2005). Delineating the suitable environmental conditions for a threatened species within a given landscape enable to focus planning efforts on worthy areas. Nevertheless, observed species spatial distributions are sometimes not directly superposable on the local habitat suitability (e.g. Pulliam and Danielson 1991, Dunning *et al.* 1992, Pulliam 2000, Lane *et al.* 2001, Schlaepfer *et al.* 2002). As a consequence, focusing conservation efforts on geographically delineating and protecting the suitable habitats may not suffice. Understanding the decoupling between habitat occupancy and suitability in a landscape context is necessary to elucidate observed distribution patterns (e.g. Turner 1989, Opdam *et al.* 2001) and to implement effective conservation strategies.

Such dissociations between habitat suitability and occupancy may originate from a variety of community or population processes (e.g. Pulliam 2000, Stamps 2001, Schlaepfer *et al.* 2002, Shochat *et al.* 2005, Stamps and Krishnan 2005). To make optimal breeding habitat selection decisions, individuals have to rely on cues predicting local breeding habitat suitability (e.g. Stamps 2001, Doligez *et al.* 2004). The settlement of individuals in an area depends on the intrinsic quality of the environmental conditions in terms of food abundance and accessibility, breeding sites availability, protection from predators or other direct cues that enhance growth, survival or offspring production (e.g. Wiens 1989b, Muller *et al.* 1997, Stamps 2001).

Individuals may take short-cuts and use indirect cues that integrate the effects of many environmental factors on somatic condition or reproductive performance, therefore revealing the effect of habitat suitability on fitness. This should be more powerful and parsimonious than using each factor separately (Stamps 2001, Danchin *et al.* 2001, Stamps and Krishnan 2005).

In migratory birds, strong time constraints on mating and breeding constitute a selective pressure that may have contributed to the evolution of such mechanisms (Stamps 2001, Doligez *et al.* 2004). Among these integrative cues, presence or density of conspecifics could be considered by individuals as a mirror of habitat suitability (e.g. Stamps 1988, Reed and Dobson 1993). This has been highlighted for several territorial songbirds like the Black-capped Vireo (Ward and Schlossberg 2004), the House Wren (Muller *et al.* 1997), the Collared Flycatcher (Doligez *et al.* 2004) or the Loggerhead Shrike (Etterson 2003).

Conspecific attraction is the tendency for individuals of a species to establish close to each other leading to an aggregated distribution (Stamps 1988, 2001). This behaviour may occur even in territorial birds (Stamps 1988, Muller *et al.* 1997). Dispersing or migrating individuals can use conspecifics as indicators of habitat suitability during the process of habitat selection (e.g. Serrano and Tella 2003, Alonso *et al.* 2004, Ward and Schlossberg 2004) and aggregate at already occupied areas, leaving other suitable patches unoccupied (Muller *et al.* 1997, Lane *et al.* 2001). If the attractiveness of a given area is heightened by conspecific presence or density, further individuals will settle down preferentially in this occupied area, irrespective of whether alternative areas of equal suitability exist elsewhere (Stephens and Sutherland 1999, Lane *et al.* 2001, Stamps and Krishnan 2005). Clusters might then develop like growing crystals, starting at one point and attaching additional molecules to the present ones (Van Nieuwenhuyse 2000b). Late individuals or those lacking experience in selecting breeding habitat may especially use such indirect cues (Stamps and

Krishnan 2005), while experienced settlers have already a knowledge about habitat suitability (Muller *et al.* 1997, Stephens and Sutherland 1999, Alonso *et al.* 2004).

Beside the non-occupancy of suitable habitats (hereafter called meaningless absences), some proportion of individuals could select sub-optimal or even unsuitable sites for breeding (hereafter called outsiders) because attracted by conspecifics (see e.g. Stamps 2001, Stamps and Krishnan 2005 for a theoretical framework and syntheses) even though high-quality habitats exist elsewhere in the landscape (a particular case of ecological trapping, see Kokko and Sutherland 2001, Schlaepfer *et al.* 2002, Kristan 2003 for further conceptual details). Such individual decisions could result from an unbalanced trade-off between intrinsic habitat features (*direct* cues) and the vicinity of conspecifics (*indirect* cues) (e.g. Stamps 1988). Accordingly, such socially-mediated habitat selection mechanisms may lead to aggregated species distributions that are not superposable on the spatial distribution of the suitable habitats. This uncoupling between habitat suitability and attractiveness is of particular conservation concern because affecting individual fitness and hence population productivity. Stamps and Krishnan (2005), in their recent conceptual synthesis, called for additional work to determine whether indirect cueing is a common feature of habitat selection.

We propose here a niche-based spatial modelling framework integrating pairing and breeding success data that enables to (1) highlight the decoupling between habitat occupancy and suitability and to (2) relate this decoupling to conspecific proximity. We use the Red-backed Shrike (*Lanius collurio*) as a model bird species in Southern Belgium. We particularly aim at revealing the relation between conspecific proximity on the one hand, and (1) the establishment of individuals in unsuitable environmental conditions or (2) the pairing success of males, on the other hand. Results are subsequently discussed in a landscape perspective and from a conservation viewpoint. Specific recommendations are suggested for enhancing the population management effectiveness at a landscape scale.

2. Materials and methods

2.1. Study area

The reader is referred to Chapter 2 and Figure 7 for the description of the study area.

2.2. Species data

The reader is referred to Chapter 2 and 3 for the description of species data collection. The species data set of Chapter 3 was used in this section (year 2005).

2.3. Environmental data

The reader is referred to Chapter 2 (Table 5) and 3 (Table 7) for the description of environmental data collection and the computation of the Environmental Functional Descriptors (EFDs), respectively. Their underlying ecological rationale is detailed in the Foreword of Chapter 2.

2.4. Statistical methodology

A first *a priori* screening discarded those cells that were completely wooded or with neither bush nor hedge because they were clearly unsuitable for Shrike settlement.

EFDs were standardized (mean=0 and variance=1) and normalized using the Box-Cox algorithm (Sokal and Rohlf 1998). Departure from normality after transformation was detected for some EFDs but the adopted ordination technique is not too sensitive to this assumption (Hirzel *et al.* 2002a).

The following steps enabled to (1) model the species-habitat relationship, (2) identify unpaired males and occurrences in unsuitable conditions, (3) delineate the environmental conditions occupied by these outsiders and (4) evaluate the role of conspecific proximity as a putative incitement during the settling and pairing phases (spatial analyses).

2.4.1. Ecological Niche Factor Analysis

The pattern of settlement of a Red-backed Shrike population can change considerably between successive years, with suitable sites not occupied each year (Van Nieuwenhuyse 2000b, Söderström 2001). Consequently, we assumed the population could not saturate all suitable sites within the landscape and absences were deemed unreliable indicators of unsuitable conditions, leading to severe limitations and biases for fitting classical predictive 'discrimination models' relying on presence-absence data (e.g. Guisan *et al.* 2002, Gu and Swihart 2004, Guisan and Thuiller 2005).

Methods based on presence-only data, like the Ecological Niche Factor Analysis (ENFA – Hirzel *et al.* 2002a), are particularly advisable in such a case (Hirzel *et al.* 2001, Pearce and Boyce 2005, Elith *et al.* 2006). This recent factorial approach relies on the ecological niche paradigm (Hutchinson 1957) and has proved to produce valuable results (e.g. Brotons *et al.* 2004, Hirzel *et al.* 2004). ENFA uses the distribution of species occurrences along the EFDs to summarise them into independent components (as Principal Component Analysis) that are related to the species ecological niche. The first component explains the 'marginality' of the species regarding the EFDs, describing how far the species optimum is from the mean environmental conditions within the whole study area. Next factors are then extracted orthogonally in order to explain the 'specialization' of the species, describing the narrowness of its niche (see Hirzel *et al.* 2002a and Chapter 2 for a detailed description of the method). ENFA was performed with BIOMAPPER 3.1 (Hirzel *et al.* 2002b).

2.4.2. Habitat Suitability modelling

The few first factors of the ENFA, gathering the majority of the information, were used to compute Habitat Suitability (HS[1]) values between 0 and 1 for all sites within the study area. The distance geometric-mean algorithm was used and the reader is referred to Chapter 2 for a thorough description of this distance-based algorithm.

2.4.3. Model evaluation

The lack of reliable absence data makes the evaluation of presence-only models difficult with classical methods related to confusion matrix (Pearce and Ferrier 2000, Boyce *et al.* 2002, Ottaviani *et al.* 2004). Two evaluation indices were computed. The Absolute Validation Index (AVI) was the proportion of occupied cells for which HS value was higher than 0.5 (Hirzel and Arlettaz 2003b). The Contrast Validation Index (CVI) was equal to the difference between the AVI and A_g, where A_g was the same as AVI but for the whole study area (Hirzel and Arlettaz 2003b). CVI ranges from 0 to 1 – A_g and reflects model accuracy (Hirzel *et al.* 2004), values near 0 indicating that the model does not outperform a random one (Hirzel and Arlettaz 2003b).

AVI and CVI were estimated through a cross-validation procedure (Manly 1997, Fielding and Bell 1997, Sokal and Rohlf 1998). The data set was partitioned into 20 subsets. In turn, 19 of these were used for model calibration and the remaining one for model evaluation (AVI and CVI computation). This procedure allowed calculating mean and standard deviation for both evaluation indices.

2.4.4. Outsiders – Designation rules

Pulliam (2000) underlined the importance of identifying individuals occurring outside the environmental bounds of the species ecological niche

---

[1] This acronym is equivalent to ENFA-HS used in Chapter 2.

(*sensu* Grinnell 1917), therefore not conveying reliable information about habitat suitability (outsiders). Based on this same Red-backed Shrike breeding success data set, Titeux *et al.* (2006c, Chapter 3) showed that 10% of occurrences with the lowest HS values (most distant from the species optimum) should be regarded as outsiders, but only towards the average global conditions along the marginality axis to remedy an artefact of the distance geometric-mean algorithm. Indeed, the disparity in the ratio of availability-to-occupancy of environmental conditions between both sides of the marginality factor was not appropriately taken into account by distance-based algorithms. They showed that, whatever their low HS values, uncommon environmental conditions on the positive end of the marginality axis were actually highly suitable. Occurrences in such conditions were therefore not considered as outsiders.

2.4.5. Outsiders – Environmental conditions

The environmental conditions occupied by the q designated outsiders were modelled in the hyperspace created above by the ENFA. The distance harmonic-mean algorithm (Hirzel and Arlettaz 2003a) was adopted and outsiders alone were used as occurrence data. Like the distance geometric-mean algorithm, it uses density of occurrences in the hyperspace to increase the influence of those that are close to each other, but gives more weight to individual occurrences and fits more closely to occurrence records. For each point P in the ENFA hyperspace, the density of outsiders was modelled by the harmonic mean H_H of its distances δ to the q outsiders O_i (Hirzel and Arlettaz 2003a), creating a multidimensional H_H field (see equation 2).

$$H_H(P) = \frac{1}{\frac{1}{q}\sum\limits_{\substack{i=1 \\ P \neq O_i}}^{q} \frac{1}{\delta(P, O_i)}} \qquad (2)$$

High values of H_H reflected low outsiders density in the hyperspace. Envelopes were then delineated by circumscribing all points in the hyperspace that had a lower H_H than a certain threshold. An 'outsiders-field' was thus created in the ENFA hyperspace. Several envelopes were defined enclosing different proportions of outsiders, from the innermost to the outermost part of the outsiders cloud. The enclosed cells constituted multiple data sets representing more or less closely the environmental conditions occupied by the outsiders. As stated by Hirzel and Arlettaz (2003a), this algorithm proves useful when 'the paucity of the sampling advocates for extracting information from each occurrence'. This was undoubtedly the case here due to the low amount of outsiders (10% of all occurrences) and this justifies its use.

Nevertheless, Titeux *et al.* (2006c, Chapter 3) showed that there was not necessarily a straightforward relationship between the HS values given to a cell by a distance-based algorithm and its departure from the average conditions regarding each EFD individually. Due to non-substitutable resources – like hunting areas, prey density or nesting sites availability – the Red-backed Shrike may be very sensitive to such univariate departure during the settling phase. Therefore, from the above created multiple data sets, univariate screening discarded cells for which at least one of the EFDs was farther away from the extreme values in the species distributions.

Finally, from the remaining cells, those that enclosed suitable conditions were discarded, i.e. either cells with HS values higher than the highest HS value among outsiders or cells on the positive end of the marginality axis whatever their HS values (to remedy algorithm artefact, see above).

After these successive screening processes, the ensuing multiple data sets were considered as representative of the environmental conditions occupied by the outsiders and not used by the reliable occurrences. This representativeness of the data sets depends on the envelope considered along the outsiders-field.

2.4.6. Outsiders and unpaired males – Spatial analysis

In the Red-backed Shrike, one potential incentive for settling down in unsuitable environmental conditions is the proximity of conspecifics. Moreover, the pairing failure for males could have been induced by their geographical isolation in the landscape. These hypotheses were here regarded as alternatives to the null hypothesis of randomness.

First, the territory density in the neighbourhood of each cell was quantified by computing several autocovariate terms (*Autcov*). The classically adopted method consists in counting, for a given cell, the number of occupied cells within a circular neighbourhood (e.g. Smith 1994, Augustin *et al.* 1996). However, as some territories could straddle two or three cells, this approach was somewhat adapted here to avoid overestimating *Autcov*. For each cell, they were calculated by counting the number of territories intersected by a circular area of given radius starting from the centre of the cell. As we aimed at quantifying the neighbouring density around each cell, a value of 0 or 1 was deducted from this count for unoccupied or occupied cells respectively (see Figure 16). This operation enabled not overestimating *Autcov* for occupied cells.

Because we could not determine *a priori* which radius was the most appropriate, multiple distances were used (150, 300, 450 and 600m). Larger distances were not investigated to avoid edge effect at borders of the study area. Actually, the species was surveyed within a 600m external buffer area but not further away. Such terms quantified the multi-scale conspecific density in the neighbourhood of any given cell.

Second, for the different radii, the mean values of *Autcov* among the *q* outsiders were computed for quantifying conspecific density in their vicinity. These mean values were hereafter called $Autcov_{o,d}$ (*o* referred to outsiders and *d* identified the radius).

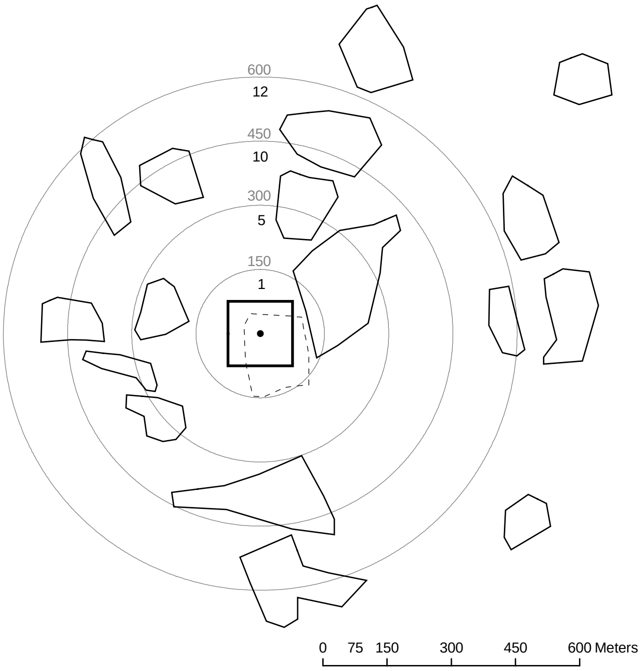


Figure 16. Illustration of the procedure used for calculating the autocovariates (*Autcov*). The focal cell and its centre are indicated by black square and point, respectively. The boundaries of Red-backed Shrike territories are shown in black. The concentric circular areas (and their corresponding radius values) are depicted in grey and centred on the focal cell. The number of territories intersected by each circular area is indicated. The territory covering the focal cell (dotted territory) was ignored for the computation, because we aimed at describing the conspecific density in the neighbourhood of the cell. This was achieved for each occupied and unoccupied cell of the study area.

Third, from each data set representing environmental conditions occupied by the outsiders, q cells were randomly sampled. The data sets differed in the envelope chosen along the outsiders-field. For each random sample of q cells and for the different radii, the mean values of *Autcov* were computed. These mean values were hereafter called $Autcov_{r,d,e}$ (r referred to random, d and e identified the radius and the envelope, respectively) but vary from one random sample to another within each data set. Therefore, this step was achieved for 999 random samples of q cells in every data set. For each combination of radius *d* and envelope *e*, these samples were used for describing the randomization distribution of $Autcov_{r,d,e}$.

Fourth, for each combination of radius *d* and envelope *e*, the $Autcov_{o,d}$ were compared with the distributions of $Autcov_{r,d,e}$. The proportion $\hat{p}_{d,e}$ of $Autcov_{r,d,e}$ that were as high as or higher than $Autcov_{o,d}$ in the distribution was calculated using the equation 3 (randomization tests, Manly 1997):

$$\hat{p}_{d,e} = \frac{m_{Aut\,\mathrm{cov}_{r,d,e} \geq Aut\,\mathrm{cov}_{o,d}}}{m_{Aut\,\mathrm{cov}_{r,d,e}} + 1} \qquad (3)$$

where *m* is the number of random samples complying with the condition stated in its index.

This proportion $\hat{p}_{d,e}$ can be regarded as a measurement of the significance level of classical one-sided null hypothesis tests. Therefore, randomization tests told whether the observed pattern in the data (here, such a high conspecific density in the vicinity of outsiders) was likely to have arisen by chance (Manly 1997).

Similarly, for the different radii, the mean values of *Autcov* among the *k* unpaired males were computed for quantifying conspecific density in their vicinity. These mean values were hereafter called $Autcov_{u,d}$ (*u* referred to unpaired males and *d* identified the radius) and were compared with the distribution of $Autcov_{ro,d}$ obtained by randomly selecting 999 samples of *k* cells among all occurrences (*ro* referred to random occurrences and *d* identified the radius).

In this case, null hypothesis was that unpaired males were randomly distributed among occurrences, irrespective of conspecific proximity. For each radius distance, $\hat{p}_d$ were the proportions of $Autcov_{ro,d}$ that were as small as or smaller than $Autcov_{u,d}$ in the distributions, using the equation 4:

$$\hat{p}_d = \frac{m_{Aut\,\mathrm{cov}_{ro,d} \leq Aut\,\mathrm{cov}_{u,d}}}{m_{Aut\,\mathrm{cov}_{ro,d}} + 1} \tag{4}$$

where m is the number of random samples complying with the condition stated in its index.

3. Results

The *a priori* screening of cells where land-use was thought to be incompatible whit the Red-backed Shrike's settlement and breeding discarded 480 cells (Figure 11). The remaining 1184 were used for subsequent analyses.

In 2005, 74 males settled down within the study area, of which 67 were paired with females. Territories were aggregated at the cell level, reaching a total of 110 occupied cells because of straddling territories. From the 67 pairs, 44 (65%) produced at least one fledgling at first (n=36) or second (replacement clutch, n=8) attempt (63 cells of the study area – Figure 11).

3.1. Ecological Niche Factor Analysis

ENFA computed global marginality and tolerance coefficients of 1.32 and 0.72 respectively, indicating that the Red-backed Shrikes settled down in environmental conditions highly different from the average of the global ones, but with a rather large niche breadth (Hirzel *et al.* 2002b). By comparing the eigenvalues to Mac-Arthur's broken-stick distribution (Jackson 1993, Hirzel *et al.* 2002a), 4 significant factors of the ENFA were retained for HS computation. They altogether explained about 72% of the information (100% of the marginality and 44% of the specialization). This means that 4 factors were sufficient to describe the niche of the species. The projection of the 1184 cells in the first two dimensions of the environmental hyperspace is illustrated on Figure 17A. From these factors, the first one was not the most important for explaining the specialization of the species; besides 100% of the marginality, this factor explained only 8% of the specialization (Table 8). This low fraction means that the combination of EFDs explaining the species marginality did not explain its specialization adequately.

The cross-validation procedure provided a mean AVI of 0.53 (SD = 0.20) and a mean CVI of 0.35 (SD = 0.20). These results indicated that the model prediction power was good but the difference between AVI and CVI showed that part of model performance could be attributable to randomness (Hirzel *et al.* 2004).

Positive coefficients on the marginality axis (Table 8) indicated that the Red-backed Shrikes were found in areas where these EFDs were higher than the average conditions (by decreasing order of importance, nest carriers abundance, distance to urban areas, suitable foraging areas and prey density). Negative coefficients indicated that the Red-backed Shrikes were established in areas where the soil dryness and the amount of very intensive pastures were slightly lower than average conditions. Niche breadth of the Red-backed Shrike appeared to be mainly restricted by nest carriers' abundance and by distance to human settlements.

3.2. Spatial analysis of outsiders

According to the designation rules, 12 cells were deemed as outsiders towards the global mean conditions (see Figure 17A). This discarding rate was statistically founded on the one hand (see Chapter 2) and based on the observed dramatic decline of breeding success for these occurrences on the other hand (see Figure 14, Chapter 3), actually revealing unsuitable environmental condition for the reproduction of the species (see breeding success on Figure 17A).

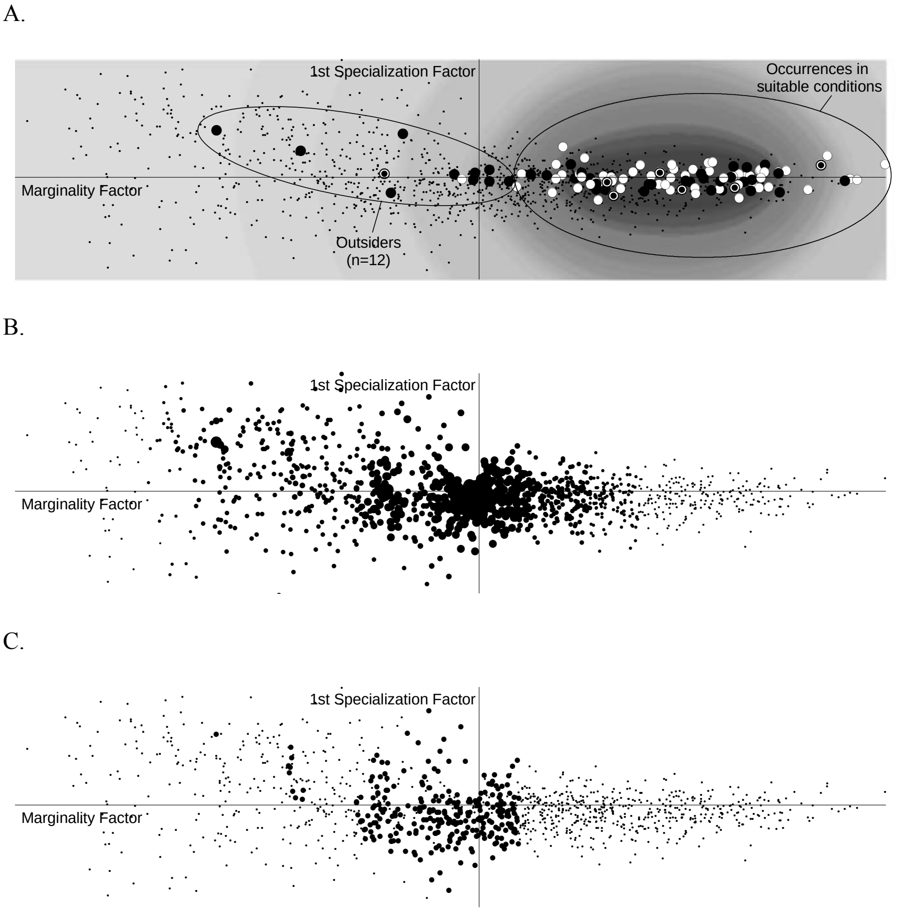


Figure 17. (A) Occurrences plotted in the 2D environmental space represented by the first two factors of the ENFA. Successful-, unsuccessful pairs and unpaired males are represented by white dots, black dots and encircled black points, respectively. Little black points are other cells of the study area. The HS field modelled by the distance geometric-mean algorithm is represented in 2D (the darker, the higher the HS values). (B) 'Outsiders-field' modelled by the distance harmonic-mean algorithm. The size of the black bubbles is proportional to the similarity with environmental conditions occupied by outsiders. (C) Distribution of the cells representing environmental conditions occupied by outsiders and not by reliable occurrences, after the successive screening processes (bold black points), for a moderately restrictive envelope (e=4, Table 10). Other cells of the study area are depicted by little black points.

Multiple data sets were created by enclosing cells that displayed similar environmental conditions than the outsiders (see ‘outsiders-field’ on Figure 17B). Nine concentric envelopes delineated within this field were created by the distance harmonic-mean algorithm; from envelope 1 to 9, they were increasingly restrictive regarding these conditions, so that they enclosed decreasing numbers of cells (see an example on Figure 17C).

These data sets were subsequently used for describing the ‘randomization distribution’ of *Autcov* among cells enclosing those environmental conditions ($Autcov_{r,d,e}$). Randomization tests allowed computing a measurement of the strength of evidence (equation 3) that the observed conspecific density in the vicinity of the outsiders ($Autcov_{o,d}$ in Table 10) was not so high by mere chance. The lower the proportion $\hat{p}_{d,e}$, the higher the evidence that the null hypothesis (chance) was not true. Results (Table 10) show that there was some increasing evidence against null hypothesis from small to large radius distances, indicating that outsiders were settled down in unsuitable environmental conditions but where the conspecific density in the non-direct vicinity was higher than by chance only. Randomization tests indicate that the value of $Autcov_o$ for the 600 m radius (4 neighbouring conspecifics in average) has most probably not arisen by mere chance (alternative hypothesis). Results are more equivocal for the 450 m radius. On the other hand, there is no evidence that conspecific presence in the most direct vicinity (within a radius of 150 or 300 m) of outsiders has not arisen by chance. While the effect of radius distance on p did not really differ according to the chosen envelope, the just above mentioned global conclusions did not hold for envelopes 8 and 9, probably because they were too restrictive and did not capture the environmental conditions occupied by most of the outsiders.

Table 10. Results of the randomization tests for the outsiders (n=12 / 999 random samples).

