Universitat de Lleida

Species distribution models for birds How useful are their outcomes for conservation applications?

Olatz Aizpurua San Roman

http://hdl.handle.net/10803/667947

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

Universitat de Lleida

TESI DOCTORAL

Species distribution models for birds How useful are their outcomes for conservation applications?

Olatz Aizpurua San Roman

Memòria presentada per optar al grau de Doctor per la
Universitat de Lleida
Programa de Doctorat en Gestió forestal i medi natural

Directors
Nicolas Titeux
Lluís Brotons

Tutor
Jose Antonio Bonet

Calendars and clocks exist to measure time, but that signifies little because we all know that an hour can seem as eternity or pass in a flash, according to how we spend it.

Michael Ende, *Momo*

Table of content

Summary

Species distribution models are increasingly used to guide biodiversity conservation actions. These models predict the probability of species occurrence in locations where the species presence is unknown based on the link between species presence and environmental conditions. Probability of species occurrence is often considered as indicator of habitat quality. Correctly interpreting the outcomes of this modelling technique is of paramount importance before using the models in conservation applications. The main objective of this thesis is to contribute to providing evidence of the usefulness and applicability of species distribution models for some conservation and management applications.

Despite the growing research interest about species distribution models, many conservation practitioners remain sceptical about their usefulness in biodiversity conservation projects. We show that model outcomes are as able as local bird experts to identify unknown presence areas for a nearly threatened bird species. We also present an innovative analytical framework using data from breeding bird atlases to help in the initial design of monitoring projects. The data generated through this monitoring projects would be appropriate to produce accurate species distribution models and maps. We evaluated the reliability of species distribution models using measures of reproductive performance with data obtained from constant effort bird ringing sites. We showed that models may help to predict habitat quality but not for all species at any spatial scale. We also showed that the outcomes of species distribution models might provide misleading information to guide the spatial prioritization of management or conservation options when species switch or expand to a novel habitat.

With our results, we encourage the use of species distribution models among practitioners as an accepted tool to support biodiversity conservation and management. However, caution is needed when interpreting model outcomes, especially in the areas where a species occupies several habitat types and when novel conditions are emerging as a result of human-induced rapid environmental changes.

Resum

Els models de distribució d'espècies s'utilitzen cada vegada més per orientar les accions de conservació de la biodiversitat. Aquests models prediuen la probabilitat d'aparició d'espècies en llocs on la seva presència es desconeix i són funció del vincle conegut entre la presència de l’espècie i les condicions ambientals. La probabilitat d'aparició d'espècies sovint es considera com un indicador de la qualitat de l'hàbitat. La interpretació correcta dels resultats d'aquesta tècnica de modelatge és de primordial importància abans d'utilitzar els models en aplicacions de conservació. L'objectiu principal d'aquesta tesi és contribuir a evidenciar la utilitat i aplicabilitat dels models de distribució d'espècies en la conservació i gestió de la biodiversitat.

Malgrat l'interès creixent de la recerca sobre models de distribució d'espècies, molts professionals de la conservació segueixen sent escèptics sobre la seva utilitat en projectes de conservació de la biodiversitat. En aquesta tesi es demostra que els resultats dels models són igual de capaços que els ornitòlegs experts locals per identificar àrees potencials de presència per a una espècie d'ocell gairebé amenaçada. També es presenta un marc analític innovador que utilitza dades d'atles d’ocells per ajudar en el disseny inicial de projectes de monitoratge. Aquests projectes podrien generar dades apropiades per produir models i mapes de distribució d'espècies d’alta precisió. Per altre costat, s’ha avaluat la fiabilitat dels models de distribució d’espècies utilitzant mesures de rendiment reproductiu procedents de dades d’estacions d’anellament. S’ha demostrat que els models poden ajudar a predir la qualitat de l'hàbitat però no per a totes les espècies i depenent de l‘escala espacial. També s’ha demostrat que els resultats obtinguts dels models de distribució d'espècies poden proporcionar informació espacial enganyosa a l’hora de prioritzar les opcions de gestió o conservació quan les espècies canvien o s'expandeixen a un hàbitat nou.

Amb aquests resultats, es pretén fomentar l'ús de models de distribució d'espècies entre els professionals com a eina acceptada per donar suport a la conservació i gestió de la biodiversitat. Tanmateix, cal tenir precaució a l'hora d'interpretar els resultats dels models, especialment en aquelles zones on una espècie ocupa diversos tipus d'hàbitats i quan sorgeixen noves condicions com a conseqüència dels canvis ambientals ràpids induïts per l'humà.

Resumen

Los modelos de distribución de especies se utilizan cada vez más para guiar las acciones de conservación de la biodiversidad. Estos modelos predicen la probabilidad de ocurrencia de especies en lugares donde se desconoce, basándose en la relación conocida entre la presencia de la especie y las condiciones ambientales. La probabilidad de ocurrencia de especies a menudo se considera como un indicador de la calidad del hábitat. Interpretar correctamente los resultados de esta técnica de modelización es de suma importancia antes de utilizar los modelos en aplicaciones de conservación. El objetivo principal de esta tesis es contribuir a proporcionar evidencia de la utilidad y aplicabilidad de los modelos de distribución de especies en la conservación y gestión de la biodiversidad.

A pesar del creciente interés de la investigación sobre los modelos de distribución de especies, muchos profesionales de la conservación siguen siendo escépticos sobre su utilidad en proyectos de conservación de la biodiversidad. En la presente tesis se demuestra que los resultados de los modelos son tan capaces como los ornitólogos expertos de identificar áreas de presencia desconocidas para una especie de ave en declive. También presentamos un marco analítico innovador que utiliza datos de atlas de aves reproductoras para ayudar en el diseño inicial de proyectos de monitoreo, los cuales podrían generar datos apropiados para producir mapas y modelos de distribución de especies de alta precisión. Se ha evaluado la fiabilidad de los modelos de distribución de especies utilizando medidas de productividad reproductiva con datos obtenidos en estaciones de anillamiento de aves. A partir de estos datos se ha demostramos que los modelos pueden ayudar a predecir la calidad del hábitat, pero no para todas las especies ni a cualquier escala espacial. También demostramos que los resultados de los modelos de distribución de especies podrían proporcionar información engañosa en el momento de orientar la priorización de las opciones de gestión o conservación a nivel espacial cuando las especies cambian o se expanden a un hábitat nuevo.

Estos resultados ayudan a fomentar el uso de modelos de distribución de especies entre los profesionales como una herramienta aceptada para apoyar la conservación y gestión de la biodiversidad. Sin embargo, se debe tener precaución al interpretar los resultados del modelo, especialmente en las áreas donde una especie ocupa varios tipos de hábitat y cuando surgen condiciones nuevas como resultado de cambios ambientales rápidos inducidos por el humano.

Introduction

The knowledge on where species occur in space, i.e. their geographical distribution, and the changes in these distribution over time, are prerequisite for the efficient conservation and management of a variety of taxa (e.g. Araújo & Williams 2000; Brotons, Herrando, & Pla 2007). Many biodiversity conservation decision-making projects, such as those related to reserve selection (e.g. Cabeza & Moilanen 2001), management of biological invasions (e.g. Gormley *et al.* 2011) or identification of key habitats for threatened species (e.g. Brotons, Mañosa, & Estrada 2004), require this type of information.

Depending on the objective of the study and the resources available, different sampling designs and methods can be used to gain knowledge on the distribution of species. For instance, if a study focuses on a single species, a targeted sample design would be most appropriate (Anadón *et al.* 2007; Hollander *et al.* 2017). On the other hand, the objective many studies have is to track the distribution of many taxa (e.g. Kremen *et al.* 2008; Titeux, Moes, & Hoffmann 2009) or a variety of species within a taxon (e.g. Van der Wal *et al.* 2015, Luoto *et al.* 2006, Titeux *et al.* 2009); in these cases a more general purpose sampling design would be more adequate.

Collecting reliable and comprehensive information on the distribution of a species can be costly in terms of manpower and may necessitate a considerable amount of time. Very often, management decisions need, however, to be made within a relatively short period. An interesting alternative to obtain knowledge on the whereabouts of a species is to use models and predict its likely distribution based on existing and potentially incomplete information.

SPECIES DISTRIBUTION MODELS

Species distribution models (SDMs) have become a very popular tool in ecological research to predict the probability or likelihood of occurrence of a species based on the statistical link between distribution data and variables describing the environmental conditions (e.g. climate, vegetation, land use; Guisan & Zimmermann 2000; Elith & Leathwick 2009; Franklin 2013). SDMs relate species occurrence records with environmental variables to estimate species response curves and predict the distribution of the species in areas where its occurrence in unknown (Fig.1). Most techniques analyse the environmental conditions that are used by the species relative to those that are either available to them or not used (Warton & Aarts 2013; Lele et al. 2013; Guillera-Arroita et al. 2015). The derived continuous distribution maps are tools increasingly used in many different aspects related to biodiversity conservation and management (Chefaoui, Hortal, & Lobo 2005; Schmolke et al. 2010; Barbosa, Real, & Vargas 2010; Guisan et al. 2013; Kearney & Porter 2014; Acevedo et al. 2014). For instance, SDMs are used to inform on potentially suitable habitats in areas where the species presence has not been sampled, which can be useful e.g. in conservation prioritization, planning and reserve design. SDMs are also frequently used as tools to project species distribution in other regions or in the future (Martin et al. 2013). Examples include environmental impact assessments, predicting the effects of global change on ecosystems (Ackerly 2003; Peterson et al. 2004), predicting the risk of pathogens and exotic species spread into new regions (Peterson & Vieglais 2001), and ecological restoration and species reintroductions (Wright et al. 2006; Franklin 2009). Despite the wide range of applications, some practitioners remain poorly inclined to rely on modelling outcomes for on-ground interventions (Jeltsch et al. 2013) and a stronger linkage between modelling science and conservation practice has been advocated (Guisan et al. 2013).

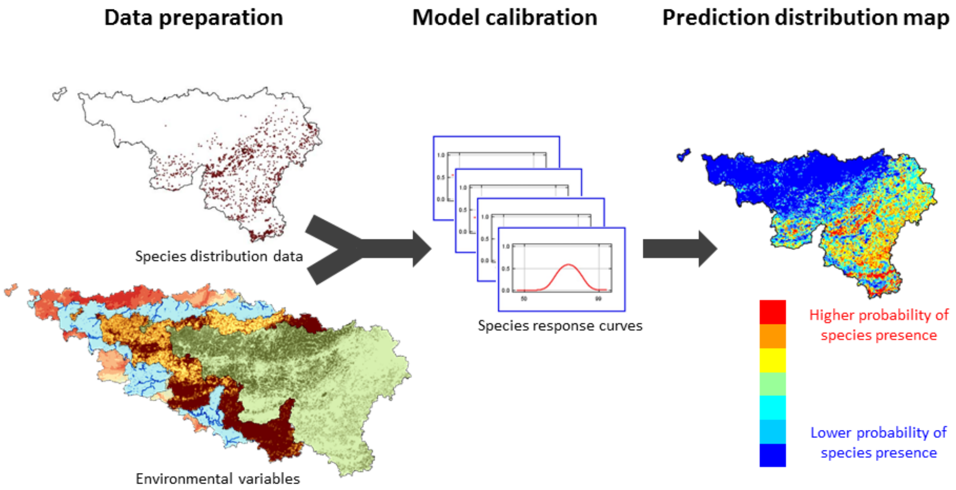
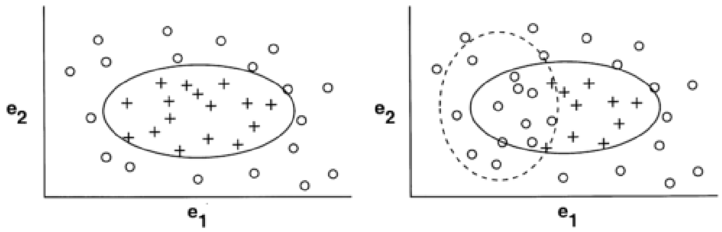


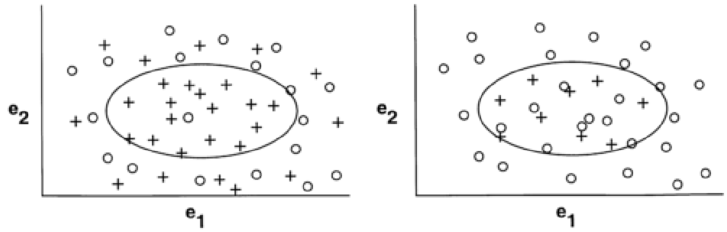
Figure 1. Simplified scheme of the main steps involved in predicting the potential distribution of a species using species distribution models (SDMs).

Niche theory

Ecological niche theory is essential to understand species distributions and is the basis to the development of species distribution models (Austin 2002; Guisan & Thuiller 2005; Araújo & Guisan 2006; Hirzel & Le Lay 2008; Peterson 2011). The ecological niche is a central concept in ecology and even though it was first described in the mid-20th century, it keeps constantly modernising and driving debates in the scientific literature until today (Chase & Leibold 2003; Soberón & Nakamura 2009).



A. Grinellian Niche B. Hutchinsonian Realized Niche



C. Source-Sink Dynamics D. Dispersal Limitation

Figure 2. Four bi-dimensional graphs representing the relation between niche and species distribution (from Pulliam 2000).

Grinnell (1917) defined the species niche as the “environmental requirements of the species” (Fig. 2-A). A decade later, Elton (1927) defined niche as “the role of the species in the community”, integrating the interactions among species. These two definitions (geographical and functional) were merged by Hutchinson (1957), who described the species niche as “ the coordinates of the species with n-dimensional resources axes”. He made the distinction between the “fundamental niche” where “the conditions allow the species to exist indefinitely” and the “realized niche”, a subset of the fundamental niche that corresponds to “the conditions where the species can persist in interaction with other organisms”.

One of the most important assumptions of SDMs, based on Hutchinson’s definition, is that a species is present wherever the local environmental conditions are within the species demographic niche (i.e. wherever the population is at equilibrium) (Peterson 2011). However, these models frequently ignore possible deviations from this equilibrium, such as source-sink dynamics and dispersal limitation (Pulliam 2000) (Fig.2) or ecological traps (Robertson & Hutto 2006; Robertson *et al.* 2017), where individuals occur under certain environmental conditions beyond the boundaries of their demographic niche.

This important assumption of species distribution modelling directly challenges the outcomes of these models and their usefulness for predicting habitat quality to support the persistence of the species. This issue has drawn our attention to the need for a better understanding of the link between model predictions and habitat quality.

Types of data used in SDMs

There are numerous methods that combine species occurrence data with environmental variables to estimate species distribution in space. The type and quality of the available data (e.g. sample size, sample bias, spatial resolution, geographical extent) and the ecological questions to be addressed are key aspects when selecting a modelling method (Segurado & Araújo 2004; Lobo, Jiménez-Valverde, & Real 2008; Elith & Leathwick 2009; Guillera-Arroita *et al.* 2015).

Species distribution data

There are mainly two types of species distribution data (see Franklin 2009 for details): those that enclose only information on the presence of a species (presence-only data) and those that provide information on the presence and the absence of a species (presence-absence data). Many data collection methods, especially those involving many species and large geographical extents, can only afford recording the species presence rather than distinguishing between presences and absences.

Some of the modelling methods which use presence-only data are: BIOCLIM (Busby, Margules, & Austin 1991); HABITAT (Walker & Cocks 1991); DOMAIN (Carpenter, Gillison, & Winter 1993); ecological niche factor analysis (ENFA) (Hirzel *et al.* 2002); genetic algorithm for rule-set production (GARP) (Stockwell & Peters 1999); maximum entropy (MaxEnt) (Phillips, Dudík, & Schapire 2004) and Mahalanobis distance (Farber & Kadmon 2003). Examples of modelling methods that use presence and absence data include generalised linear models (GLM) (Guisan, Edwards, & Hastie 2002), generalised additive models (GAM) (Guisan, Edwards, & Hastie 2002; Elith *et al.* 2006), artificial neural network (ANN) (Pearson *et al.* 2002) and boosted regression trees (BRT).

Many scientific studies are dedicated to the comparison between species distribution modelling methods and algorithms (Segurado & Araújo 2004; Tsoar *et al.* 2007; Jiménez-Valverde, Lobo, & Hortal 2008; Elith & Graham 2009). Nowadays, platforms such as BIOMOD have been developed to make use of different modelling algorithms and predict the species distribution based on an ensemble of methods (Thuiller *et al.* 2009). We do not provide detailed information on mathematical algorithms and modelling methodologies, as this is not the scope of this introduction.

Environmental variables

The choice of environmental predictors has a large influence on SDM performance (Austin *et al.* 2006). The modeller needs to identify environmental information that represents resource gradients determining the species distribution and to select and assemble appropriate environmental data to be used in the modelling. Conditioned to what is available in the study area, issues such as data quality, spatial resolution or variable selection are not trivial problems to solve (Franklin 2009).

The typically used environmental variables can be grouped in different types: climate (e.g. temperature, precipitation), topography (e.g. elevation, orientation), substrate (e.g. soil types, geology), land cover and vegetation, remote sensing-based land surface characterisations (such as vegetation indexes) and measures of landscape pattern (Austin 2002; Synes & Osborne 2011).

What about the name?

Ecological niche-models, habitat suitability models or species distribution models? Much literature is dedicated to explaining the differences among these terms and the appropriateness of using one or the other. This literature, albeit interesting, may lead to confusion (Austin 2007; Elith & Leathwick 2009; Peterson 2011; Peterson & Soberón 2012; Warren 2012; McInerny & Etienne 2013; Lele et al. 2013). In this PhD-thesis, we mainly used the more general term “species distribution model” as we think that it is the term that best encloses all the other terms, but we take them all as synonyms.

Interpretation: understanding modelling outcomes

Currently, the more complex models are most frequently used, even though this implies the interpretation of complex outcomes (Elith, Kearney, & Phillips 2010). The nature of the data and the model algorithm used, with the underlining assumptions, affects the interpretation of the resulting SDM and spatial prediction (Franklin 2009).

Most SDMs are built under the assumption that they directly inform on the quality of the habitats for the species. Generally, the model predictions correlate well with species abundance (Weber *et al.* 2016), but not much quantitative information is available regarding the link with population growth (Thuiller *et al.* 2014). Outcomes of these models are often – sometimes blindly – used as proxies for demographic parameters to differentiate between areas with high individual fitness (high survival of the individual and high offspring production) and areas that are not suitable for viable populations (Pulliam 2000; Guisan & Thuiller 2005; VanDerWal *et al.* 2009; Pellissier *et al.* 2013).