	Radius d (m)			
	150	300	450	600
$Autcov_{o,d}$	0.25	1.08	2.25	4
$\hat{p}_{e=1}$	0.318	0.105	0.043 *	0.016 *
$\hat{p}_{e=2}$	0.271	0.104	0.034 *	0.009 **
$\hat{p}_{e=3}$	0.265	0.093	0.022 *	0.013 *
$\hat{p}_{e=4}$	0.302	0.104	0.046 *	0.009 **
$\hat{p}_{e=5}$	0.370	0.175	0.066	0.018 *
$\hat{p}_{e=6}$	0.360	0.151	0.056	0.024 *
$\hat{p}_{e=7}$	0.351	0.156	0.046 *	0.023 *
$\hat{p}_{e=8}$	0.453	0.297	0.134	0.109
$\hat{p}_{e=9}$	0.414	0.271	0.143	0.140

Tests were achieved for various combinations of radius and envelopes (from e = 1 to 9, envelopes are increasingly restrictive regarding environmental conditions occupied by outsiders). The Red-backed Shrike's mean territory area corresponds to a radius of about 80 m. For each radius d, $\hat{p}$ is the proportions of values in the 'randomization distribution' that were as high as or higher than $Autcov_{o,d}$ (mean $Autcov_d$ for outsiders, see equation 3) and reflects the significance level of a classical one-sided null hypothesis tests (* significant, ** highly significant).

3.3. Spatial analysis of unpaired males

A total of 7 unpaired males (about 10%) were found within the study area in 2005 (Figure 11). They were homogeneously distributed in the ENFA hyperspace, most of which in suitable environmental conditions (Figure 17A). Randomization tests allowed computing a measurement of the strength of evidence that unpaired males were not randomly distributed among occurrences, regarding the conspecific density in their vicinity. The proportions $\hat{p}_d$ of $Autcov_{ro,d}$ that were as small as or smaller than $Autcov_{u,d}$

in the distributions were computed for each radius distance *d* (see equation 4). Results clearly show that there was strong evidence against null hypothesis (Table 11), indicating that unpaired males were established where the conspecific density in the direct and non-direct vicinity was lower than by chance only, as compared with the 'randomization distribution' of neighbouring conspecific density for all occurrences.

Table 11. Results of the randomization tests for the unpaired males (n=7 / 999 random samples).

	Radius *d* (m)			
	150	300	450	600
$Autcov_{u,d}$	0	0.14286	0.85714	1.71429
$\hat{p}$	0.001 **	0.001 **	0.002 **	0.001 **

The Red-backed Shrike's mean territory area corresponds to a radius of about 80 m. For each radius *d*, $\hat{p}$ is the proportions of values in the 'randomization distribution' that were as small as or smaller than $Autcov_{u,d}$ (mean $Autcov_d$ for unpaired males, see equation 4) and reflects the significance level of a classical one-sided null hypothesis tests (* significant, ** highly significant).

4. Discussion

Ecological Niche Factor Analysis revealed the importance of suitable nesting sites availability, suitable foraging areas, distance to urban areas and prey density for the settlement and the breeding of the Red-backed Shrike. As pure habitat requirements for this species are not the main the scope of this paper, they are not discussed here in more details and we refer the interested reader to Titeux *et al.* 2006c (Chapter 3) for a thorough description of the species habitat requirements.

In light of breeding success information, ENFA illustrated that the observed pattern of the species distribution departed from that of the suitable habitats distribution; some individuals most probably occurred outside the environmental bounds of the species niche and were designated as outsiders (see their breeding success on Figure 17A). Furthermore, most of the unpaired males were established in suitable environmental conditions (see Figure 17A), revealing that the settlement in locally suitable habitats was not straightforwardly rewarded by pairing and mating success. This global pattern highlighted a decoupling between habitat suitability, habitat occupancy and pairing or breeding success.

Randomization testing showed here that the existence of outsiders and unpaired males in the Red-backed Shrike is partly related to the spatial distribution of the population. In particular, the conspecific density in a given area seems to be a significant incitement for settling and pairing decisions. Indeed, individuals that occurred in unsuitable environmental conditions (outsiders) were established where neighbouring conspecific density was higher than expected by chance only (this does not hold for the close conspecific proximity). Furthermore, whatever the suitability of the environmental conditions, unpaired males were established where the neighbouring conspecific density was lower than by mere chance (as compared with all occurrences). Nevertheless, these results need to be confirmed from one year to another.

Neighbouring conspecific density beyond a 600 meters radius was not explored here to avoid edge effects because the species was not surveyed outside a 600m external buffer area. However, future works should investigate larger radius distances and the ensuing behaviour of the randomization tests outcomes because a question remained unanswered about up to which distance the neighbouring conspecific density was influent. This could have been achieved with this dataset by using only the innermost part of the study area. However, numerous occurrences located at the edge of the study area (see Figure 11) would have been lost for these spatial analyses.

Hereafter, we discuss the general observed pattern in light of breeding ecology of the Red-backed Shrike and in a landscape context. We suggest some proximate and ultimate factors that could explain this pattern and we emphasize their implications for the conservation of this species at the landscape scale. Nevertheless, the genuine understanding of the mechanisms involved in settling and pairing decisions is out of the scope of this study. Formally identifying critical processes underlying the distribution of a rare species is challenging and requires a huge experimental design (see Doligez *et al.* 2004 for the Collared Flycatcher). Experimental approaches such as habitat, breeding success or conspecific attraction manipulation are needed because they can provide straight insights into processes affecting some populations (but see Wiens *et al.* 1986). Accordingly, the following discussion is partly speculative in comparison to the results of this study, but stresses some needed additional research directions.

4.1. Indirect consequences of conspecific proximity

A first explanation for conspecific proximity is the use of conspecific presence or density as indicators of habitat suitability – hence potential breeding success – during the process of habitat selection (conspecific attraction, see e.g. Stamps 1988, Reed and Dobson 1993, Serrano and Tella 2003, Alonso *et al.* 2004, Ward and Schlossberg 2004). Because dispersing or migrating individuals may aggregate at already occupied areas, a possible consequence is the settlement of some late or inexperienced individuals in sub-optimal or unsuitable areas near conspecifics (see Stamps 2001), inducing a spatially-structured decoupling between habitat suitability and occupancy. Our results indicated that outsiders were established in globally occupied areas, but not necessarily in the direct proximity of conspecifics. The breeding success of these individuals was affected by unsuitable environmental conditions (see Figure 17A), reducing the global productivity of the population (ecological trapping – see Schlaepfer *et al.* 2002, Shochat *et al.* 2005, Stamps and Krishnan 2005). Titeux *et al.* (2006c, Chapter 3) showed that these occupied but unsuitable areas lacked one or few essential components of habitat requirements for the reproduction of the species. As these diverse requirements are probably non-substitutable resources (see also Chapter 5), shortage of one of these could have induced the observed breeding failure.

While the proximate mechanisms remain to be explored, our results stress the need to account for their repercussion when implementing conservation initiatives. Indeed, owing to the current anthropogenic rural landscape alteration (especially since the foundation of the Common Agricultural Policy (CAP) in the 1950s), suitable sites with increasingly high edge-to-area ratio should especially suffer from the ecological trapping (see also Chapter 5). The fragmentation of suitable habitats in a landscape should expose individuals to settle down in unsuitable environmental conditions (ecological trapping), therefore involving a global decline in the population growth rate. A spatially-explicit modelling approach remains to be designed for identifying the discrepancies between habitat suitability and likelihood of occupancy within a landscape (see Chapter 5).

In the face of widespread loss of suitable habitats, populations within extensively fragmented rural landscapes may become non self-sustaining below a given threshold of suitable areas (Donovan *et al.* 1995, Trine 1998). It is not utopian to hypothesize their persistence may then depend on the immigration from landscape sources through a regional source-sink dynamics (metalandscape connectivity – With *et al.* 2006). Nevertheless, future works should integrate the temporal dimension within a multi-scale modelling framework to gauge the veracity of such speculations.

4.2. Direct consequences of conspecific proximity

A second explanation for conspecific proximity is the direct research of an ultimate fitness advantage (Muller *et al.* 1997, Stephens and Sutherland 1999, Stamps 2001, Alonso *et al.* 2004, Stephens *et al.* 2005). Fitness of

Red-backed Shrike could increase with conspecific density (*Allee* effect) through several ways.

First, males could sing and call in slack flocks for increasing their chances of attracting and finding a mate, because facilitating the detection of occupied areas by prospecting females (Van Nieuwenhuyse 1998a, 2000b). Consequently, the relative – but not necessarily direct – proximity of conspecifics might be a very important cue for potential breeding performance assessment in males of Red-backed Shrike (Van Nieuwenhuyse 2000b) and might explain the non pairing of too isolated males in the landscape, as observed here.

Second, aggregation of males could enhance extra-pair copulations (EPC) chances (Van Nieuwenhuyse 1998a, 2000b), which is evolutionarily advantageous for both sexes (Griffith *et al.* 2002). In this specific case, the assumption that EPC probability enhancement could stimulate individuals to occupy unsuitable conditions is less supported by our results, because outsiders are not particularly settled down in the most direct vicinity of conspecifics, which would be required for achieving EPC.

5. Conclusion

Simply focusing conservation efforts on local habitat requirements may prove to be insufficient for an effective management of the Red-backed Shrike due to the socially-induced dissociation between the suitability and the attractiveness of a given site. Actually, the designed niche-based modelling framework and spatial analyses showed that the aggregated species distribution was not a straight mirror of the spatial variation in intrinsic habitat suitability, most probably because (1) males are attracted by conspecifics as well as by habitat features or resources while settling down in an area and (2) females are more likely to be attracted by aggregated males when prospecting for pairing. Nevertheless, an experimental approach is needed for providing empirical support to such interpretations of the observed pattern. Proximate behavioural mechanisms (processes) driving the settling and pairing decisions should be further investigated. Their consequences (decoupling between habitat suitability and occupancy) on the species distribution should be explicitly considered when planning protection or restoration. This integration is necessary for adequately managing the pattern of future landscapes in order to sustain the required ecological processes (see also Turner 1989, Opdam *et al.* 2001). More attention needs to be paid to the wider environment than by creating small and isolated protected reserves (Van Nieuwenhuyse 1999). The maintenance of traditional agro-grazing practices is not only of paramount importance, but conservation strategies in fragmented landscapes should also focus on alleviating ecological trapping, by maintaining or restoring a network of large and compacted suitable areas, particularly around traditionally occupied areas as a priority.

Acknowledgements

We thank Jean-Baptiste Schneider and Julien Radoux for their technical help on the field and with GIS software, respectively. We are grateful to Alexandre H. Hirzel and Marie-Aline Wibrin for useful discussions and for commenting on a draft of the manuscript, respectively. We thank the Institut Géographique National (Belgium) and the Walloon Region for providing us with the vector 1:10,000 maps and the aerial colour orthophotographs (Licence number 031205-1000), respectively.

Chapter 5 – Integrative Spatial Modelling of Species Distribution when Habitat Occupancy Departs from Suitability

Note – This study is to be submitted to Ecological Modelling (Titeux, N., Dufrêne, M. and Defourny, P. Identifying breeding habitat requirements and selecting conservation-interest sites for the Red-backed Shrike (*Lanius collurio*) in Southern Belgium – A functional and spatial habitat modelling approach). The text presented here is slightly modified from the paper to be submitted for layout and terminology harmonization. Some sections of 'Materials and method' are partly – but not fully – redundant with Chapter 2, 3 and 4, and we therefore apologize to the reader.

Abstract

Identifying the key habitat requirements for a threatened species and delineating the boundaries of the habitats is increasingly needed for conservation. Fine-scale habitat-based predictive models are valuable tools for providing such information and for guiding protection, management or restoration planning. Nevertheless, in many specific cases, spatially-structured intrinsic population-based processes may induce a considerable decoupling between habitat suitability and occupancy. Evaluating and locating such distortions is of paramount importance and overlooking them may be counter-productive as far as conservation is concerned. The Red-backed Shrike is a depleted species of man-shaped farming landscape throughout its Western European range. Currently, conservation professionals need precise and quantified information about the main forces driving the distribution of this species for implementing effective conservation initiatives. We developed here a phased approach for assessing habitat suitability-occupancy discrepancy for this species in Southern Belgium. Occurrences not conveying reliable information about habitat suitability were first identified and discarded. Multi-model inference was then used for building fine-scale functional habitat suitability (HS) model.

Discarded occurrences were subsequently reintegrated and sites that were the most likely to be occupied were defined by accounting for the spatial structure in the data, thanks to a spatially-explicit modelling approach. The HS model allowed identifying and geographically delineating the fine-scale key resources required by the Red-backed Shrike, indicating that its main habitat requirements were related to (1) both composition and spatial configuration of the farming landscape and to (2) multiple functional resources mainly regarding nesting sites availability, suitable foraging areas, prey density and predation pressure. This stressed the importance of accounting for every aspect of habitat requirement when planning management or restoration. Furthermore, drawing a parallel between suitability and likelihood of occupancy enabled to locate unsuitable sites that were still expected to be occupied. Such inconsistency assessment allowed prioritizing sites that require protection and/or restoration measures. Finally, besides specific conservation implications, the proposed integrative methodological framework could be transferable to any biological data that are collected for inventory purpose and gathered haphazardly.

1. Introduction

Fine-grained information on the distribution of species and their suitable habitats is increasingly needed for conservation and management planning (e.g. Rushton *et al.* 2004, Guisan and Thuiller 2005, Seoane *et al.* 2005). Direct observations of species occurrences provide part of this information but are not sufficient (Scott *et al.* 2002, Elith *et al.* 2006), especially for most of needed management decisions (e.g. Seoane *et al.* 2005). Predictive models can be valuable tools for supplying more useful and precise information (see Scott *et al.* 2002, Swihart and Moore 2004, Rhodes *et al.* 2006, Elith *et al.* 2006), which in turn allows making decisions about restoration or management strategies (Johnson *et al.* 2004, Vanreusel *et al.* 2006), but also designing reserves (Araujo and Williams 2000, Cabeza *et al.* 2004, Tole 2006) or implementing strategic environmental impact assessment programs (Gates *et al.* 1993, Chamberlain and Fuller 2001). Most commonly adopted methods investigate the relationships between the species and its environment (Heglund 2002) and generate 'Resource Selection Functions' (e.g. Boyce and McDonald 1999, Guisan and Zimmermann 2000, Zaniewski *et al.* 2002, Johnson *et al.* 2004, Guisan and Thuiller 2005). RSF enable to identify the key habitat requirements of the individuals and to geographically delineate the most suitable conditions within a given landscape (Habitat Suitability maps).

The Red-backed Shrike (*Lanius collurio*) is a depleted species throughout its Western European range (BirdLife International 2004). This global decline is mainly attributable to recent human development and agricultural intensification (review by Lefranc and Worfolk 1997). In Southern Belgium, the Red-backed Shrike occupies a rather flexible range of semi-open biotopes mainly in farming areas (see Van Nieuwenhuyse and Vandekerkhove 1992 for a typology attempt). While the habitat requirements for this passerine are quite qualitatively understood, we still need more precise information about the relative roles of various resources in driving its fine-scale spatial distribution. Nevertheless, due to the flexibility in the habitat selection pattern of this species, detecting its key resource requirements becomes challenging. A predictive modelling approach could therefore provide valuable directions for implementing efficient conservation initiatives for this species.

Besides environmental forcing, the spatial distribution of a species may originate from intrinsic population-related processes (Legendre 1993). Such processes are often spatially-structured and should not be left out of consideration (Lichstein *et al.* 2002). Instead, explicitly introducing the spatial structures into ecological models is necessary for assessing and quantifying the relative roles of environment and intrinsic processes in structuring distribution patterns (Keitt *et al.* 2002, Legendre *et al.* 2002). In the Red-backed Shrike, conspecific attraction may lead to social aggregation

(Van Nieuwenhuyse 2000b), that could subsequently induce a considerable distortion between local habitat suitability and occupancy (e.g. Schlaepfer *et al.* 2002, Shochat *et al.* 2005, Stamps and Krishnan 2005). Individuals established in unsuitable environmental conditions were shown to be located near other individuals, highlighting the probable importance of conspecific proximity during the settlement (Titeux *et al.* 2006a, Chapter 4). On the one hand, such individuals (hereafter called outsiders) are unreliable indicators of habitat suitability and should be considered as such before delineating pure Habitat Suitability (HS) maps. On the other hand, they still convey vital information about the spatial distribution pattern of the species and accounting for them when evaluating the likelihood of occupancy within a given site should lead to build more plausible Expected Occupancy (EO) maps. Moreover, drawing a parallel between HS and EO maps and localizing the main discrepancies should be highly informative for designing efficient conservation strategies, because reflecting the decoupling between habitat suitability and occupancy.

Most of the predictive models relate species occurrences to coarse-scaled environmental variables (e.g. Gates *et al.* 1993, Atkinson *et al.* 2002, Engler *et al.* 2004, Brotons *et al.* 2004), most often including topography, climate or surrogate predictors without any direct ecological foundation. Such approaches are useful for describing large-scale distribution pattern but suffer from multiple criticism (Guisan and Thuiller 2005). First, such predictors are not easy to handle because not under the influence of local management or restoration actions (Guisan and Thuiller 2005, Vanreusel *et al.* 2006), leading the models to be of poor direct applicability for managers. Second, these surrogates have no straight ecological relation to the causal factors and provide little insight into the functional processes underlying species-environment patterns (Austin 2002, Guisan and Thuiller 2005). Third, correlations between surrogate and causal factors may be area-specific, leading models to be spuriously overfitted to the local conditions

(Luoto *et al.* 2002) and hence to be weakly transferable to other areas (Van Horne 2002, Vanreusel *et al.* 2006).

At a fine spatial scale it seems essential to focus on ecological resources determining the functional interactions between organisms and their environment (Tischendorf 2001, Vos *et al.* 2001, Vaughan and Ormerod 2003). Besides direct applications regarding management and restoration planning, resource-based models are more likely to be transferable to other areas in the same eco-region where resource availability and use are similar, because not revealing spurious correlations (Vanreusel *et al.* 2006). Furthermore, investigating the causal species-environment relationships should allow dealing with the flexibility of the habitat selection pattern of the species. It becomes therefore necessary to compute ecologically-founded functional predictors reflecting the fine-scale resources that are common to this variety of occupied habitats. On the other hand, the information about the distribution of functional resources is rarely readily available and its acquisition may be time-consuming (Seoane *et al.* 2005), hence limiting the spatial extent of the study area. However, Vanreusel *et al.* (2006) showed that such an ecological approach allows building accurate predictive models that are transferable to larger areas and with wide conservation implications.

Finally, the geographical pattern of settlement of a Red-backed Shrike population can change considerably between successive years (Van Nieuwenhuyse 2000b, Söderström 2001). Consequently, we assumed the population does not saturate all suitable sites within the landscape and absences were deemed meaningless, leading to severe limitations and biases for fitting classical predictive 'discrimination models' relying on presence-absence data (like Generalized Linear models, GLM) because they assume that they respectively indicate suitable and unsuitable environmental conditions (e.g. Hirzel *et al.* 2001, Guisan *et al.* 2002, Gu and Swihart 2004, Engler *et al.* 2004, Brotons *et al.* 2004, Guisan and Thuiller 2005).

The first objective of the present study is to identify and to geographically delineate the fine-scale key functional resources required by the Red-backed Shrike, a flexible species for which absences are meaningless and occurrences may be spatially-structured and unreliable indicators of habitat suitability. The second objective is to account for the spatial heterogeneity of the environment and the spatial structure of the species distribution to define the sites that are the most likely to be occupied within a given landscape. The third objective is to locate the main discrepancies between suitability and likelihood of occupancy and to assess their implications for the conservation of the species. Obviously, the specific issues detailed above regarding the Red-backed Shrike have a wider range of application than for this specific-case only. Accordingly, the integrative modelling framework was designed in order to be applicable for any biological data that are concerned by such issues, particularly for those that are collected for inventory purpose and gathered without ad-hoc sampling design.

2. Materials and method

2.1. Study area

The reader is referred to Chapter 2 and Figure 7 for the description of the study area.

2.2. Species data

The reader is referred to Chapter 2 for the technical details about description of species data collection. Unlike in previous Chapters, we here used species data gathered during May-July 2003 (non-exhaustive survey), 2004 and 2005 (exhaustive surveys).

According to the procedure detailed in Chapter 2, a value of 0 or 1 was assigned to each cell for absence or presence during 2003-2005. Because of breeding site fidelity in the Red-backed Shrike (Van Nieuwenhuyse 2000b), the occupancy frequency over years is not a direct mirror of habitat suitability. Therefore, cells that were occupied during two or three years did not count double or triple in the occurrence set.

2.3. Environmental data

The reader is referred to Chapter 2 (Table 5) and 3 (Table 7) for the description of environmental data collection and the computation of the Environmental Functional Descriptors (EFDs), respectively. Their underlying ecological rationale is detailed in the Foreword of Chapter 2.

2.4. Statistical methodology

A first *a priori* screening discarded those cells that were completely wooded or with neither bush nor hedge because they were clearly unsuitable for Shrike settlement. The remaining ones (n=1184) were used for subsequent analyses.

All EFDs were standardized (mean=0 and variance=1) and normalized using the Box-Cox algorithm (Sokal and Rohlf 1998). Departure from normality after transformation was detected for some EFDs but subsequently adopted ordination technique is not too sensitive to such violation (Hirzel *et al.* 2002a).

The occurrence set was randomly split into two subsets. A *training set* (two third of occupied cells during 2003-2005) and an *evaluation set* (one third of occupied cells during 2003-2005) were respectively used as basis for models calibration and evaluation. The following steps of the modelling procedure are reported on the flow chart in Figure 18. This procedure is divided into two main parts: Habitat Suitability (HS, steps 1 to 8) and Expected Occupancy (EO, steps 9 to 16) modelling. HS modelling intended to delineate the spatial distribution of suitable environmental conditions (ENFA-GLM combination, Chapter 2) by accounting for the existence of outsiders (outsiders' designation, Chapter 3). EO modelling integrated the spatial structure of the data (see Chapter 4) to delineate the likely to be occupied cells across the study area.

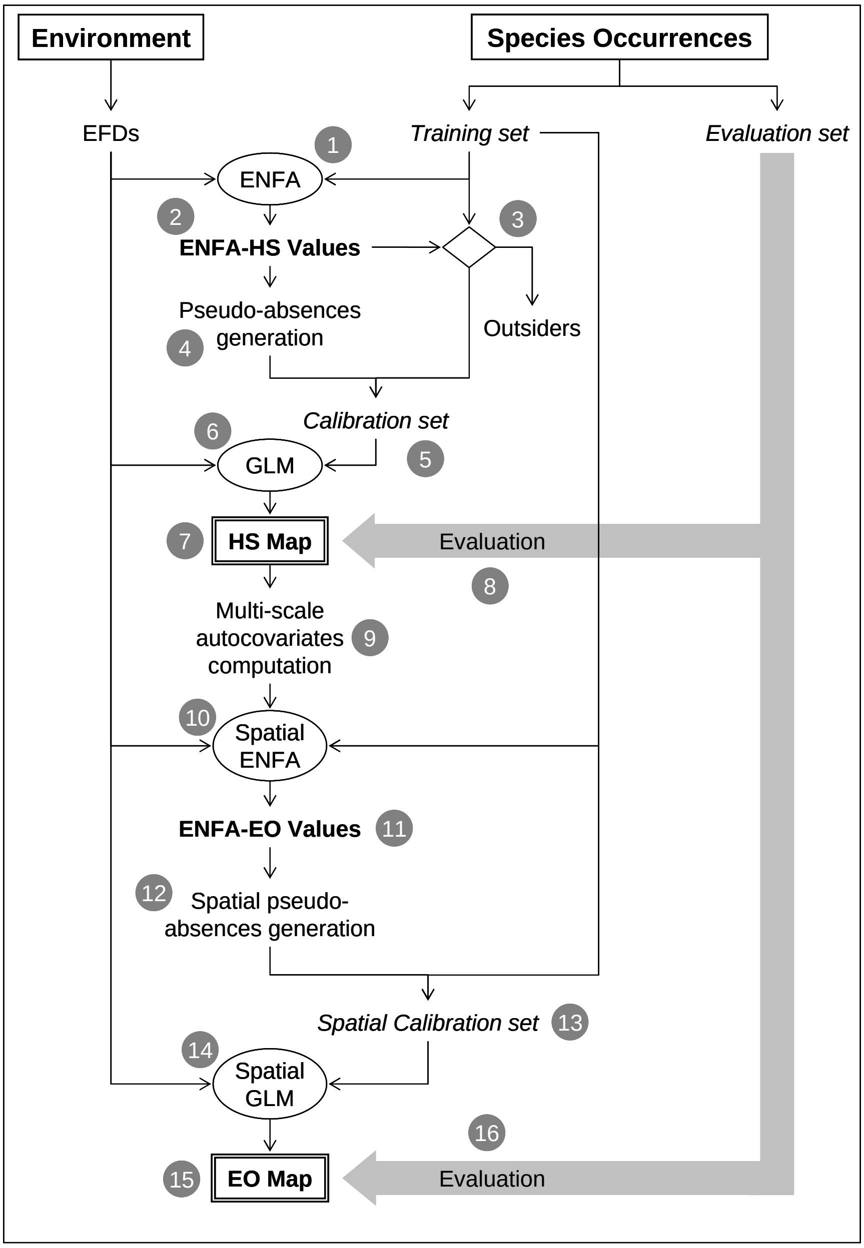


Figure 18. Flow chart of the Habitat Suitability (HS) and Expected Occupancy (EO) modelling framework. Steps 1 to 16 are described in details in the following sections.

2.4.1. Habitat Suitability modelling

2.4.1.1. Ecological Niche Factor Analysis (ENFA)

The lack of reliable absence data advocates here using ‘envelope methods’ based on presence-only data (Hirzel *et al.* 2001, Pearce and Boyce 2005, Elith *et al.* 2006).

The Ecological Niche Factor Analysis (ENFA – Hirzel *et al.* 2002a) is a recently developed approach providing improvements on other similar techniques (see Pearce and Boyce 2005). ENFA relies on the ecological niche paradigm (Hutchinson 1957). This ‘envelope approach’ is like a Principal Component Analysis (PCA) of cells where principal components are ecologically meaningful. The first one explains the marginality of the species regarding the EFDs, describing how far the species optimum is from the mean environmental conditions within the whole study area. Next factors are then extracted orthogonally in order to explain the specialization of the species, describing the narrowness of its niche (see Hirzel *et al.* 2002a and Chapter 2). ENFA was performed on the training set[1] using BIOMAPPER 3.1 (Hirzel *et al.* 2002b) (Figure 18, step 1).

The few first factors of the ENFA generally gather the majority of the information and are used to compute Habitat Suitability values (ENFA-HS) between 0 and 1 (Figure 18, step 2). Each cell obtained an ENFA-HS value that was proportional to the distance between its position in the factorial space and the position of the species optimum (distance-based algorithms). The distance geometric-mean algorithm was used as a good trade-off between precision and generality for computing ENFA-HS values (Hirzel and Arlettaz 2003b and see Chapter 2).

---

[1] ENFA were also performed on 2004 and 2005 data sets separately to check for year effect on ENFA-HS models. These yearly-fitted models were not used in the subsequent modelling procedure.

2.4.1.2. Identification of outsiders and generation of pseudo-absences

For several reasons, ENFA tends to provide over-optimistic ENFA-HS values and less accurate predictions than classical 'discrimination approach' like GLM (e.g. Hirzel *et al.* 2001, Engler *et al.* 2004, Brotons *et al.* 2004, Olivier and Wotherspoon 2005, Pearce and Boyce 2005 and see Titeux *et al.* 2006b for the Red-backed Shrike specifically). Overestimating areas of suitable habitats may obviously have highly undesirable consequences when these HS maps are used for specific resource management or for designing protection measures (Loiselle *et al.* 2003).

When absence data are unavailable or unreliable, one can still build 'discrimination models' by generating pseudo-absences (Zaniewski *et al.* 2002, Engler *et al.* 2004, Pearce and Boyce 2005, Elith *et al.* 2006). The followed methodology for selecting pseudo-absences is important and can affect the quality of the final model (Boyce *et al.* 2002, Olivier and Wotherspoon 2005).

Engler *et al.* (2004) proposed a strategy of ENFA-weighted pseudo-absences generation that reduces the chance of selecting absences where the species really is (but not detected) or where the conditions are suitable even if the species is absent. According to their methodology, an envelope enclosing all occurrences in the ENFA factorial space is delineated. Pseudo-absences are then randomly chosen outside this envelope and combined with real occurrences for GLM calibration. In order to give the same importance to occurrences and pseudo-absences in GLM, the number of randomly-chosen pseudo-absences is the same as the total number of real occurrences (Liu *et al.* 2005).

Pulliam (2000) underlined the importance of accounting for occurrences not conveying reliable information about local habitat suitability. Titeux *et al.* (2006b,c – Chapter 2 and 3) showed that such outsiders could hugely skew the pseudo-absences generation. Based on breeding success data on the Red-backed Shrike, they suggested considering outsiders as 10% of occurrences with the lowest ENFA-HS value (most distant from the species optimum), but only towards the average global conditions along the marginality axis to remedy an artefact of the distance geometric-mean algorithm. The reader is referred to Chapter 3 for technical considerations. Outsiders were deemed uninformative and were temporarily discarded from the training set (Figure 18, step 3). Pseudo-absences were subsequently randomly chosen among cells below the same ENFA-HS threshold as that used to identify the outsiders (Figure 18, step 4, and see Chapter 2 for similar approach). They were combined with and chosen in equal number than the remaining informative occurrences, providing the *calibration set* for GLM (Figure 18, step 5).

2.4.1.3. Model building – Information-theoretic approach

A binomial distribution and a logit link-function was specified (logistic regression model, Hosmer and Lemeshow 2000) to allow describing the relationship between the response variable (presence/pseudo-absence) and the EFDs (Figure 18, step 6).

Automatic statistical procedures for selecting explanatory variables during model building have been widely criticized and are currently viewed as heretical by numerous wildlife biologists, notably because they are naive and create spurious effects (e.g. Mac Nally 2000, Burnham and Anderson 2002). Recently, an Information-Theoretic Approach has been suggested as an interesting alternative offering many advantages, especially when multiple hypotheses are plausible (e.g. Burnham and Anderson 2002, Johnson and Omland 2004, Stephens *et al.* 2005, Welch and MacMahon 2005, Greaves *et al.* 2006). This recommended approach allows avoiding pitfalls of sequential null-hypothesis tests that lead to unsound inference and controversial interpretation of results (Welch and MacMahon 2005, Greaves *et al.* 2006).

Following this approach, a 3-steps procedure was first designed so as to select the most informative EFDs regarding each functional type (see

Table 7). Then, a model scaling and model averaging procedure enabled inference to be drawn from several models simultaneously.

1. Each EFD was related separately to the calibration set using logistic models and the presence of curvilinear relationships was assessed by incorporating its quadratic term (Gaussian logit curve). The best function was retained using the small sample version of *Akaike's Information Criterion* (AIC_c, recommended by Burnham and Anderson 2002), calculated for each model with the equation 5:

$$AIC_c = -2\log(L) + 2K + \frac{2K(K+1)}{n-K-1} \tag{5}$$

   where $\log(L)$ is the maximized log-likelihood, n is the number of observations in the calibration set (i.e. 'informative' occurrences and pseudo-absences) and K is the number of parameters. This criterion compromises between model fit and complexity (Burnham and Anderson 2002). For each comparison, the model (logit or Gaussian logit curve) with the lowest AIC_c was retained.

2. For each multi-scale EFD (75, 150 or 300 meters, see Table 7), the functions retained above for each scale were compared using the same procedure and the function with the lowest AIC_c was retained.

3. Subsets of models were constructed for each EFD type (see different types/functions in Table 7). All EFDs belonging to a given type and retained by the two previous steps were combined to formulate 2^x-1 models, where x is the number of EFDs for this type. From these 'functional subsets', we eliminated models where correlated EFDs (spearman $|\rho| > 0.7$) simultaneously appeared. For each model, we subsequently calculated AIC_c. Within a given functional subset, we then calculated ΔAIC_c ($AIC_c - AIC_{c\ min}$) for each model ($AIC_{c\ min}$ is the lowest AIC_c among all models of the functional subset). The larger this difference, the less plausible it is that the model is the best one among

those of the functional subset. Models with $\Delta AIC_c > 4$ (see Burnham and Anderson 2002) were discarded and EFDs appearing in the remaining ones were retained for the subsequent analysis.

These steps allowed retaining the most important EFDs for each functional type. We used all possible combinations of the y retained EFDs to formulate 2^y-1 models. From this set of plausible models, we discarded those with $\Delta AIC_c > 4$ (differences are this time calculated against the best model of the whole set) and we calculated an *Akaike* weight (w_i) for each r_i of the R remaining candidates, using the weighted mean (Burnham and Anderson 2002)

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta AIC_{c\,i}\right)}{\sum_{r=1}^{R}\exp\left(-\frac{1}{2}\Delta AIC_{c\,r}\right)} \qquad (6)$$

These weights serve to quantify the evidence for or against the different models, given the set of candidate ones, providing a ranking of alternatives models. A multi-model inference was realized using model averaging for calculating parameter estimates (see Burnham and Anderson 2002 for details).

First, the relative importance (w_{+j}) of each EFD j was calculated by adding up w_i for all r_i models containing this EFD using the formulas 7 and 8:

$$w_{+j} = \sum_{i=1}^{R} w_i\, I_j(r_i) \qquad (7)$$

$$I_j(r_i) = \begin{cases} 1 & \text{if EFD } j \text{ is in model } r_i \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Second, for each EFD *j*, we computed the averaged parameter estimate $\hat{\bar{\beta}}_j$ by multiplying the parameter estimates $\hat{\beta}_{j,i}$ by w_i for each r_i of the *R* models and by summing the ensuing products:

$$\hat{\bar{\beta}}_j = \sum_{i=1}^{R} w_i I_j(r_i)\hat{\beta}_{j,i} \tag{9}$$

In order to avoid biasing averaged parameter estimates $\hat{\bar{\beta}}_j$ away from zero, we considered that each EFD was present in every model but that the corresponding parameter estimate $\hat{\beta}_{j,i}$ was set to zero in some models, rather than unknown. Consequently, we did not subsequently divide the averaged parameter estimate $\hat{\bar{\beta}}_j$ by w_{+j}, for each EFD.

These averaged parameter estimates were finally used for providing the relative Habitat Suitability (GLM-HS) value in each of the 1184 cells of the study area (*HS Map*, Figure 18, step 7). The Receiver Operating Characteristic (ROC) curve was used to select a threshold of GLM-HS above which the cell was considered as suitable. The same cost was assigned to false-negative and false-positive classifications among the calibration set. The optimum GLM-HS threshold was thus found by reading the point on the curve at which the sum of sensitivity and specificity was maximized (Zweig and Campbell 1993, Fielding and Bell 1997, Manel *et al.* 2001). For the 480 cells *a priori* masked out, GLM-HS was set to zero.

2.4.2. Expected Occupancy modelling

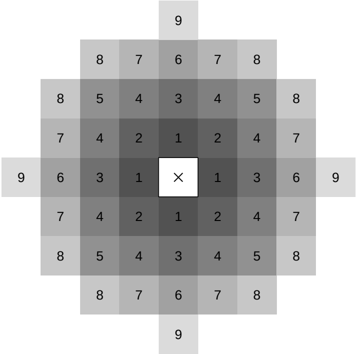
2.4.2.1. Autocovariate terms

In addition to environmental forcing, the likelihood of occupancy within a given cell may depend on whether the species occurs in neighbouring ones. Adding a term for autocorrelation (Autocovariate, see Chapter 4) in the logistic models allows accounting for this issue (autologistic models, see e.g. Smith 1994, Augustin *et al.* 1996). Nevertheless, when data are collected for inventory purpose, occurrences are often gathered without ad-hoc sampling design and the presence/absence of the species is unknown in several cells of the study area. This is the case here since we merged occurrences acquired during three different years and we used a subset of these occurrences for building the models. In those cases, the spatial structure of the population is partly occulted and the autocovariate risks to be unsound if directly calculated. One solution is to calculate the autocovariate terms for each cell using the relative GLM-HS estimated above, instead of the direct presence/absence data, using the formula 10 (Augustin *et al.* 1996)

$$\mathrm{Autcov}_i = \frac{\sum_{j=1}^{k_i} w_{ij} HS_j}{\sum_{j=1}^{k_i} w_{ij}} \tag{10}$$

For a given cell i, Autcov_i was a weighted average of the estimated relative GLM-HS (HS_j in formula 10) among a set of k_i neighbouring cells (Figure 18, step 9). The weight w_{ij} associated to a cell j was equal to $1/h_{ij}$, where h_{ij} was the Euclidian distance between the centres of cells i and j. Such autocovariates reflected the suitability of environmental conditions in the vicinity of the focal cell, but not directly the neighbouring occupancy.

Because we could not *a priori* determine the most adequate number of k_i neighbouring cells and the optimal distance h_{ij}, we used multiple order of neighbours (see Figure 19), from Autcov1 with $k_i = 4$ ($h_{ij} = 150$m, close vicinity) to Autcov9 with $k_i = 48$ ($h_{ij\ \max} = 600$m, more distant vicinity), following the procedure of Smith (1994) or Augustin *et al.* (1996). The response variable was thought to be uncorrelated beyond this maximal distance (600m).



				9				
		8	7	6	7	8		
	8	5	4	3	4	5	8	
	7	4	2	1	2	4	7	
9	6	3	1	×	1	3	6	9
	7	4	2	1	2	4	7	
	8	5	4	3	4	5	8	
		8	7	6	7	8		
				9				

Figure 19. Illustration of the procedure used for calculating the autocovariates. The focal cell is indicated by a cross. The order is indicated for each of the neighbouring cells, from direct (order 1, darker, h=150m) to more distant (order 9, lighter, $h_{\max}$=600m) vicinity. For a given order q, the autocovariate (Autcov$_q$) was calculated using all cells with order $\leq q$, with formula 10. As a consequence, from order 1 to 9, the number of neighbours (k) was equal to 4, 8, 12, 20, 24, 28, 36, 44 and 48.

2.4.2.2. Spatial ENFA and spatial pseudo-absences generation

Several ENFA were performed on the training set (including outsiders, Figure 18, step 10), using the EFDs retained by HS modelling with, in turn, one of the nine computed autocovariates (Autcov1 to 9). For each 'spatial ENFA' (see Olivier and Wotherspoon 2006 for a similar approach), the informative axes were retained by comparing the eigenvalues to Mac-Arthur's broken-stick distribution (see Jackson 1993, Hirzel *et al.* 2002a for details). The distance geometric-mean algorithm was then used for computing an Expected Occupancy (ENFA-EO, Figure 18, step 11) value for each cell of the study area (n=1184) in the reduced environmental hyperspace. This ENFA-EO value is similar to the ENFA-HS value provided by a classical ENFA, further accounting for the spatial structure in the data. For each cell, we subsequently computed the weighted average ENFA-EO value through the nine spatial ENFA runs, weights being the marginality coefficients (see Hirzel *et al.* 2002a) for each spatial ENFA. The minimal average ENFA-EO value among occurrences of the training set was used as cut-off value and spatial pseudo-absences were randomly chosen among cells of the global data with a lower average ENFA-EO value, in equal number than the occurrences in the training set (Figure 18, step 12). They were finally combined with the training set for providing a *spatial calibration set* (Figure 18, step 13)[1].

2.4.2.3. Autologistic models

We applied the following protocol for building a set of candidate autologistic models (Figure 18, step 14). We first use the reduced set of candidate logistic models retained at the end of the HS modelling procedure. Second, we included one of the nine autocovariates to each model. We therefore obtained nine (q=9) subsets of identical non-spatial models improved by the inclusion of Autcov1 to Autcov9. Third, within each of these nine 'spatial subsets', we computed AIC_c for each model (using the spatial calibration set), we discarded those with $\Delta AIC_{c,q} > 4$ and we calculated an *Akaike* weight ($w_{i\,q}$) for each of the remaining R_q candidate ones (rescaled for adding up to 1). For each of the q subsets, we then estimated the parameters using model averaging, following the same procedure as described above. Fourth, we computed an $AIC_{c,q}$ for each of the nine ensuing averaged models (each containing Autcov1 to 9). Using these nine averaged models, we subsequently performed exactly the same procedure again and obtained one global averaged model. Briefly, this model averaging in two steps allowed not overestimating the spatial component in the data because of the presence of nine partially redundant autocovariates that did not count for nine different spatial predictors.

The final averaged parameter estimates were used for providing the relative likelihood of Red-backed Shrike occurrence (GLM-EO) in each of the 1184 cells of the study area (*Expected Occupancy Map*, Figure 18, step

---

[1] This procedure is similar to that proposed by Engler *et al.* (2004) but within a spatially-explicit framework here.

15). The ROC curve was used to select a threshold of likelihood above which the cell was considered likely to be occupied by the species. For the 480 cells discarded *a priori*, GLM-EO was set to zero.

2.4.3. Model evaluation

Two evaluation indices were computed for each model, inspired from Hirzel and Arlettaz (2003b) (Figure 18, steps 8 and 16). The Absolute Validation Index (AVI) was the proportion of occupied cells in the evaluation set for which GLM-HS/GLM-EO value was higher than the ROC-based thresholds used for binarizing HS/EO maps. The Contrast Validation Index (CVI) was equal to the difference between the AVI and A_g (A_g was the same as AVI but for the whole study area). CVI ranges from 0 to $1 - A_g$ and reflects model accuracy (Hirzel *et al.* 2004), values near 0 indicating that the model does not outperform a random one (Hirzel and Arlettaz 2003b).

3. Results

The numbers of territories found in 2003 (non-exhaustive survey), 2004 and 2005 were respectively 43, 79 and 74. Because of straddling territories, these numbers respectively corresponded to 52, 96 and 110 occupied cells, reaching a total of 162 occupied cells during 2003-2005. The training and evaluation sets were therefore respectively constituted by 108 and 54 cells (Figure 20). The ENFA outputs for 2004 and 2005 separately are not detailed here but did not exhibit considerable discrepancies regarding (1) the coefficients of each EFD on the marginality and specialization factors and (2) the ensuing HS maps. These preliminary results strengthened our intention of gathering species data from multiple sampling years.

No strong correlation (spearman $|\rho| < 0.7$) was detected between EFDs belonging to different types (see ‘functions’ in Table 7), while few ones were quite correlated (spearman $|\rho| > 0.7$) within some of these types (especially among ‘prey detectability’ EFDs) or between spatial scales for multi-scale EFDs (Extensive, Arable, Hunt).

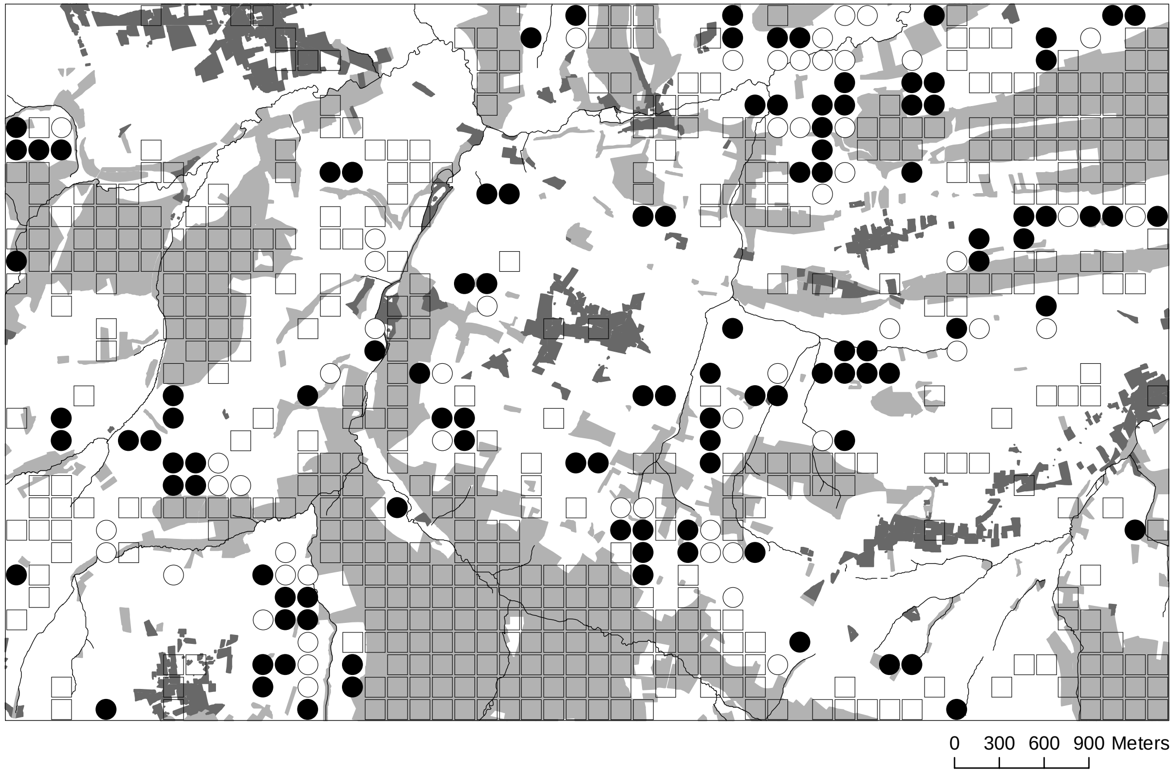


Figure 20. Localisation of occupied cells of the training (●) and evaluation (○) sets within the study area. Squares represent *a priori* masked out cells (n=480). See Figure 7 for background features.

3.1. HS modelling

From the training set, 10% of occurrences (n=11) were deemed unreliable indicators of habitat suitability following the preliminary ENFA and were discarded. The remaining 97 occurrences constituted half the calibration set. The subsequent ENFA-weighted pseudo-absences generation provided a pool of 581 potential pseudo-absences, from which 97 were randomly selected to complete the calibration set (n=194).

Using the Information-Theoretic Approach, we only detected the presence of a curvilinear relationship between the species occurrence and the distance to forest. The retained scale of computation for each EFD is mentioned in Table 13.

The HS candidate models for which $\Delta AIC_c < 4$ (R=15) are shown in Table 12. The *Akaike* weights of these models indicated that (1) there was no strong evidence that a single model was overwhelmingly ($w_i > 0.9$ following Burnham and Anderson 2002) supported by the data and (2) it was therefore risky to base prediction on the 'best' model only (model 1, $AIC_{c\ min}$).

Eight EFDs and the square term for DiForest were finally retained out of the 24 computed ones. The relative importance of each of them (Table 13) was calculated on the basis of the 15 candidate models and showed that resources from all functional types are required for adequately describing the suitability of the habitats.

Table 12. Set of candidate models ($\Delta AIC_c < 4$) retained for Habitat Suitability modelling, ranked by increasing ΔAIC_c.

	EFDs and associated scales	AIC_c	ΔAIC_c *	w_i †	K ††
1	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	40.26	0.00	0.12410	7
2	NestDist150 NestStd150 Extensive300 DiUrban Interface150 Contrast300	40.33	0.07	0.11956	7
3	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	40.34	0.08	0.11924	6
4	NestDist150 NestStd150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	40.49	0.23	0.11057	8
5	NestDist150 NestStd150 Extensive300 DiUrban Hunt150 Contrast300	40.67	0.41	0.10105	7
6	NestDist150 NestStd150 Extensive300 DiUrban Contrast300	40.88	0.62	0.09083	6
7	NestDist150 NestStd150 Extensive300 DiForest DiForest² DiUrban Hunt150 Contrast300	41.88	1.63	0.05506	9
8	NestDist150 Extensive300 DiForest DiForest² DiUrban Hunt150 Contrast300	41.93	1.67	0.05379	8
9	NestDist150 NestStd150 Extensive300 DiForest DiForest² DiUrban Contrast300	42.42	2.16	0.04216	8
10	NestDist150 NestStd150 Extensive300 DiForest DiForest² DiUrban Hunt150 Interface150 Contrast300	42.62	2.36	0.03812	10
11	NestDist150 NestStd150 Extensive300 DiForest DiForest² DiUrban Interface150 Contrast300	42.63	2.37	0.03790	9
12	NestDist150 Extensive300 DiForest DiForest² DiUrban Hunt150 Interface150 Contrast300	42.76	2.50	0.03559	9
13	NestDist150 Extensive300 DiUrban Interface150 Contrast300	43.08	2.82	0.03026	6
14	NestDist150 Extensive300 DiUrban Contrast300	43.80	3.55	0.02107	5
15	NestDist150 DiUrban Hunt150 Contrast300	43.84	3.58	0.02071	5

* $\Delta AIC_c = AIC_c$ of the corresponding model minus the lowest AIC_c among all models ('best' model)

† *Akaike* weights (w_i), see Equation 6

†† K is the number of parameters in the model (intercept included).

Table 13. Relative importance of EFDs and Autcov (w_{+j}) and averaged parameter estimates ($\hat{\beta}_j$) for final HS and EO models.

EFD/Autcov	HS model		EO model		
	Relative Importance	Averaged Parameter Estimate	Relative Importance		Averaged Parameter Estimate
Nest150	0		0		
NestDist150	1	3.736	1		1.908
NestStd150	0.59	0.807	0.01		0.045
Arable150	0		0		
Arable300 *	0		0		
IntPast300	0		0		
SemiNat300	0		0		
SemiNatWe150	0		0		
SemiNatWe300 *	0		0		
Extensive150	0		0		
Extensive300 *	0.98	1.699	1		3.472
SoilDry	0		0		
SoilDepth300	0		0		
DiForest	0.26	-0.040	0.01		0.001
DiForest²	0.26	-0.409	0.01		-0.051
DiUrban	1	2.235	1		6.024
Hunt75	0		0		
Hunt150 *	0.66	1.380	0.99		11.339
HuntVg75	0		0		
HuntVg150 *	0		0		
NbHunt150	0		0		
Interface150 *	0.50	0.541	0.32		0.190
Interface300	0		0		
Contrast150	0		0		
Contrast300 *	1	1.183	1		3.600
Autcov1			0.14	1	1.698
Autcov2			0.04		0.677
Autcov3			0.25		3.480
Autcov4			0.10		1.240
Autcov5			0.08		0.856
Autcov6			0.08		0.910
Autcov7			0.12		1.451
Autcov8			0.09		1.020
Autcov9			0.09		0.980

The retained scale of computation for each EFD following univariate analyses is marked by *. The retained EFDs following the HS modelling are in *italic*.

The averaged parameter estimates (Table 13) were used for computing the HS map (Figure 21A). These parameters may be directly compared because EFDs were brought to some common scale (standardisation). The Red-backed Shrike was mainly associated with sites where available nest carriers were numerous (and diversified) and that were distant from human settlements (NestDist, NestStd, DiUrban EFDs). Furthermore, low-intensity farming areas were particularly looked for (Extensive), especially where (ligneous) perches and interfaces between contrasted open-biotopes patches (Interface and Contrast EFDs) were adequately located to optimize the 'usable areas' for hunting (Hunt EFD). Finally, the proximity of the forest was slightly positively associated with occurrences but the species did not settle down in the direct vicinity of forest edges (DiForest and DiForest^2 EFDs).

Each cell of the HS map was allocated to suitable (n_s=466, 28% of the study area) or unsuitable (n_u=718 + 480 discarded *a priori*), following the ROC-based defined GLM-HS threshold, providing the binary HS map on Figure 21B. On this basis, the evaluation procedure provided an AVI of 0.8333 and a CVI of 0.5532. These measures were just slightly lower than those computed on the training set (AVI=0.8796 / CVI=0.5960), indicating that the final HS model was not overfitted to the training set. Because of the random splitting of the initial occurrence set between training and evaluation sets, outsiders were likely to occur in the same proportions within both sets, which largely explained the fact that AVI was not close to 1 for any of both sets. Keeping this in mind, these results indicated that the model performance is good but the difference between AVI and CVI showed that part of model performance could be attributable to randomness.