Mechanistic approaches that integrate demographic, physiological, evolutionary or behavioural processes into the modelling framework have been proposed as valuable alternatives or counterparts to SDMs (e.g. Kearney & Porter 2009; Buckley *et al.* 2010; Sánchez-Clavijo, Hearns, & Quintana-Ascencio 2016). However, the development of models based on mechanisms remains strongly constrained by the availability of appropriate input data at relevant scales (Thuiller et al. 2013; Urban et al. 2016). For this reason, SDMs are expected to be often used for conservation and management applications in the future (Elith, Kearney, & Phillips 2010; Dormann *et al.* 2012; Morán-Ordoñez, Briscoe, & Wintle 2016). Therefore, there is a strong need to better understand situations in which these statistical approaches produce useful outcomes for this purpose (Bean *et al.* 2014)

USING BIRDS AS MODEL ORGANISMS

Birds are worldwide recognized as useful indicators of the state of the environment because they occupy a high trophic level and they occur in a wide range of habitats (Bibby 1999; BirdLife International 2004; Gregory & Strien 2010).

Birds are relatively easy to census as they are well known, easily recognisable and simpler to locate than other taxonomic groups (Bibby *et al.* 1992). Western Europe counts with a great number of amateur ornithologist and has a strong ornithological culture, which allows the implementation and accomplishment of many different bird survey projects.

In this PhD-thesis we present 4 articles, dealing with bird distribution data and species distribution modelling. More details about the specific questions addressed are presented in the next chapter (Objectives). For articles 1 and 4, we worked with the same bird species, the Red-backed Shrike (*Lanius collurio*), a passerine bird included in the Annex 1 of the European “Birds Directive”. Articles 2 and 3 are dedicated to groups of bird species (20 and 19 species, respectively), which, in both cases, needed to fulfil some specific criteria necessary to answer the specific objectives of each article.

Bird mapping projects

Based on the collection of observations covering large areas (Fig.3), atlas projects require substantial effort and budget. For this reason, they are generally completed during considerable time periods (3-6 years) and repeated at long-time intervals (15 - 20 years), which prevent them from being suitable to detect changes in species distribution with an appropriate time scale for decision-making.

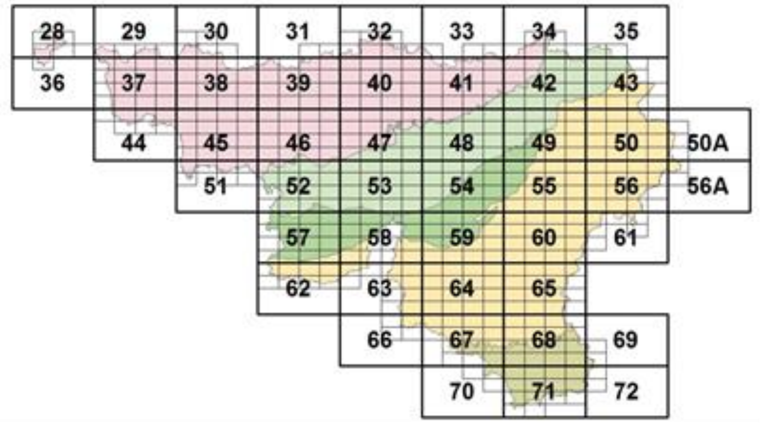


Figure 3. This figure, from the Methods chapter of the Atlas of Breeding Birds of Wallonia, represents the grid division of the region to be sampled, with the ecological regions of the study area in the background (figure from Jacob *et al.* 2010).

In articles 2, 3 and 4 of this PhD-thesis, we used data collected for breeding bird atlas projects for calibrating the SDMs. For articles 2 and 4 we used the Atlas of Breeding Birds of Wallonia (Jacob *et al.* 2010) and for article 3 we used the Catalan Breeding Bird Atlas (Estrada *et al.* 2004) (Fig 4).

Figure 4. Cover pages of the Atlas of Breeding Birds of Wallonia (Jacob et al. 2010) and the Catalan Breeding Bird Atlas (Estrada et al. 2004), major data sources of this PhD-thesis.

Bird monitoring projects

Long-term monitoring projects are increasingly conducted in several areas to track overall changes in species composition and/or abundance with a finer temporal resolution (intervals of ca. 1-5 years) than large-scale mapping projects. Sampling sites are normally scarcely distributed in space (Fig. 5). Monitoring projects are mostly dedicated to inform on temporal trends of abundance and occurrence of common and widespread species (Gregory *et al.* 2005), with an emphasis on understanding the underlying causes of population changes in the countryside and on evaluating and improving the efficiency of management policies (Yoccoz, Nichols, & Boulinier 2001). The main aim of monitoring projects is to provide a scientific evaluation of the changes in the conservation status of species over time. The information gathered in a series of small and geographically scattered sampling locations is not readily usable to map the distribution of species in a continuous way across the region of interest, but it may provide us with a sufficient spatial data that could be used for mapping species distributions through modelling (Brotons, Herrando, & Pla 2007; Honrado, Pereira, & Guisan 2016).

Data collection

Large-scale mapping and long-term monitoring projects usually need a large group of devoted volunteers and/or professionals who collect data in the field. The data collection itself can be done following different sampling procedures, such as point counts, transects or territory mapping (Fig.5). These procedures are chosen according to the specific objectives of the project, but also depend on the human and financial resources.

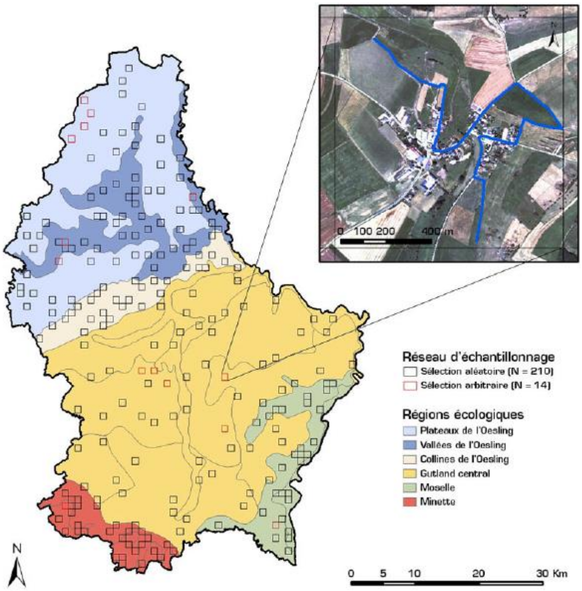


Figure 5. A transect serving as sampling unit for birds in the Global biodiversity long-term monitoring project of Luxembourg. The background represents the ecological regions of Luxembourg (figure from Titeux, Moes, & Hoffmann 2009).

This PhD-thesis focuses on the uses of SDMs for direct conservation issues: 1) identifying the most promising areas for the conservation of bird species (articles 1, 3 and 4) and 2) guiding the establishment of monitoring projects (article 2). With our results, we hope to encourage the use of models among sceptical practitioners involved in decision-making related to conservation and management issues (articles 1 and 2). We also focus on the ecological interpretation of SDM outcomes, analysing their link with the quality of habitat conditions for the reproduction of the species (articles 3 and 4).

REFERENCES

Anadón, J.D., Giménez, A., Martínez, M., Palazón, J.A. & Esteve, M.A. (2007) Assessing changes in habitat quality due to land use changes in the spur-thighed tortoise Testudo graeca using hierarchical predictive habitat models. *Diversity and Distributions*, **13**, 324–331.

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

Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.

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

Austin, M.P. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

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

Bean, W.T., Prugh, L.R., Stafford, R., Butterfield, H.S., Westphal, M. & Brashares, J.S. (2014) Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. *Journal of Applied Ecology*, **51**, 1116–1125.

Bibby, C.J. (1999) Making the most of birds as environmental indicators. *Ostrich*, **70**, 81–88.

Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S.H. (1992) *Bird Census Techniques*. Academic Press, London.

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

Brotons, L., Herrando, S. & Pla, M. (2007) Updating bird species distribution at large spatial scales: Applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions*, **13**, 276–288.

Brotons, L., Mañosa, S. & Estrada, J. (2004) Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodiversity and Conservation*, **13**, 1039–1058.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J. & Sears, M.W. (2010) Can mechanism inform species' distribution models ? *Ecology Letters*, **13**, 1041–1054.

Busby, J.R., Margules, C.R. & Austin, M.P. (1991) BIOCLIM - a bioclimate analysis and prediction system, 64–68.

Cabeza, M. & Moilanen, A. (2001) Design of reserve networks and the persistence of biodiversity. *Trends in Ecology & Evolution*, **16**, 242–248.

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

Chase, J.M. & Leibold, M.A. (2003) *Ecological Niches - Linking Classical and Comtemporary Approaches*. University of Chicago Press, Chicago.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M. & Singer, A. (2012) Correlation and process in species distribution models : bridging a dichotomy, 2119–2131.

Elith, J. & Graham, C.H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32**, 66–77.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, 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.P., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Williams, S.E., Wisz, M.S., Zimmermann, N.E., Graham, H., Anderson, P., Hijmans, J., Leathwick, R., Lohmann, G., Loiselle, A., Overton, C.M., Townsend Peterson, A., Phillips, J., Schapire, E., Soberón, J.M., Wisz, S. & Zimmermann, E. (2006) Novel methods improve prediction of species ’ distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J., Kearney, M.R. & Phillips, S.J.P. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677–697.

Elton, C. (1927) Animal Ecology, 209.

Estrada, J., Pedrocchi, V., Brotons, L. & Herrando, S. (2004) *Atles Dels Ocells Nidificants de Catalunya (1999-2002)* (eds J Estrada, V Pedrocchi, L Brotons, and S Herrando). Institut Català d’Ornitologia (ICO) / Lynx Edicions, Barcelona.

Farber, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160**, 115–130.

Franklin, J. (2009) *Mapping Species Distributions: Spatial Inference and Predictions*. Cambridge University Press, Cambridge, UK.

Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S., Scroggie, M.P. & Woodford, L. (2011) Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, **48**, 25–34.

Gregory, R.D. & Strien, A. V. (2010) Wild bird indicators: Using composite population trends of birds as measures of environmental health. *Ornithological Science*, **9**, 3–22.

Gregory, R.D.N., van Strien, A.J., Vorísek, P., Gmelig Meyling, A.W., Noble, D.G., Foppen, R.P.B., Gibbons, D.W., Strien, A. Van & Meyling, A.W.G. (2005) Developing indicators for European birds. *Philosophical Transactions of The Royal Society B*, **360**, 269–288.

Grinnell, J. (1917) The niche relationship of the Californian Thrasher. *Auk*, **34**, 427–433.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015) Is my species distribution model fit for purpose ? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.

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

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

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., Irzel, A.H.H., Ausser, J.H. & Hessel, D.C. (2002) Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, **83**, 2027–2036.

Hirzel, A.H. & Le Lay, G. (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**, 1372–1381.

Hollander, F.A., Titeux, N., Holveck, M.-J. & Van Dyck, H. (2017) Timing of breeding in an ecologically trapped bird. *The American Naturalist*, **189**, 515–525.

Honrado, P., Pereira, H.M. & Guisan, A. (2016) Fostering integration between biodiversity monitoring and modelling. *Journal of Applied Ecology*, **53**, 1299–1304.

Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–427.

Jacob, J.-P., Dehem, C., Burnel, A., Dambiermort, J.L., Fasol, M., Kinet, T., van der Elst, D. & Paquet, J.-Y. (2010) *Atlas Des Oiseaux Nicheurs de Wallonie*. Publication d'Aves et du Département de l'Etude du Milieu Naturel et Agricole.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2008) Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions*, **14**, 885–890.

Kearney, M.R. & Porter, W.P. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, **12**, 334–50.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S.J.P., Thomas, C.D., Beentje, H., Dransfield, J., Fisher, B.L., Glaw, F., Good, T.C., Harper, G.J., Hijmans, R.J., Lees, D.C., Louis Jr., E., Nussbaum, R.A., Raxworthy, C.J., Razafimpahanana, A., Schatz, G.E., Vences, M., Vieites, D.R., Wright, P.C. & Zjhra, M.L. (2008) Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, **320**, 222–226.

Lele, S.R., Merrill, E.H., Keim, J. & Boyce, M.S. (2013) Selection, use, choice, and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology*, **82**, 1183–1191.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

Luoto, M., Heikkinen, R.K., Poyry, J. & Saarinen, K. (2006) Determinants of the biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography*, **33**, 1764–1778.

McInerny, G.J. & Etienne, R.S. (2013) 'Niche' or 'distribution' modelling? A response to Warren. *Trends in Ecology and Evolution*, **28**, 191–192.

Morán-Ordoñez, A., Briscoe, N.J. & Wintle, B.A. (2016) Modelling species responses to extreme weather provides new insights into constraints on range and likely climate change impacts for Australian mammals. *Ecography*, 10.1111/ecog.02850.

Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. (2002) SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling*, **154**, 289–300.

Pellissier, L., Meltofte, H., Hansen, J., Schmidt, N.M., Tamstorf, M.P., Maiorano, L., Aastrup, P., Olsen, J., Guisan, A. & Wisz, M.S. (2013) Suitability, success and sinks: how do predictions of nesting distributions relate to fitness parameters in high arctic waders? *Diversity and Distributions*, **19**, 1496–1505.

Peterson, A.T. (2011) Ecological niche conservatism: A time-structured review of evidence. *Journal of Biogeography*, **38**, 817–827.

Peterson, A.T. & Soberón, J.M. (2012) Species Distribution Modeling and Ecological Niche Modeling : Getting the Concepts Right. *Natureza & Conservação*, **10**, 102–107.

Phillips, S.J.P., Dudík, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st. international conference on machine learning*, 655–662.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Robertson, B.A., Campbell, D.-R., Durovich, C., Hetterich, I., Les, J. & Horváth, G. (2017) The interface of ecological novelty and behavioral context in the formation of ecological traps. *Behavioral Ecology*, **28**, 1166–1175.

Robertson, B.A. & Hutto, R.L. (2006) A framework for understanding ecological traps and an evolution of existing evidence. *Ecology*, **87**, 1075–1085.

Sánchez-Clavijo, L.M., Hearns, J. & Quintana-Ascencio, P.F. (2016) Modeling the effect of habitat selection mechanisms on population responses to landscape structure.

Ecological Modelling, **328**, 99–107.

Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.

Soberón, J.M. & Nakamura, M. (2009) Niches and distributional areas : Concepts , methods ,. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19644–19650.

Stockwell, D.R.. B. & Peters, D. (1999) The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143–158.

Synes, N.W. & Osborne, P.E. (2011) Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. *Global Ecology and Biogeography*, **20**, 904–914.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.

Thuiller, W., Münkemüller, T., Schiffers, K.H., Georges, D., Dullinger, S., Eckhart, V.M., Edwards, T.C., Gravel, D., Kunstler, G., Merow, C., Moore, K., Piedallu, C., Vissault, S., Zimmermann, N.E., Zurell, D. & Schurr, F.M. (2014) Does probability of occurrence relate to population dynamics? *Ecography*, **37**, 1155–1166.

Titeux, N., Biver, G., Lorgé, P. & Hoffmann, L. (2009) Implementation of the Common Bird Monitoring scheme in Luxembourg. *Bird Census News*, **22**, 42–50.

Titeux, N., Moes, M. & Hoffmann, L. (2009) *Élaboration D’un Programme de Surveillance et de Monitoring de La Biodiversité Au Luxembourg*. Ministère de l’Environnement & Centre de Recherche Public - Gabriel Lippmann, Luxembourg.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, **13**, 397–405.

VanDerWal, J., Shoo, L.P., Graham, C. & Williams, S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220**, 589–594.

Van der Wal, R., Anderson, H., Robinson, A., Sharma, N., Mellish, C., Roberts, S., Darvill, B. & Siddharthan, A. (2015) Mapping species distributions : A comparison of skilled naturalist and lay citizen science recording. *Ambio*, **44**, 584–600.

Walker, P.A. & 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.

Warren, D.L. (2012) In defense of ‘niche modeling’. *Trends in Ecology and Evolution*, **27**, 497–500.

Weber, M.M., Stevens, R.D., Diniz-Filho, J.A.F. & Grelle, C.E. V. (2016) Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography*, doi: 10.1111/ecog.02125.

Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, **16**, 446–453.

Objectives

As species distribution models (SDMs) are expected to be increasingly used to guide conservation decision-making (Guisan *et al.* 2013), there is a strong need to better understand situations in which these models can be confidently used as tools to support the different steps leading to the achievement of conservation goals. The overall objective of this PhD-thesis is to contribute to providing evidence of the usefulness and applicability of SDMs for some of these steps.

Across the different chapters of the thesis, we will examine the usefulness of SDMs to improve our knowledge on the distributions of the species and on the spatial arrangement of their suitable habitats. To do so, the first half of the thesis (articles 1 and 2) will focus on the power of the SDMs to predict the distribution of the species and the second half (articles 3 and 4) will explicitly distinguish species distribution from reproductive performance that directly reflects habitat quality for the species. Identifying suitable habitats for the species across an area of interest is a crucial issue that can impact the spatial prioritization of management or conservation options.

This general objective is further divided into four questions, which are addressed in separate articles, as described below (Fig.1).

<u>Question 1</u>: Are SDMs as useful as experts to improve our knowledge on the distribution of a species of conservation concern?

SDMs are an increasingly used technique for identifying suitable sites for the species, but in many cases conservation practitioners remain sceptical about their usefulness and mostly rely on expert knowledge when making decisions. Here, we explicitly compare both approaches to illustrate the potential of SDMs for maximizing the detection of new presence areas for a conservation-

concern bird species (Red-backed shrike, *Lanius collurio*) at a national scale (Luxembourg).

We designed three separate sampling strategies to identify new shrike records across the country: 1) a sampling strategy based on ornithological expert knowledge; 2) a sampling strategy based on the predictions of SDMs; and 3) a random sampling strategy. We conducted ground validation according to the three different strategies to evaluate and compare their effectiveness in detecting previously unknown presence areas for the shrikes.

We believe that this first article may contribute to encouraging the use of SDMs among practitioners as an accepted tool to support biodiversity conservation and management.

<u>Question 2</u>**: How to optimise sampling design in long-term bird monitoring projects so that the data collected in the field could be used to map the distribution of the species?**

Long-term monitoring projects have as a prime objective to estimate temporal trends in population sizes. Nowadays, integrated with SDM techniques, monitoring schemes can also be used to obtain large-scale species distribution maps. These projects are often carried out in a network of sites where data is collected repeatedly over time according to established procedures. The sampling design often results from a trade-off between the number of sampling sites and the number of repeated surveys that can be conducted in these sites to track the abundance of the species with an acceptable level of precision. Mapping the distribution of the species is, therefore, seldom integrated when preparing the sampling design for monitoring projects.

We developed an analytical framework based on the data from a breeding bird atlas project in Wallonia (Belgium) to guide the establishment of a bird monitoring project that would have the potential to produce appropriate information to build useful SDMs and maps of species distribution. We manipulated breeding bird atlas data to imitate a broad gradient of possible numbers of sampling sites in a monitoring project. We built and estimated the precision of SDMs with this varying amount of sampling sites.