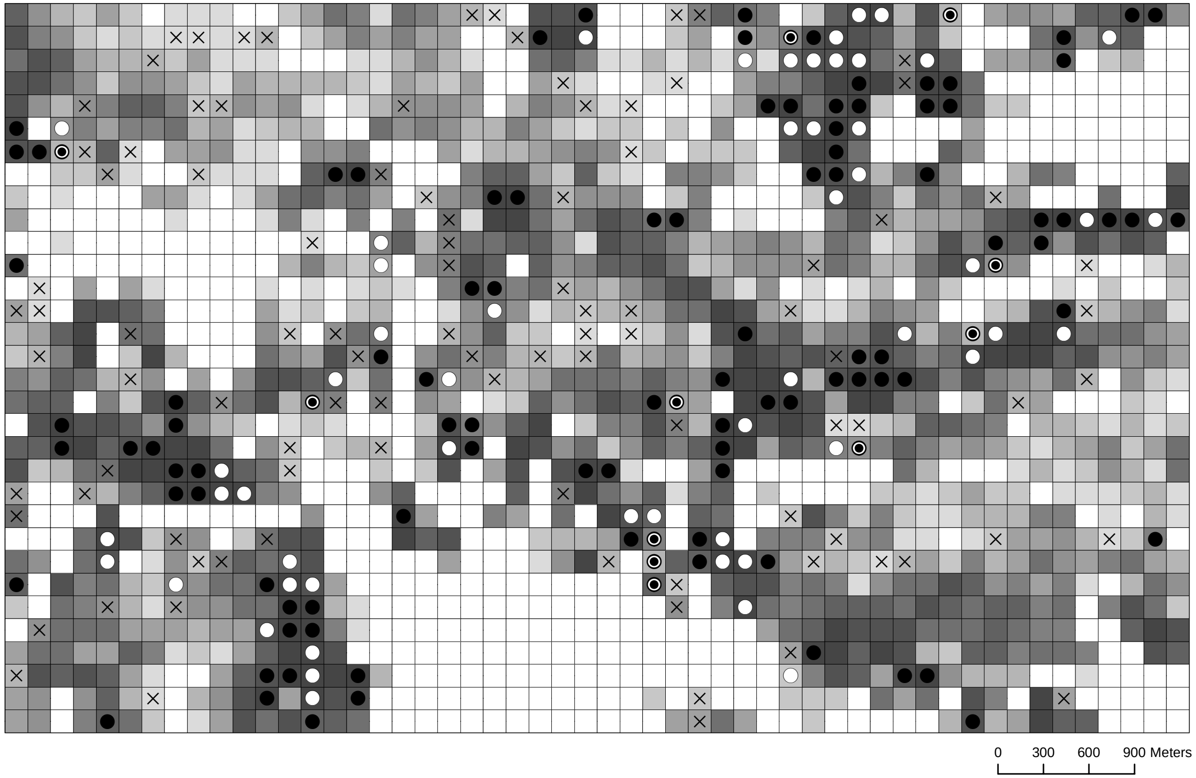


Figure 21A. Habitat Suitability map with values between 0 and 1, from unsuitable (lighter and masked out cells) to suitable (darker) showing the spatial distribution of suitable areas (grey shade scale; 10th percentile intervals). Occupied cells – training set (● informative occurrences (n=97), ⦿ outsiders (n=11)) / evaluation set (○, n=54) – and pseudo-absences (×, n=97) are represented.

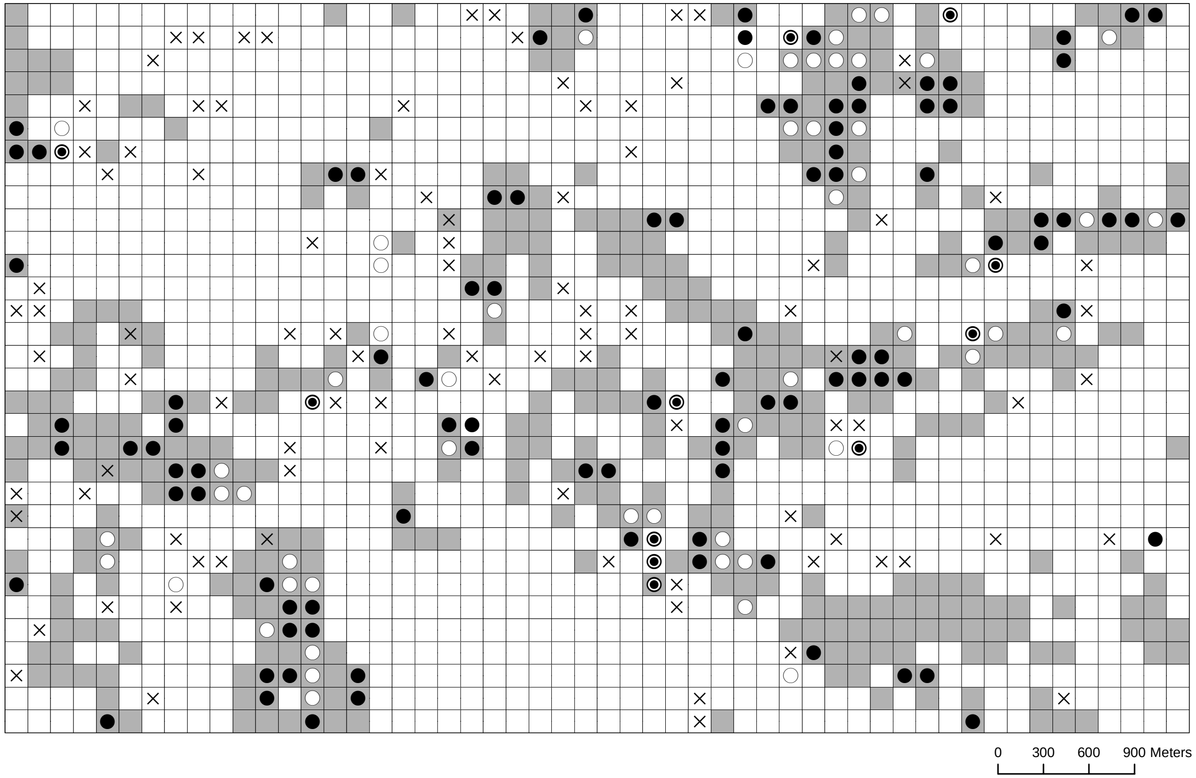


Figure 21B. Distribution of suitable (grey) and unsuitable (white) cells, following the ROC-based defined GLM-HS threshold. Occupied cells – training set (● informative occurrences (n=97), ⦿ outsiders (n=11)) / evaluation set (○, n=54) – and pseudo-absences (×, n=97) are represented.

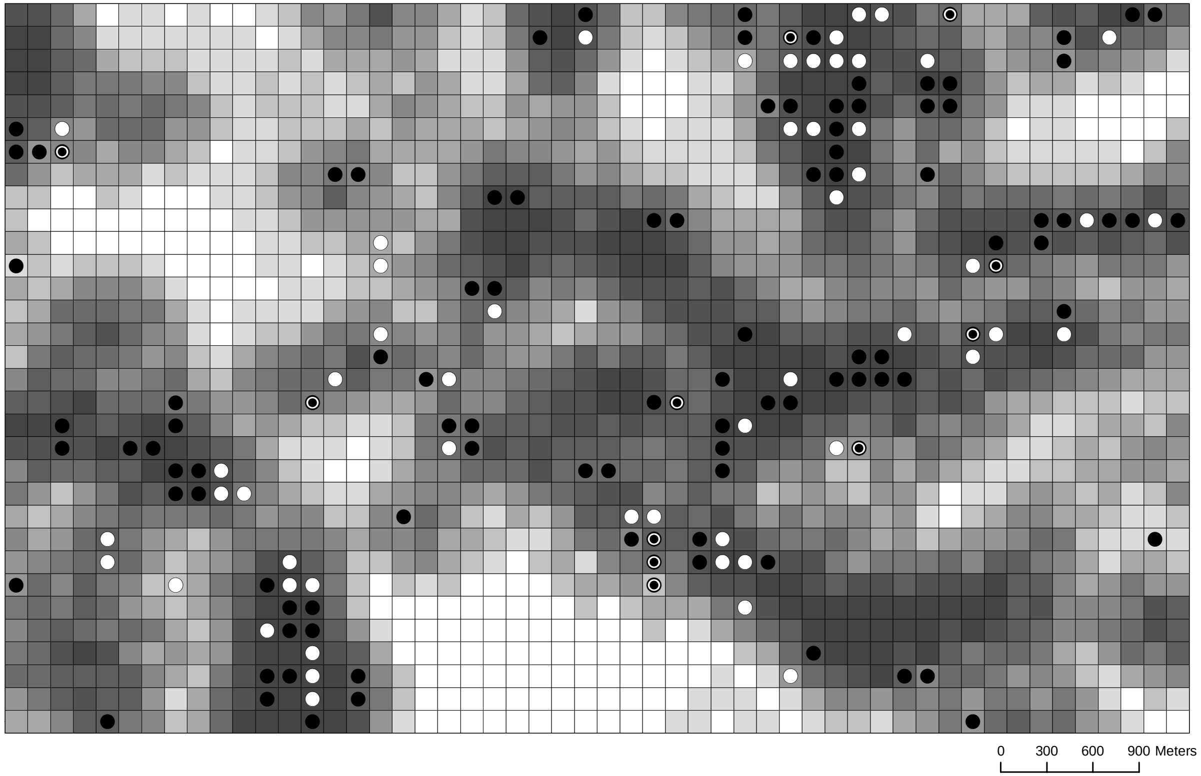


Figure 22. Representation of the 3rd order autocovariate (Autcov3) calculated following the HS map of Figure 21A (grey shade scale; 10th percentile intervals; the darker, the higher the Autcov3). Occupied cells of the training set are represented by black dots (●) for informative occurrences (n=97) and by encircled black points (⦿) for outsiders (n=11). The evaluation set is depicted by white dots (○, n=54).

3.2. EO modelling

The resulting HS map (Figure 21A) was used as a starting point for computing autocovariates and elaborating an Expected Occupancy (EO) map, accounting for the spatial structure in the data. While the outsiders designated during the HS modelling were associated with low local GLM-HS values (see Figure 21A,B), they were generally located where autocovariate values were high (see Figure 22, third-order autocovariate), thus surrounded by suitable cells. The ENFA-weighted spatial pseudo-absences generation provided a pool of 290 potential spatial pseudo-absences, from which 108 were randomly selected and combined with the training set to from the spatial calibration set (n=216).

The EO candidate models for which $\Delta AIC_{c,q} < 4$ ($R_q = 2$ or 3) and their relative *Akaike* weights within each spatial subset are shown in Table 14. With the exception of the third model associated to Autcov2 (model 5), the same EFDs and models were selected whatever the scale of autocovariate computation. This indicated a great stability in the local environmental role explaining the pattern of the occurrence data, when considering various ‘spheres’ of neighbouring suitability.

The successive weighting procedures using the 19 candidate models allowed adding up the weights of all autocovariates to 1 (see Table 13), in such a way that the global influence of autocorrelation on model calibration was not overestimated compared with other environmental influence. Consequently, since an autocovariate term appeared in every candidate model, the global relative importance of autocorrelation amounted 1, as NestDist150 for instance. Furthermore, within the global influence of autocorrelation, we disentangled the relative importance of each scale (Table 13), revealing a slight predominance of Autcov3 and then of Autcov1 in a lesser extent. However, no single scale was excessively supported by the data.

Table 14. Set of candidate models (ΔAICc,q) retained for Expected Occupancy modelling, ranked by increasing ΔAICc within each 'spatial subset'.

	Autcov	q^{*}	EFDs and associated scales	AIC_c	$\Delta AIC_{c\,q}$ **	$w_{i\,q}$ †	K ††
1	Autcov1	1	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	25.22	0	0.67608	7
2	Autcov1	1	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	26.70	1.47	0.32392	8
3	Autcov2	2	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	26.06	0	0.61516	7
4	Autcov2	2	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.89	1.83	0.24618	8
5	Autcov2	2	NestDist150 NestStd150 Extensive300 DiForest DiForest^2 DiUrban Interface150 Contrast300	29.04	2.98	0.13865	10
6	Autcov3	3	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	24.03	0	0.68559	7
7	Autcov3	3	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	25.59	1.56	0.31441	8
8	Autcov4	4	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	25.76	0	0.69650	7
9	Autcov4	4	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.42	1.66	0.30350	8
10	Autcov5	5	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	26.37	0	0.68625	7
11	Autcov5	5	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.94	1.57	0.31375	8
12	Autcov6	6	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	26.34	0	0.67348	7
13	Autcov6	6	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.79	1.45	0.32652	8
14	Autcov7	7	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	25.64	0	0.67331	7
15	Autcov7	7	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.09	1.45	0.32669	8

(Table 14 continued)

16	Autcov8	8	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	26.06	0	0.68796	7
17	Autcov8	8	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.64	1.58	0.31204	8
18	Autcov9	9	NestDist150 Extensive300 DiUrban Hunt150 Contrast300	26.12	0	0.68263	7
19	Autcov9	9	NestDist150 Extensive300 DiUrban Hunt150 Interface150 Contrast300	27.65	1.53	0.31737	8

[*] q refers to the spatial subset (models including Autcov1 to Autcov9).

[**] $\Delta AIC_{c\,q} = AIC_c$ of the corresponding model minus the lowest AIC_c among all models within the spatial subset q.

[†] *Akaike* weights ($w_{i\,q}$) were computed within each spatial subset q using Equation 6.

[††] K is the number of parameters in the model (intercept included).

The relative importance of the EFDs and the values of the associated averaged parameter estimates (Table 13) revealed that all functional resource types were still required for adequately describing the species occurrence pattern, when accounting for the habitat suitability in the neighbourhood. Furthermore, the influence of nesting sites availability and of forest proximity appeared to be less important than in the HS model. The 'usable areas' for hunting, the distance to urban settlements and the extensive areas were the main local EFDs driving the species occurrence pattern, when accounting for the neighbouring suitability. The Figure 23 illustrates (A) the EO map and (B) the geographical variations of the discrepancies between GLM-HS and GLM-EO values. One can particularly observe that the EO values of unsuitable cells situated at the margins of suitable areas are higher than their HS values.

Each cell of the EO map was allocated to likely to be occupied (n=677, 40% of the study area) or not, following the ROC-based defined GLM-EO threshold, providing a binary EO map (not shown). On this basis, the evaluation procedure provided an AVI of 0.9444 and a CVI of 0.5376, for both training and evaluation sets, because the proportions of misclassified occurrences were exactly the same. These similarities indicated that the final EO model was not overfitted to the training set. These results pointed out that the model performance was good, but the quite large difference between AVI and CVI drew attention on the fact that part of model performance could be attributable to randomness.

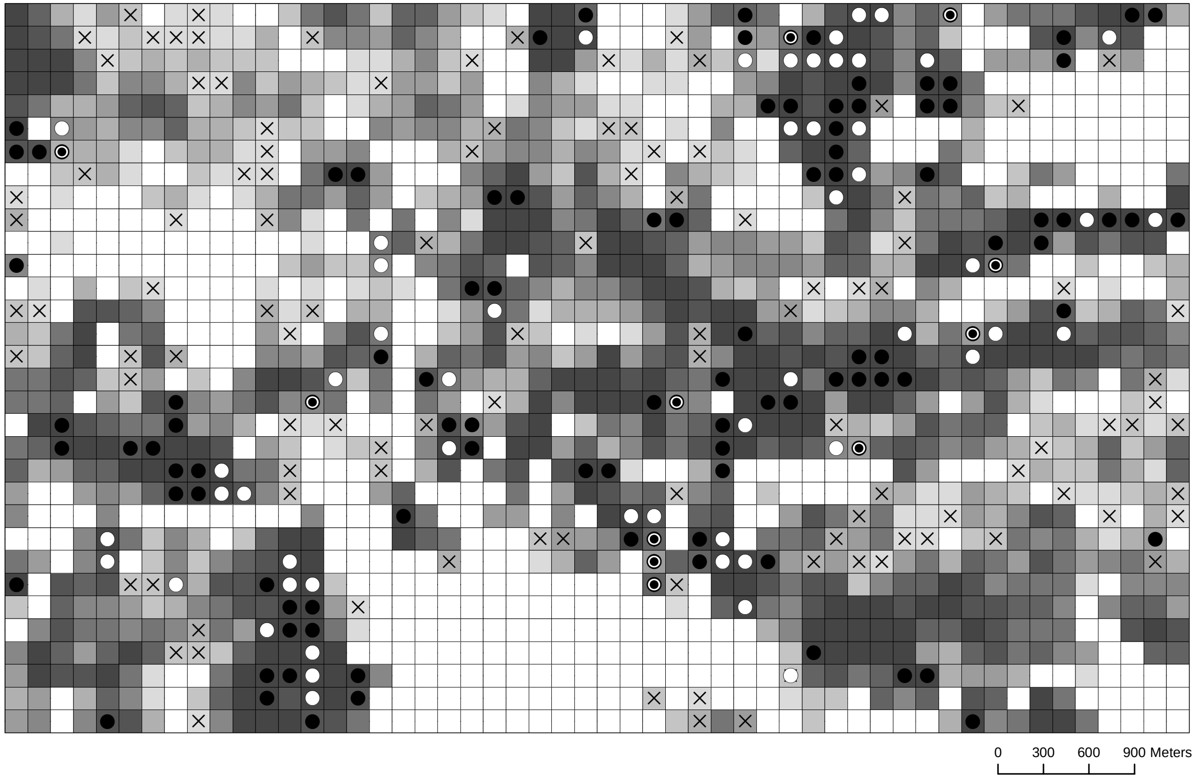


Figure 23A. Expected Occupancy map showing the spatial distribution of the likelihood of occupancy (grey shade scale; 10th percentile intervals), from unlikely (lighter and masked out cells) to most likely (darker) to be occupied. Occupied cells – training set (● informative occurrences (n=97), ⦿ outsiders (n=11)) / evaluation set (○, n=54) – and spatial pseudo-absences (×, n=108) are represented.

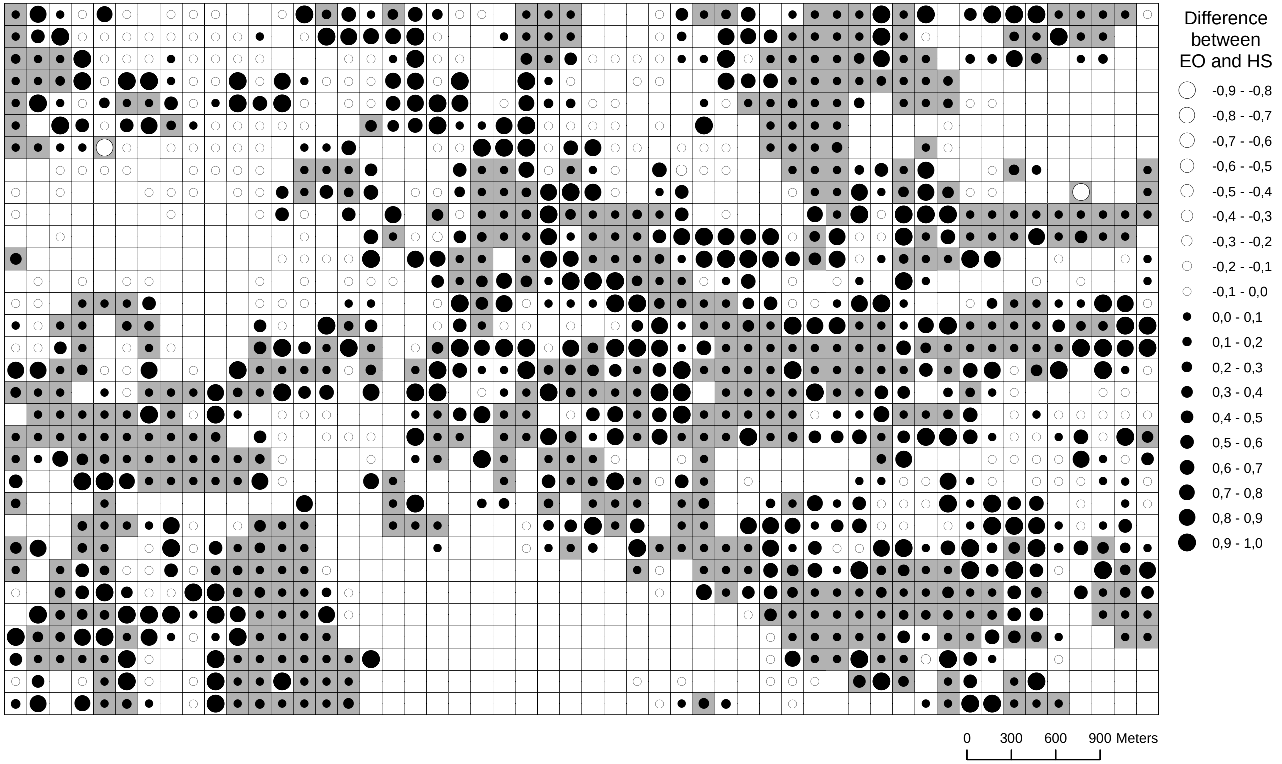


Figure 23B. Distribution of suitable (grey) and unsuitable (white) cells (see Figure 21B). Bubbles represent the discrepancies between GLM-EO and GLM-HS values for each cell. The size of the bubbles is directly proportional (equal intervals) to the absolute difference between GLM-EO and GLM-HS. Black and white bubbles represent cells for which GLM-EO was respectively higher and lower than GLM-HS.

4. Discussion

The designed ENFA-GLM approach carefully accounted for the existence of occurrences not conveying reliable information about habitat suitability before predicting the distribution of suitable sites throughout the study area (see Figure 21). This was already shown to be of paramount importance in the Red-backed Shrike by Titeux *et al.* (2006b,c) for not spuriously overestimating the extent of suitable sites across the study area.

The model scaling procedure revealed that no single biologically candidate HS model was best supported by the data, which prompted to use model averaging for estimating the parameters of the final HS model. This multi-model inference enabled to retain the most informative EFDs (Burnham and Anderson 2002) and not to overfit the final model to the data, which is reflected by the similarities of AVI and CVI for both training and evaluation sets.

4.1. Usefulness of habitat suitability model

Every species may respond to a combination of rather different sets of environmental forces in different parts of its distribution range (e.g. Gibson *et al.* 2004, Vanreusel *et al.* 2006). This is undoubtedly the case with the Red-backed Shrike in the forested Ardenne region in Southern Belgium for instance, where the species is typically found in recently felled areas or early-stages plantations (Jacob 1999). However, the ENFA-GLM combination used here was based on the comparison between the species occurrences and the available environmental conditions within the study area. Even if the HS model performed well within the study area, its efficiency was evaluated using occurrences in this same area and not in an independent one (Fielding and Bell 1997, Whittingham *et al.* 2003). Consequently, generalizing the model and transferring it to other areas should be achieved very cautiously (Fielding and Haworth 1995, Whittingham *et al.* 2003). While this issue is still under debate (e.g. Seoane *et al.* 2005, Vanreusel *et al.* 2006), focalizing on ecological resources that determine the functional interactions between the organisms and their environment – instead of computing environmental surrogates – allows more confidently transferring our main findings to populations established in similar landscapes (Vanreusel *et al.* 2006), at least at the Southern Belgian scale and most probably to other Western European areas. However, boundaries of application of the models remain to be assessed rigorously.

Moreover, the fine-grained resolution of the information combined with the direct relations to functional resources (e.g. nesting sites availability, foraging areas, prey density) represent the 'operational scale of conservation and restoration management' (Vanreusel *et al.* 2006). First, this approach provided relevant insight about the relative roles of various resources in driving the breeding habitat selection of the Red-backed Shrike. Second, most of these resources are directly in the scope of local or regional management or restoration planning, which is a prerequisite for effective implementation of conservation strategies.

Besides identifying the key habitat requirements for the species, geographically delineating the suitable parts of the study area (Figure 21) allowed prioritizing sites that require protection measures based on their GLM-HS values. In particular, the HS map revealed the existence of wide suitable but unoccupied areas, probably indicating that the population size did not reach the carrying capacity of the environment. Such identification is obviously not achievable when relying on direct occurrence records. In spite of species absence (that could be explained either by habitat-independent or by neglected factors), such unoccupied but suitable areas are not worthless but require conservation efforts just like occupied ones, especially in terms of habitat preservation and management.

4.2. Local habitat requirements

The final HS model showed that the Red-backed Shrike is mainly associated to areas where suitable nesting sites were abundant (Nest EFD[1]) and, on a lesser extent, structurally diversified (NestStd EFD). Nest site selection and structural characteristics of the immediate nest environment have been shown having a huge effect on the breeding performances of the species (Tryjanowski *et al.* 2000, Muller *et al.* 2005). Locally, the probability of finding an adequate nest site and of suitably concealing the nest increases where the bushes or hedges density is high and their physiognomy is heterogeneous (Van Nieuwenhuyse 1998a, Muller *et al.* 2005).

The species did not settle down in the close proximity of human settlements (DiUrban EFD, see also Kuzniak and Tryjanowski 2000), maybe because this EFD act as a surrogate for unmeasured resources, but also probably due to direct disturbance and to a potentially higher predation pressure by Black-billed Magpies (*Pica pica*) or cats (Muller *et al.* 2005). These should be minimized by controlling the extension of many villages and infrastructure installations in rural areas.

Moreover, the Red-backed Shrike was shown to be associated to extensive farming techniques, not only within the territories but also in their vicinity (Extensive300), most probably due to its diet (Lefranc and Worfolk 1997, Tryjanowski *et al.* 2003, Karlsson 2004). Intensive grazing pressure as well as nitrogen fertilizer affect the composition and the structure of the vegetation, inducing a dramatic decline in prey richness and abundance, whatever the taxonomic group (e.g. Andrews and Rebane 1994, Morris 2000, Holland 2002, Kruess and Tscharntke 2002a,b). In turn, this shortage of preys could induce the production of fewer or lighter nestlings, thus a lower survival probability (Leugger-Eggiman 1997). Besides offspring

---

[1] We refer in the text to the EFDs retained by the HS model, which allow describing the habitat requirements of the Red-backed Shrike in the study area.

consequences, parental-expenditure in high-intensity farming areas could be higher in terms of time devoted to hunting flight activity and energy allocation (Leugger-Eggiman 1997). Both consequences most probably constitute significant determinisms in shaping the habitat selection pattern of the species regarding this aspect. In areas where the conservation of the species is focused, pastures and hay meadows should then be preserved and fertilizer application or grazing intensity limited.

Furthermore, the species was associated to sites where suitable perches for hunting (i.e. compatible with the 'sit-and-wait' strategy, 1-4m high) were abundant, but particularly where their spatial arrangement provided a maximal 'usable area' (Hunt150 EFD). A distance of about 15-20 meters between suitable hunting perches was suggested by Van Nieuwenhuyse *et al.* (1999). This distance creates an optimal network of perching sites that increases the accessible foraging area and the foraging efficiency by minimizing the energy expenditure (see also Van Nieuwenhuyse 1998a). This exhibited the importance of any ligneous feature in the rural landscape, along field margins or inside pastures, even those that seem to be insignificant. The usable area was already shown as a key factor for habitat occupancy in the case of the Great Grey Shrike (*Lanius excubitor*) by Rothhaupt and Klein (1998), and Yosef (1993) highlighted the effect of hunting perches location on territory size in this same species.

The importance of a high total length of interfaces between open-vegetation patches within the territory was revealed in the final HS model (Interface150). In addition, contrasted vegetation heights were important within and around the territories (Contrast300). Field margins density and mosaic of patches of high and low vegetation types favour (1) the global abundance of preys (e.g. Meek *et al.* 2002, Holland 2002, Backman and Tiainen 2002, Pywell *et al.* 2005) and (2) their local accessibility (Van Nieuwenhuyse 1998a, Van Nieuwenhuyse *et al.* 1999, Lefranc 2004). Boundaries between different herbaceous vegetation heights are particularly attractive for foraging (Van Nieuwenhuyse *et al.* 1999) and the interspersion of tight and tiny patches of different vegetation heights allows maximizing the heterogeneity on a minimal total surface (Van Nieuwenhuyse 1999). Within mowed and grazed complexes, such spatial heterogeneity could be enhanced via phased and rotational mowing and grazing, as suggested by Van Nieuwenhuyse (1998a). This requirement undoubtedly leads to the exclusion of the species from landscapes affected by re-allotment schemes in Belgium (Jacob 1999) and in other Western European countries (Lefranc and Worfolk 1997), because they induce a structural homogenization of the farming areas.

Finally, even if this effect was less marked, the proximity of the forest was slightly preferred (DiForest EFD) but the species did not settle down in the direct vicinity of forest edges (DiForest² EFD). Prey density and predation pressure could have antagonistic effects during breeding site selection and explain this pattern. On the one hand, the positive association of forest proximity was probably attributable to highest density of preys like carabids (Burel 1989, Magura *et al.* 2001, 2002) or cockchafers (Couturier and Robert 1955). On the other hand, the predation pressure was very high in the direct proximity of forests and attributable to Eurasian Jay (*Garrulus glandarius*) (Söderström *et al.* 1998).

The required resources were related to both composition and spatial configuration of the farming landscape and were probably non-substitutable for the reproduction of the Red-backed Shrike. This statement is strengthened by the fact that the importance of each function is revealed by the final HS model, while no strong correlation was detected between EFDs belonging to different functions. Hence, the HS model allowed to prioritize the most important resources (see relative importance, Table 13) and showed the complementarity of multiple resource types (e.g. nesting sites, prey density, suitable foraging areas, predators distance). This stressed the importance of accounting for every aspect of habitat requirement when planning management or restoration.

4.3. Implications of spatial considerations for conservation

The second phase of the modelling procedure accounted for the spatial structure in the data and allowed quantifying the relative likelihood of occupancy for each site throughout the study area, depending on its local and neighbouring environmental conditions (EO model, Table 13).

The relative importance of the EFDs and the value of the associated averaged parameter estimates in the EO model (Table 13) are not directly biologically interpretable, because conditional on the previous HS modelling outcomes. These values reflected the relative local importance of each EFD while accounting for the GLM-HS of neighbouring cells that were themselves described by the HS model.

On the other hand, the values of averaged parameter estimates associated to autocovariates in the final EO model revealed the influence of the spatially-structured environmental conditions on the distribution pattern of the species. In particular, they indicated that, besides local environmental forcing, the species is more likely to occur in areas that are surrounded by suitable sites in the direct to nearby vicinity (see relative importance of autocovariates in Table 13). This outcome is mainly explained by the existence of individuals established in unsuitable conditions (outsiders, low GLM-HS values, Figure 21), because these were generally close to suitable conditions (outsiders, high Autcov values, Figure 22). They could thus have a high influence in autocovariates parameter estimation (autologistic models), leading to increase the relative likelihood of occupancy (GLM-EO) for unsuitable cells that are surrounded by or at the margins of suitable cells (see Figure 23B for a spatial comparison between GLM-HS and GLM-EO).

As suggested by Van Nieuwenhuyse (2000b) and Titeux *et al.* (2006a), these outsiders could have established near suitable sites that were themselves occupied by other attractive individuals or pairs (conspecific attraction, Stamps 1988). Titeux *et al.* (2006a, Chapter 4) discussed the direct and indirect importance of conspecific proximity. As we gathered multi-temporal occurrence records from whose we randomly chose a subset for building the models, the conspecific proximity was not strictly apprehended in the spatial models. This was first justified by the fact that we aimed at designing an approach that could be applied to any multi-date and non-exhaustive biological data set resulting from inventories and gathering occurrences collected with or without ad-hoc sampling design. The resulting training set thus partly occulted the yearly spatial distribution pattern of the species and fitting the autocovariates directly to this set would have been doubly unsound. Secondly, the direct computation of the autocovariates with exhaustive data acquired during a single year survey would have resulted in a 'yearly overfitted' representation of the neighbouring occupancy, hence useless in a predictive perspective. Indeed, the pattern of settlement of a Red-backed Shrike population changes considerably from one year to another. A phased procedure using the preliminary HS model for computing the autocovariates and then building the EO model was thus justified (Augustin *et al.* 1996). Another advantage of this approach is that the final EO model could be transferred (but still cautiously) to other similar farming areas with no information about the distribution of the species, unlike basic autoregressive models (e.g. Klute *et al.* 2002).

Iterative procedures for refining the EO model are conceivable (see Augustin *et al.* 1996) but they require more computation investment, especially within this step by step model averaging approach. Anyway, fitting an autologistic model like here provides better results than a logistic one in terms of predicting the spatial distribution of a species (see Augustin *et al.* 1996).

Drawing a parallel between HS and EO maps and localizing the main discrepancies was highly informative (Figure 23B), because revealing the spatially-structured decoupling between habitat suitability and occupancy (potential species distribution). The two steps approach revealed that 28% of the study area was suitable while 40% was likely to be occupied. The differences between HS and EO were particularly marked on the edges of suitable areas. As far as conservation is concerned, overlooking such discrepancies may be counter-productive. Indeed, this distortion was most probably due to social attraction that drove some individuals to settle down in unsuitable conditions near suitable sites (Titeux *et al.* 2006a, Chapter 4), inducing negative consequences in terms of individual fitness and population productivity (Schlaepfer *et al.* 2002, Shochat *et al.* 2005, Stamps and Krishnan 2005, Titeux *et al.* 2006c). Suitable sites with high 'edge-to-area' ratio (see centre of the study area, Figure 21) should especially suffer from such ecological trapping, compared with less fragmented suitable areas (northeast and southwest of the study area). Local restoration measures proposed above should thus be undertaken to tackle this perverse effect by focussing on preventing the fragmentation of suitable areas, particularly where (1) the highest discrepancies between GLM-EO and GLM-HS were revealed by the adopted procedure and (2) near traditionally occupied sites as a priority.

In conclusion, when investigating the fine-grained species-habitat relationship with a conservation purpose, we suggest (1) focussing on ecological resources that determine the functional interactions between the organism and its environment and that could be affected by management or restoration, (2) cautiously considering some occurrences as unreliable indicators of habitat suitability and (3) assessing the decoupling between habitat suitability and occupancy by the use of a spatially-explicit modelling framework. The approach proposed here is likely to be of considerable value in many situations where the complete distribution of a species is unknown or when the resources to conduct a survey are limited, ruling out any exhaustiveness.

Acknowledgements

We thank Jean-Baptiste Schneider and Julien Radoux for their technical help on the field and with GIS software, respectively. We are grateful to Alexandre H. Hirzel for useful discussions. We thank the Institut Géographique National (Belgium) and the Walloon Region for providing us with the vector 1:10,000 maps and the aerial colour orthophotographs (Licence number 031205-1000), respectively.

Conclusions

This thesis intended to enhance species-environment modelling techniques for untangling bird-environment relationships and was conducted in Southern Belgium (Central Famenne *sensu lato*). With this aim, we conducted a dichotomous approach, using first a coarse-scale, synecological and descriptive gradient analysis, and then a fine-scale, auto-ecological and functional niche-based modelling framework.

When investigating species-environment relations, a trade-off between data quantity and quality is regularly the heart of the problem regarding the ecological questions that can be asked (Van Horne 2002). This is hence of paramount importance to be aware of the usefulness and the interest of these contrasted approaches, as well as the limitations in the issues they respectively allow to deal with. Therefore, the first part of this conclusion aims to clarify the relative ins and outs of both conducted approaches, especially regarding their respective conservation implications.

Second, we emphasize the general poverty of available spatial land cover data that are ecologically relevant for building species-habitat models at the landscape level, especially in agricultural areas. We call for tuning future researches in spatial data acquisition to ecological needs in order to achieve ecologically sound estimation and monitoring of biophysical variables in such landscapes. We plead for financing complementary ground-based investigations to counterbalance the inherent limitations of airborne technologies.

Third, we call particular attention to the necessity of focussing future modelling studies on properly dissecting the distortion between species niches and distributions (as claimed by Pulliam 2000, Guisan and Thuiller 2005) rather than assuming a straightforward relation between habitat suitability, attractiveness and occupancy. More generally, modellers should fully integrate ecological theory into the modelling process (especially during the far-reaching conceptualization phase of model building), which is paradoxically neglected in the abounding literature. This constitutes the main technical and conceptual 'take-home message'.

Finally, we synthesize the ecological requirements for the Red-backed Shrike in Southern Belgian farming areas and underline the importance of implementing spatially-thoughtful conservation strategies for preserving the ecological processes involved in the species breeding habitat selection.

In this closing part of the present thesis, we hope the reader will not only focus his attention on the debate between the coarse and fine approaches, but will also get caught up in the improvements and the applied perspectives of the single-species habitat-based spatial modelling framework.

1. Ins and outs of coarse- and fine-scale approaches

The first part of the present thesis (Chapter 1) was devoted to the description of the major environmental and spatial gradients related to the variation in bird assemblages on a quite large spatial extent (about 300 km²). For this purpose, the approach relied on readily available data bases regarding both biological and environmental information. We based our work on a breeding bird atlas survey and on existing spatial environmental data. While the spatial resolution of the 'Atlas de Lesse et Lomme' was very fine (1 km²) for such a survey (data collection for inventory purpose, see Jacob and Paquay 1992), it was rather coarse for matching the far-reaching ecological processes underlying the distribution patterns for most of the species. Furthermore, the available digital cartographic information was devoid of robust ecological foundation merely because not initially gathered for ecological purpose, hence rendering a fine description of the habitats unachievable for most of the species. Instead, the broad environmental conditions – not the habitats of all species – were quantified within each sampling unit. A subordinate aim of this approach was to assess the conservation interest and usefulness of such fashionable and increasingly available atlas data (Hagemeijer and Blair 1997, Donald and Fuller 1998, Schmid *et al.* 1998, Tobalske and Tobalske 1999, Pasinelli *et al.* 2001) when investigating species-environment relations.

A finer approach was elaborated (Chapters 2 to 5) with the aim of identifying the most important resources and of accurately mapping the suitable environmental conditions of a single farmland species whose needs encapsulate those of many other ones (Padoa-Schioppa *et al.* 2006). This fine-grained identification at the landscape level demanded to design a spatial modelling framework that paid a particular and explicit attention to (1) the decoupling between resources availability and use and to (2) the spatial structure of this decoupling. This approach required ad-hoc field data collection for both species and resource information. The adopted spatial resolution was tuned to the key ecological processes thought to drive the species distribution at the landscape level. Accordingly, this data collection restricted both the number of species (auto-ecological approach in this case) and the spatial extent of the study (40 km^2).

Both approaches were based on correlations between the species responses (presence, density or breeding success) and environmental variables, therefore (1) *a priori* providing little insight about the genuine causal factors creating the distribution patterns and (2) questioning about model generalisation (see Van Horne 2002). Beside scale considerations, one of the main discrepancies between both approaches lied in the quality and the ecological relevance of the computed environmental variables. On the one hand, the synecological approach used coarse-scaled environmental metrics not describing the specific habitat requirements but instead most probably acting as surrogate environmental descriptors to which causal factors were partly correlated. This gradient analysis framework was therefore not suited to reveal the important resources and to investigate the finest processes involved in habitat selection for the many species, but rather to provide an overview of the main environmentally-related spatial arrangement of bird assemblages in a heterogeneous landscape. The outcomes of such an approach might be poorly transferable to other areas due to the possible distortion or even breaking down in space of the link between proximate and surrogate variables (Fielding and Haworth 1995, Van Horne 2002, Mitchell 2005). On the other hand, the functional approach was based on available specific ecological knowledge for computing environmental descriptors that reflected the diverse resources required by the organisms. This framework hence looked into the fundamental interactions – causing the observed patterns – between the individuals and their surrounding environment (Mackey and Lindenmayer 2001, Dennis *et al.* 2003). Nevertheless, we are conscious that manipulative experiments are required if we are to fully prove the cause-and-effect relationships. But they still do not constitute the panacea for testing all hypotheses (see Wiens *et al.* 1986, Van Horne 2002). Anyway, while limited in their spatial extent, the resource-based models were probably rather confidently transferable to other areas in the same eco-region, where resources availability and use are similar (Maki-Petays *et al.* 2002, Vanreusel *et al.* 2006). Nevertheless, these transferability considerations remain an unsolved issue in species-habitat modelling that is currently widely debated (e.g. Whittingham *et al.* 2003, Seoane *et al.* 2005, Mitchell 2005, Vanreusel *et al.* 2006) and further work is needed in order to better circumscribe models boundaries and domains of applications (Van Horne 2002), which constitutes the final step of model building according to Guisan and Zimmermann (2000).

Following the synecological approach, a rather high proportion (66%) of spatial variation in bird assemblages remained unexplained after partialling out the spatially-, environmentally- and mixed-explained components. In such cases, ‘noise’ in biological data is classically and excessively castigated to relieve one’s conscience. However, the decoupling between species niches and observed distributions (see Pulliam 2000, Guisan and Thuiller 2005) and the multiple inherent methodological biases of atlas data (especially regarding heterogeneous survey effort in space or time and species-, biotope- or birdwatcher-dependent detection probability, see Titeux 2002 for more details) inevitably contribute to generate this so-called ‘background noise’.

More specifically, the compositional axes primarily dissected the forest- and urban-species assemblages and almost not the farmland ones (but see the influence of arable surface). This is first due to the fact that the farming areas were not accurately described by the computed environmental variables. Secondly, farmland birds seemed to be finely spatially structured as compared with forest or urban birds whose distribution patterns were quite well captured by this coarse-grained approach. Actually, most of the farmland species require precise environmental conditions (Jacob and Paquay 1992) not reflected by the generic environmental variables and their spatial structuring is blurred at this non-flexible spatial resolution (1 km²). As a consequence, most of the farmland species that are of conservation concern at the regional level end up buried in some main and finally quite trivial environmental gradients (see e.g. Common Kestrel *Falco tinnunculus*, Northern Lapwing *Vanellus vanellus*, Common Stonechat *Saxicola torquata*, Common Grasshopper-Warbler *Locustella naevia*, Melodious Warbler *Hippolais polyglotta*, Red-backed Shrike *Lanius collurio* or Reed Bunting *Emberiza schoeniclus* – but the Great Grey Shrike *Lanius excubitor* and the Skylark *Alauda arvensis* are counterexamples).

This double issue (environmental description and spatial accuracy) inevitably went towards raising the unexplained variation for those species. For instance, only 8% of the spatial distribution in the Red-backed Shrike was environmentally explained in spite of its detection easiness. This species requires precise and non-substitutable conditions and is more finely spatially-structured. Other conservation interest farmland species are merely expressly discarded from the analysis owing to their scarcity (e.g. Corn Crake *Crex crex*, Wood Lark *Lullula arborea* or Corn Bunting *Miliaria calandra*). Accordingly, the coarseness of the grain and the ecological irrelevance of the currently available environmental data limit the conservation implications of such an approach at the landscape scale, especially for farmland birds. We nevertheless agree that this global descriptive approach (Chapter 1) was not as thoroughly conducted as the functional one (Chapters 2 to 5). We could for instance have focussed gradient analyses on farmland bird assemblages only for further dissecting their spatial arrangement (see Paquet *et al.* 2006 for such an approach). But anyway this would not have solved the critical issues regarding the resolution and the relevance of the available biological and environmental information.

We do not intend to denigrate the merits of this multi-species and coarse-scale approach, but simply to warn the reader about the many inherent limitations and drawbacks, especially regarding methodological biases, distal environmental descriptors and conservation implications at the landscape level. This holistic method was original in that it was the first one attempting to dissect the observed spatial variations of bird assemblage on the basis of an atlas survey, by partialling out the spatial components. Nevertheless, we think this approach could prove more useful at broader (regional or continental) spatial scales for coarsely highlighting the overall geographical and environmental factors related to species assemblages zonation (see e.g. Pasinelli *et al.* 2001, Storch *et al.* 2003), because the resolution of the information and the available land cover data (e.g. the 25ha minimum mapping unit of Corine Land Cover 2000 database, Büttner *et al.* 2004) better match such kind of broad addressed questions. We therefore advise future atlas databases users to exploit this variation partitioning technique when coarsely disentangling species-environment relationships.

Nevertheless, as far as conservation is concerned, further information is crucially required about the habitat requirements for many – especially farmland – species. Available and appealing atlas databases, as they are collected until now, are not the most adequate ones for reaching such objectives owing to the diverse biases and limitations mentioned above (see also Donald and Fuller 1998, Dennis *et al.* 1999, Dennis and Thomas 2000, Dennis and Hardy 2001, Dennis *et al.* 2002). Atlas data are currently gathered in a survey perspective first and not necessarily for species-environment assessment. They are thus useful for providing basic overview of species status and trend, as well as coarse environmental gradient related to distributions at various spatial scales (e.g. Robbins *et al.* 1989, Robertson *et al.* 1995, Greenwood *et al.* 1997, Bircham and Jordan 1997, Vaisanen 1998, Pasinelli *et al.* 2001, Telfer *et al.* 2002). Nonetheless, atlases should be regarded as an essential complement and not a substitution to other population monitoring schemes and fine-scale studies of bird-habitat relationships.

In this study, tuning the functional approach to the far-reaching specific ecological processes thought to govern the spatial distribution pattern (for instance, social attraction, hunting or nesting strategy) enabled to focus on the fundamental needs of the Red-backed Shrike. Ensuing ecological models properly explained the observed species spatial variations (see evaluation indices in Chapters 3 and 5), even with a quite low number of explanatory variables (9 retained resource-based variables in Chapter 5). Ultimately, such models allowed suggesting effective management strategies. Obviously, as mentioned at the beginning of Chapter 2, single-species management based on such an approach would unavoidably conflict with the management of other species (e.g. Simberloff 1998, Lindenmayer and Fischer 2003). On this account, this species-specific approach proved to be limited regarding implications for wider biodiversity conservation (but see Perspectives for further nuanced considerations).

2. A mismatch between available spatial land cover data and ecological processes

Nowadays, there is an ongoing trend to exploit available spatial land cover data for building species-habitats models at various spatial scales (e.g. Mack *et al.* 1997, Corsi *et al.* 2000). This is explained by the fact that the emergence of new tools like geographical information systems (GIS) and remote sensing (RS) was naturally followed by a period of euphoria among ecologist (Van Horne 2002). GIS and RS offer many possibilities and provide a profusion of interesting spatial land cover data. However we would like here to call for cautiousness to avoid inadequately exploiting this available information. In particular, we wish to (1) briefly warn the landscape ecologist about the main risks incurred by the abuse of such spatial data without being aware of their limitations and (2) subsequently advocate more closely connecting researches in these technological fields and spatial ecology.

The main problem linked to spatial land cover data lies either in the content or in the spatial accuracy of the information they enclose (Schmit *et al.* 2006). While useful for many regional or broader applications, the environmental patterns they highlight often prove to be of limited assistance in understanding ecological processes at the landscape scale. In their recent study in Belgium, Schmit *et al.* (2006) showed that readily available spatial land cover data do not fully and accurately provide the required information for many environmental analyses (see also Bodin *et al.* 2006). This a fortiori holds for resource-based modelling. In particular, minor land-uses such as unmanaged areas in farmland – that are fundamental for ecological processes – are frequently under-represented or overlooked in such spatial data (for instance 25ha minimum mapping unit in the Corine Land Cover 2000 update, see Büttner *et al.* 2004).

More accurate cartographic maps are available but suffer from ecological irrelevance. For instance, the digital 1:10,000 vector topographic maps of the Institut Géographique National (edited in 2004) give information about the main and large hedges in rural areas. This is very useful in a regionally perspective but pretty not for finer applications like resource-based modelling, as illustrated by the habitat requirements of the Red-backed Shrike. There is also a crucial lack of information about the ecological quality of farming or forested areas (e.g. phytosociological association, presence of useful indicator species) (De Blust *et al.* 1994, Bodin *et al.* 2006). Remote sensing is for instance intrinsically limited in its ability to provide accurate information on the composition of hedges or the intensification level of pastures, which are determining elements driving the Red-backed Shrike distribution pattern in the landscape. As a consequence, the correlation between the landscape patterns extracted from readily available spatial land cover data on the one hand and the true ecological value of those landscapes in terms of resources for species on the other hand could prove to be highly unstable either in space or in time, or both. This mismatch between the relevance of the (potentially) available spatial information and the far-reaching ecological processes involved at the landscape scale unavoidably hampers the ecological soundness and the direct applications of species-environment models which rely on such available spatial data only.

Having said that, RS and GIS are invaluable tools for ecologists because they allow to describe the spatial arrangement of landscape features at various spatial scales (e.g. Burrough 1992, DeMers 1997, Johnston 1998) and to efficiently plan field work (Bodin *et al.* 2006). Recent technological advancements are promisingly bridging the gap between needs and availability of spatial information (see e.g. Carleer 2005, Bodin *et al.* 2006 and Kayitakire 2006), so that we strongly stimulate RS researchers for increasingly focussing a significant part of their future investigations in this direction. Nevertheless, owing to the inherently limited spatial, temporal and spectral resolutions of the sensors, the gap will probably never be totally filled without complementary ground-based data acquirement (see De Blust *et al.* 1994).

3. A call to account for the decoupling between species niche and distribution

While the relationships between species niches and observed distribution patterns are often far from obvious (see e.g. Van Horne 1983, Dunning *et al.* 1992, Robinson *et al.* 1995, Donovan *et al.* 1995, Pulliam 2000, Donovan and Thompson 2001, Delibes *et al.* 2001a, Stamps and Krishnan 2005), most habitat-based modelling techniques largely overlook this issue by assuming a direct link between habitat suitability and occupancy (occurrence or density).

Within a multi-species framework like the one designed here, the decoupling between species niches and distributions may inevitably involve an increase in the part of the chi-square statistics – measuring the total inertia in the species assemblage data – that can not be explained by the environmental or spatial gradients defining the compositional axes of the direct gradient analysis. We agree that explicitly dealing with this concern for every species of the assemblages is logistically unachievable, but both modeller and model-user should be conscious of these species-specific deterministic distortions instead of excessively incriminating the so-called ‘stochastic noise’ in the biological data sets, especially if further (specific) investigations are envisaged.

Most of the time, discrimination approaches like Generalized Linear Models (GLM) – logistic regressions for instance – are used for predicting species occurrence or for untangling specific species-habitat relations within a modelling framework (e.g. Manel *et al.* 1999a, Pearce and Ferrier 2000, Scott *et al.* 2002, Guisan *et al.* 2002, Rushton *et al.* 2004). While largely violated in many cases, discrimination approaches though assume that species presence and absence respectively indicate suitable and unsuitable environmental conditions. It is strictly required to carefully validate such a huge assumption – and more generally to better root the modelling approach in ecological theory – during the initial conceptualization phase of model building (see Guisan and Zimmermann 2000, Huston 2002), before blindly fitting any discriminant function. Should this not happen, alternative approaches are to be adopted or designed to deal with the observed decoupling.

The frequent lack of valid and reliable absences for artefactual (e.g. Anderson 2003) or biological (e.g. Pulliam 2000) reasons has led to the development of ‘presence-only’ approaches (Walker and Cocks 1991, Busby 1991, Carpenter *et al.* 1993, Robertson *et al.* 2001, Hirzel *et al.* 2002a, Robertson *et al.* 2004). They compare the species occupancy with the available conditions in the environmental hyperspace and therefore do not rely on absence data. These methods – recently thoroughly reviewed by Pearce and Boyce (2005) and Elith *et al.* (2006) – do not explicitly and *a priori* account for individuals occurring outside the environmental boundaries of the species niche and hence assume that species presence is a reliable indicator of habitat suitability (but see Hirzel and Arlettaz 2003a, 2003b).

We showed here thanks to a sensitive analysis that such outsiders could though have a huge and spurious effect on the ensuing models and we therefore proposed a statistical (Chapter 2) – and ecologically sound (Chapter 3) – phased approach for identifying and discarding such misleading occurrences before building reliable habitat suitability models (as suggested by Guisan and Thuiller 2005). We suggest subsequently looking into the cause(s) of this kind of decoupling. This step could be time-consuming and might require manipulative experiments if the aim is to definitely identify the underlying processes (which was not thoroughly achieved in Chapter 4). Finally, the gathered information should be explicitly integrated into the modelling framework (Chapter 5) to (1) identify what truly constitute suitable environmental conditions and to (2) assess the discrepancies between the suitability, the attractiveness and the relative likelihood of occupancy for any given site. This phased approach could be viewed as disheartening or daunting but is undoubtedly worthwhile for implementing effective conservations plans.

We finally call attention here that models based on presence-only data do not accurately predict the probability of species presence, merely because the proportions of presences within the calibration sets do not represent the true prevalence of the species in the landscape (Pearce and Boyce 2005). Rather, they are useful in their ability to rank sites' suitability or likelihood of occupancy on a relative scale (Elith *et al.* 2006). As a consequence, such models neither depict the real state of the populations nor can be used for estimating their sizes.

4. The Red-backed Shrike conservation in Southern Belgian farming areas

The required resources for the Red-backed Shrike in Southern Belgian farming areas are related to both landscape composition and spatial configuration. Habitat Suitability models (see Chapters 3 and 5) allowed ranking the importance of several types of resources (in decreasing order of importance, abundance of adequate nesting sites, remoteness of urban areas, prey density, suitable foraging areas). However, they showed that the different types are complementary and hence all required for the achievement of the species breeding cycle. These findings stress the need to account for every component of the habitat when planning management or ecological restoration.