We believe that this second article provides an interesting analytical framework to aid in the initial design of monitoring projects if their aim is to document the distribution of the species in addition to trends in population sizes.

<u>Question 3</u>: Are SDMs reliable to inform on habitat quality for breeding birds at the regional scale?

Predictions of SDMs are often used under the assumption that they directly inform on the quality of the habitats for the species. In other words, SDM outcomes are frequently used as proxies for demographic parameters to differentiate between areas of high importance for the persistence of the populations and areas that are not suitable for viable populations.

We assessed how the predictions derived from SDMs are related to measures of reproductive performance obtained from data collected in Constant Effort Sites (CES) mist-netting stations in Catalonia (Spain). We used bird presence records collected in the frame of a breeding bird atlas project to build SDMs. The proportion of juveniles captured in each CES was used as a measure of reproductive performance for each bird species. We tested at multiple spatial scales if the geographical variations in reproductive performance as measured in the CES matched the variations in habitat quality as derived from the SDMs.

We believe that this third article may contribute to the understanding of the relationship between the predictions obtained from the SDMs and the production of juveniles in breeding areas.

<u>Question 4</u>**: Are SDMs reliable to inform on habitat quality for an ecologically trapped bird?**

Ecological traps result from behavioural maladaptation during habitat selection: they are poor-quality habitats that attract organisms more than higher-quality options available in the landscape. This may challenge the relevance of SDM outcomes if they are used to document habitat quality weigh the importance of different areas for the conservation of the species. In direct line with article 3, we examined the relevance of using SDM outcomes to reflect how habitat quality varies across the landscape for ecologically trapped organisms.

Previous work conducted in Wallonia (Belgium) showed that Red-backed shrikes (*Lanius collurio*) are attracted to harvested areas in coniferous plantation forests more than to traditional farmland habitat, but their reproductive performance is markedly higher in farmland than in forest. Here, we built SDMs with shrike distribution data obtained from the breeding bird atlas of Wallonia and estimate how the SDM outcomes varied across the landscape. We compare the outcomes of the SDMs in farmland and forest habitats to parameters of reproductive performance that proximately inform on the quality of these two breeding habitats for shrikes.

We believe that this fourth article provides a step further in the understanding of the relationship between the SDMs outcomes and the reproduction parameters, as this is done for an ecologically trapped bird, where the habitat preference of the species is the habitat where its reproductive performance is lower.

These four questions were addressed through specific studies developed as scientific articles following similar methodological approaches and using similar datasets (Fig. 2). This allowed us to explore the usefulness of SDMs for conservation practice and to better understand the ecological interpretation of SDM outcomes.

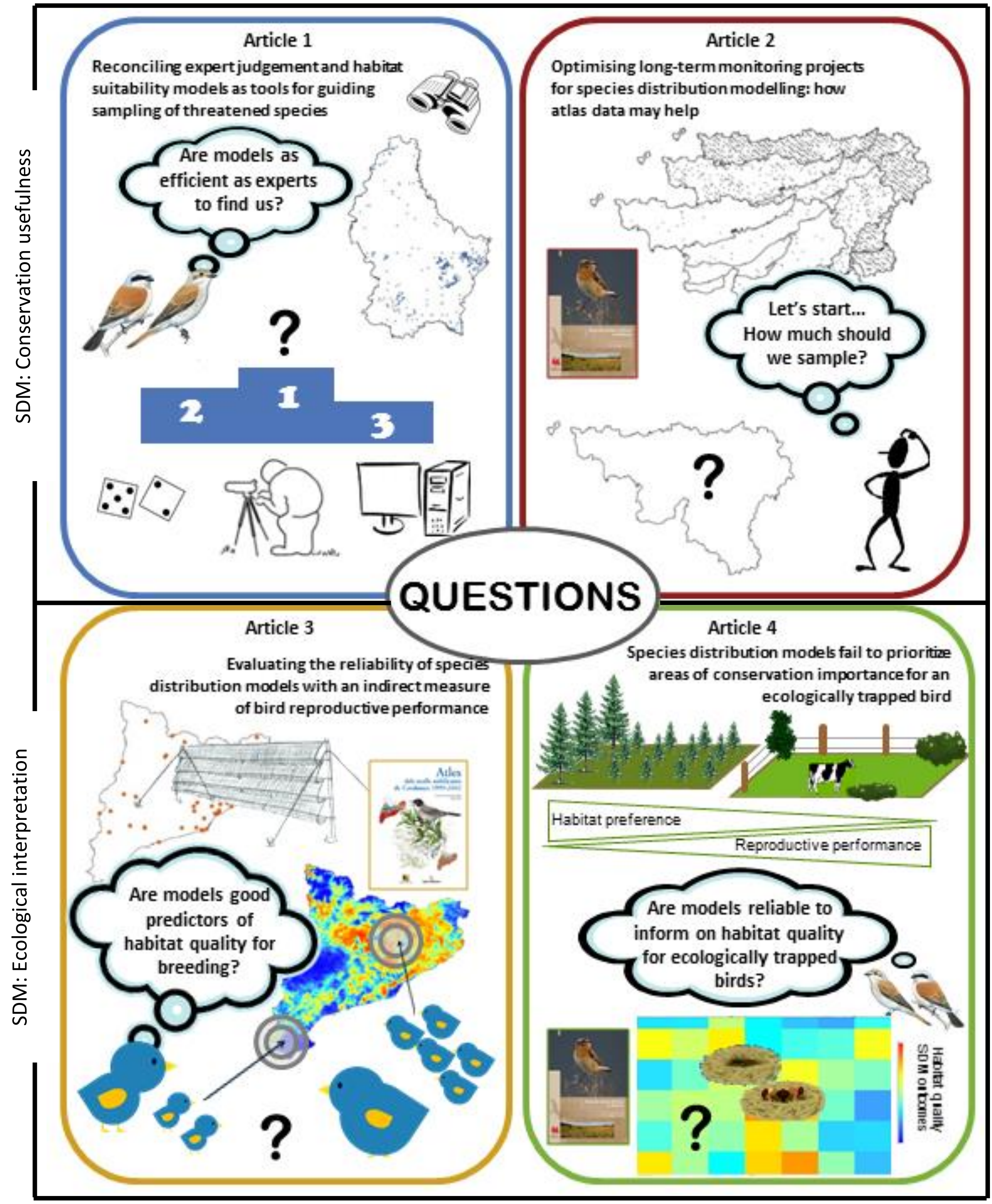


Figure 1. Graphic summary of each of the articles presented in this PhD-thesis.

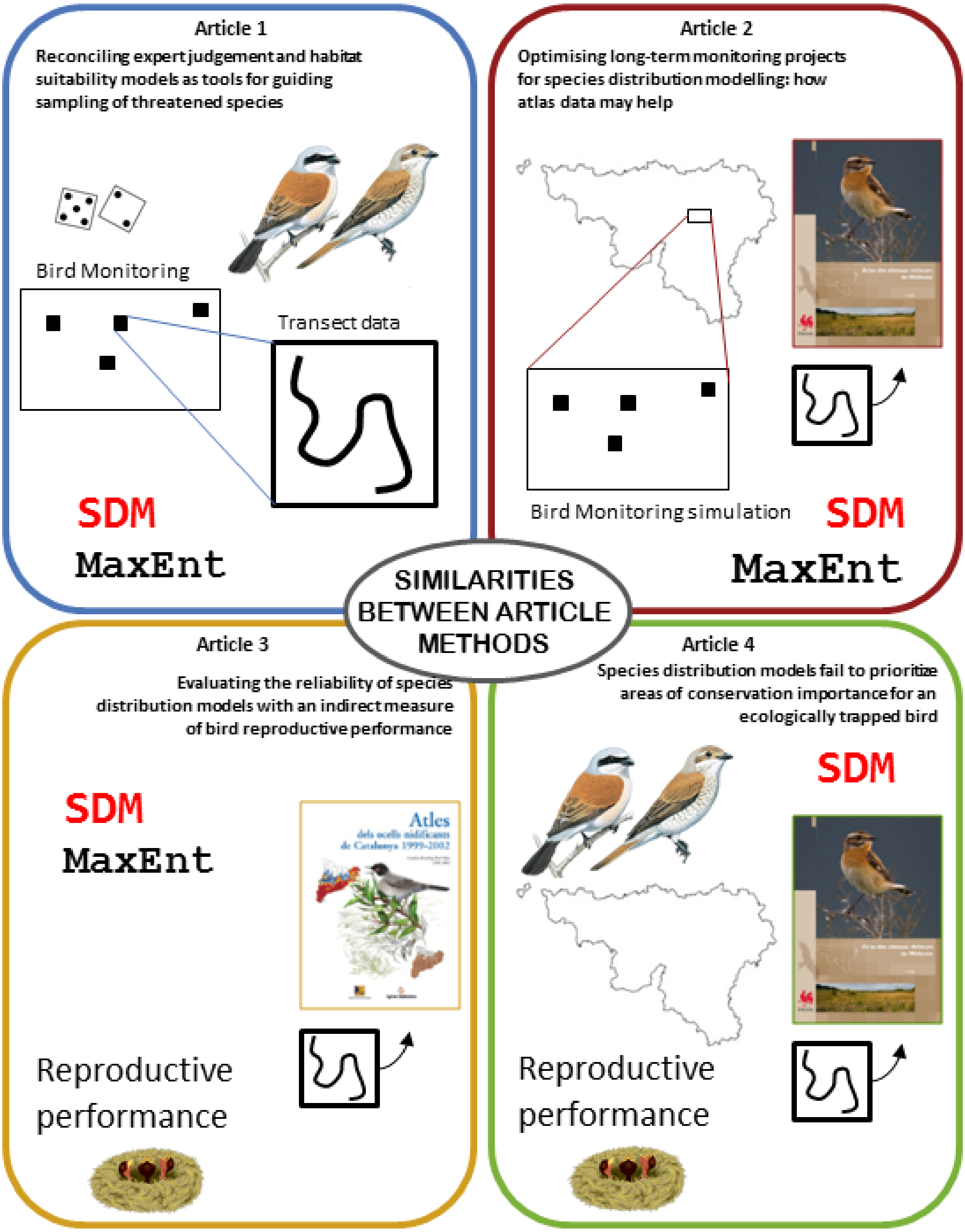


Figure 2. Graphic summary of the common methods used in the articles presented in this PhD-thesis.

REFERENCES

Austin, M.P. (2007) Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, **200**, 1–19.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics*, **40**, 677–697.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Hollander, F.A., Van Dyck, H., San Martin, G. & Titeux, N. (2011) Maladaptive habitat selection of a migratory passerine bird in a human-modified landscape. *PLoS one*, **6**, e25703.

Lele, S.R., Merrill, E.H., Keim, J. & Boyce, M.S. (2013) Selection, use, choice, and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology*, **82**, 1193-1191.

McInerny, G.J. & Etienne, R.S. (2013) ‘Niche’ or ‘distribution’ modelling? A response to Warren. *Trends in Ecology and Evolution*, **28**, 191–192.

Peterson, A.T. (2011) Ecological niche conservatism: A time-structured review of evidence. *Journal of Biogeography*, **38**, 817–827.

Peterson, A.T. & Soberón, J.M. (2012) Species Distribution Modeling and Ecological Niche Modeling : Getting the Concepts Right. *Natureza & Conservação*, **10**, 102–107.

Warren, D.L. (2012) In defense of ‘niche modeling’. *Trends in Ecology and Evolution*, **27**, 497–500.

Article 1

Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species

The content of this chapter has been published*:

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Biver, G., Brotons, L., Titeux, N. (2015). Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species. *Journal of Applied Ecology* **52,** 1608–1616**.**

*text in the thesis has been marginally been modified compared to the published version

ABSTRACT

1. Up-to-date knowledge on species distribution is needed for efficient biodiversity conservation and management decision-making. Implementing efficient sampling strategies to identify previously unknown locations of species of conservation-concern is therefore a key challenge. Both structured expert judgement and habitat suitability models may help target sampling towards areas where chances to find the species are highest. However, practitioners often object to the use of models and believe they do not result in better decisions than the subjective opinion of experts, thus potentially constraining an optimal use of available methods and information.
2. To illustrate the potential of habitat suitability models for guiding sampling strategies, we evaluated and compared the ability of experts and models to identify important areas for the conservation of a bird species (*Lanius collurio*) in Luxembourg. We conducted extensive fieldwork to find as many unknown bird territories as possible according to three independent sampling strategies: (i) a sampling strategy based on structured expert judgement, (ii) a sampling strategy based on the predictions of a habitat suitability model and (iii) a general-purpose stratified random sampling strategy used as a baseline reference.
3. Both the expert-based and the model-based sampling strategies substantially outperformed the general-purpose sampling strategy in identifying new species records. In addition, the model-based sampling strategy performed significantly better than the expert-based sampling strategy.
4. *Synthesis and applications*. This study explicitly shows that habitat suitability models can efficiently guide field data collection towards suitable areas for species of conservation-concern. Results may facilitate the involvement of practitioners in the development of habitat suitability models with the objective of maximizing the robustness of modelling applications in conservation practice and management decision-making.

INTRODUCTION

Accurate knowledge on species occurrence is a prerequisite for appropriate biodiversity conservation decision-making, such as reserve selection (e.g. Cabeza & Moilanen 2001), management of biological invasions (e.g. Gormley *et al.* 2011) or identification of key habitats for threatened species (e.g. Brotons, Mañosa, & Estrada 2004). Such information often consists of opportunistically collected data available as museum records or from web-based biodiversity data-gathering portals (Sardà-Palomera *et al.* 2012). Field data are also increasingly collected during structured field sampling, such as biodiversity mapping (e.g. atlas projects) or monitoring programmes (e.g. Robertson, Cumming, & Erasmus 2010). For rare species or species of conservation-concern, information is often lacking or incomplete; finding new presence areas is critical because increased knowledge on their distribution may provide key guidance on their conservation and management (Guisan *et al.* 2006). However, the collection of additional field data can be costly in terms of manpower, time and budget. It is therefore highly important to define the most efficient sampling strategies to minimize costs and maximize gains in knowledge (Aizpurua *et al.* 2015). The distribution of species of conservation-concern may be geographically limited due to their restricted habitat requirements or population sizes. Hence, identifying new presence areas for those species might be challenging and sometimes inefficient using general-purpose sampling designs (Le Lay *et al.* 2010).

To find new presence areas for species of conservation-concern, information on their habitat requirements is needed (e.g. Anadón *et al.* 2009). One option to obtain such information is the application of methods that aim to elicit information from experts (Franklin 2009). Experts have achieved high knowledge on a particular topic through their life experience (Kuhnert, Martin,

& Griffiths 2010; Burgman *et al.* 2011) and are classically defined by their qualifications, track record and professional standing (Burgman *et al.* 2011). One advantage of expert elicitation is the possibility of obtaining high-quality and structured information on species distributions with a relatively low cost (Murray *et al.* 2009; Cerqueira *et al.* 2013). This may prove useful when available information on species distribution is insufficient to implement more quantitative methods (Doswald, Zimmermann, & Breitenmoser 2007; Cerqueira *et al.* 2013; Turvey *et al.* 2015). For instance, reliable information on local distribution and abundance of the spur-thighed tortoise *Testudo graeca* L. was easily obtained by interviewing local shepherds about the number of encounters with the species (Anadón *et al.* 2009). Eliciting expert information involves dealing with multiple expert judgements, with different sources of biases in the elicited information and with uncertainty around expert estimates (Martin *et al.* 2012; McBride *et al.* 2012). For example, expertise may be restricted to the region of interest of the experts (Murray *et al.* 2009). Hence, a careful pre-elicitation analysis of expert availability and the preparation of a structured elicitation design are needed to account for such potential biases and to obtain the highest quality of information (Martin *et al.* 2012; McBride *et al.* 2012). Sampling design based on structured expert judgement may then prove to be cost-efficient for identifying important presence areas for the conservation of threatened species (Murray *et al.* 2009; Cerqueira *et al.* 2013).

Habitat suitability modelling is a more recent tool that uses existing data and may assist in identifying sites where additional sampling is to be conducted. Here, a statistical link is established between the locations where the target species has been observed and a series of variables describing the environmental conditions in those sites (Guisan & Zimmermann 2000; Franklin 2009; Elith *et al.* 2011). Such predictive models may be used to inform on potentially suitable habitats in areas where the species presence is unknown, which may constitute an efficient data-driven approach to guide further sampling (Guisan *et al.* 2006; Crall *et al.* 2013). Models also suffer from important limitations including geographical biases, data availability and uncertainties in their predictions (Barry & Elith 2006). A variety of statistical methods exist to evaluate their ability to predict species distributions accurately (e.g. Vaughan & Ormerod 2005).

Although predictive models have the potential to play a key role in supporting conservation and management decision-making, practitioners are often not easily inclined to rely on their outcomes for on-the-ground interventions (Jeltsch *et al.* 2013). Addison *et al.* (2013) provided evidence of common objections to the use of models in environmental decision-making and reported that practitioners often believe that models do not result in better decisions than those supported by the subjective opinion of experts. Alternatively, managers may object to the use of such approaches as they consider that models fail to capture the different factors influencing conservation and management options (Hajkowicz 2007), or provide outcomes that are uncertain and poorly communicated (Borowski & Hare 2007). An additional objection relates to the need for a considerable level of conceptual and technical expertise or to the amount of resources and time needed to implement such procedures and to obtain enough input data (Borowski & Hare 2007).

A stronger linkage between modelling science and conservation practice has been recently advocated to help modellers improve the effectiveness, relevance and usefulness of their work in supporting conservation and management decision-making (Guisan *et al.* 2013). In the last few decades, effort has been invested on integrating structured expert judgement into modelling approaches to improve model predictions (Krueger *et al.* 2012). Such expert-informed modelling can contribute to bridging the gap between modellers and practitioners. Structured expert judgement may be incorporated in predictive models at different stages of the modelling procedure (Pearce *et al.* 2001), e.g. for the preparation of input data, the selection of relevant variables or the refinement of model predictions.

An alternative option to illustrate the potential of predictive models is to compare the ability of such quantitative approaches with that of an approach based on expert elicitation to guide on conservation decisions (Drolet *et al.* 2015). Rather than integrating structured expert judgement into the modelling procedure, we compared the capability of models and experts to optimize the detection of previously unknown presence areas for a bird species of conservation-concern. First, we designed three separate sampling strategies: a sampling strategy based on structured expert judgement without the aid of modelling approaches; a sampling strategy based on the predictions of a habitat suitability model independent of expert judgement and a general-purpose strategy based on a stratified random sampling design. Second, we conducted ground validation according to the different sampling strategies to evaluate and compare their effectiveness to update our knowledge on the distribution of the target species (Williams *et al.* 2009; Rebelo & Jones 2010). We hypothesized that predictive models are useful to guide sampling if a model-based sampling strategy performs better than a general-purpose strategy (Le Lay *et al.* 2010) and as good as an expert-based sampling strategy (Drolet *et al.* 2015). To the best of our knowledge, using structured expert judgement to illustrate the effectiveness of habitat suitability models to achieve a management and conservation objective in an explicit and straightforward way, as we propose here, has not been reported to date. We believe this may contribute to encouraging the use of models among practitioners as an accepted tool to support biodiversity conservation and management decision-making.