The modelling framework designed through Chapters 2 to 5 clearly showed that the Red-backed Shrike is linked to farming areas dedicated to a combination of hay production and livestock rearing. The suitability of such areas for the settlement and the reproduction of the species could be maintained or improved through several main conservation measures that should be taken all together (see particularly Chapters 3 and 5). The pervasive and ongoing human settlement in Southern Belgian rural landscapes should be alleviated, because the species was shown to occupy sites faraway from urban areas. Models indicate that the Red-backed Shrike is associated to extensive farming areas, most probably because the high-intensity of farming techniques are insidiously turning down the density and biomass of preyed invertebrates, as shown by Kruess and Tscharntke (2002a,b). Punctual or linear thorny features inside or along pastures or hay meadows were shown to be of high importance, because their local structural diversity offers a beneficial network of nesting and perching sites, therefore respectively allowing individuals to conceal their nests and to enjoy suitable foraging areas. Finally, habitat suitability modelling revealed the importance of tight, tiny and imbricated open-vegetation patches of contrasted vegetation heights, in the direct vicinity of perching and nesting sites. This spatial configuration of farming landscapes simultaneously improves prey density and detectability. This latter requirement entails the local maintenance of diversified farming parcels dedicated to various uses. This habitat component should lead to the exclusion of the species from areas affected by re-allotment schemes, because they induce a structural trivialization of the farming areas.

Furthermore, merely focussing conservation efforts on improving local habitat suitability without any spatial consideration may be poorly productive. Actually, a socially-mediated decoupling between habitat suitability and attractiveness induces a species distribution pattern that is not a straight mirror of suitable conditions. This distortion is spatially-structured (Chapter 5) and most probably due to the influence of conspecific attraction during the settlement and pairing phases (Chapter 4). As shown by Habitat Suitability and Expected Occupancy modelling, the discrepancies between suitability and likelihood of occupancy are particularly marked on the edges of suitable areas (Chapter 5), indicating that unsuitable sites near suitable ones are still attractive and occupied by individuals that are rewarded by a low breeding success (Chapter 3). In our particular study area, 12% of the landscape was likely to be occupied by the species while covered by unsuitable environmental conditions. A spatially-reasoned management strategy should therefore focus on reducing this ecological trapping because it induces a decrease in the population productivity. This could be achieved by maintaining large and compacted suitable habitat patches in the rural landscape, so as to reduce the edge-to-area ratio of these patches.

Conspecific attraction has probably been evolutionarily shaped since Neolithic times (appearance of extensive farming techniques and rural landscapes beneficial to shrikes – see Lefranc and Worfolk 1997) because ultimately adaptive. But this mechanism may currently prove to be maladaptive in highly fragmented habitats due to a sudden modification of selective regime (Stamps 2001, Etterson 2003). As mentioned by Schlaepfer *et al.* (2002) and Kristan (2003), this might involve a dissociation between cues that organisms use to make any decision (*proximate* factors, here conspecific presence or density) and expected outcomes following that decision (*ultimate* factors, here suitable breeding conditions), some organisms being 'trapped' by their evolutionary responses to these cues. Nevertheless, empirical support to this speculative discussion is needed.

Managing rural landscapes is obviously a more complex task than shaping ‘shriky-landscapes’, which constitute an over-simplistic conservation perspective. Measures to be implemented should integrate more than specific prescriptions and are to be pondered within a socio-economic context. Nevertheless, the Red-backed Shrike habitat modelling outcome illustrated that a particular attention needs to be paid to the wider environment. Financially-supported measures implemented in European agri-environment and set-aside schemes may prove to be effective to incite farmers modifying their agricultural practices for enhancing the perpetuation of far-reaching ecological functions at the landscape scale (see e.g. Donald and Evans 2006 but equivocal appraisal by Kleijn and Sutherland 2003, Kleijn *et al.* 2006). Schemes such as livestock grazing pressure reduction, permanent grasslands maintenance, extensive herbaceous field margins or hedge preservation are to be particularly promoted for this specific case. Moreover, the Common Agricultural Policy should be further restructured in order to reduce the economic vulnerability of the traditional farmers, remnants of a vanishing agriculture that sustains farmland bird diversity.

The attentive reader could finally notice that the Red-backed Shrike population density seems to be quite below the apparent carrying capacity in our study area. Indeed, 6-7% of the landscape is yearly occupied (see Chapters 2 to 5) while 28% was declared as suitable following our modelling approach (Chapter 5). This inevitably questions on the actual threat of habitat loss or degradation for the species within the study area. At first glance, one could believe that the availability of suitable sites does not constitute a limiting factor for this population that would be held below carrying capacity by some external factors. Nevertheless, when breeding site fidelity and conspecific attraction are synergistically operating, populations could be particularly prone to suffer from the degradation and the fragmentation of their habitats, because more individuals are likely to be trapped by their evolutionary response to conspecific attraction and small or isolated suitable sites are less likely to be occupied. As a result, populations may become non self-sustaining below a given (but currently unknown) threshold of suitable areas. Their persistence may then depend on the immigration from landscape sources through a regional source-sink dynamics (see With *et al.* 2006). In addition to habitat suitability, further analyses at broader spatial scales are therefore required for assessing such 'landscape sustainability'. A multi-scale approach would enable to relevantly evaluate the actual carrying capacity given the proportion and the spatial configuration of the suitable sites within the landscape and in a regional context. It is not utopian to believe that this relation is non-linear (see Delibes *et al.* 2001b) and that landscapes could be deserted even if they still enclose some significant proportions of suitable sites, merely due to an overwhelming proportion of sink habitats. This threshold was probably reached in Belgian Condroz where suitable sites are locally remaining but being only sporadically occupied by the species. These speculative considerations lead us to the perspectives of this thesis.

Perspectives

Punctual perspectives were already mentioned throughout this document, such as all-scale decortication of the spatial structures of bird assemblages with PCNM, variation partitioning within a RDA framework, improvement of distance-based HS algorithms, enhancement of evaluation methods when species distribution is not known exactly or when niche-distribution distortion results from ecological processes, or assessment of models domains and boundaries. Some of them are technically challenging but are worth looking into. We would like here to present some general and more conceptual perspectives.

1. Issues of spatial scales in atlases

The 'Principal Coordinates of Neighbour Matrices' approach (see postscript of Chapter 1) enables to dissect the spatial variation of bird assemblages data set at all scales (Borcard *et al.* 2004). Yet this method is not a magic one and achieves a spectral decomposition of the spatial relationships among the sampling sites that remains limited by the basic resolution of the data at hand (Borcard and Legendre 2002). Since many ecological processes and patterns at the landscape level are more finely grained than atlases resolutions and if we could enjoy available ecologically sound environmental information at finer resolution, it should be possible to estimate the finer spatial species distributions with downscaling methods (see Verburg *et al.* 2006, Dendoncker *et al.* 2006) that are founded on the Bayesian theorem. The development of such probabilistic methods is currently emerging in diverse scientific fields and these have so far rarely been applied for ecological purpose. They still could prove to be very helpful for refining the resolution of coarse atlas databases (see Araujo *et al.* 2005), but their ecological reliability should be investigated. This way, it could be conceivable to (1) match more closely the ecological processes creating patterns at the landscape level and the local management requirements and (2) rigorously assess the rightfulness of the focal species used in subsequent fine-scale single- or few-species modelling framework, hence bridging the gap between both conducted approaches.

Should this proposed approach reveal unsuccessful or unsound, we would plead for a radical change in the classically adopted atlas data acquirement. Targeting data collection on a subset of sites a priori determined using a stratified and multi-scale sampling strategy, instead of dispersing sampling effort, should allow considerably improving the quality and the spatial accuracy of the gathered biological data. Such fine-scale information could subsequently be used for inferring the species distribution patterns in unsampled areas, based on habitat modelling.

Anyway, these alternative approaches and their respective limitations have to be properly evaluated and tested before giving founded advice and further guidelines. In both cases, the quality of the spatial land cover or other environmental data should be critical.

2. Community-based modelling

Community-based distribution modelling (see Elith *et al.* 2006 and cited works therein) or multi-objective response surfaces (e.g. Dehhaoui 1996) are other promising ways for reconciling single and multi-species approaches. Nevertheless, their limitations should be carefully assessed as detailed below.

- The former is an emerging approach using information of several species to guide the environmental variable selection during the ‘single-species’ presence-only modelling, and for this reason could prove to be very helpful for rare species. We think that (1) this community signalling could further inform on the selection of pseudo-absences among the background data but that (2) such methods should be used very cautiously owing to the possible disrupted pattern of species co-occurrence in time and space.

- The latter is a more dated but underused multiple criteria decision method that allows simultaneously optimizing several response surfaces such as habitat models that have been built independently. We think carefulness is nevertheless required when using such an approach because the 'solution' is a trade-off between several specific needs that could be antagonistic or exclusive. In this case, it does not make sense wanting to have one's cake and eat it too, because the ensuing risk is to manage the landscape in such a 'compromise' way that it would become unsuitable for all focal species, hence missing the boat.

3. Temporal dimension and multi-scale approach

The breeding success-supported modelling framework we designed here to cope with the spatially-structured niche-distribution decoupling is indubitably technically perfectible, notably by considering a fuzzier distinction between individuals occurring in- and outside the environmental bounds of the species niche (especially for pseudo-absences random generation). More conceptually, it should integrate the temporal dimension (long-time data series analysis) by explicitly accounting for both site fidelity and conspecific attraction for (1) modelling the uncoupling between habitat suitability and attractiveness producing an hypothetical source-sink dynamic and ecological trapping, (2) evaluating the mal-adaptedness of indirect cues used in habitat selection within fragmented landscapes and (3) providing accurate predictions of their consequences at the population level (see Donovan and Thompson 2001, Delibes *et al.* 2001a, Delibes *et al.* 2001b, Kristan 2003 for theoretical frameworks). Experimental works are absolutely needed for rigorously quantifying the relative contributions and interactions of the different processes involved in habitat selection (habitat features, site fidelity, conspecific attraction or others not investigated here, see e.g. Doligez *et al.* 1999, 2003, 2004, Danchin *et al.* 2001).

Furthermore, multi-scale investigations should be undertaken to assess (1) the importance of maintaining a sufficient proportion of suitable sites at the landscape level ('landscape sustainability') and (2) the existence of an eventual regional dynamics (metalandscape analysis, see With *et al.* 2006). The inclusion of such parameters related to population dynamics into the modelling framework is challenging and requires long-term population surveys. But it would enable to throw off the shackles of fundamental Grinnellian or realized Hutchinsonian niches (see Figure 1A and B) for better understanding the species spatial distribution patterns and, in turn, supporting population studies (see Guisan and Thuiller 2005).

4. Enhancing findings impact through appropriate communication

Finally, one of the most important perspectives of such a species-environment study is its presentation to decision-makers in a non-technical way (as claimed by BirdLife International 2004), by converting the scientific outcomes of the models in straight management and conservation suggestions (McCracken and Bignal 1998). For this reason, thresholds regarding vital resources should be further investigated for intentionally simplifying these outcomes into policy-relevant messages. This way, these management prescriptions would be more easily integrated into conservation schedules like Natura 2000 network or through agri-environmental schemes. Providing some key figures to decision-makers is though highly challenging because its effectiveness depends on a well thought-out simplification of habitat models that are themselves low-dimensional abstractions of infinite-dimensional forces acting on individuals. Finally, the geographical distribution of currently lacking specific resources should be mapped to precisely identifying '*where to do what*' from a management or ecological restoration standpoint.

Publications

Maes, D., Gilbert, M., Titeux, N., Goffart, P. and Dennis, R.L.H. 2003. Prediction of butterfly diversity hotspots in Belgium: a comparison of statistically-focused and land use-focused models. Journal of Biogeography, 30: 1907-1920.

Titeux, N., Dufrêne, M., Jacob, J.P., Paquay, M. and Defourny, P. 2004. Multivariate analysis of a fine-scale breeding bird atlas using a geographical information system and partial canonical correspondence analysis: environmental and spatial effects. Journal of Biogeography, 31: 1841-1856.

Titeux, N., Dufrêne, M., Hirzel, A.H. and Defourny, P. (submitted). Modelling habitat suitability with presence-only data: we need to account for unreliable species occurrences. Ecography.

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. and Defourny, P. (submitted). Identifying the Habitat Requirements of the Red-backed Shrike (Lanius collurio) – The Importance of Breeding Success. Journal of Applied Ecology.

Titeux, N., Dufrêne, M. and Defourny, P. (to be submitted). The Importance of Conspecific Proximity for the Conservation of Territorial Migratory Bird Species – The Case of the Red-backed Shrike. Landscape Ecology.

Titeux, N., Dufrêne, M. and Defourny, P. (to be submitted). Identifying breeding habitat requirements and selecting conservation-interest sites for the Red-backed Shrike (Lanius collurio) in Southern Belgium – A functional and spatial habitat modelling approach. Ecological Modelling.

References

Agresti, A. 2002. Categorical Data Analysis. John Wiley and Sons, Hoboken, New Jersey.

Allen, T.F.H. 1998. The landscape level is dead: persuading the family to take it off the respirator. - In: Ecological Scale: Theory and Applications. (eds Peterson, D.L. and Parker, V.T.), Columbia University Press, New York, pp. 35-54.

Alonso, J.C., Martin, C.A., Alonso, J.A., Palacin, C., Magana, M. and Lane, S.J. 2004. Distribution dynamics of a great bustard metapopulation throughout a decade: influence of conspecific attraction and recruitment. Biodiversity and Conservation, 13: 1659-1674.

Anderson, R.P. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia : Muridae) in Venezuela. Journal of Biogeography, 30: 591-605.

Anderson, R.P. and Martinez-Meyer, E. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador. Biological Conservation, 116: 167-179.

Andren, H. 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: A review. Oikos, 71: 355-366.

Andrews, J. and Rebane, M. 1994. Farming and Wildlife. RSPB Royal Society for the Protection of Birds, 1st Edition, Bedfordshire.

Angelstam, P. 1997. Landscape analysis as a tool for the scientific management of biodiversity. Ecological Bulletins, 46: 140-170.

Araujo, M.B., Thuiller, W., Williams, P.H. and Reginster, I. 2005. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. Global Ecology and Biogeography, 14: 17-30.

Araujo, M.B. and Williams, P.H. 2000. Selecting areas for species persistence using occurrence data. Biological Conservation, 96: 331-345.

Atkinson, P.W., Fuller, R.J. and Vickery, J.A. 2002. Large-scale patterns of summer and winter bird distribution in relation to farmland type in England and Wales. Ecography, 25: 466-480.

Augustin, N.H., Mugglestone, M.A. and Buckland, S.T. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology, 33: 339-347.

Austin, G.E., Thomas, C.J., Houston, D.C. and Thompson, D.B.A. 1996. Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a geographical information system and remote sensing. Journal of Applied Ecology, 33: 1541-1550.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 157: 101-118.

Avois-Jacquet, C., Legendre, P. and Louis, M. 2005. Multiscale spatial variability of zooplankton biomass and environmental forcing in a coastal reef lagoon. Journal of Plankton Research, submitted.

Backman, J.P.C. and Tiainen, J. 2002. Habitat quality of field margins in a Finnish farmland area for bumblebees (Hymenoptera : *Bombus* and *Psithyrus*). Agriculture Ecosystems & Environment, 89: 53-68.

Baguette, M. 2004. The classical metapopulation theory and the real, natural world: a critical appraisal. Basic and Applied Ecology, 5: 213-224.

Baguette, M. and Mennechez, G. 2004. Resource and habitat patches, landscape ecology and metapopulation biology: a consensual viewpoint. Oikos, 106: 399-403.

Balcom, B.J. and Yahner, R.H. 1996. Microhabitat and landscape characteristics associated with the threatened Allegheny woodrat. Conservation Biology, 10: 515-525.

Bani, L., Massimino, D., Bottoni, L. and Massa, R. 2006. A multiscale method for selecting indicator species and priority conservation areas: A case study for broadleaved forests in Lombardy, Italy. Conservation Biology, 20: 512-526.

Baudry, J., Bunce, R.G.H. and Burel, F. 2000. Hedgerows: An international perspective on their origin, function and management. Journal of Environmental Management, 60: 7-22.

Bayliss, J.L., Simonite, V. and Thompson, S. 2005. The use of probabilistic habitat suitability models for biodiversity action planning. Agriculture Ecosystems & Environment, 108: 228-250.

Bellehumeur, C. and Legendre, P. 1998. Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs. Landscape Ecology, 13: 15-25.

Benayas, J.M.R. and de la Montana, E. 2003. Identifying areas of high-value vertebrate diversity for strengthening conservation. Biological Conservation, 114: 357-370.

Berry, J.K. and McGarigal, K. 1998. FRAGSTATS*ARC: Integrating ARC/INFO with the FRAGSTATS Landscape Analysis Program. ESRI 1998 User Conference, San Diego, CA.

Bibby, C.J. 1999. Making the most of birds as environmental indicators. Ostrich, 70: 81-88.

Bibby, C.J., Burgess, N.D. and Hill, D.A. 1992. Bird Census Techniques. Academic Press, Londres.

Bio, A.M.F., Alkemade, R. and Barendregt, A. 1998. Determining alternative models for vegetation response analysis: a non-parametric approach. Journal of Vegetation Science, 9: 5-16.

Bircham, P.M.M. and Jordan, W.J. 1997. A consideration of some of the changes in distribution of "common birds" as revealed by The New Atlas of Breeding Birds in Britain and Ireland. Ibis, 139: 183-186.

BirdLife International 2004. Birds in Europe: population estimates, trends and conservation status. BirdLife Conservation Series No. 12, Cambridge, UK.

Bissonette, J.A. 1997. Wildlife and Landscape Ecology: Effects of Pattern and Scale. Springer Verlag, Wildlife Society Meeting, Portland.

Block, W.M. and Brennan, L.A. 1993. The habitat concept in ornithology: Theory and applications. - In: Current Ornithology. (ed. Power, D.M.), Plenum Press, New York, pp. 35-91.

Blondel, J. 1995. Biogéographie. Approche écologique et évolutive. Masson, Paris.

Bodin, N., Guissard, V., Giot, P., Devillez, F. and Defourny, P. 2006. Inventaire des habitats dans le cadre du suivi de la biodiversité en Région wallonne: apport de la géomatique et des systèmes d'information géographique. - In: Biodiversité: Etat, enjeux et perspectives, Chaire Tractebel-Environnement 2004. Commission de l'Environnement et du Développement Durable, Université catholique de Louvain, pp. 137-158.

Böhning-Gaese, K. 1997. Determinants of avian species richness at different spatial scales. Journal of Biogeography, 24: 49-60.

Böhning-Gaese, K. and Bauer, H.G. 1996. Changes in species abundance, distribution, and diversity in a central European bird community. Conservation Biology, 10: 175-187.

Bonn, A. and Schroder, B. 2001. Habitat models and their transfer for single and multi species groups: a case study of carabids in an alluvial forest. Ecography, 24: 483-496.

Borcard, D. and Legendre, P. 1994. Environmental control and spatial component in ecological communities: an example using Oribatid mites (*Acari*, *Oribatei*). Environmental and Ecological Statistics, 1: 37-53.

Borcard, D. and Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecological Modelling, 153: 51-68.

Borcard, D., Legendre, P., Avois-Jacquet, C. and Tuomisto, H. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology, 85: 1826-1832.

Borcard, D., Legendre, P. and Drapeau, P. 1992. Partialling out the spatial component of ecological variation. Ecology, 73: 1045-1055.

Boren, J.C., Engle, D.M., Palmer, M.W., Masters, R.E. and Criner, T. 1999. Land use change effects on breeding bird community composition. Journal of Range Management, 52: 420-430.

Box, G.E.P. 1976. Science and statistics. Journal of the American Statistical Association, 71: 791-799.

Boyce, M.S. and McDonald, L.L. 1999. Relating populations to habitats using resource selection functions. Trends in Ecology & Evolution, 14: 268-272.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. and Schmiegelow, F.K.A. 2002. Evaluating resource selection functions. Ecological Modelling, 157: 281-300.

Brind'Amour, A., Boisclair, D., Legendre, P. and Borcard, D. 2005. Multiscale spatial distribution of a littoral fish community in relation to environmental variables. Limnology and Oceanography, 50: 465-479.

Brotons, L., Thuiller, W., Araujo, M.B. and Hirzel, A.H. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 27: 437-448.

Burel, F. 1989. Landscape structure effects on carabid beetles spatial patterns in Western France. Landscape Ecology, 2: 215-226.

Burel, F. 1996. Hedgerows and their role in agricultural landscapes. Critical Reviews in Plant Sciences, 15: 169-190.

Burke, V.J. 2000. Landscape ecology and species conservation. Landscape Ecology, 15: 1-3.

Burnham, K.P. and Anderson, D.R. 2002. Model Selection and Multi-model inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York.

Burrough, P.A. 1992. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press, Oxford.

Burrough, P.A. and McDonnell, R.A. 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford.

Busby, J.R. 1991. BIOCLIM - a bioclimate analysis and prediction system. - In: Nature Conservation: Cost effective biological surveys and data analysis (eds Margules, C.R. and Austin, M.P.), CSIRO, Melbourne, pp. 64-68.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., and Soukup, L. 2004. The European Corine Land Cover 2000 Project. XXth Congress of International Sociey for Photogrammetry and Remote Sensing, Istanbul, Turkey.

Cabeza, M., Araujo, M.B., Wilson, R.J., Thomas, C.D., Cowley, M.J.R. and Moilanen, A. 2004. Combining probabilities of occurrence with spatial reserve design. Journal of Applied Ecology, 41: 252-262.

Canova, L. 2006. Protected areas and landscape conservation in the Lombardy plain (northern Italy): an appraisal. Landscape and Urban Planning, 74: 102-109.

Capen, D.E. 1981. The use of multivariate statistics in studies of wildlife habitat. General Technical Report RM-87, Fort Collins, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.

Carleer, A. 2005. Region-based classification potential for land-cover classification with very high spatial resolution satellite data. Thèse de Doctorat, Université Libre de Bruxelles.

Carpenter, G., Gillison, A.N. and Winter, J. 1993. DOMAIN - a flexible modeling procedure for mapping potential distributions of plants and animals. Biodiversity and Conservation, 2: 667-680.

Chamberlain, D.E. and Fuller, R.J. 2000. Local extinctions and changes in species richness of lowland farmland birds in England and Wales in relation to recent changes in agricultural land-use. Agriculture, Ecosystems & Environment, 78: 1-17.

Chamberlain, D.E. and Fuller, R.J. 2001. Constrating patterns of change in the distribution and abundance of farmland birds in relation to farming system in lowland Britain. Global Ecology and Biogeography, 10: 399-409.

Chamberlain, D.E., Fuller, R.J., Bunce, R.G.H., Duckworth, J.C. and Shrubb, M. 2000. Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. Journal of Applied Ecology, 37: 771-788.

Chamberlain, D.E., Fuller, R.J., Garthwaite, D.G. and Impey, A.J. 2001. A comparison of farmland bird density and species richness in lowland England between two periods of contrasting agricultural practice. Bird Study, 48: 245-251.

Chamberlain, D.E. and Gregory, R.D. 1999. Coarse and fine scale habitat associations of breeding Skylarks *Alauda arvensis* in the UK. Bird Study, 46: 34-47.

Chase, J.M. and Leibold, M.A. 2003. Ecological Niches: Linking Classical and Contemporary Approaches. University of Chicago Press, Chicago.

Clobert, J., Danchin, E., Dhondt, A. and Nichols, J.D. 2001. Dispersal. Oxfor University Press, New-York.

Coppedge, B.R., Engle, D.M., Masters, R.E. and Gregory, M.S. 2001. Avian response to landscape change in fragmented southern Great Plains grasslands. Ecological Applications, 11: 47-59.

Corsi, F., De Leeuw, J. and Skidmore, A. 2000. Modelling species distribution with GIS. - In: Landscape ecology and geographical information systems (eds Haines-Young, R.H., Green, D.R. and Cousins, G.S.), pp. 389-434.

Couturier, A. and Robert, P. 1955. Recherches sur les migrations du hanneton commun (*Melolontha melolontha*, L.). Annales des Epiphyties, pp. 19-60.

Couvreur, J.M. and Godeau, J.F. 2000. Atlas des Orthoptères de la Famenne (Criquets, sauterelles et grillons). Jeunes & Nature a.s.b.l., Centre de Recherche de la Nature, des Forêts et du Bois.

Dale, M.R.T., Dixon, P.M., Fortin, M.J., Legendre, P., Myers, D. and Rosenberg, M.S. 2002. Conceptual and mathematical relationships among methods for spatial analysis. Ecography, 25: 558-577.

Danchin, E., Heg, D. and Doligez, B. 2001. Public information and breeding habitat selection. - In: Dispersal (eds Clobert, J., Danchin, E., Dhondt, A. and Nichols, J.), Oxford University Press, New-York, pp. 230-242.

Davies, C.E. and Moss, D. 2002. EUNIS Habitat Classification. Final Report to the European Topic Centre on Nature Conservation, European Environment Agency., 125 pp.

De Blust, G., Paelinckx, D. and Kuijken, E. 1994. Up-to-date information on nature quality for environmental management in Flanders. - In: Ecosystem classification for environmental management. (ed. Klijn, F.), Kluwer Academic Publishers, Dordrecht, Boston, London, pp. 223-249.

de Heer, M., Kapos, V. and ten Brink, B.J.E. 2005. Biodiversity trends in Europe: development and testing of a species trend indicator for evaluating progress towards the 2010 target. Philosophical Transactions of the Royal Society B-Biological Sciences, 360: 297-308.

De Klerk, H.M., Fjeldsa, J., Blyth, S. and Burgess, N.D. 2004. Gaps in the protected area network for threatened Afrotropical birds. Biological Conservation, 117: 529-537.

de Sadeleer, N. and Born, C.-H. 2004. Droit international et communautaire de la biodiversité. Dalloz, Paris.

De'Ath, G. 1999. Principal curves: A new technique for indirect and direct gradient analysis. Ecology, 80: 2237-2253.

De'Ath, G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology, 83: 1105-1117.

Dehhaoui, M. 1996. Analyse multivariée des surfaces de réponse en relation avec les plans expérimentaux. Thèse de doctorat, Université catholique de Louvain.

Delibes, M., Ferreras, P. and Gaona, P. 2001a. Attractive sinks, or how individual behavioural decisions determine source-sink dynamics. Ecology Letters, 4: 401-403.

Delibes, M., Gaona, P. and Ferreras, P. 2001b. Effects of an attractive sink leading into maladaptive habitat selection. American Naturalist, 158: 277-285.

DeMers, M.N. 1997. Fundamentals of Geographic Information Systems. John Wiley and Sons, Inc.

Dendoncker, N., Bogaert, P. and Rounsevell, M. 2006. A statistical method to downscale aggregated landuse data and scenarios. Journal of Land Use Science, submitted.

Dennis, R.L.H. and Hardy, P.B. 1999. Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. Global Ecology and Biogeography, 8: 443-454.

Dennis, R.L.H. and Hardy, P.B. 2001. Loss rates of butterfly species with urban development. A test of atlas data and sampling artefacts at a fine scale. Biodiversity and Conservation, 10: 1831-1837.

Dennis, R.L.H., Shreeve, T.G., Sparks, T.H. and Lhonor, J.E. 2002. A comparison of geographical and neighbourhood models for improving atlas databases. The case of french butterfly atlas. Biological Conservation, 108: 143-159.

Dennis, R.L.H., Shreeve, T.G. and Van Dyck, H. 2003. Towards a functional resource-based concept for habitat: a butterfly biology viewpoint. Oikos, 102: 417-426.

Dennis, R.L.H., Sparks, T.H. and Hardy, P.B. 1999. Bias in butterfly distribution maps: The effects of sampling effort. Journal of Insect Conservation, 3: 33-42.

Dennis, R.L.H. and Thomas, C.D. 2000. Bias in butterfly distribution maps: The influence of hot spots and recorder's home range. Journal of Insect Conservation, 4: 73-77.

Doligez, B., Cadet, C., Danchin, E. and Boulinier, T. 2003. When to use public information for breeding habitat selection? The role of environmental predictability and density dependence. Animal Behaviour, 66 Part 5: 973-988.

Doligez, B., Danchin, E., Clobert, J. and Gustafsson, L. 1999. The use of conspecific reproductive success for breeding habitat selection in a non-colonial, holenesting species, the collared flycatcher. Journal of Animal Ecology, 68: 1193-1206.

Doligez, B., Part, T., Danchin, E., Clobert, J. and Gustafsson, L. 2004. Availability and use of public information and conspecific density for settlement decisions in the collared flycatcher. Journal of Animal Ecology, 73: 75-87.

Donald, P.F. and Evans, A.D. 2006. Habitat connectivity and matrix restoration: the wider implications of agri-environment schemes. Journal of Applied Ecology, 43: 209-218.

Donald, P.F., Evans, A.D., Muirhead, L.B., Buckingham, D.L., Kirby, W.B. and Schmitt, S.I.A. 2002a. Survival rates, causes of failure and productivity of Skylark *Alauda arvensis* nests on lowland farmland. Ibis, 144: 652-664.

Donald, P.F. and Fuller, R.J. 1998. Ornithological atlas data: a review of uses and limitations. Bird Study, 45: 129-145.

Donald, P.F., Green, R.E. and Heath, M.F. 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. Proceedings of the Royal Society of London Series B-Biological Sciences, 268: 25-29.

Donald, P.F., Pisano, G., Rayment, M.D. and Pain, D.J. 2002b. The Common Agricultural Policy, EU enlargement and the conservation of Europe's farmland birds. Agriculture, Ecosystems & Environment, 89: 167-182.

Donovan, T.M. and Thompson, F.R. 2001. Modeling the ecological trap hypothesis: A habitat and demographic analysis for migrant songbirds. Ecological Applications, 11: 871-882.

Donovan, T.M., Thompson, F.R., Faaborg, J. and Probst, J.R. 1995. Reproductive success of migratory birds in habitat sources and sinks. Conservation Biology, 9: 1380-1395.

Dray, S., Legendre, P. and Peres-Neto, P.R. 2005. Spatial modeling : a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). Ecological Modelling, Accepted.

Driscoll, M.J.L., Donovan, T., Mickey, R., Howard, A. and Fleming, K.K. 2005. Determinants of wood thrush nest success: A multi-scale, model selection approach. Journal of Wildlife Management, 69: 699-709.

Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.J., Jakomulska, A., Miriti, M. and Rosenberg, M.S. 2002. A balanced view of scale in spatial statistical analysis. Ecography, 25: 626-640.

Dunning, J.B., Danielson, B.J. and Pulliam, H.R. 1992. Ecological processes that affect populations in complex landscapes. Oikos, 65: 169-175.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29: 129-151.

Elton, C. 1927. Animal Ecology. Sidgwick and Jackson, London.

Engler, R., Guisan, A. and Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology, 41: 263-274.

Etterson, M.A. 2003. Conspecific attraction in loggerhead shrikes: implications for habitat conservation and reintroduction. Biological Conservation, 114: 199-205.

Fabricius, C., Burger, M. and Hockey, P.A.R. 2003. Comparing biodiversity between protected areas and adjacent rangeland in xeric succulent thicket, South Africa: arthropods and reptiles. Journal of Applied Ecology, 40: 392-403.

Fahrig, L. 2001. How much habitat is enough ? Biological Conservation, 100: 65-74.

Fielding, A.H. and Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24: 38-49.

Fielding, A.H. and Haworth, P.F. 1995. Testing the generality of bird-habitat models. Conservation Biology, 9: 1466-1481.

Fleishman, E., MacNally, R., Fay, J.P. and Murphy, D.D. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. Conservation Biology, 15: 1674-1685.

Fleishman, E., Thomson, J.R., Mac Nally, R., Murphy, D.D. and Fay, J.P. 2005. Using indicator species to predict species richness of multiple taxonomic groups. Conservation Biology, 19: 1125-1137.

Forman, R.T.T. 1995. Some general principles of landscape and regional ecology. Landscape Ecology, 10: 133-142.

Forman, R.T.T. and Godron, M. 1986. Landscape ecology. Wiley, New York.

Fuisz, T.I., Mosket, C., and Park, J.Y. 1998. Nest site selection and habitat use in the Red-Backed Shrike (*Lanius collurio*) in Hungary. International Birdwatching Center in Eilat, Israel, pp. 30-33.

Fuller, R.J., Chamberlain, D.E., Burton, N.H.K. and Gough, S.J. 2001. Distributions of birds in lowland agricultural landscapes of England and Wales: How distinctive are bird communities of hedgerows and woodland? Agriculture, Ecosystems & Environment, 84: 79-92.

Fuller, R.J., Gregory, R.D., Gibbons, D.W., Marchant, J.H., Wilson, J.D., Baillie, S.R. and Carter, N. 1995. Population declines and range contractions among lowland farmland birds in Britain. Conservation Biology, 9: 1425-1441.

Gates, S., Gibbons, D.W., Lack, P.C. and Fuller, R.J. 1993. Declining farmland bird species: modelling geographical patterns of abundance in Britain. - In: Large-scale Ecology and Conservation Biology. (eds Edwards, P.J., May, R.M. and Webb, N.R.), Blackwell Scientific Publications, pp. 153-177.

Gibson, L.A., Wilson, B.A., Cahill, D.M. and Hill, J. 2004. Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. Journal of Applied Ecology, 41: 213-223.

Grand, J. and Cushman, S.A. 2003. A multi-scale analysis of species-environment relationships: breeding birds in a pitch pine-scrub oak *(Pinus rigida-Quercus ilicifolia*) community. Biological Conservation, 112: 307-317.

Greaves, R.K., Sanderson, R.A. and Rushton, S.P. 2006. Predicting species occurrence using information-theoretic approaches and significance testing: An example of dormouse distribution in Cumbria, UK. Biological Conservation, 130: 239-250.

Greenwood, J.J.D., Fuller, R.J. and Gibbons, D.W. 1997. Assessing changes in distribution from atlas data: Reply to Bircham and Jordan. Ibis, 139: 186-189.

Gregory, R.D., van Strien, A., Vorisek, P., Meyling, A.W.G., Noble, D.G., Foppen, R.P.B. and Gibbons, D.W. 2005. Developing indicators for European birds. Philosophical Transactions of the Royal Society B-Biological Sciences, 360: 269-288.

Griffith, S.C., Owens, I.P.F. and Thuman, K.A. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. Molecular Ecology, 11: 2195-2212.

Grinnell, J. 1917. The niche relationship of the Californian Thrasher. Auk, 34: 427-433.

Gu, W. and Swihart, R.K. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation, 116: 195-203.

Guisan, A. 2002. Semiquantitative Response Models for Predicting the Spatial Distribution of Plant Species. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 315-326.

Guisan, A., Edwards, T.C. and Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling, 157: 89-100.

Guisan, A. and Harrell, F.E. 2000. Ordinal response regression models in ecology. Journal of Vegetation Science, 11: 617-626.

Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8: 993-1009.

Guisan, A., Weiss, S.B. and Weiss, A.D. 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecology, 143: 107-122.

Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135: 147-186.

Gustafson, E.J. 1998. Quantifying landscape spatial pattern: What is the state of the art? Ecosystems, 1: 143-156.

Hagemeijer, W.J.M. and Blair, M.J. 1997. The EBCC atlas of european breeding birds: their distribution and abundance. Poyser, London/San Diego, 960 pp.

Hanski, I. and Gaggiotti, O.E. 2004. Ecology, Genetics and Evolution of Metapopulations. Elsevier Academic Press, Oxford.

Hanski, I.A. and Gilpin, M.E. 1997. Metapopulation Biology. Academic Press, New York.

Hansson, L. and Angelstam, P. 1991. Landscape ecology as a theoretical basis for nature conservation. Landscape Ecology, 5: 191-201.

Hastie, T., Tibshirani, R. and Friedman, J.H. 2001. The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, New York.

Heglund, P.J. 2002. Foundations of Species-Environment Relations. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 35-41.

Henle, K., Lindenmayer, D.B., Margules, C.R., Saunders, D.A. and Wissel, C. 2004. Species survival in fragmented landscapes: where are we now? Biodiversity and Conservation, 13: 1-8.

Hinsley, S.A. 2000. The costs of multiple patch use by birds. Landscape Ecology, 15: 765-775.

Hinsley, S.A. and Bellamy, P.E. 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.H. 2001. When GIS come to life. Linking landscape- and population ecology for large population management modelling: the case of Ibex (*Capra ibex*) in Switzerland. Thèse de doctorat, Faculté des Sciences, Université de Lausanne.

Hirzel, A.H. and Arlettaz, R. 2003a. Environmental-envelope based habitat-suitability models., Laramie, USA, pp. 67-76.

Hirzel, A.H. and Arlettaz, R. 2003b. Modeling habitat suitability for complex species distributions by environmental-distance geometric mean. Environmental Management, 32: 614-623.

Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N. 2002a. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data ? Ecology, 83: 2027-2036.

Hirzel, A.H., Hausser, J. and Perrin, N. 2002b. Biomapper 3.1.Division of Conservation Biology, University of Bern. URL: http://www.unil.ch/biomapper.

Hirzel, A.H., Helfer, V. and Metral, F. 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling, 145: 111-121.

Hirzel, A.H., Posse, B., Oggier, P.A., Crettenand, Y., Glenz, C. and Arlettaz, R. 2004. Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture. Journal of Applied Ecology, 41: 1103-1116.

Hobson, K.A., Kirk, D.A. and Smith, A.R. 2000. A multivariate analysis of breeding bird species of western and central Canadian boreal forests: Stand and spatial effects. Ecoscience, 7: 267-280.

Holdgate, M.W. 1994. Protected Areas in the Future - the Implications of Change, and the Need for New Policies. Biodiversity and Conservation, 3: 406-410.

Holland, J.M. 2002. The Agroecology of Carabid Beetles. The Game Conservancy Trust, Fordingbridge, Hampshire, UK.

Horvath, R., Kovacs, K., and Farkas, R. 1998. Reproductive biology of the Red-Backed Shrike (*Lanius collurio*) in the Aggteleki Nemzeti Park, Hungary. International Birdwatching Center in Eilat, Israel, pp. 49-50.

Hosmer, D.W. and Lemeshow, S. 2000. Applied logistic regression. John Wiley and Sons, New York.

Hustings, M.F.H., Kwak, R.G.M., Opdam, P.F.M. and Reijnen, M.J.S.M. 1985. Natuurbeheer in Nederland. Deel 3: vogelinventarisatie. Achtergronden, richtlijnen en verlaglegging. Pudoc, Wageningen & Nederlandse Vereniging tot Bescherming van Vogels, Zeist.

Huston, M.A. 2002. Critical Issues for Improving Predictions. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 7-21.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology, 22: 415-427.

I.R.S.I.A. 1966. Carte des sols de la Belgique. Laboratoire de Geopedologie, Faculte universitaire des Sciences agronomiques, Gembloux, Belgique.

Jackson, D.A. 1993. Stopping rules in principal component analysis: A comparison of heuristical and statistical approaches. Ecology, 74: 2204-2214.

Jackson, S.F., Kershaw, M. and Gaston, K.J. 2004. The performance of procedures for selecting conservation areas: waterbirds in the UK. Biological Conservation, 118: 261-270.

Jacob, J.P. 1999. La situation des pies-grièches écorcheur (*Lanius collurio*) et grise (*Lanius excubitor*) en Wallonie (Belgique). Aves, 36: 7-30.

Jacob, J.P. and Paquay, M. 1992. Oiseaux nicheurs de Famenne. Atlas de Lesse et Lomme. Aves, Liège, 360 pp.

Jakober, H. and Stauber, W. 1994. Copulations and mate-guarding in the red-backed shrike *Lanius collurio*. Journal fur Ornithologie, 135: 535-547.

James, F.C., Johnston, R.F., Wamer, N.O., Niemi, G.J. and Boecklen, W.J. 1984. The Grinnellian niche of the Wood Thrush. American Naturalist, 124: 17-47.

Johnson, C.J., Seip, D.R. and Boyce, M.S. 2004. A quantitative approach to conservation planning: using resource selection functions to map the distribution of mountain caribou at multiple spatial scales. Journal of Applied Ecology, 41: 238-251.

Johnson, J.B. and Omland, K.S. 2004. Model selection in ecology and evolution. Trends in Ecology & Evolution, 19: 101-108.

Johnston, C.A. 1998. Geographic Information System in Ecology. Blackwell Science Ltd, Oxford.

Jongman, R.H.G., ter Braak, C.J.F. and van Tongeren, E.F.R. 1995. Data Analysis in Community and Landscape Ecology. Cambridge Univ Press.

Karlsson, S. 2004. Season-dependent diet composition and habitat use of red-backed shrikes *Lanius collurio* in SW Finland. Ornis Fennica, 81: 97-108.

Kayitakire, F. 2006. Forest stand characterization using very high resolution remote sensing. Thèse de doctorat, Université catholique de Louvain.

Keitt, T.H., Bjornstad, O.N., Dixon, P.M. and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography, 25: 616-625.

King, A.W. 1991. Translating models across scales in the landscape. - In: Quantitative methosd in landscape ecology. (eds Turner, M.G. and Gardner, R.H.), Springer-Verlag, New York, pp. 479-517.

Kleijn, D., Baquero, R.A., Clough, Y., Diaz, M., De Esteban, J., Fernandez, F., Gabriel, D., Herzog, F., Holzschuh, A., Johl, R., Knop, E., Kruess, A., Marshall, E.J.P., Steffan-Dewenter, I., Tscharntke, T., Verhulst, J., West, T.M. and Yela, J.L. 2006. Mixed biodiversity benefits of agri-environment schemes in five European countries. Ecology Letters, 9: 243-254.

Kleijn, D. and Sutherland, W.J. 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? Journal of Applied Ecology, 40: 947-969.

Klute, D.S., Lovallo, M.J. and Tzilkowski, W.M. 2002. Autologistic regression modeling of American Woodcock habitat use with spatially dependent data. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 335-343.

Koenig, W.D. 1999. Spatial autocorrelation of ecological phenomena. Trends in Ecology & Evolution, 14: 22-26.

Kokko, H. and Sutherland, W.J. 2001. Ecological traps in changing environments: Ecological and evolutionary consequences of a behaviourally mediated Allee effect. Evolutionary Ecology Research, 3: 537-551.

Kristan, W.B. 2003. The role of habitat selection behavior in population dynamics: source-sink systems and ecological traps. Oikos, 103: 457-468.

Kruess, A. and Tscharntke, T. 2002a. Contrasting responses of plant and insect diversity to variation in grazing intensity. Biological Conservation, 106: 293-302.

Kruess, A. and Tscharntke, T. 2002b. Grazing intensity and the diversity of grasshoppers, butterflies, and trap-nesting bees and wasps. Conservation Biology, 16: 1570-1580.

Kuzniak, S. and Tryjanowski, P. 2000. Distribution and breeding habitat of the red-backed shrike (*Lanius collurio*) in an intensively used farmland. Ring, 22: 89-93.

Kyrkos, A., Wilson, J.D. and Fuller, R.J. 1998. Farmland habitat change and abundance of Yellowhammers *Emberiza citrinella*: an analysis of common birds census data. Bird Study, 45: 232-246.

Lambeck, R.J. 1997. Focal species: A multi-species umbrella for nature conservation. Conservation Biology, 11: 849-856.

Landres, P.B., Verner, J. and Thomas, J.W. 1988. Ecological uses of vertebrate indicator species: a critique. Conservation Biology, 2: 316-328.

Lane, S.J., Alonso, J.C. and Martin, C.A. 2001. Habitat preferences of great bustard Otis tarda flocks in the arable steppes of central Spain: are potentially suitable areas unoccupied? Journal of Applied Ecology, 38: 193-203.

Laymon, S.A. and Barrett, R.H. 1986. Developing and testing habitat-capability models: Pitfalls and recommendations. - In: Wildlife 2000: Modeling habitat relationships of terrestrial vertebrates. (eds Verner, J., Morrison, M.L. and Ralph, C.J.), University of Wisconsin Press, Madison, pp. 87-92.

Lee, M., Fahrig, L., Freemark, K.E. and Currie, D.J. 2002. Importance of patch scale vs landscape scale on selected forest birds. Oikos, 96: 110-118.

Lefranc, N. 1993. Les pie-grièches d'Europe, d'Afrique du Nord et du Moyen-Orient. Delachaux et Niestlé S.A., Paris.

Lefranc, N. 2004. La pie-grièche écorcheur. Eveil Nature, Approche.

Lefranc, N. and Worfolk, T. 1997. Shrikes : a guide to the shrikes of the World. Pica Press, Sussex, England.

Legendre, P. 1993. Spatial autocorrelation - Trouble or new paradigm ? Ecology, 74: 1659-1673.

Legendre, P. and Borcard, D. 2006. Quelles sont les échelles spatiales importantes dans un écosystème ? - In: Analyses statistiques de données spatiales. (eds Droesbeke, J.-J., Lejeune, M. and Saporta, G.), Paris.

Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J., Hohn, M. and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography, 25: 601-615.

Legendre, P. and Legendre, L. 1998. Numerical Ecology. Elsevier Health Sciences B.V., Amsterdam.

Leugger-Eggiman, U. 1997. Parental expenditure of Red-backed Shrike in habitats of varying farming intensity. PhD Thesis, Université de Bâle, Allschwill.

Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology, 73: 1943-1967.

Lichstein, J.W., Simons, T.R., Shriner, S.A. and Franzreb, K.E. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs, 72: 445-463.

Lindenmayer, D.B. and Fischer, J. 2003. Sound science or social hook - a response to Brooker's application of the focal species approach. Landscape and Urban Planning, 62: 149-158.

Liu, C., Berry, P.M., Dawson, T.P. and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28: 385-393.

Lobo, J.M., Lumaret, J.P. and Jay-Robert, P. 2002. Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. Global Ecology and Biogeography, 11: 265-277.

Lobo, J.M. and Martin-Piera, F. 2002. Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. Conservation Biology, 16: 158-173.

Lobo, J.M., Verdu, J.R. and Numa, C. 2006. Environmental and geographical factors affecting the Iberian distribution of flightless Jekelius species (Coleoptera: Geotrupidae). Diversity & Distributions, 12: 179-188.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. and Williams, P.H. 2003. Avoiding Pitfalls of Using Species Distribution Models in Conservation Planning. Conservation Biology, 17: 1591-1600.

Lovejoy, T.E. 2006. Protected areas: a prism for a changing world. Trends in Ecology & Evolution, 21: 329-333.

Luoto, M., Kuussaari, M. and Toivonen, T. 2002. Modelling butterfly distribution based on remote sensing data. Journal of Biogeography, 29: 1027-1037.

Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of 'predictive' and 'explanatory' models. Biodiversity and Conservation, 9: 655-671.

MacFaden, S.W. and Capen, D.E. 2002. Avian habitat relationships at multiple scales in a New England forest. Forest Science, 48: 243-253.

Mack, E.L., Firbank, L.G., Bellamy, P.E., Hinsley, S.A. and Veitch, N. 1997. The comparison of remotely sensed and ground-based habitat area data using species-area models. Journal of Applied Ecology, 34: 1222-1228.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Hines, J.E. and Bailey, L.L. 2005. Occupancy Estimation and Modeling. Inferring Patterns and Dynamics of Species Occurrence. Elsevier, San Diego, USA.

Mackey, B.G. and Lindenmayer, D.B. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. Journal of Biogeography, 28: 1147-1166.

Maes, D., Gilbert, M., Titeux, N., Goffart, P. and Dennis, R.L.H. 2003. Prediction of butterfly diversity hotspots in Belgium: a comparison of statistically-focused and land use-focused models. Journal of Biogeography, 30: 1907-1920.

Magura, T. 2002. Carabids and forest edge: spatial pattern and edge effect. Forest Ecology and Management, 157: 23-37.

Magura, T., Tothmeresz, B. and Molnar, T. 2001. Forest edge and diversity: carabids along forest-grassland transects. Biodiversity and Conservation, 10: 287-300.

Maki-Petays, A., Huusko, A., Erkinaro, J. and Muotka, T. 2002. Transferability of habitat suitability criteria of juvenile Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquatic Sciences, 59: 218-228.

Manel, S., Dias, J.M., Buckton, S.T. and Ormerod, S.J. 1999a. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology, 36: 734-747.

Manel, S., Dias, J.M. and Ormerod, S.J. 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecological Modelling, 120: 337-347.

Manel, S., Williams, H.C. and Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38: 921-931.

Manly, B.F.J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Chapman and Hall, London.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L. and Erickson, W.P. 2002. Resource selection by animals: Statistical design and analysis for field studies. Kluwer Academic Publishers, Dordrecht.

Marsden, S.J., Whiffin, M., Galetti, M. and Fielding, A.H. 2005. How well will Brazil's system of atlantic forest reserves maintain viable bird populations? Biodiversity and Conservation, 14: 2835-2853.

Marzluff, J.M. and Ewing, K. 2001. Restoration of fragmented landscapes for the conservation of birds: A general framework and specific recommendations for urbanizing landscapes. Restoration Ecology, 9: 280-292.

McCracken, D.I. and Bignal, E.M. 1998. Applying the results of ecological studies to land-use policies and practices. Journal of Applied Ecology, 35: 961-967.

McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models. Chapman & Hall, London.

McGarigal, K. and Marks, B. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. General technical report PNW 351, U.S. Forest Service, Corvallis, Oregon.

Meek, B., Loxton, D., Sparks, T., Pywell, R., Pickett, H. and Nowakowski, M. 2002. The effect of arable field margin composition on invertebrate biodiversity. Biological Conservation, 106: 259-271.

Meffe, G.K. 2001. The context of conservation biology. Conservation Biology, 15: 815-816.

Meffe, G.K. and Carroll, C.R. 1997. Principles of Conservation Biology. Sinauer Associates, Sunderland, Massachusetts.

Meot, A., Legendre, P. and Borcard, D. 1998. Partialling out the spatial component of ecological variation: questions and propositions in the linear modelling framework. Environmental and Ecological Statistics, 5: 1-27.

Mitchell, M.S., Lancia, R.A. and Gerwin, J.A. 2001. Using landscape-level data to predict the distribution of birds on a managed forest: Effects of scale. Ecological Applications, 11: 1692-1708.

Mitchell, S.C. 2005. How useful is the concept of habitat? a critique. Oikos, 110: 634-638.

Monti, D., Legendre, L., Therriault, J.C. and Demers, S. 1996. Horizontal distribution of sea-ice microalgae: Environmental control and spatial processes (southeastern Hudson Bay, Canada). Marine Ecology-Progress Series, 133: 229-240.

Morris, M.G. 2000. The effects of structure and its dynamics on the ecology and conservation of arthropods in British grasslands. Biological Conservation, 95: 129-142.

Morrison, M.L. and Hall, L.S. 2002. Standard Terminology: Towards a Common Language to Advance Ecological Understanding and Application. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 43-52.

Morrison, M.L., Marcot, B.G. and Mannan, R.W. 1998. Wildlife-habitat relationships: Concepts and applications. Madison, University of Wisconsin Press.

Muller, K.L., Stamps, J.A., Krishnan, V.V. and Willits, N.H. 1997. The Effects of Conspecific Attraction and Habitat Quality on Habitat Selection in Territorial Birds (*Troglodytes aedon*). American Naturalist, 150: 650-661.

Muller, M., Pasinelli, G., Schiegg, K., Spaar, R. and Jenni, L. 2005. Ecological and social effects on reproduction and local recruitment in the red-backed shrike. Oecologia, 143: 37-50.

Natuhara, Y. and Imai, C. 1996. Spatial structure of avifauna along urban-rural gradients. Ecological Research, 11: 1-9.

Naugle, D.E., Higgins, K.F., Nusser, S.M. and Johnson, W.C. 1999. Scale-dependent habitat use in three species of prairie wetland birds. Landscape Ecology, 14: 267-276.

O' Neill, K.M., Olson, B.E., Rolston, M.G., Wallander, R., Larson, D.P. and Seibert, C.E. 2003. Effects of livestock grazing on rangeland grasshopper (Orthoptera : Acrididae) abundance. Agriculture Ecosystems & Environment, 97: 51-64.

O'Dea, N., Araujo, M.B. and Whittaker, R.J. 2006. How well do important bird areas represent species and minimize conservation conflict in the tropical Andes? Diversity and Distributions, 12: 205-214.

Oldfield, T.E.E., Smith, R.J., Harrop, S.R. and Leader-Williams, N. 2004. A gap analysis of terrestrial protected areas in England and its implications for conservation policy. Biological Conservation, 120: 303-309.

Olivier, F. and Wotherspoon, S.J. 2005. Modelling habitat selection using presence-only data: Case study of a colonial hollow nesting bird, the snow petrel. Ecological Modelling, in press: Corrected Proof-.

Olivier, F. and Wotherspoon, S.J. 2006. Modelling habitat selection using presence-only data: Case study of a colonial hollow nesting bird, the snow petrel. Ecological Modelling, 195: 187-204.

Olson, G.S., Glenn, E.M., Anthony, R.G., Forsman, E.D., Reid, J.A., Loschl, P.J. and Ripple, W.J. 2004. Modeling demographic performance of northern spotted owls relative to forest habitat in Oregon. Journal of Wildlife Management, 68: 1039-1053.

Opdam, P.F.M., Foppen, R. and Vos, C.C. 2001. Bridging the gap between ecology and spatial planning in landscape ecology. Landscape Ecology, 16: 767-779.