METHODS

Study area and species

The study was conducted in Luxembourg (2586 km2, Fig.1a). *Lanius collurio* L., a passerine bird categorized as nearly threatened in this country (Lorgé & Melchior 2010), was chosen as a model species of conservation-concern. This bird breeds in semi-open areas under a management regime of extensive farming with scattered and thorny hedges and bushes for nesting (Titeux et al. 2007). Individuals arrive to the breeding sites from late April to late May and the breeding period extends until late July. Their sit-and-wait hunting strategy and their territory-defence behaviour make *L. collurio* easily detectable (Titeux et al. 2007).

Sampling strategies

A total of 737 known *L. collurio* territories recorded during the period 2000–2009 were made available from the national data set managed by the bird conservation association in Luxembourg (BirdLife Luxembourg). These known territories were used as a common source of basic information to design the expert-based and model-based sampling strategies as described below (Fig.1b). Data included presence-only records with varying spatial precision, but only the records with a precision ranging from 10 m to 100 m were retained for subsequent analyses.

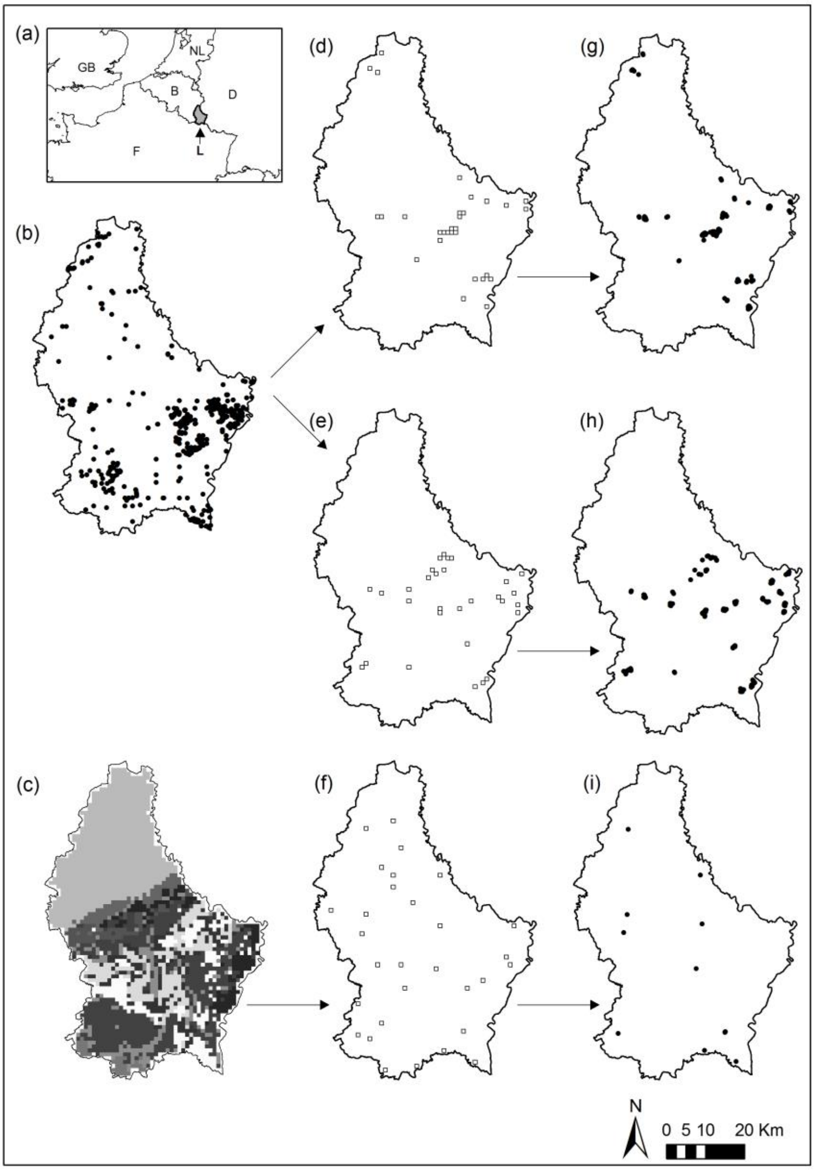


Figure 1. Overview of the different sampling strategies used to identify new territories of *L. collurio* in Luxembourg. (a) Luxembourg in north-west Europe. (b) Known *L. collurio* territories in Luxembourg (2000–2009). (c) Environmentally homogeneous strata in Luxembourg (Titeux et al. 2009). Sampling squares selected using (d) expert-based, (e) model-based and (f) stratified random sampling strategies. Location of the new *L. collurio* territories found according to (g) expert-based, (h) model-based and (i) stratified random sampling strategies.

Expert-based sampling strategy

We interviewed nationally recognized bird experts who agreed to participate in this study. Judgement was elicited from seven experts to obtain a reliable selection of sampling sites and to decrease the possible geographical biases (McBride, Fidler, & Burgman 2012). These experts were considered to have the best knowledge on *L. collurio* in Luxembourg (see Appendix S1 in Supporting Information). Elicitation sessions were conducted individually and independently to enhance the diversity of knowledge elicited and to avoid experts being unduly influenced by group pressures (Martin et al. 2012). Experts were provided with the locations of the known *L. collurio* territories to produce their guidance on the sampling areas. Using the 1-km resolution grid system in Luxembourg, each expert was asked to select 30 squares with the potential to find as many new shrike territories as possible during ground validation. Experts were informed that a new territory would be considered as such during ground validation if previously known territories were absent within a 200-m distance. Once provided with such information, they were left to select 1-km resolution squares with or without previously known territories. Each expert was asked to allocate the 30 selected squares to three classes: 10 squares classified as high priority, 10 squares as medium priority and 10 squares as low priority for further sampling. In order to aggregate the elicited expert judgements, we used a simple mathematical equal-weighted opinion pooling (Martin et al. 2012) that did not involve any interaction among experts and where they were viewed as equivalent (Clemen & Winkler 1999). A simple spatial overlay rule in a GIS environment (e.g. Sugurmaran & Degroote 2010) allowed us to identify those squares selected by at least two experts. We considered them for sampling as they reflected among-expert agreement on potentially suitable squares. Then, we randomly sampled additional squares among the remaining ones classified as high priority by the experts so as to obtain an expert-based set of 30 sampling squares for ground validation (Fig.1d).

Model-based sampling strategy

Habitat suitability models were developed using a Maximum Entropy procedure (Maxent) (Phillips, Anderson, & Schapire 2006). Maxent is a machine-learning and user-friendly technique based on the principle of maximum entropy that is recommended when using presence-only species data (Phillips, Anderson, & Schapire 2006; Franklin 2009; Elith et al. 2011). Maxent was used to build a habitat suitability map for *L. collurio* based on the link between the known territories of the species and the environmental conditions in those sites (Elith et al. 2011). Shrike records were allocated to 100-m resolution grid cells nested in the same 1-km resolution grid system as the one used during expert elicitation. We selected 10 environmental variables considered to characterize the most important habitat conditions for *L. collurio* (Titeux et al. 2007) (Table 1). All environmental data, available at various resolutions, were resampled to correspond to the 100-m resolution grid with the species presence data and values were derived for each cell. The quadratic terms of the continuous environmental variables were included in addition to the linear functions. We used a five-fold cross-validation approach to define the training and test data sets to fit the models and to statistically evaluate their performance using the area under the receiver operating characteristics (ROC) curve (AUC). AUC values reflected the ability of the model to discriminate between shrike presence records and randomly selected grid cells (Phillips & Dudík 2008).

The modelling outputs at 100-m resolution were aggregated at the scale of the 1-km resolution squares, by adding up the habitat suitability values predicted in the 100-m grid cells enclosed within each square. Aggregated habitat suitability values were then used to rank the squares in decreasing order of suitability for *L. collurio* across Luxembourg. Squares with five or more known shrike territories were eliminated as we considered that the chances of finding additional territories beyond a 200-m distance around the previously known ones were low. From the remaining squares, we selected the top-ranked, most suitable ones (Williams et al. 2009) to create the model-based set of 30 sampling squares for ground validation (Fig.1e).

Table 1. Environmental variables used in a habitat suitability model to identify suitable areas for the red-backed shrikes *L. collurio* in Luxembourg

Variable	Source	Year	Units
Predominant soil type	Soil map	1970	–
Mean percentage slope	Digital elevation model	2001	%
Topographic moisture index†	Digital elevation model	2001	–
Annual crops	Land cover map	2007	m^2
Meadows and pastures	Land cover map	2007	m^2
Urbanized areas	Land cover map	2007	m^2
Distance to closest urbanized	Land cover map	2007	m
Forests	Land cover map	2007	m^2
Distance to closest forest	Land cover map	2007	m
Hedges	Topographic map	1998	m

† Topographic moisture index was calculated following Beven & Kirkby (1979)
Soil map : adapted from 'Carte pédologique du Luxembourg'
Digital elevation model: 'Modèle numérique de terrain du Luxembourg'
Land cover map : 'Occupation biophysique du sol'
Topographic map : 'Base de données topo-cartographique du Luxembourg'

Stratified random sampling strategy

We also selected a set of sampling squares according to a stratified random sampling strategy recently implemented in the common bird monitoring programme in Luxembourg (Titeux et al. 2009). Based on a series of environmental variables known to influence biodiversity (see Table S1), the whole set of 1-km resolution squares in Luxembourg was divided into 10 environmental strata (Fig.1c). In order to cover the main environmental conditions in the country, a stratified random sampling procedure was applied to select a number of squares within each stratum in proportion to their spatial extent. For the common bird monitoring programme, a set of 30 squares was randomly generated and is used for yearly sampling of breeding birds. This set was used here as a baseline reference to reflect a general-purpose sampling strategy (Fig.1f).

Ground validation

Fieldwork was conducted to detect and count *L. collurio* territories in the 1-km resolution squares selected according to each sampling strategy. In the squares selected based on the expert- and model-based sampling strategies, transects with a length of 2.5 km were delineated in potentially suitable open land for shrikes. For the squares selected according to the stratified random sampling strategy, 2.5-km long transects were randomly delineated across all habitat types in the squares, as they constituted the sampling units reflecting a general-purpose sampling strategy. All transects were sampled on foot at a walking speed. Shrike territories were searched with the aid of binoculars and based on auditory cues and they were georeferenced with the highest possible spatial accuracy.

To maximize the probability of finding shrike territories during the breeding season, the selected squares were surveyed once in June and once in July during two consecutive years (2010 and 2011). The squares were sampled by different observers (n=7) within the same dates and using the same field procedure. After the two breeding seasons, field data were integrated with previously known territories to identify the new *L. collurio* territories found in each 1-km square.

Data analysis

The number of new territories found during ground validation was used as a measure of efficiency of the three sampling strategies. This measure was compared among sampling strategies to evaluate if the model-based sampling strategy performed better than by chance (stratified random sampling strategy) and if it was as useful as experts (expert-based sampling strategy).

A likelihood ratio test (LRT) within a generalized linear modelling (GLM) framework with a Poisson distribution was used to compare the efficiency of the three sampling strategies. Year of sampling and observer identity were included as factors in the GLM to account for their effect on the response variable. Interaction terms were not considered, as there was no biologically relevant reason to do so. A post-hoc analysis with multiple comparisons and Bonferroni correction was used to compare the efficiency of the different sampling strategies with each other.

We also tested whether the elicitation of structured expert judgement led to the identification of new *L. collurio* territories closer to the network of established protected areas than methods that explicitly ignored such features. The distance between each new territory and the closest protected area designated under the European Union Directive on the Conservation of Wild Birds (Directive 2009/147/EC) was calculated. Distances were square-root transformed and compared among sampling strategies within a linear modelling (LM) framework with a normal distribution. Year of sampling was included as a factor in the analysis. We also performed a post-hoc analysis to test if there were differences in the mean distance to protected areas between each of the sampling strategies.

RESULTS

A total of 87 1-km resolution squares were sampled during ground validation to evaluate the efficiency of the different sampling strategies to identify new shrike territories. Among the squares selected by the experts, 27 squares were identified by at least two of them and three additional ones were randomly chosen among the rest of the squares classified as high priority by the experts. None of the squares selected based on the general-purpose stratified random sampling strategy were selected according to the expert- or model-based strategies. Only three squares overlapped between expert- and model-based sampling strategies, indicating a high level of discrepancy between the areas identified as with the highest probability of finding new *L. collurio* territories by the experts and the models.

The average AUC value obtained from the five-fold cross-validation in the modelling procedure was 0.85 ± 0.021. This means that there is an 85% probability that a grid cell occupied by the shrike receives a habitat suitability value higher than that of a randomly selected grid cell. Based on this statistical evaluation, model outcomes can be considered as potentially useful (Phillips & Dudík 2008).

A total of 95 new shrike territories were found during ground validation in 2010–2011 when using habitat suitability models, while only 11 new territories were discovered in the squares selected based on the stratified random sampling strategy (Table 2). The average number of territories per km2 found according to the model-based strategy was 2.73 in 2010 and 2.66 in 2011 (range: 0–7), while 0.23 new territories per km2 were found in 2010 and 0.2 in 2011 (range: 0–2) based on the stratified random sampling strategy (Figs 1 & 2). The species was not detected in only three squares selected based on the model-based sampling strategy (in both 2010 and 2011) and this was the case in 24 (in 2010) and 25 (in 2011) squares selected with the stratified random sampling strategy. According to the expert-based sampling strategy, 72 new territories were found during ground validation in 2010–2011 (Table 2), with an average number of territories per km2 of 2.03 in 2010 and 1.73 in 2011 (range: 0–8) (Figs 1 & 2). *Lanius collurio* was not detected in six (in 2010) and eight (in 2011) squares selected according to the expert-based sampling strategy.

Table 2. Total number of new *L. collurio* territories found during 2010 and 2011 according to the expert-based, model-based and stratified random sampling strategies

Sampling strategy	2010	2011	Total in 2010–2011†
Expert-based	61	52	72
Model-based	82	80	95
Stratified random	7	6	11

† Some shrike territories were observed during both years (2010 and 2011)

The GLM analysis indicated that sampling strategy was the only significant ($P < 0.05$) factor explaining the variation in the number of new *L. collurio* territories within the sampling squares (Table 3). The post-hoc multiple comparisons showed that a significantly higher number of new shrike territories were found when using the sampling strategies targeting on the shrike (Table 4): there was a 7.5-fold and a 5.3-fold increase in the number of new territories found per km2 when using the model-based and expert-based sampling strategies respectively, compared to the stratified random sampling strategy. The model-based sampling strategy performed also significantly better than the expert-based sampling strategy in guiding sampling towards areas with a higher number of unknown territories (Table 4): there was a 1.4-fold increase in the number of new territories per km2 when using the model-based sampling strategy compared to the expert-based sampling strategy.

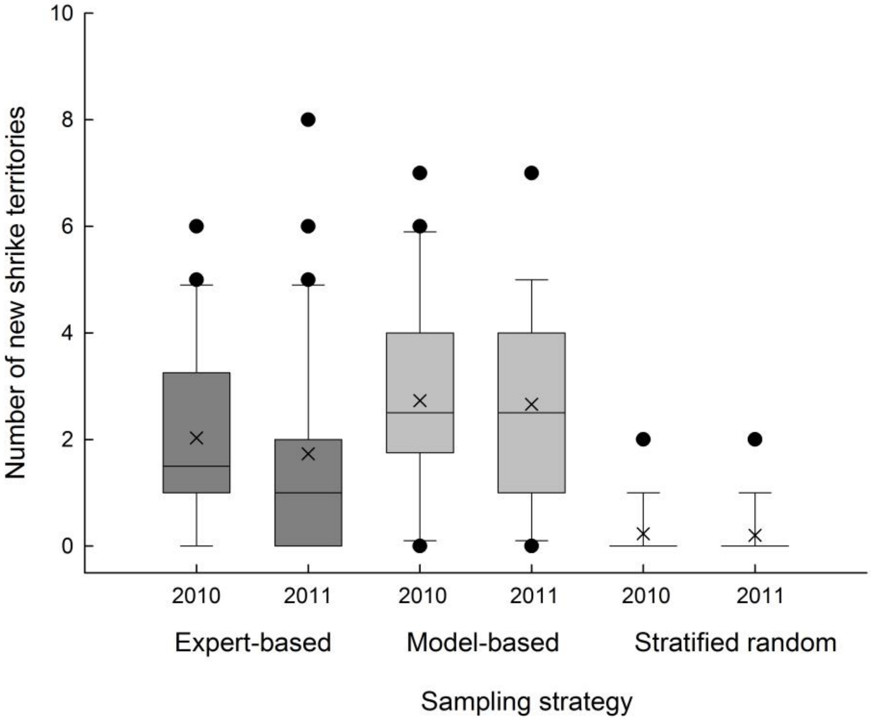


Figure 2. Box-and-whisker plots (⊥ and ⊤: 5th and 95th percentiles, •: outlying values, x: mean value, −: median value) for the number of new *L. collurio* territories found per km2 during 2010 and 2011 according to the expert-based (dark grey), the model-based (light grey) and the stratified random (white) sampling strategies.

New *L. collurio* territories found according to the expert-based sampling strategy were on average closer to protected areas for birds than those found using model-based or stratified random sampling strategies (F2,284 = 5.29, P = 0.005). The post-hoc multiple comparisons showed that new territories found using the expert-based sampling strategy were significantly closer to protected areas than those found according to the model-based sampling strategy (t = -2.96, P = 0.008). The results of the post-hoc comparisons with the stratified random sampling strategy are uncertain due to the low number of new territories found during ground validation when using this sampling strategy.

Table 3. Results of the generalized linear model (GLM) and likelihood ratio tests (LRT) used to determine which factors explained the number of new *L. collurio* territories observed during ground validation.

Factor dropped	d.f.	Deviance	LRT	*P* (>Chi)
(full model)		215		
Sampling strategy	2	241.33	26.337	1.91E-06
Year	1	218.83	3.837	0.051
Observer	6	225.83	10.834	0.093

Table 4. Results of the post-hoc multiple pairwise comparisons with Bonferroni correction used to compare the efficiency of the different sampling strategies to find new *L. collurio* territories during ground validation

Sampling strategy	Estimate†	SE	z value	*P* (>|z|)
Model-based – Stratified random	2.022	0.516	3.921	<0.001
Expert-based – Stratified random	1.666	0.524	3.177	0.003
Model-based – Expert-based	0.356	0.124	2.861	0.009

† Estimates are provided using a log scale and have to be inverse transformed using exp() to compare the relative efficiency of different sampling strategies on a linear scale

DISCUSSION

Decision-making for biodiversity conservation and management often involves dealing with alternative options when ecological knowledge is incomplete and outcomes are uncertain (Regan *et al.* 2005). In day-to-day practice, practitioners work with short timelines and limited resources (Cook, Hockings, & Carter 2010). Hence, they frequently use expert opinion to support conservation and management decision-making (Fazey *et al.* 2006; Addison *et al.* 2013). The subjective opinion of experts may induce opaque or ill-informed management decisions due to psychological and/or motivational biases (Burgman *et al.* 2011). It is expected that the use of quantitative data and scientific tools by managers and practitioners to support their decisions will improve the overall efficiency of conservation and management interventions (Sutherland *et al.* 2004; Drolet *et al.* 2015).

Among other available scientific tools that use quantitative data, habitat suitability models have been proposed to play a key role in supporting conservation decision-making (e.g. Guisan & Thuiller 2005). With the limited funds available for biodiversity conservation and management, the implementation of predictive modelling approaches is often considered costly and resource intensive (e.g. hardware, technical requirements, need for in-house expertise) in comparison with the experience and knowledge of practitioners (Borowski & Hare 2007). However, one of the main advantages of such approaches is that, once they are operational, they can be applied routinely to a large number of species and outcomes may be repeatedly updated in a cost-efficient way as new data are collected (Guisan *et al.* 2006). Yet, despite the demonstrated performance and benefits of predictive models, practitioners may remain sceptical about their usefulness and sometimes object to their use for conservation practice, as they often believe models do not outperform expert opinion or consider models to be wrong, inaccurate or inappropriate (Jeltsch *et al.* 2013; Addison *et al.* 2013). As a consequence, model outcomes are rarely translated into actions and decisions that actually contribute to biodiversity conservation and management (Guisan *et al.* 2013). Among the few examples of the successful application of models in a management decision-making framework, Brotons, Mañosa, & Estrada (2004) used habitat suitability models to identify critical habitats for endangered bird species and this information was used in a legal decree to guide land-use decisions in a farmland area affected by a large-scale irrigation plan.

As other authors have stressed, we also believe that there is a need for a stronger linkage between practitioners and modellers to improve the relevance of models as tools to support conservation and management decision-making. Involving experts in the modelling procedure might be one way to reinforce the link between the two communities. Such integration within an expert-informed modelling framework is expected to reduce the reluctance that some practitioners may show for model-based approaches and to increase their relevance and field of application (Krueger *et al.* 2012). Another way to contribute to convincing practitioners of the usefulness of habitat suitability models is by confronting the efficiency of such tools with that of structured expert judgement to guide conservation decision-making. McConnachie & Cowling (2013) even go a step ahead and examine the ability of practitioners to learn and update their beliefs after being provided with the outcomes from model-based approaches.

Here, we used a structured ground validation procedure to evaluate and compare the ability of experts and models to achieve a clearly defined conservation objective, i.e. optimizing the detection of new presence areas and improving the current knowledge on the distribution of a bird species of conservation-concern. A stratified random sampling strategy was first used as a baseline reference to evaluate the outcomes of the other sampling strategies targeting on the focal species. As expected, these sampling strategies performed much better than the stratified random sampling strategy. Guisan *et al.* (2006) and Le Lay *et al.* (2010) also showed that model-based sampling strategies considerably increase the discovery rates of new populations of rare plant species compared to random sampling designs. Stratified random sampling approaches are general-purpose designs classically implemented in biodiversity mapping or monitoring projects. However, they remain poorly suited to detect rare or threatened species, either because of the low probability of finding the species by chance across the study area or because the species may be restricted to particular habitat types that have been overlooked in the stratification approach (Le Lay *et al.* 2010).

In contrast with most studies that assessed the efficiency of model-based sampling strategies by comparing it to the results obtained according to a random sampling procedure (Guisan *et al.* 2006; Le Lay *et al.* 2010), we also directly challenged the performance of a model-based sampling strategy with the outcomes of a structured expert-based approach using the same baseline presence data (Clevenger *et al.* 2002; Drolet *et al.* 2015). We implemented extensive fieldwork and we showed that the model-based sampling strategy significantly outperformed the expert-based strategy, increasing the number of new shrike territories found per km^2 in Luxembourg by a factor 1.4. If we are to advocate on the usefulness of model-based approaches to address a management objective, providing such evidence that models may guide the prospective sampling of species of conservation-concern as good as, and even better than structured expert judgement, is really needed for two reasons. First, objection to the use of models often comes from the fact that decision-makers consider that model outcomes do not result in better predictions than those provided by the subjective opinion of experts (Addison *et al.* 2013). Second, modelling outcomes alone might be insufficient for practitioners to change their beliefs (McConnachie & Cowling 2013).

As we used a single-species approach due to the extensive fieldwork needed during ground validation (see also Guisan *et al.* 2006), we acknowledge the limitations associated with the overall conclusions that may be derived from this study. Structured expert judgment may prove to perform better than models in the case of rare or elusive species due to insufficient or low-quality data to build reliable models (Doswald, Zimmermann, & Breitenmoser 2007; Turvey *et al.* 2015). Hence, it is now warranted to make such comparisons using a number of species across a range of scales because it is still open for discussion whether the observed pattern is actually representative of a larger sample of species, experts and regions. It would also be needed to examine alternative procedures to deal with the multiple judgements of several experts in the identification of the priority squares for further sampling as this might influence the performance of the expert-based sampling strategy. We regard the results of the present study as an incentive to test further the usefulness of habitat suitability models through a direct comparison with structured expert judgement. We anticipate that the outcomes of such an extensive comparison will help to reduce the scepticism and prejudice against information derived from modelling procedures and will contribute to convincing practitioners of the usefulness of such tools to improve on the management of species of conservation-concern.

Interestingly, the new territories found according to the expert-based sampling strategy in our study were on average closer to protected areas designated for bird conservation than those found using the model-based sampling strategy.

These results indicate that eliciting expert judgement may guide sampling strategy towards protected but potentially less suitable areas for the target species, whereas models ignore information on protected areas and have the potential to identify unprotected but highly suitable areas. This probably reflects some geographical, psychological or motivational biases in expert judgement (Burgman *et al.* 2011). Although sophisticated elicitation procedures are available to mitigate such biases and could be further implemented in this context, they remain among the most important limitations of structured expert judgement. Cowling *et al.* (2003) also showed some differences between expert-based and systematic approaches when identifying important conservation areas for biodiversity and highlighted the importance of considering these two approaches as complementary instead of mutually exclusive. Based on our results, we also suggest that expert-based methods may be best suited to guide possible extensions or enlargements of already existing protected areas, while predictive models may contribute to guiding the creation of additional protected areas when data, time and resources are available. Thus, even though the modelling process and expert judgement elicitation were carried out independently in our study for comparison purposes, our results suggest the importance of moving forward with integrated model- and expert-based approaches for conservation and management decision-making, rather than emphasizing the dichotomy between both (Guisan *et al.* 2013; Drolet *et al.* 2015). More generally, we encourage managers and modellers to work hand in hand to help bridge the research–implementation gap between conservation science and real-world action (Knight *et al.* 2008; Sutherland & Freckleton 2012).

ACKNOWLEDGEMENTS

We thank experts and fieldworkers for their participation in this study. BirdLife Luxembourg provided the data on bird distribution. Digital elevation model, land cover, topographic, soil and geological maps were provided by the Administration du cadastre et de la topographie, the Administration des services techniques de l'agriculture, the Ministère du développement durable et des infrastructures and the Ministère des Travaux Publics (Luxembourg). O.A. was funded by the National Research Fund, Luxembourg (AFR-PHD-08-63). N.T. and L.B. were funded through the EU BON project (contract no. 308454; FP7-ENV-2012, European Commission).

REFERENCES

Addison, P.F.E., Rumpff, L., Bau, S.S., Carey, J.M., Chee, Y.E., Jarrad, F.C., McBride, M.F. & Burgman, M.A. (2013) Practical solutions for making models indispensable in conservation decision-making. *Diversity and Distributions*, **19**, 490–502.

Aizpurua, O., Paquet, J.-Y., Brotons, L. & Titeux, N. (2015) Optimising long-term monitoring projects for species distribution modelling: how atlas data may help. *Ecography*, **38**, 29–40.

Anadón, J.D., Giménez, A., Ballestar, R. & Pérez, I. (2009) Evaluation of local ecological knowledge as a method for collecting extensive data on animal abundance. *Conservation Biology*, **23**, 617–625.

Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.

Beven, K.J. & Kirkby, M.J. (1979) Physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, **24**, 43–69.

Borowski, I. & Hare, M. (2007) Exploring the gap between water managers and researchers: Difficulties of model-based tools to support practical water management. *Water Resources Management*, **21**, 1049–1074.

Brotons, L., Mañosa, S. & Estrada, J. (2004) Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodiversity and Conservation*, **13**, 1039–1058.

Burgman, M.A., Carr, A., Godden, L., Gregory, R., McBride, M.F., Flander, L. & Maguire, L. (2011) Redefining expertise and improving ecological judgment. *Conservation Letters*, **4**, 81–87.

Cabeza, M. & Moilanen, A. (2001) Design of reserve networks and the persistence of biodiversity. *Trends in Ecology & Evolution*, **16**, 242–248.

Cerqueira, M.C., Cohn-Haft, M., Vargas, C.F., Nader, C.E., Andretti, C.B., Costa, T.V.V., Sberze, M., Hines, J.E. & Ferraz, G. (2013) Rare or elusive? A test of expert knowledge about rarity of Amazon forest birds. *Diversity and Distributions*, **19**, 710–721.

Clevenger, A.P., Wierzchowski, J., Chruszcz, B. & Gunson, K. (2002) Identifying wildlife habitat linkages and planning mitigation passages. *Conservation Biology*, **16**, 503–514.

Cook, C.N., Hockings, M. & Carter, R.W. (2010) Conservation in the dark? The information used to support management decisions. *Frontiers in Ecology and the Environment*, **8**, 181–186.

Cowling, R.M., Pressey, R.L., Sims-Castley, R., le Roux, A., Baard, E., Burgers, C.J. & Palmer, G. (2003) The expert or the algorithm? - comparison of priority conservation areas in the Cape Floristic Region identified by park managers and reserve selection software. *Biological Conservation*, **112**, 147–167.

Crall, A.W., Jarnevich, C.S., Panke, B., Young, N., Renz, M. & Morisette, J. (2013) Using habitat suitability models to target invasive plant species surveys. *Ecological Applications*, **23**, 60–72.

Doswald, N., Zimmermann, F. & Breitenmoser, U. (2007) Testing expert groups for a habitat suitability model for the lynx Lynx lynx in the Swiss Alps. *Wildlife Biology*, **13**, 430–446.

Drolet, D., Locke, A., Lewis, M.A. & Davidson, J. (2015) Evidence-based tool surpasses expert opinion in predicting probability of eradication of aquatic nonindigenous species. *Ecological Applications*, **25**, 441–450.

Elith, J., Phillips, S.J.P., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Fazey, I., Fazey, J.A., Salisbury, J.G., Lindermayer, D.B. & Dovers, S. (2006) The nature and role of experiential knowledge for environmental conservation. *Environmental Conservation*, **33**, 1–10.

Franklin, J. (2009) *Mapping Species Distributions*. Cambridge University Press, Cambridge, UK.

Gormley, A.M., Forsyth, D.M., Griffioen, P., Lindeman, M., Ramsey, D.S., Scroggie, M.P. & Woodford, L. (2011) Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, **48**, 25–34.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A. & Zimmermann, N.E. (2006) Using niche-based models to improve the sampling of rare species. *Conservation Biology*, **20**, 501–511.

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

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hajkowicz, S. (2007) A comparison of multiple criteria analysis and unaided approaches to environmental decision making. *Environmental Science & Policy*, **10**, 177–184.

Jeltsch, F., Blaum, N., Brose, U., Chipperfield, J.D., Clough, Y., Farwig, N., Geissler, K., Graham, C.H., Grimm, V., Hickler, T., Huth, A., May, F., Meyer, K.M., Pagel, J., Reineking, B., Rillig, M.C., Shea, K., Schurr, F.M., Schröder, B., Tielbörger, K., Weiss, L., Wiegand, K., Wiegand, T., Wirth, C. & Zurell, D. (2013) How can we bring together empiricists and modellers in functional biodiversity research? *Basic and Applied Ecology*, **14**, 93–101.

Knight, A.T., Cowling, R.M., Rouget, M., Balmford, A., Lombard, A.T. & Campbell, B.M. (2008) Knowing but not doing: selecting priority conservation areas and the research-implementation gap. *Conservation biology*, **22**, 610–617.

Krueger, T., Page, T., Hubacek, K., Smith, L. & Hiscock, K. (2012) The role of expert opinion in environmental modelling. *Environmental Modelling & Software*, **36**, 4–18.

Kuhnert, P.M., Martin, T.G. & Griffiths, S.P. (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, **13**, 900–914.

Le Lay, G., Engler, R., Franc, E. & Guisan, A. (2010) Prospective sampling based on model ensembles improves the detection of rare species. *Ecography*, **33**, 1015–1027.

Lorgé, P. & Melchior, E. (2010) Neuntöter. *Vögel Luxemburgs* (ed L.N. Vulleschutzliga), p. 152. Saint-Paul Publishing - Sacha Heck, Luxembourg.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M.F. & Mengersen, K. (2012) Eliciting expert knowledge in conservation science. *Conservation Biology*, **26**, 29–38.

McBride, M.F., Garnett, S.T., Szabo, J.K., Burbidge, A.H., Butchart, S.H.M., Christidis, L., Dutson, G., Ford, H.A., Loyn, R.H., Watson, D.M. & Burgman, M.A. (2012) Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. *Methods in Ecology and Evolution*, **3**, 906–920.

McConnachie, M.M. & Cowling, R.M. (2013) On the accuracy of conservation managers' beliefs and if they learn from evidence-based knowledge: a preliminary investigation. *Journal of Environmental Management*, **128**, 7–14.

Murray, J. V., Goldizen, A.W., O'Leary, R.A., Mcalpine, C.A., Possingham, H.P. & Choy, S.L. (2009) How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies Petrogale penicillata. *Journal of Applied Ecology*, **46**, 842–851.

Pearce, J.L., Cherry, K., Drielsma, M., Ferrier, S. & Whish, G. (2001) Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. *Journal of Applied Ecology*, **38**, 412–424.

Rebelo, H. & Jones, G. (2010) Ground validation of presence-only modelling with rare species: A case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). *Journal of Applied Ecology*, **47**, 410–420.

Regan, H.M., Ben-Haim, Y., Langford, B., Wilson, W.G., Lundberg, P., Andelman, S.J. & Burgman, M.A. (2005) Robust decision-making under severe uncertainty for conservation management. *Ecological Apllications*, **15**, 1471–1477.

Robertson, M.P., Cumming, G.S. & Erasmus, B.F.N. (2010) Getting the most out of atlas data. *Diversity and Distributions*, **16**, 363–375.

Sardà-Palomera, F., Brotons, L., Villero, D., Sierdsema, H., Newson, S.E. & Jiguet, F. (2012) Mapping from heterogeneous biodiversity monitoring data sources. *Biodiversity and Conservation*, **21**, 2927–2948.

Sutherland, W.J. & Freckleton, R.P. (2012) Making predictive ecology more relevant to policy makers and practitioners. *Philosophical transactions of the Royal Society of London. Series B-Biological sciences*, **367**, 322–30.

Sutherland, W.J., Pullin, A.S., Dolman, P.M. & Knight, T.M. (2004) The need for evidence-based conservation. *Trends in Ecology & Evolution*, **19**, 305–308.

Titeux, N., Biver, G., Lorgé, P. & Hoffmann, L. (2009) Implementation of the Common Bird Monitoring scheme in Luxembourg. *Bird Census News*, **22**, 42–50.

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. & Defourny, P. (2007) Fitness-related parameters improve presence-only distribution modelling for conservation practice: The case of the Red-backed shrike. *Biological Conservation*, **138**, 207–223.

Turvey, S.T., Trung, C.T., Quyet, V.D., Nhu, H. Van, Thoai, D. Van, Tuan, V.C.A., Hoa, D.T., Kacha, K., Sysomphone, T., Wallate, S., Hai, C.T.T., Thanh, N. Van & Wilkinson, N.M. (2015) Interview-based sighting histories can inform regional conservation prioritization for highly threatened cryptic species. *Journal of Applied Ecology*, **52**, 422–433.

Vaughan, I.P. & Ormerod, S.J. (2005) The continuing challenges of testing species distribution models. *Journal of Applied Ecology*, **42**, 720–730.

Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brien, J.M. & Schwartz, M.W. (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, **15**, 565–576.

SUPPORTING INFORMATION

Appendix S1. Further explanation about expert selection and qualification

The conventional approach to defining experts is by their qualifications, track record, professional standing, and experience. However, Burgman *et al.* 2011) emphasizes that such an approach may sometimes exclude people with useful knowledge and lead to biases of expert judgments linked to the social status of the selected experts. For that reason, we decided to identify experts based on the qualifications and experience allowing them to provide us with 'contributory expertise' (i.e. fully developed and internalized skills and knowledge, *sensu* Burgman *et al.* 2011) on bird ecology and conservation at the national scale. The small population of the country (Luxembourg: 543,000 inhabitants) made it possible to identify easily the people working in this field. They were identified and asked to participate in the study following a snowball sampling method (Ntshotsho *et al.* 2015; Turvey *et al.* 2015), where we directly asked them if they knew other people fitting in the study requirements. A total of 9 experts were identified and invited to participate in the study: 7 of them accepted the invitation and 2 of them declined it due to a lack of time. All of them were selected because of their 'lay' knowledge (Burgman *et al.* 2011), obtained with many years (around 15-20 years for each of them) of experience in the field. They have been collecting bird distribution data in the field across the whole country for many years on a volunteer (3 experts) or professional (4 experts) basis. Two of them are particularly known for their personal interest for shrikes. Due to the small size of the country, we assumed that the expertise selected for this study covered the whole distribution of the target species at the national level.

References

Burgman, M.A., Carr, A., Godden, L., Gregory, R., McBride, M.F., Flander, L. & Maguire, L. (2011) Redefining expertise and improving ecological judgment. Conservation Letters, 4, 81–87.

Ntshotsho, P., Prozesky, H.E., Esler, K.J. & Reyers, B. (2015) What drives the use of scientific evidence in decision making? The case of the South African Working for Water program. Biological Conservation, 184, 136–144.