Orians, G.H. and Wittenberger, J.F. 1991. Spatial and Temporal Scales in Habitat Selection. American Naturalist, 137: S29-S49.

Osborne, P.E. and Tigar, B.J. 1992. Interpreting bird atlas data using logistic-models - A example from Lesotho, Southern Africa. Journal of Applied Ecology, 29: 55-62.

Ottaviani, D., Lasinio, G.J. and Boitani, L. 2004. Two statistical methods to validate habitat suitability models using presence-only data. Ecological Modelling, 179: 417-443.

Ozesmi, U., Tan, C.O., Ozesmi, S.L. and Robertson, R.J. 2006. Generalizability of artificial neural network models in ecological applications: Predicting nest occurrence and breeding success of the red-winged blackbird Agelaius phoeniceus. Ecological Modelling, 195: 94-104.

Padoa-Schioppa, E., Baietto, M., Massa, R. and Bottoni, L. 2006. Bird communities as bioindicators: The focal species concept in agricultural landscapes. Ecological Indicators, 6: 83-93.

Pain, D.J., Hill, D. and McCracken, D.I. 1997. Impact of agricultural intensification of pastoral systems on bird distributions in Britain 1970-1990. Agriculture, Ecosystems & Environment, 64: 19-32.

Pain, D.J. and Pienkowski, M.W. 1997. Farming and Birds in Europe: The Common Agricultural Policy and Its Implications for Bird Conservation. Academic Press, San Diego.

Palmer, M.W. 1993. Putting things in even better order - The advantages of canonical correspondence-analysis. Ecology, 74: 2215-2230.

Paquet, J.-Y., Laudelout, A., Titeux, N., Vandevyvre, X. and Rondeux, J. 2006. The relative roles of management, environment and spatial structure in explaining variation of early-successional breeding bird community in plantation forest. in prep.

Pasinelli, G., Naef-Daenzer, B., Schmid, H., Keller, V., Holzang, O., Graf, R. and Zbinden, N. 2001. An avifaunal zonation of Switzerland and its relation to environmental conditions. Global Ecology and Biogeography, 10: 261-274.

Peach, W.J., Siriwardena, G.M. and Gregory, R.D. 1999. Long-term changes in over-winter survival rates explain the decline of reed buntings *Emberiza schoeniclus* in Britain. Journal of Applied Ecology, 36: 798-811.

Pearce, J. and Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecological Modelling, 128: 127-147.

Pearce, J.L. and Boyce, M.S. 2005. Modelling distribution and abundance with presence-only data. Journal of Applied Ecology, in press: doi: 10.1111/j;1365-2664.2005.01112.x-.

Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S. and Whish, G. 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. Journal of Applied Ecology, 38: 412-424.

Pearman, P.B., Penskar, M.R., Schools, E.H. and Enander, H.D. 2006. Identifying potential indicators of conservation value using natural heritage occurrence data. Ecological Applications, 16: 186-201.

Pearson, R.G. and Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful ? Global Ecology and Biogeography, 12: 361-371.

Pearson, R.G., Dawson, T.P. and Liu, C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography, 27: 285-298.

Pedrini, P. and Sergio, F. 2002. Regional conservation priorities for a large predator: golden eagles (Aquila chrysaetos) in the Alpine range. Biological Conservation, 103: 163-172.

Penteriani, V., Balbontin, J. and Ferrer, M. 2003. Simultaneous effects of age and territory quality on fecundity in Bonelli's Eagle Hieraaetus fasciatus. Ibis, 145: E77-E82.

Pereira, H.M. and Cooper, H.D. 2006. Towards the global monitoring of biodiversity change. Trends in Ecology & Evolution, 21: 123-129.

Peres-Neto, P.R., Dray, S., Legendre, P. and Borcard, D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology, Accepted.

Perry, J.N., Liebhold, A.M., Rosenberg, M.S., Dungan, J., Miriti, M., Jakomulska, A. and Citron-Pousty, S. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. Ecography, 25: 578-600.

Petit, S. and Firbank, L.G. 2006. Predicting the risk of losing parcels of semi-natural habitat to intensive agriculture. Agriculture, Ecosystems & Environment, 115: 277-280.

Pienkowski, M.W. 1991. Using Long-Term Ornithological Studies in Setting Targets for Conservation in Britain. Ibis, 133: 62-75.

Pinelalloul, B., Niyonsenga, T. and Legendre, P. 1995. Spatial and environmental components of fresh-water zooplankton structure. Ecoscience, 2: 1-19.

Pino, J., Rodà, F., Ribas, J. and Pons, X. 2000. Landscape structure and bird species richness: implications for conservation in rural areas between natural parks. Landscape and Urban Planning, 49: 35-48.

Pope, S.E., Fahrig, L. and Merriam, N.G. 2000. Landscape complementation and metapopulation effects on leopard frog populations. Ecology, 81: 2498-2508.

Pulliam, H.R. 1988. Sources, sinks and population regulation. American Naturalist, 132: 652-661.

Pulliam, H.R. 2000. On the relationship between niche and distribution. Ecology Letters, 3: 349-361.

Pulliam, H.R. and Danielson, B.J. 1991. Sources, Sinks, and Habitat Selection - A Landscape Perspective on Population-Dynamics. American Naturalist, 137: S50-S66.

Pywell, R.F., Warman, E.A., Carvell, C., Sparks, T.H., Dicks, L.V., Bennett, D., Wright, A., Critchley, C.N.R. and Sherwood, A. 2005. Providing foraging resources for bumblebees in intensively farmed landscapes. Biological Conservation, 121: 479-494.

Railsback, S.F., Stauffer, H.B. and Harvey, B.C. 2003. What can habitat preference models tell us? Tests using a virtual trout population. Ecological Applications, 13: 1580-1594.

Reed, J.M. and Dobson, A.P. 1993. Behavioral Constraints and Conservation Biology - Conspecific Attraction and Recruitment. Trends in Ecology & Evolution, 8: 253-256.

Rhodes, J.R., Wiegand, T., McAlpine, C.A., Callaghan, J., Lunney, D., Bowen, M. and Possingham, H.P. 2006. Modeling Species' Distributions to Improve Conservation in Semiurban Landscapes: Koala Case Study. Conservation Biology, 20: 449-459.

Ricklefs, R.E. 1987. Community diversity: relative roles of local and regional processes. Science, 235: 167-171.

Robbins, C.S., Droege, S. and Sauer, J.R. 1989. Monitoring bird populations with Breeding Bird Survey and atlas data. Annales Zoologici Fennici, 26: 297-304.

Roberge, J.M. and Angelstam, P. 2004. Usefulness of the umbrella species concept as a conservation tool. Conservation Biology, 18: 76-85.

Robertson, A., Simmons, R.E., Jarvis, A.M. and Brown, C.J. 1995. Can bird atlas data be used to estimate population-size - A case-study using Namibian endemics. Biological Conservation, 71: 87-95.

Robertson, M.P., Caithness, N. and Villet, M.H. 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. Diversity and Distributions, 7: 15-27.

Robertson, M.P., Villet, M.H. and Palmer, A.R. 2004. A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. Diversity and Distributions, 10: 461-474.

Robinson, S.K., Thompson, F.R., Donovan, T.M., Whitehead, D.R. and Faaborg, J.R. 1995. Regional forest fragmentation and the nesting success of migratory birds. Science, 267: 1987-1990.

Rodrigues, A.S.L., Akcakaya, H.R., Andelman, S.J., Bakarr, M.I., Boitani, L., Brooks, T.M., Chanson, J.S., Fishpool, L.D.C., da Fonseca, G.A.B., Gaston, K.J., Hoffmann, M., Marquet, P.A., Pilgrim, J.D., Pressey, R.L., Schipper, J., Sechrest, W., Stuart, S.N., Underhill, L.G., Waller, R.W., Watts, M.E.J. and Yan, X. 2004a. Global gap analysis: Priority regions for expanding the global protected-area network. Bioscience, 54: 1092-1100.

Rodrigues, A.S.L., Andelman, S.J., Bakarr, M.I., Boitani, L., Brooks, T.M., Cowling, R.M., Fishpool, L.D.C., da Fonseca, G.A.B., Gaston, K.J., Hoffmann, M., Long, J.S., Marquet, P.A., Pilgrim, J.D., Pressey, R.L., Schipper, J., Sechrest, W., Stuart, S.N., Underhill, L.G., Waller, R.W., Watts, M.E.J. and Yan, X. 2004b. Effectiveness of the global protected area network in representing species diversity. Nature, 428: 640-643.

Roos, S. 2002. Functional response, seasonal decline and landscape differences in nest predation risk. Oecologia, 133: 608-615.

Roos, S. and Part, T. 2004. Nest predators affect spatial dynamics of breeding red-backed shrikes (*Lanius collurio*). Journal of Animal Ecology, 73: 117-127.

Rothhaupt, G. and Klein, H. 1998. Usable area as a key factor for habitat occupancy in the Great Grey Shrike. International Birdwatching Center in Eilat, Israel, pp. 64-67.

Rushton, S.P., Ormerod, S.J. and Kerby, G. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology, 41: 193-200.

SAS Institute 1990. SAS/STAT ® User's guide. Version 6, 4th edn. SAS Institute Inc., Cary, NC.

Schlaepfer, M.A., Runge, M.C. and Sherman, P.W. 2002. Ecological and evolutionary traps. Trends in Ecology & Evolution, 17: 474-480.

Schmid, H., Luder, R., Naef-Daenzer, B., Graf, R. and Zbinden, N. 1998. Atlas des oiseaux nicheurs de Suisse 1993-1996. Station ornithologique suisse, Sempach, Suisse, 574 pp.

Schmit, C., Rounsevell, M.D.A. and La Jeunesse, I. 2006. The limitations of spatial land use data in environmental analysis. Environmental Science & Policy, 9: 174-188.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B. 2002. Predicting species occurrences: issues of accuracy and scale. Island Press, Colelo, Washington.

Scozzafava, S. and De Sanctis, A. 2006. Exploring the effects of land abandonment on habitat structures and on habitat suitability for three passerine species in a highland area of Central Italy. Landscape and Urban Planning, 75: 23-33.

Segurado, P. and Araujo, M.B. 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography, 31: 1555-1568.

Seoane, J., Bustamante, J. and Diaz-Delgado, R. 2005. Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conservation Biology, 19: 512-522.

Seoane, J., Justribo, J.H., Garcia, F., Retamar, J., Rabadan, C. and Atienza, J.C. 2006. Habitat-suitability modelling to assess the effects of land-use changes on Dupont's lark Chersophilus duponti: A case study in the Layna Important Bird Area. Biological Conservation, 128: 241-252.

Serrano, D. and Tella, J.L. 2003. Dispersal within a spatially structured population of lesser kestrels: the role of spatial isolation and conspecific attraction. Journal of Animal Ecology, 72: 400-410.

Shochat, E., Patten, M.A., Morris, D.W., Reinking, D.L., Wolfe, D.H. and Sherrod, S.K. 2005. Ecological traps in isodars: effects of tallgrass prairie management on bird nest success. Oikos, 111: 159-169.

Simberloff, D. 1998. Flagships, umbrellas, and keystones: Is single-species management passe in the landscape era ? Biological Conservation, 83: 247-257.

Simek, J. 2001. Patterns of breeding fidelity in the red-backed shrike (*Lanius collurio*). Ornis Fennica, 78: 61-71.

Siriwardena, G.M., Baillie, S.R., Buckland, S.T., Fewster, R.M., Marchant, J.H. and Wilson, J.D. 1998. Trends in the abundance of farmland birds: a quantitative comparison of smoothed Common Birds Census indices. Journal of Applied Ecology, 35: 24-43.

Siriwardena, G.M., Baillie, S.R., Crick, H.Q.P. and Wilson, J.D. 2001. Changes in agricultural land-use and breeding performance of some granivorous farmland passerines in Britain. Agriculture, Ecosystems & Environment, 84: 191-206.

Siriwardena, G.M., Crick, H.Q.P., Baillie, S.R. and Wilson, J.D. 2000a. Agricultural habitat-type and the breeding performance of granivorous farmland birds in Britain. Bird Study, 47: 66-81.

Siriwardena, G.M., Crick, H.Q.P., Baillie, S.R. and Wilson, J.D. 2000b. Agricultural land-use and the spatial distribution of granivorous lowland farmland birds. Ecography, 23: 702-719.

Smallwood, K.S. 2002. Habitat Models Based on Numerical Comparisons. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 83-95.

Smith, P.A. 1994. Autocorrelation in logistic regression modelling of species' distributions. Global Ecology and Biogeography Letters, 4: 47-61.

Söderström, B. 2001. Seasonal change in red-backed shrike *Lanius collurio* territory quality - the role of nest predation. Ibis, 143: 561-571.

Söderström, B., Part, T., Erikson, S., and Kavanagh, R.P. 1998. The effect of breeding habitat and nest site choice on Red-Backed Shrike (*Lanius collurio*) reproductive succes : a comparison of clearcuts and grasslands. International Birdwatching Center in Eilat, Israel, pp. 74-78.

Sokal, R.R. and Rohlf, F.J. 1998. Biometry : the principles and practice of statistics in biological research. W.H. Freeman and Company, New York, USA, 3rd edition.

Soulé, M.E. 1986. Conservation biology, the science of scarcity and diversity. Sinauer Associates, Sunderland, Massachussets.

Southwood, T.R.E. 1977. Habitat, the templet for ecological strategies. Journal of Animal Ecology, 46: 337-365.

Sparks, T.H., Parish, T. and Hinsley, S.A. 1996. Breeding birds in field boundaries in an agricultural landscape. Agriculture, Ecosystems & Environment, 60: 1-8.

Stamps, J.A. 1988. Conspecific attraction and aggregation in territorial species. American Naturalist, 131: 329-347.

Stamps, J.A. 2001. Habitat selection by dispersers: integrating proximate and ultimate approaches. - In: Dispersal (eds Clobert, J., Danchin, E., Dhondt, A. and Nichols, J.), Oxford University Press, New-York, pp. 230-242.

Stamps, J.A. and Krishnan, V.V. 2005. Nonintuitive cue use in habitat selection. Ecology, 86: 2860-2867.

Stauffer, D.F. 2002. Linking Populations and Habitats: Where have we been? Where Are We Going? - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 53-61.

Stephens, P.A., Buskirk, S.W., Hayward, G.D. and Del Rio, C.M. 2005. Information theory and hypothesis testing: a call for pluralism. Journal of Applied Ecology, 42: 4-12.

Stephens, P.A. and Sutherland, W.J. 1999. Consequences of the Allee effect for behaviour, ecology and conservation. Trends in Ecology & Evolution, 14: 401-405.

Stevens, V.M. 2006. Dispersal in the Natterjack toad, a species with a fragmented distribution - Spatial ecology and conservation of a pond-breeding amphibian. Thèse de doctorat, Université catholique de Louvain, Faculté des Sciences, Ecole doctorale en biodiversité.

Stoate, C., Boatman, N.D., Borralho, R.J., Rio Carvalho, C., de Snoo, G.R. and Eden, P. 2001. Ecological impacts of arable intensification in Europe. Journal of Environmental Management, 63: 337-365.

Stoner, K.J.L. and Joern, A. 2004. Landscape vs. local habitat scale influences to insect communities from tallgrass prairie remnants. Ecological Applications, 14: 1306-1320.

Storch, D., Konvicka, M., Benes, J., Martinkova, J. and Gaston, K.J. 2003. Distribution patterns in butterflies and birds of the Czech Republic : separating effects of habitat and geographical position. Journal of Biogeography, 30: 1195-1205.

Stowe, T.J., Newton, A.V., Green, R.E. and Mayes, E. 1993. The decline of the corncrake *Crex crex* in Britain and Ireland in relation to habitat. Journal of Applied Ecology, 30: 53-62.

Swihart, R.K. and Moore, J.E. 2004. Conserving Biodiversity in Agricultural Landscapes: Model-Based Planning Tools. Purdue University Press, West Lafayette, IN, USA.

Telfer, M.G., Preston, C.D. and Rothery, P. 2002. A general method for measuring relative change in range size from biological atlas data. Biological Conservation, 107: 99-109.

ter Braak, C.J.F. 1986. Canonical correspondance analysis: a new eigen vector technique for multivariate direct gradient analysis. Ecology, 67: 1167-1179.

ter Braak, C.J.F. and Smilauer, P. 1998. CANOCO Reference Manual and User's Guide to Canoco for Windows : Software for Canonical Community Ordination (version 4). Microcomputer Power, Ithaca, NY, USA.

Thill, A. 1964. La Flore et la Végétation du Parc National de Lesse et Lomme., Ardenne et Gaume, Monographie 5.

Timothy, J. and Sharrock, R. 1974. Minutes of the second meeting of the European Ornithological Committee. Acta Ornithologica, 14: 404-411.

Tischendorf, L. 2001. Can landscape indices predict ecological processes consistently ? Landscape Ecology, 16: 235-254.

Titeux, N. 2002. Apport d'un atlas ornithologique à l'évaluation de la structuration des peuplements et à la modélisation spatiale de la sélection de l'habitat par l'avifaune. Mémoire de Diplôme d'Etudes Approfondies en Biologie, Université catholique de Louvain.

Titeux, N., Dufrêne, M. and Defourny, P. 2006a. The Importance of Conspecific Proximity for the Conservation of Territorial Migratory Bird Species - The Case of the Red-backed Shrike. Landscape Ecology, to be submitted.

Titeux, N., Dufrêne, M., Hirzel, A.H. and Defourny, P. 2006b. Modelling habitat suitability with presence-only data: we need to account for unreliable species occurrences. Ecography, submitted.

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. and Defourny, P. 2006c. Identifying the Habitat Requirements of the Red-backed Shrike (*Lanius collurio*) – The Importance of Breeding Success. Journal of Applied Ecology, submitted.

Tobalske, C. and Tobalske, B.W. 1999. Using atlas data to model the distribution of woodpecker species in the Jura, France. Condor, 101: 472-483.

Tole, L. 2006. Choosing reserve sites probabilistically: A Colombian Amazon case study. Ecological Modelling, 194: 344-356.

Trine, C.L. 1998. Wood thrush population sinks and implications for the scale of regional conservation strategies. Conservation Biology, 12: 576-585.

Tryjanowski, P., Karg, M.K. and Karg, J. 2003. Diet composition and prey choice by the red-backed shrike *Lanius collurio* in western Poland. Belgian Journal of Zoology, 133: 157-162.

Tryjanowski, P., Kuzniak, S. and Diehl, B. 2000. Does breeding performance of red-backed shrike *Lanius collurio* depend on nest site selection ? Ornis Fennica, 77: 137-141.

Trzcinski, M.K., Fahrig, L. and Merriam, G. 1999. Independent effects of forest cover and fragmentation on the distribution of forest breeding birds. Ecological Applications, 9: 586-593.

Tucker, G.M. and Evans, M.I. 1997. Habitats for Birds in Europe - A conservation Strategy for the Wider Environment. BirdLife International (BirdLife Conservation Series no. 6), Cambridge, UK.

Turner, M.G. 1989. Landscape ecology: the effect of pattern on process. Annual Review of Ecology and Systematics, 20: 171-197.

Turner, M.G., Gardner, R.H. and O'Neill, R.V. 2001. Landscape Ecology in Theory and Practice: Pattern and Process. Springer Verlag.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. and Possingham, H.P. 2003. Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. Ecological Applications, 13: 1790-1801.

Urban, D.L., O'Neill, R.V. and Shugart, H.H. 1987. Landscape ecology: A hierarchical perspective can help scientist understand spatial patterns. Bioscience, 37: 119-127.

Vaisanen, R.A. 1998. Changes in the distribution of species between two bird atlas surveys: the analysis of covariance after controlling for research activity. Ornis Fennica, 75: 53-67.

Van Horne, B. 1983. Density as a misleading indicator of habitat quality. Journal of Wildlife Management, 47: 893-901.

Van Horne, B. 2002. Approaches to Habitat Modelling : the Tensions between Pattern and Processes and between Specificity and Generality. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 7-21.

Van Nieuwenhuyse, D. 1998a. Conservation opportunities for the Red-Backed Shrike (*Lanius collurio*). International Birdwatching Center in Eilat, Israel, pp. 79-82.

Van Nieuwenhuyse, D. 1999. Global Shrike Conservation : problems, methods and opportunities. Aves, 36: 193-204.

Van Nieuwenhuyse, D. 2000b. Dispersal patterns of the Red-backed Shrike (*Lanius collurio*) in Gaume, Belgium. Ring, 22: 65-78.

Van Nieuwenhuyse, D., Nollet, F. and Evans, A.D. 1999. The ecology and conservation of the Red-backed Shrike *Lanius collurio* breeding in Europe. Aves, 36: 179-192.

Van Nieuwenhuyse, D. and Vandekerkhove, K. 1992. Caractéristiques et typologie des territoires de la pie-grièche écorcheur (*Lanius collurio*) en Lorraine belge. Aves, 29: 137-154.

Van Teeffelen, A.J.A., Cabeza, M. and Moilanen, A. 2006. Connectivity, probabilities and persistence: Comparing reserve selection strategies. Biodiversity and Conservation, 15: 899-919.

Vanhinsbergh, D.P. and Chamberlain, D.E. 2001. Habitat associations of breeding Meadow Pipits *Anthus pratensis* in the British uplands. Bird Study, 48: 159-172.

Vanreusel, W., Maes, D. and Van Dyck, H. 2006. Transferability of species distribution models: a functional habitat approach for two regionally threatened butterflies. Conservation Biology, in press.

Vaughan, I.P. and Ormerod, S.J. 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conservation Biology, 17: 1601-1611.

Verburg, P.H., Schulp, C.J.E., Witte, N. and Veldkamp, A. 2006. Downscaling of land use change scenarios to assess the dynamics of European landscapes. Agriculture, Ecosystems & Environment, 114: 39-56.

Vos, C.C., Verboom, J., Opdam, P.F.M. and ter Braak, C.J.F. 2001. Toward ecologically scaled landscape indices. American Naturalist, 157: 24-41.

Walker, P.A. and Cocks, K.D. 1991. HABITAT - a procedure for modeling a disjoint environmental envelope for a plant or animal species. Global Ecology and Biogeography, 1: 108-118.

Ward, M.P. and Schlossberg, S. 2004. Conspecific attraction and the conservation of territorial songbirds. Conservation Biology, 18: 519-525.

Watson, J. and Whitfield, P. 2002. A conservation framework for the Golden Eagle (Aquila chrysaetos) in Scotland. Journal of Raptor Research, 36: 41-49.

Welch, N.E. and MacMahon, J.A. 2005. Identifying habitat variables important to the rare Columbia spotted frog in utah (USA): An information-theoretic approach. Conservation Biology, 19: 473-481.

Whittaker, R.H., Levin, S.A. and Root, R.B. 1973. Niche, habitat, and ecotope. American Naturalist, 107: 321-338.

Whittingham, M.J., Wilson, J.D. and Donald, P.F. 2003. Do habitat association models have any generality? Predicting skylark Alauda arvensis abundance in different regions of southern England. Ecography, 26: 521-531.

Wiens, J.A. 1989a. Spatial scaling in ecology. Functional Ecology, 3: 385-397.

Wiens, J.A. 1989b. The ecology of bird community. Volume1: Foundation and patterns. Cambridge University Press, Cambridge.

Wiens, J.A. 1989c. The ecology of bird community. Volume2: Processes and variations. Cambridge University Press, Cambridge.

Wiens, J.A. 1992. What is landscape ecology, really. Landscape Ecology, 7: 149-150.

Wiens, J.A. 2001. The landscape context of dispersal. - In: Dispersal (eds Clobert, J., Danchin, E., Dhondt, A. and Nichols, J.), Oxford University Press, New-York, pp. 96-109.

Wiens, J.A. 2002. Predicting Species Occurrences: Progress, Problems, and Prospects. - In: Predicting species occurrences: issues of accuracy and scale (eds Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. and Samson, F.B.), Island Press, Colelo, Washington, pp. 739-749.

Wiens, J.A. and Milne, B.T. 1989. Scaling of 'landscapes' in in landscape ecology, or, landscape ecology from a beetle's perspective. Landscape Ecology, 3: 87-96.

Wiens, J.A. and Rotenberry, J.T. 1981. Habitat associations and community structure of birds in shrubsteppe environments. Ecological Monographs, 51: 21-41.

Wiens, J.A., Rotenberry, J.T. and Van Horne, B. 1986. A lesson in the limitation of field experiments: shrubsteppe birds and habitat alternation. Ecology, 67: 365-376.

Wiens, J.A., Stenseth, N.C., Horne, B.V. and Ims, R.A. 1993. Ecological mechanisms and landscape ecology. Oikos, 66: 369-380.

Wilcox, B.A. and Murphy, D.D. 1985. Conservation strategy: the effects of fragmentation on extinction. American Naturalist, 125: 879-887.

Williams, B.K. and Nichols, J.D. 1984. Optimal timing in biological processes. American Naturalist, 123: 1-19.

Wintle, B.A., McCarthy, M.A., Parris, K.M. and Burgman, M.A. 2004. Precision and bias of methods for estimating point survey detection probabilities. Ecological Applications, 14: 703-712.

With, K.A., Gardner, R.H. and Turner, M.G. 1997. Landscape connectivity and population distributions in heterogeneous environments. Oikos, 78: 151-169.

With, K.A., Schrott, G.R. and King, A.W. 2006. The implications of metalandscape connectivity for population viability in migratory songbirds. Landscape Ecology, 21: 157-167.

Woodward, A.A., Fink, A.D. and Thompson, F.R. 2001. Edge effects and ecological traps: Effects on shrubland birds in Missouri. Journal of Wildlife Management, 65: 668-675.

Wu, J. 2004. Effects of changing scale on landscape patterns analysis: scaling relations. Landscape Ecology, 19: 125-138.

Yosef, R. 1993. Influence of observation posts on territory size of Northern Shrikes. Wilson Bulletin, 105: 180-183.

Yosef, R. 1994. Evaluation of the Global Decline in the True Shrikes (Family *Laniidae*). Auk, 111: 228-233.

Young, T.P. 2000. Restoration ecology and conservation biology. Biological Conservation, 92: 73-83.

Zaniewski, A.E., Lehmann, A. and Overton, J.M.C. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling, 157: 261-280.

Zimmermann, N.E. and Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. Journal of Vegetation Science, 10: 469-482.

Zweig, M.H. and Campbell, G. 1993. Receiver-operating characteristic (ROC) plots - A fundamental evaluation tool in clinical medicine. Clinical Chemistry, 39: 561-577.