Turvey, S.T., Trung, C.T., Quyet, V.D., Nhu, H. Van, Thoai, D. Van, Tuan, V.C.A., Hoa, D.T., Kacha, K., Sysomphone, T., Wallate, S., Hai, C.T.T., Thanh, N. Van & Wilkinson, N.M. (2015) Interview-based sighting histories can inform regional conservation prioritization for highly threatened cryptic species. Journal of Applied Ecology, 52, 422–433.

Table S1. Environmental variables calculated in the 1-km resolution squares and used to identify the environmental strata in Luxembourg

Variable	Source	Year	Units
Mean elevation	Digital elevation	2001	m
Mean orientation	Digital elevation	2001	°
Mean percentage slope	Digital elevation	2001	%
Annual mean temperature	WorldClim	2005	°C
Temperature seasonality	WorldClim	2005	°C
Maximum temp. of warmest month	WorldClim	2005	°C
Minimum temp. of coldest month	WorldClim	2005	°C
Temperature annual range	WorldClim	2005	°C
Annual precipitation	WorldClim	2005	mm
Precipitation of wettest month	WorldClim	2005	mm
Precipitation of driest month	WorldClim	2005	mm
Precipitation seasonality	WorldClim	2005	mm
Urban areas	Land cover map	2007	m2
Annual crops	Land cover map	2007	m2
Vineyards	Land cover map	2007	m2
Meadows and pastures	Land cover map	2007	m2
Broad-leaved forests	Land cover map	2007	m2
Coniferous forests	Land cover map	2007	m2
Mixed forests	Land cover map	2007	m2
Dry grasslands	Land cover map	2007	m2
Heathlands	Land cover map	2007	m2
Shrub-covered areas	Land cover map	2007	m2
Rocky areas	Land cover map	2007	m2
Wetlands	Land cover map	2007	m2

Standing water	Land cover map	2007	m2
Running water	Land cover map	2007	m2
Orchards	Land cover map	2007	m2
Alluvial substrate	Geological map	1992	m2
Calcareous substrate	Geological map	1992	m2
Sandstone substrate	Geological map	1992	m2
Silty substrate	Geological map	1992	m2
Impermeable marly substrate	Geological map	1992	m2
Semi-permeable marly substrate	Geological map	1992	m2
Schisteous substrate	Geological map	1992	m2
Loamy soils	Soil map	2007	m2
Clayey-loamy stony (quartz pebble) soils	Soil map	2007	m2
Clayey and stony (dolomite) soils	Soil map	2007	m2
Clayey and stony (limestone) soils	Soil map	2007	m2
Sandy-loamy soils	Soil map	2007	m2
Sandy-loamy to loamy soils	Soil map	2007	m2
Colluvial soils	Soil map	2007	m2
Seeping areas	Soil map	2007	m2
Soils modified by human activities	Soil map	2007	m2
Alluvial soils	Soil map	2007	m2

Digital elevation model: '*Modèle numérique de terrain du Luxembourg*' (spatial resolution 5 m)

Worldclim: WorldClim Global Climate Data (http://www.worldclim.org)

Land cover map: '*Occupation biophysique du so du Luxembourgl*' (scale 1:15,000)

Geological map: adapted from '*Carte géologique du Luxembourg*' (scale 1:100,000)

Soil map: adapted from '*Carte pédologique du Luxembourg*' (scale 1:100,000)

Article 2

Optimising long-term monitoring projects for species distribution modelling: how atlas data may help

The content of this chapter has been published*:

Aizpurua, O., Paquet, J.Y., Brotons, L., Titeux, N. (2015). Optimising long-term monitoring projects for species distribution modelling: how atlas data may help. *Ecography* **38**, 29-40.

*text in the thesis has been marginally been modified compared to the published version

ABSTRACT

Long-term biodiversity monitoring data are mainly used to estimate changes in species occupancy or abundance over time, but they may also be incorporated into predictive models to document species distributions in space. Although changes in occupancy or abundance may be estimated from a relatively limited number of sampling units, small sample size may lead to inaccurate spatial models and maps of predicted species distributions. We provide a methodological approach to estimate the minimum sample size needed in monitoring projects to produce accurate species distribution models and maps. The method assumes that monitoring data are not yet available when sampling strategies are to be designed and is based on external distribution data from atlas projects. Atlas data are typically collected in a large number of sampling units during a restricted timeframe and are often similar in nature to the information gathered from long-term monitoring projects. The large number of sampling units in atlas projects makes it possible to simulate a broad gradient of sample sizes in monitoring data and to examine how the number of sampling units influences the accuracy of the models. We apply the method to several bird species using data from a regional breeding bird atlas. We explore the effect of prevalence, range size and habitat specialization of the species on the sample size needed to generate accurate models. Model accuracy is sensitive to particularly small sample sizes and levels off beyond a sufficiently large number of sampling units that varies among species depending mainly on their prevalence. The integration of spatial modelling techniques into monitoring projects is a cost-effective approach as it offers the possibility to estimate the dynamics of species distributions in space and over time. We believe our innovative method will help in the sampling design of future monitoring projects aiming to achieve such integration.

INTRODUCTION

Long-term wildlife monitoring is generally considered as an essential tool for biodiversity management and for research studies on biodiversity conservation (Gitzen *et al.* 2012). Monitoring projects primarily aim at delivering information on the changing status of key features of biodiversity (Lindenmayer *et al.* 2012). State variables are used to characterise the status of these features at different points in time with a view to assessing system state and inferring changes in state over time (Gitzen *et al.* 2012). State variables include, among others, species occupancy (MacKenzie *et al.* 2005; Kéry *et al.* 2009) or species abundance (Royle & Nichols 2003). In such projects, field data are often repeatedly collected over time in a network of sampling units according to standardised procedures (Gitzen *et al.* 2012). Previous studies have reported that monitoring projects have also the potential to provide an appropriate source of data to document the distribution of species in space (Brotons, Herrando, & Pla 2007; Braunisch & Suchant 2010; Rodhouse *et al.* 2012). Mapping species distributions in space and documenting how they change over time may provide key information to guide effective landscape and conservation planning. Dynamic species distribution mapping may, therefore, be considered as an essential component of a biodiversity monitoring project (Brotons, Herrando, & Pla 2007; Kéry, Guillera-Arroita, & Lahoz-Monfort 2013). In any monitoring project, sampling units are, however, sparsely distributed over the region of interest, which is inconvenient for a straightforward mapping of species distributions.

Species distribution modelling is an increasingly used technique (Rodríguez *et al.* 2007) that can produce distribution maps based on monitoring data (Brotons *et al.* 2006). With these models, environmental variables describing the habitat conditions in the sampling units are related to records of species presence. These models are used to predict the species distribution beyond the sampling units in areas where species occurrence is unknown (Araújo & Guisan 2006; Elith, Kearney, & Phillips 2010). The use of models to predict species distributions is of key significance for biodiversity conservation (Guisan *et al.* 2013). Among several applications, models may be used to identify the most important environmental conditions that influence species distributions or to guide the prioritization of management options amongst areas that vary in their suitability for the species (Titeux *et al.* 2007). Species distribution models are also often built to explore the impacts of environmental changes on future species distributions (Elith, Kearney, & Phillips 2010). Previous studies examined the use of monitoring data to generate species distribution models (Brotons *et al.* 2006; Brotons, Herrando, & Pla 2007; Braunisch & Suchant 2010) and showed that the integration of monitoring data into modelling approaches may contribute to understanding how species distributions change over time (De Cáceres & Brotons 2012; Rodhouse *et al.* 2012; Kéry, Guillera-Arroita, & Lahoz-Monfort 2013).

Sampling design in a monitoring project typically results from a balance between the number of sampling units and the number of repeated surveys in these units to document the state variables with an acceptable level of precision (MacKenzie *et al.* 2005). A limited number of sampling units and a sufficient number of repeated surveys may be suited, and in some cases recommended, to derive unbiased estimates of the state variables (MacKenzie & Royle 2005; Kéry *et al.* 2009; MacKenzie 2012). This appropriate sampling design for monitoring purposes may, however, fail to produce enough spatial data to build relevant species distribution models (Brotons, Herrando, & Pla 2007), because a small number of sampling units is known to induce inaccurate spatial models

(Hernandez *et al.* 2006; Wisz *et al.* 2008; Jiménez-Valverde, Lobo, & Hortal 2009; Bean, Stafford, & Brashares 2012). This drawback can be avoided if dynamic species distribution mapping is explicitly considered when setting the objectives of the monitoring project and when making decisions about sampling design. At this stage of a project, existing monitoring data in the region of interest are, however, not yet available and other sources of information based on upfront sampling efforts are needed to help putting the monitoring project into place (Hooten, Ross, & Wikle 2012).

Atlas projects are an interesting source of spatial information that may assist in making such pilot analysis. Two-stage sampling design (Thompson 2012) is increasingly implemented in 'last-generation' atlases (Estrada *et al.* 2004; Jacob *et al.* 2010; Maes, Vanreusel, & Van Dyck 2013): species presence or abundance is recorded in 1) primary sampling units to provide a picture of the species distribution across the whole region of interest but at coarse spatial resolution and in 2) a set of secondary sampling units nested within the primary ones to explore species distribution at finer resolution. Last-generation atlases are generally completed over considerable time periods and repeated at long time intervals (Dunn & Weston 2008), which prevents them from being suited to detect changes in species distributions with time scales matching decision-making needs. Interestingly, field sampling procedures for atlas data collection in secondary sampling units (e.g. bird or butterfly counts along transects) are often similar in nature to the procedures implemented in long-term monitoring projects (Vorísek *et al.* 2008; Van Swaay *et al.* 2008). Such kind of atlas data are generally collected only once during the atlas period, but in a large number of secondary sampling units to cover an important part of the region of interest at a fine spatial resolution (Carden *et al.* 2010; Maes, Vanreusel, & Van Dyck 2013). Hence, atlas data in secondary sampling units may be manipulated to imitate a broad gradient of sample sizes in a monitoring project and to build species distribution models with varying numbers of sampling units. Such an approach may, in turn, contribute to identifying how large the number of sampling units should be at the start of a monitoring project if dynamic species distribution mapping is set as an objective.

Here, we provide an innovative analytical framework using data from last-generation atlases to aid in the initial design of monitoring projects able to generate appropriate data for the production of accurate species distribution models and maps. We draw attention to important issues that are to be addressed if we are to generate and update species distribution maps as a direct output of long-term monitoring projects. This study illustrates how datasets derived from last-generation atlas projects can contribute to the integration of spatial modelling techniques into long-term monitoring studies in order to cost-efficiently estimate biodiversity dynamics in space and over time (Rodríguez *et al.* 2007).

METHODS

An increasing number of atlas projects with two-stage sampling designs become available for different taxa worldwide (Estrada *et al.* 2004; Carden *et al.* 2010; Maes, Vanreusel, & Van Dyck 2013) and may support the integration of spatial modelling techniques into monitoring studies. The following analytical framework is of general interest as it can be applied to any dataset derived from such last-generation atlas projects. In the present study, we apply this innovative method to the ‘Breeding Bird Atlas of Wallonia’ (BBAW) data (Jacob *et al.* 2010).

Study area

Belgium is a heavily industrialized north-western European country with a high human population density. The southern part of Belgium (Wallonia, ca 16 850 km^2, Fig. 1a) is characterised by a strong gradient in landscape composition, from a densely populated and agriculture dominated landscape in the northwest to a hilly landscape with an important cover of forest and grassland in the southeast (Jacob *et al.* 2010).

Atlas data

During 2001-2007, 650 volunteer fieldworkers participated in the BBAW data collection. Data were collected across a range of spatial resolutions according to a two-stage sampling design and an additional territory-mapping procedure.

Grid-based procedure: primary sampling units

Based on regular field visits during day and night from February to August, fieldworkers were asked to report the presence, estimate the abundance and record the breeding evidence for all bird species in 40 km^2 (5 x 8 km) primary sampling units (n=514, Fig. 1b). Fieldworkers paid particular attention to survey the different habitat types present in the primary sampling units. Abundance was estimated by fieldworkers in the form of 9 abundance classes derived from a geometric progression with a common ratio set to 2 (see details in Jacob et al. 2010) and the central value of each class was used in subsequent analyses. The highest possible breeding evidence for each species was provided according to the EOAC classification, i.e. non-breeding, possible breeding, probable breeding and confirmed breeding (Timothy & Sharrock 1974).

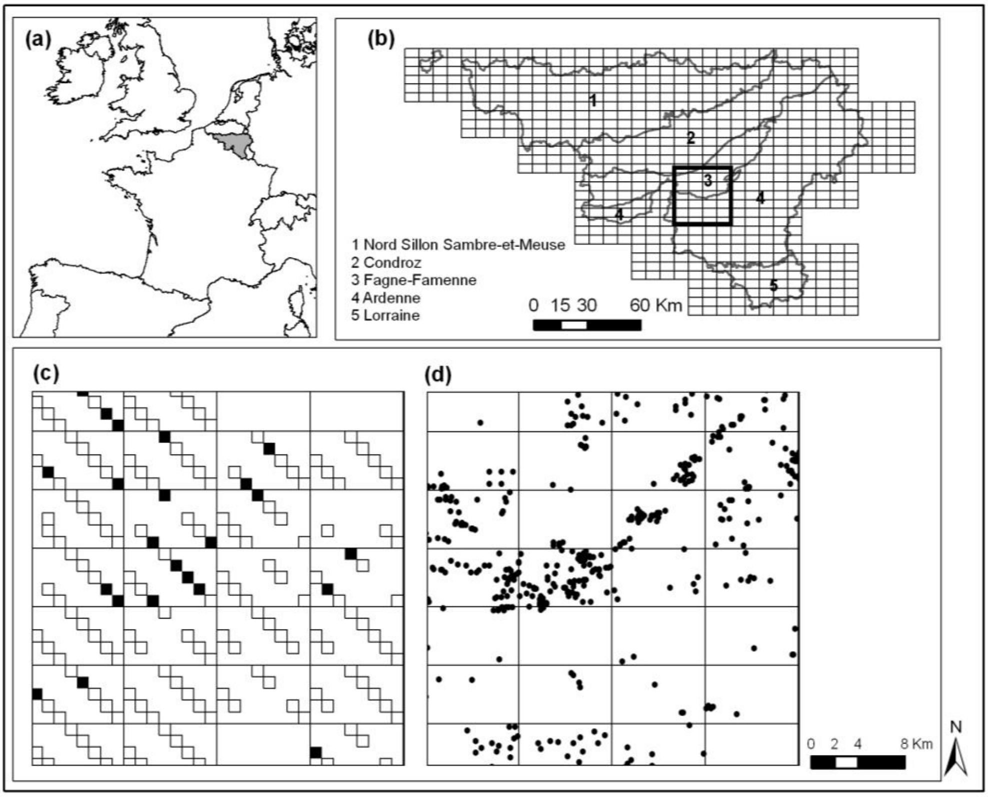


Figure 1. (a) Location of Wallonia in NW Europe. (b) Main ecological regions in Wallonia and grid system of the *Breeding Bird Atlas of Wallonia* with the 40-km^2 (5x8 km) primary sampling units. (c) Subset of the study area with the 1-km^2 secondary sampling units (black squares show an example with red-backed shrike *Lanius collurio* presence records collected during the transect-based procedure). (d) Same subset of the study area as in (c) with *L. collurio* territories (black dots) mapped during the simplified territory-mapping procedure.

Transect-based procedure: secondary sampling units

Secondary sampling units of 1-km^2 squares were selected according to a regular and systematic sampling design (see details in Jacob et al. 2010) so that all primary sampling units were geographically covered in the same way by the secondary sampling units (Fig. 1c). Within these secondary sampling units, transects were delineated by volunteer fieldworker to cover the whole diversity of habitats in the squares. Fieldworkers walked during 1 hour along these sampling routes in the first five hours after sunrise and twice a year during breeding season to record early and late breeders. Each breeding or non-breeding bird (detected either by sight or by sound) was recorded individually. In each secondary sampling unit, the transect-based procedure was conducted in only one year during the timeframe of the BBAW project. The number of secondary sampling units surveyed during the BBAW project (n=2800) covered almost 17% of the study area.

Territory-mapping procedure

At the start of the BBAW project, bird species were classified in low-, moderate- and high-abundance species according to prior knowledge of their regional abundance. Based on territorial indications collected during the regular field visits conducted in the diversity of habitat types within the primary sampling units, fieldworkers were asked to map the locations of all detected territories or colonies of low- and moderate-abundance species. These locations were considered as the centres of the territories and were associated with an accuracy ranging from 100 to 500 meters as estimated by the fieldworkers (Fig. 1d). This simplified territory-mapping procedure is a detailed and time-consuming technique and is unachievable over large areas on a regular basis.

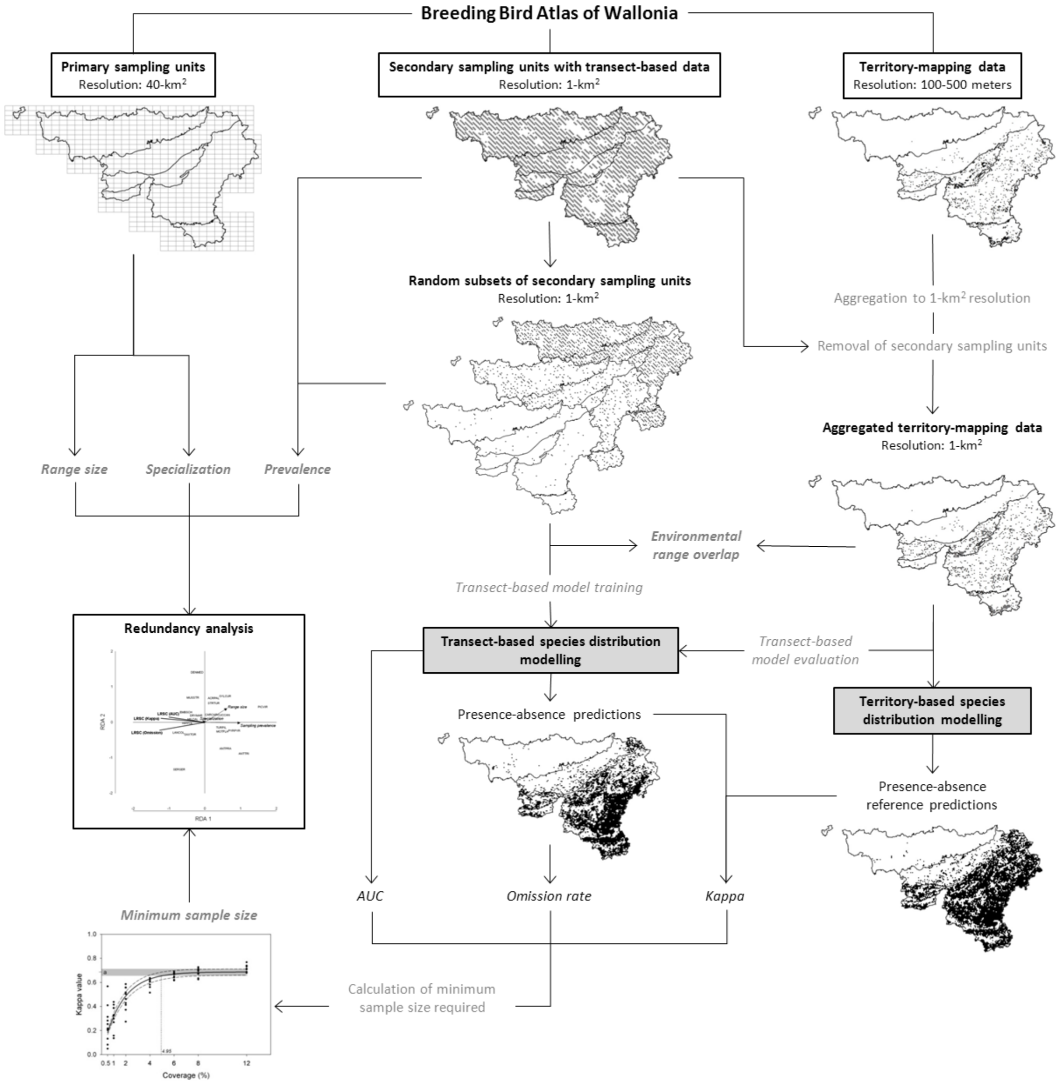


Figure 2. Overview of the modelling and analytical framework. Red-backed shrike *Lanius collurio* is used as an example.

Overview of the modelling approach

In our analytical framework (Fig. 2), we considered the data collected during the transect-based procedure in the secondary sampling units as equivalent to long-term monitoring data (Vorísek *et al.* 2008; Maes *et al.* 2012). We used these data as a basis to produce large-scale, fine-resolution species distribution models (hereafter ‘transect-based models’) and we manipulated the number of secondary sampling units in order to examine the effect of sample size on the performance of the models. The territory-mapping data covered the whole study area and provided the best available information on the distribution and habitat requirements of low- to moderate-abundance species. Therefore, we used territory-mapping data as a reference to evaluate the performance of the transect-based models. Then, we calculated the minimum sample size (i.e. minimum number of secondary sampling units) needed to reach an acceptable level of modelling performance based on three different evaluation measures. Finally, we evaluated for the whole set of species the effect of prevalence, range size and habitat specialization on the minimum sample size (redundancy analysis).

Transect-based model training

We randomly selected subsets of the available secondary sampling units to simulate a range of sample sizes in a long-term monitoring project (Jiménez-Valverde, Lobo, & Hortal 2009) 2009): 0.5% of the study area (sample size: n=83 secondary sampling units), 1% (n=166), 2% (n=332), 4% (n=664), 6% (n=996), 8% (n=1328) and 12 % (n=1992). In order for the subsets of secondary sampling units to be spread out over the whole environmental gradient in the study area, they were generated using a stratified random sampling procedure (Thompson 2012) with the main ecological regions in Wallonia as environmental strata (Jacob et al. 2010, Fig. 1b). We iterated this stratified random sampling with ten bootstrap replicates for each sample size.

We used 23 environmental variables (Table 1) that characterize the most important habitat conditions for birds (e.g. elevation, climate, land cover and soil type) in southern Belgium (Jacob *et al.* 2010) as predictors in the models. These variables were sourced from available GIS data layers and sampled in the 1-km^2 squares that are completely within the boundaries of Wallonia (n=16 600). We considered a secondary sampling unit as occupied by the species when at least one individual was recorded with breeding evidence during the transect-based procedure. The species that were included in the modelling exercise (hereafter 'focal species', Table 2) fulfilled four criteria: 1) they were recorded as present in all randomly generated subsets of secondary sampling units for model training, 2) territory-mapping data for model evaluation were available, 3) they are diurnal songbird species, and 4) their territory size or home range is on average lower than or close to the spatial resolution of the secondary sampling units.

Reliable absence data were unavailable and this issue may produce inaccurate presence-absence models (Brotons *et al.* 2004; Lobo, Jiménez-Valverde, & Hortal 2010). Hence, we applied the presence-only maximum entropy framework Maxent 3.3.1 (Phillips, Anderson, & Schapire 2006). Maxent is only moderately sensitive to sample size and outperforms other methods when sample size is small (Hernandez *et al.* 2006; Wisz *et al.* 2008; Bean, Stafford, & Brashares 2012).

For each focal species and sample size, model training was performed with the ten randomly generated subsets of secondary sampling units. The quadratic terms of the continuous environmental variables were included in addition to the linear functions. The continuous modelling outputs were converted into binary predictions by setting a threshold probability value above which the species was predicted as present. To set this value, we assumed that some presence records were located in unsuitable areas (Hirzel & Le Lay 2008) and we defined a threshold such that an omission rate of 10% was specified in the subsets of secondary sampling units used for model training (Martin *et al.* 2013). This method allows fixing a threshold that is independent of the false positive fraction, which is suitable in the case of presence-only data (Pearson *et al.* 2007).

Territory-based model training: reference distribution maps

Using the same 1-km^2 squares as for the environmental variables, we considered a square as occupied by a species when it enclosed the centre of at least one territory of the species recorded during the territory-mapping procedure. In order to avoid redundancy between the data used for model training and model evaluation (see below), we removed from the territory-mapping data the 1-km^2 squares that coincide with the set of secondary sampling units. The remaining territory-mapping data were used to build reference territory-based distribution models with the same environmental variables as for the transect-based models. Using a bootstrap approach, we fitted and averaged ten models for each focal species based on random selections of 70% of the territory-mapping data for model training. In order to create a reference distribution map for each focal species, the modelling outputs were converted into presence-absence predictions with the same threshold decision rule as for the transect-based models (10% of omission rate in the training data).

Table 1. Environmental variables calculated in the secondary sampling units and used in the modelling procedures

Variable	Source	Year	Units
Average altitude	DEM-SPW		m
Average orientation	DEM-SPW		°
Average slope	DEM-SPW		%
Topographical moisture index †	DEM-SPW		
Average spring temperature ††	Worldclim	2009	°C
Average spring precipitation rate ††	Worldclim	2009	mm
Surface of urban area	COSW	2008	ha
Surface of spring cereal and corn cultures	SIGEC	2006	ha
Surface of winter cereal cultures	SIGEC	2006	ha
Surface of forage cultures	SIGEC	2006	ha
Surface of spring-summer hoed cultures	SIGEC	2006	ha
Surface of permanent grasslands	COSW	2008	ha
Surface of temporary grasslands	COSW	2008	ha
Surface of natural (dry) grasslands	COSW	2008	ha
Surface of broadleaved forest	MRW	1993	ha
Surface of coniferous forest	MRW	1993	ha
Surface of mixed forest	MRW	1993	ha
Surface of standing water	COSW	2008	ha
Surface of orchards	COSW	2008	ha
Surface of wetlands	COSW	2008	ha
Number of isolated trees and bushes	IGN	2009	
Total length of tree lines and hedgerows	IGN	2009	m
Dominant soil type †††	CNSW	2007	

DEM-SPW: Digital Elevation Model of the '*Service Public de Wallonie*'.
CNSW: '*Cartographie Numérique des Sols de Wallonie*' – Soil map of Wallonia (scale: 1:10000).
COSW: '*Cartographie de l'Occupation du Sol en Wallonie*' – Land use map of Wallonia (scale: 1:10000).
IGN: '*Institut Géographique National*' – Land use map of Belgium (scale: 1:10000).
MRW: 'Ministère de la Région Wallonne' – Land cover map of Wallonia (pixel resolution: 20 meters).
SIGEC: '*Système Intégré de Gestion et de Contrôle*' – Agricultural land management map of Wallonia (scale: 1:10000).
Worldclim: Global climate data (http://www.worldclim.org/).
[†] Topographical moisture index was calculated following Beven and Kirkby (1979).

†† Weather conditions in spring were calculated as three-month averages fitted to the breeding period of the species.
††† Categorical variable

Transect-based model evaluation

To evaluate the performance of the transect-based models, we first calculated an omission rate to measure the percentage of presence records in the evaluation territory-mapping data that were mistakenly classified as absences. Second, the area under the curve (AUC) of a receiver operating characteristic (ROC) plot was used as a threshold-independent measure of modelling performance (Fielding & Bell 1997). ROC plots were computed using presence and background data in the evaluation dataset. AUC values reflected the ability of the transect-based models to discriminate between presence data and a randomly selected secondary sampling unit (see details in Phillips et al. 2006, Jiménez-Valverde 2012). Third, we computed misclassification matrices to calculate the agreement between the binary predictions of the transect-based and the territory-based models based on the Cohen’s kappa (Fielding & Bell 1997). The kappa value documented the extent to which the output of the transect-based models converged on those of the territory-based reference models (Hernandez *et al.* 2006).

Statistical analysis

Before analysing the modelling performance, we evaluated the extent to which the different subsets of secondary sampling units captured the range of environmental conditions used by the species. To do this, the range of all continuous variables was first normalized between 0 and 1 using a linear scaling transformation. Second, we calculated for each focal species and in each random subset of secondary sampling units, the difference between the maximum and the minimum values of the environmental variables associated with a presence record. Third, we calculated the arithmetic mean of these differences among all environmental variables to represent the width of the environmental range covered by the species in the random subsets of secondary sampling units. Fourth, we applied the same procedure for each focal species to the full set of evaluation territory-mapping data. Fifth, we computed the environmental range overlap for each focal species as the ratio between the environmental range covered by the species in the random subsets of secondary sampling units and in the territory-mapping data (Wisz *et al.* 2008; Feeley & Silman 2011).

Modelling performance was expected to increase with sample size and to level off beyond a sufficient number of secondary sampling units (Hernandez *et al.* 2006; Wisz *et al.* 2008). We plotted modelling performance measures against sample size and we fitted exponential functions to the data.

An exponential rise to maximum function was used for the AUC and kappa values:

$$y = a * (1 - e^{-bx}) \qquad (1)$$

Where y is the modelling performance measure, x is the sample size, a is the maximum asymptote y value, and b is the rise constant.

An exponential decay function was used for the omission rate:

$$y = y_0 + a * e^{-bx} \qquad (2)$$

Where y is the modelling performance measure, x is the sample size, y_0 is the minimum asymptote y value, $y_0 + a$ is the initial modelling performance measure when the sample size is equal to zero (forced to 1 in our case), and b is the decay constant.

For each modelling performance measure and each focal species separately, we calculated the minimum sample size (MSS, number of secondary sampling units) and coverage (MSC, percentage of the study area) required to achieve an acceptable level of modelling performance, defined as the lowest *x* value for which the mean predicted *y* value was within the 95% confidence limits around the asymptote value (Fig. 3).

We calculated prevalence, range size and degree of habitat specialization for each focal species (Table 2) to evaluate how these features influence the MSS. The species prevalence was calculated from the whole set of secondary sampling units as the proportion of units in which the species was present. Species range size was calculated as the number of primary sampling units in which the species was recorded with probable or confirmed breeding evidence (McPherson, Jetz, & Rogers 2004). We used a *k*-means clustering analysis (Legendre & Legendre 2012) based on the continuous environmental variables (Table 1) to allocate the primary sampling units to different habitat classes (n=10 based on an analysis of the decrease in the total error sum of squares with increasing number of classes) and we used the species abundance data in the primary sampling units to calculate the degree of habitat specialization for each focal species as the coefficient of variation (= standard deviation / average) of the average species densities among the habitat classes (see details in Julliard et al. 2006). We used a redundancy analysis (RDA) to examine how much of the among-species variation in the MSS was explained by variation in prevalence, range size and habitat specialization (Legendre & Legendre 2012). In order to present the results in a simplified manner, the set of focal species was divided in equal-size categories according to prevalence (high- and low-prevalence species), range size (wide- and restricted-range species) and degree of habitat specialization (high- and low-specialization species).

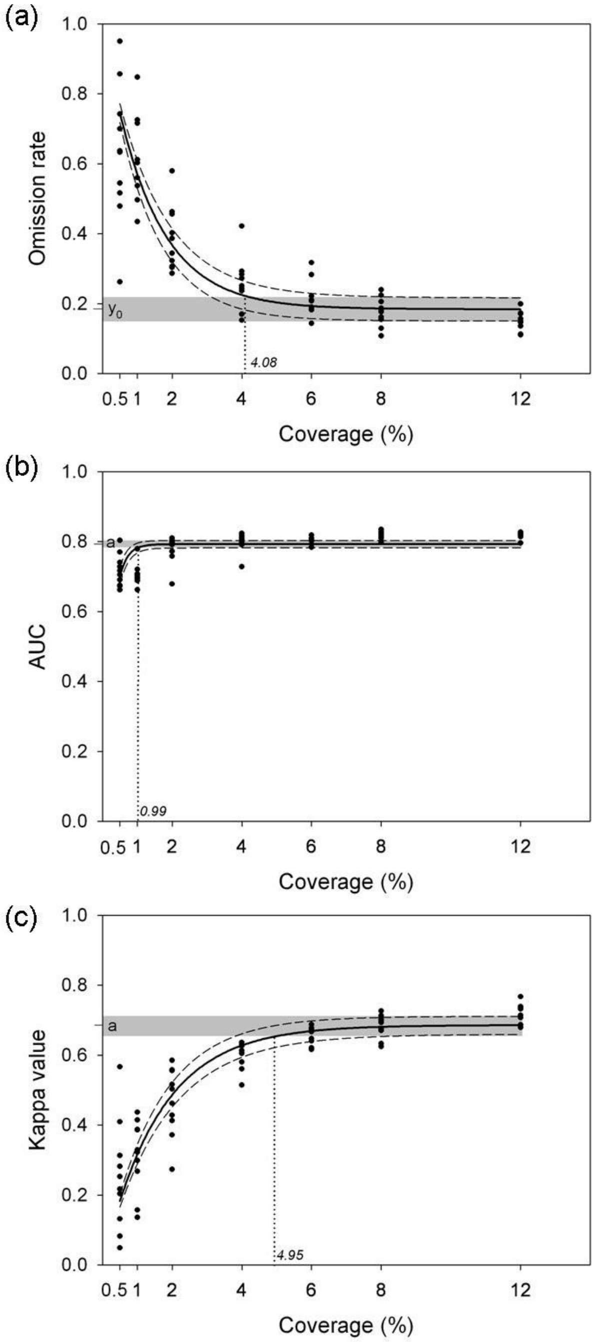


Figure 3. Identification of the minimum sampling coverage required to achieve an acceptable level of modelling performance according to (a) omission rate, (b) AUC and (c) kappa value for *Lanius collurio*. Black dots represent the modelling performance measures for the transect-based models fitted with the different subsets of secondary sampling units. Continuous and dashed black lines are the predicted average ± 95% confidence intervals after minimum square fit to the exponential function (Eq. 1 and 2). Grey areas represent the 95% confidence interval around the estimated (a) minimum or (b, c) maximum asymptote value. The dotted vertical lines indicate the minimum sampling coverage above which the modelling performance is considered to become stable.

Table 2. Minimum sampling coverage (MSC: percentage of the study area) and sample size (MSS: number of secondary sampling units) needed to achieve an acceptable level of modelling performance according to omission rate, area under the curve of a ROC plot (AUC) and kappa value for each focal species (n=20) used in this study. The species are listed by decreasing order of prevalence in secondary sampling units.

Species	Code	Prevalence	Range size
Picus viridis	PICVIR	0.37 (high)	0.88 (wide)
Anthus trivialis	ANTTRI	0.30 (high)	0.71 (wide)
Pyrrhula pyrrhula	PYRPYR	0.26 (high)	0.78 (wide)
Anthus pratensis	ANTPRA	0.23 (high)	0.73 (wide)
Sylvia curruca	SYLCUR	0.22 (high)	0.85 (wide)
Cuculus canorus	CUCCAN	0.22 (high)	0.84 (wide)
Motacilla flava	MOTFLA	0.21 (high)	0.50 (restricted)
Carduelis carduelis	CARCAR	0.20 (high)	0.83 (wide)
Turdus pilaris	TURPIL	0.19 (high)	0.56 (restricted)
Streptopelia turtur	STRTUR	0.18 (high)	0.84 (wide)
Acrocephalus palustris	ACRPAL	0.17 (low)	0.85 (wide)
Dryocopus martius	DRYMAR	0.10 (low)	0.67 (restricted)
Muscicapa striata	MUSSTR	0.09 (low)	0.75 (wide)
Saxicola torquatus	SAXTOR	0.08 (low)	0.60 (restricted)
Dendrocopos medius	DENMED	0.08 (low)	0.62 (restricted)
Lanius collurio	LANCOL	0.07 (low)	0.52 (restricted)
Hippolais polyglotta	HIPPOL	0.06 (low)	0.53 (restricted)
Miliaria calandra	MILCAL	0.05 (low)	0.27 (restricted)
Emberiza schoeniclus	EMBSCH	0.04 (low)	0.47 (restricted)

RESULTS

Range size was positively correlated with prevalence (r=0.70, p=0.0006) and negatively with habitat specialization (r=-0.69, p=0.0007), but prevalence was not related to habitat specialization (r=-0.19, p=0.4131). The training sample prevalence in the random subsets of secondary sampling units was independent of sample size (Fig. 4) and reflected the prevalence of the focal species in the whole set of sampling units (Table 2). In contrast, the proportion of the species environmental range represented in the subsets of secondary sampling units increased with sample size according to an exponential rise to maximum function (Fig. 5). This indicates that, even with the implementation of a stratified random sampling procedure, the complete range of conditions used by the species is only partly captured with very small sample sizes.

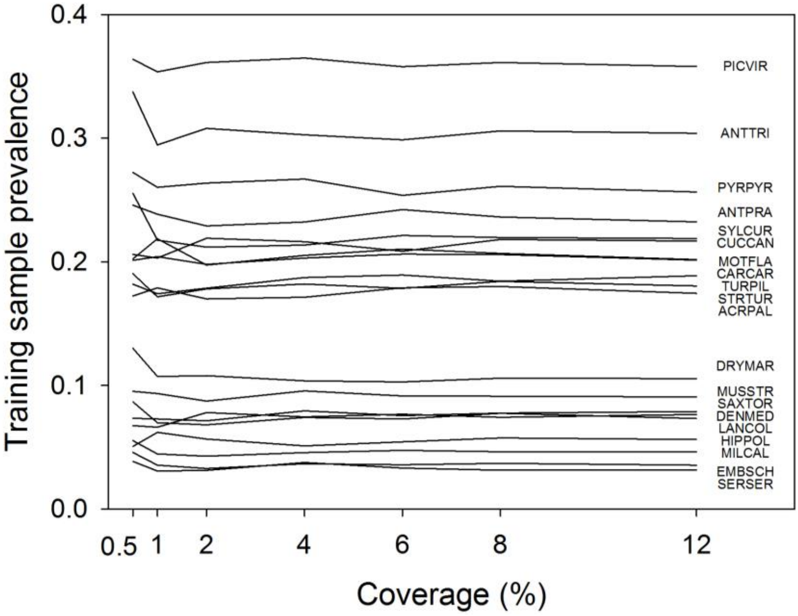


Figure 4. Average training sample prevalence of the different species along the gradient of sampling coverage. Sample prevalence was calculated as the proportion of secondary sampling units with species presence records in each individual subset of the units used to fit the models.

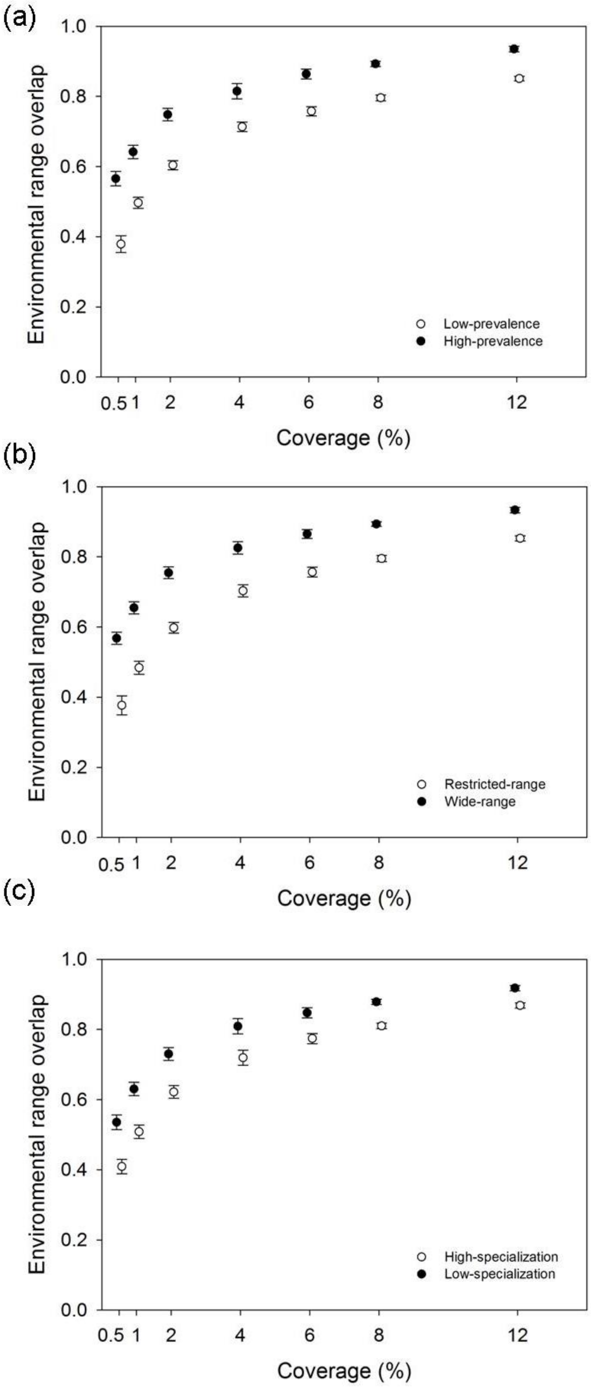


Figure 5. Average (± 95% confidence intervals) proportion of the species environmental range covered by the subsets of secondary sampling units along the gradient of sampling coverage for (a) high- and low-prevalence species, (b) wide- and restricted-range species, and (c) low- and high-specialization species.

Figures 3 and 6 show the modelling performance (omission rate, AUC and kappa value) obtained with a number of sampling units covering 0.5% to 12% of the study area. Table 2 summarizes the minimum sample size (MSS) and coverage (MSC) calculated for each focal species according to the different modelling performance measures.

The constrained axes of the redundancy analysis (RDA, Fig. 7) explained together 57% of the total variance in the data ($R^2_{adjusted}$=0.47). Only the first RDA axis was found to be statistically significant (permutation tests, $p<0.001$), accounting for more than 97% of the explained variance. The first RDA axis was for the most part related to the prevalence of the species (canonical coefficient 2.39) and, to a much lower extent, to habitat specialization (0.12) and range size (0.04). Hence, the RDA results indicated that the prevalence of the focal species in the whole set of secondary sampling units had the most prominent influence on the MSS and that the effect of range size and habitat specialization can be considered negligible.

On average, the omission rate was lower for high-prevalence species than for low-prevalence species over the entire gradient of sample size, but the difference was decreasingly pronounced with an increasing sample size (Fig. 6). The exponential functions were estimated to reach their minimal value with a smaller sample size (MSS=320 ± 29 S.E.) or sampling coverage (MSC=1.93% ± 0.18%) in high-prevalence species than in low-prevalence species (MSS =809 ± 106, MSC=4.87% ± 0.64%). The AUC was only weakly sensitive to sample size and levelled off at smaller sample size in high-prevalence species (MSS=115 ± 14, MSC=0.69% ± 0.09%) than in low-prevalence species (MSS=183 ± 15, MSC=1.10% ± 0.09%). The kappa value increased consistently with sample size, thereby indicating that the predictions of the transect-based models gradually converged on those of the reference territory-based models. Kappa values were particularly affected by small sample size in low-prevalence species: they levelled off at smaller sample size in high-prevalence species (MSS=369 ± 57, MSC=2.22% ± 0.35%) than in low-prevalence species (MSS=991 ± 111, MSC=5.97% ± 0.67%).

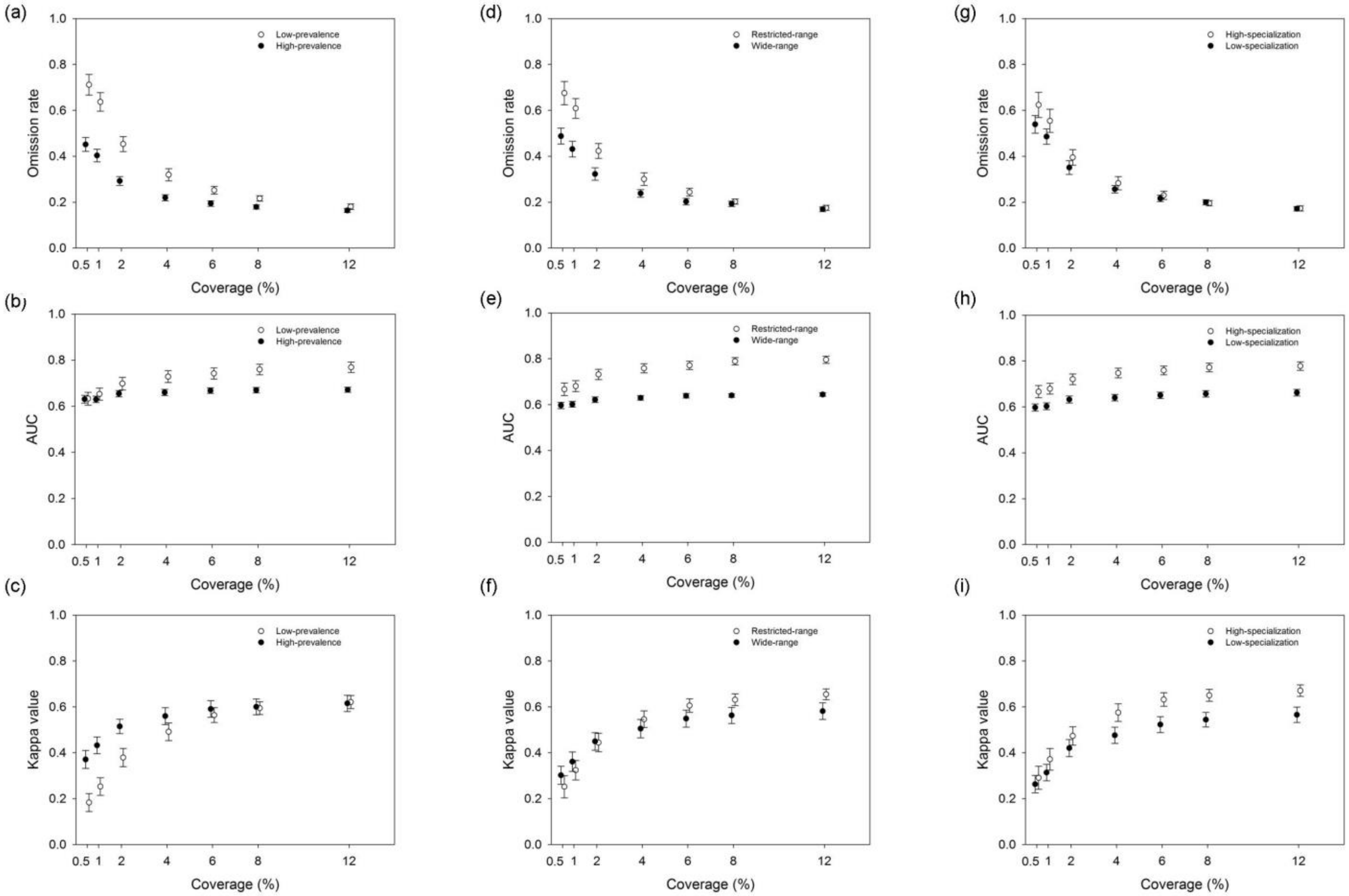


Figure 6. Average (± 95% confidence intervals) (a,d,g) omission rate, (b,e,h) AUC and (c,f,i) kappa values along the gradient of sampling coverage for high- and low-prevalence species, wide- and restricted-range species, and low- and high-specialization species.

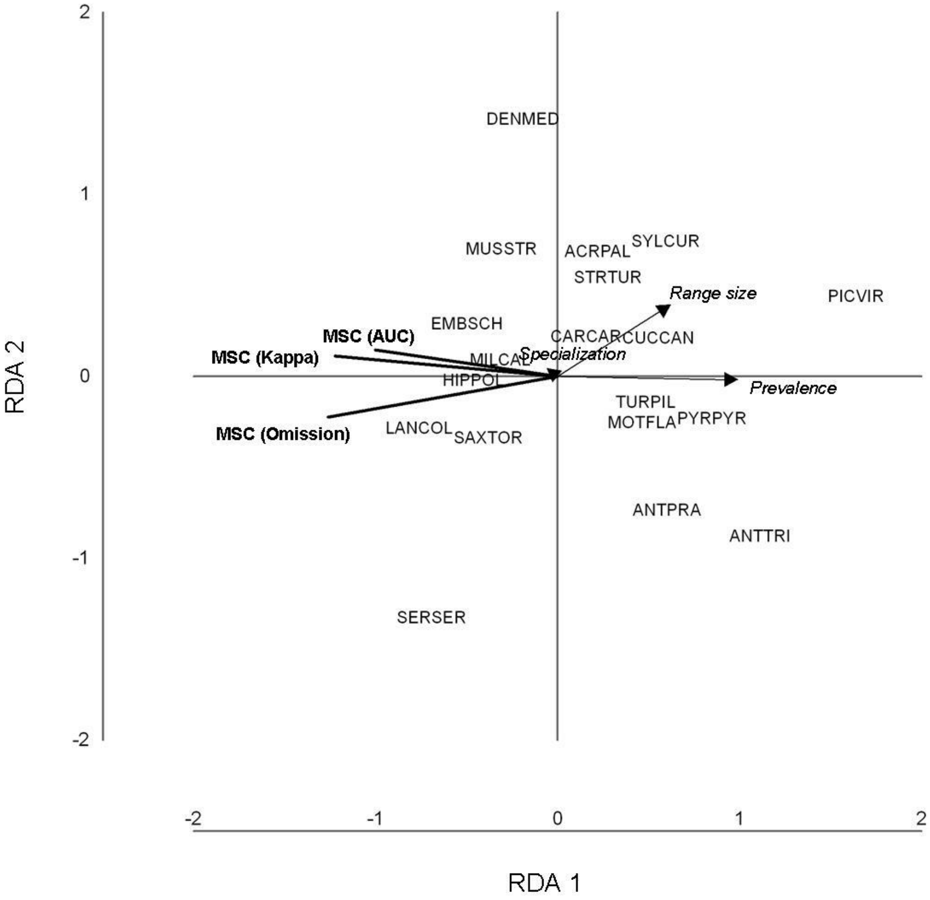


Figure 7. First two dimensions (RDA1 and RDA2) of the ordination space from the redundancy analysis (RDA, type-2 scaling). The explanatory variables (prevalence, range size and habitat specialization) are represented with arrows and the response variables (minimum sampling coverage according to omission rate, AUC and Kappa) are represented with bold black lines. Species are plotted using their code names (Table 2).

DISCUSSION

When designing a monitoring project to estimate biodiversity dynamics, a trade-off is typically made between spatial and temporal replication in the data collection strategy to minimize uncertainties associated with the estimation of changes in state variables over time (Rhodes & Jonzén 2011; Guillera-Arroita & Lahoz-Monfort 2012). In line with previous studies(De Cáceres & Brotons 2012; Rodhouse *et al.* 2012; Kéry, Guillera-Arroita, & Lahoz-Monfort 2013), we argue that monitoring data may also be cost-effectively collected and used in species distribution models to document the spatial distribution of the species.

Although the influence of sample size on the performance of species distribution models is reported in many studies (Wisz *et al.* 2008), only few have addressed this issue when models are built with data from monitoring projects (Brotons, Herrando, & Pla 2007). This is mostly due to the fact that dynamic distribution mapping is seldom explicitly addressed when setting the objectives of a project. If such an objective is integrated after the start of the project, the available data have been typically collected in a limited number of sampling units. This sampling design prevents from evaluating modelling performance over a broad gradient of sample sizes and from identifying how large the sample size should be to obtain an acceptable performance. On the other hand, monitoring data are unavailable when species distribution mapping is considered as an objective before the start of data collection. Other sources of information are therefore needed to help optimising the initial sampling design.

Here, we provide an analytical framework that makes use of data from large-scale last-generation atlases with two-stage sampling design to examine the influence of sample size on modelling performance and to identify how large the number of sampling units should be in a monitoring project to derive accurate species distribution maps. The method does not rely on existing data from already running monitoring projects and, hence, it may be applied before the start of field data collection when decisions about sampling design are made. The innovative idea was to consider part of the data collected during last-generation atlas projects as analogous to those derived from long-term monitoring projects (Vorísek *et al.* 2008; Van Swaay *et al.* 2008). In contrast with previous studies focusing on the link between monitoring projects and distribution modelling approaches (Brotons, Herrando, & Pla 2007), the manipulation of atlas data allowed us to simulate a broad gradient of sample size in order to identify an optimal number of sampling units to achieve an acceptable modelling performance. The analytical framework may be easily implemented wherever such atlas data are available and where the sampling strategies of monitoring projects need to be optimised to map species distributions. Although we used bird data to illustrate our method, it is important to note that it may also be applied to other species groups for which atlas data are collected, at least partly, in the same way as in a monitoring project, such as in butterflies (Maes, Vanreusel, & Van Dyck 2013) or bats (Carden *et al.* 2010).

We showed that modelling performance was sensitive to particularly small sample sizes and reached an asymptote level beyond a sufficiently large number of sampling units. This result is especially interesting because it is generally assumed or reported that modelling performance increases with sample size (McPherson, Jetz, & Rogers 2004; Feeley & Silman 2011), without examining how large sample size should be to obtain sufficiently well-performing models. Wintle and Bardos (2006) and Jiménez-Valverde et al. (2009) have previously studied the influence of sample size on modelling performance and also showed that the effect of sample size becomes apparent only below a certain threshold, but their studies were conducted with virtual species and may only partly reflect monitoring data.

The prevalence of the species in the random subsets of secondary sampling units remained stable along the gradient of sample sizes and reflected the prevalence of the species in the whole set of sampling units. So, the link between modelling performance and sample size was independent of the proportion of sampling units with species presence records. Hence, we avoided the confusion between the effects of sample size and training sample prevalence (McPherson, Jetz, & Rogers 2004). In contrast, the extent to which the subsets of sampling units covered the range of environmental conditions used by the species was found to decrease with sample size and this contributes to explaining why the ability of the models to capture the environmental response of the species decreased markedly below a certain sample size. This issue underlines the importance of using a well-designed sampling procedure: the stratified random sampling that we implemented (see also Jiménez-Valverde et al. 2009) maximizes the chances to sample species distribution along the whole environmental gradient of the study area even when sample size decreases (Hortal & Lobo 2005; Thompson 2012). Below a certain sample size, the number of species presence records is, however, insufficient to cover the full range of environmental conditions used by the species and the modelling performance becomes less stable and much lower (see also Wintle and Bardos 2006, Wisz et al. 2008).

The minimum sample size required to ensure an acceptable level of modelling performance was strongly related to the prevalence of the species in the sampling units. On average, the minimum sample size was larger in low-prevalence species than in high-prevalence species. In contrast, the decrease in modelling performance with increasingly smaller sample size was found to be comparable in restricted- and wide-range species and in low- and high-specialization species. A large part of the among-species variance in the minimum sample size remained, however, unexplained and may be related to additional methodological issues or ecological processes. As imperfect detection of the species may confound the link between species distribution and environmental conditions, it is for instance warranted to analyse how detection rates may influence modelling performance. Rota et al. (2011) showed that using occupancy models to account for imperfect detection may contribute to improving modelling performance and relevance, especially in situations where detection probability varies along with environmental conditions (see also Kéry et al. 2010). Such approaches are based on observation data collected during repeated surveys in the sampling units and are, therefore, only poorly suited to the context of the present study, as replicated observations are generally unavailable when setting the objectives of a monitoring project (but see Van Strien et al. 2013). Other ecological processes may have a direct or indirect influence on the minimum sample size. For instance, biotic interactions such as competition (with conspecific individuals or with other species) and predation may alter the location of the individuals in the landscape and shape the realized distribution of the species (Cadena, Loiselle, & Daniel Cadena 2007; Lima 2009). Although modelling tools become increasingly available to deal with this issue (Boulangeat, Gravel, & Thuiller 2012; Wisz *et al.* 2013), such factors are probably beyond the scope of the analyses that could be done with the available atlas data.

We also have to stress the point that further work should include additional species because the set of species used in this study had to satisfy a number of criteria for the modelling exercise, which resulted in the use of a limited number of species that may only partly reflect the entire bird species assemblage. One of the most restrictive criteria was the availability of a sufficient amount of territory-mapping data to evaluate modelling performance. Such information was collected only for low- to moderate-abundance species in the atlas project. In order to increase the number of species in the analysis, a promising approach would be to use the increasingly available information on species distribution derived from web-based encoding systems for casual observation data (Sullivan *et al.* 2009).

When applying our innovative approach to the low- to moderate-abundance bird species in southern Belgium, a minimum sampling coverage of 4-5% (n=664-830) was found to be needed in order to achieve an acceptable level of modelling performance for the majority of the studied species. Interestingly, Hoeting et al. (2000) and Wintle and Bardos (2006) obtained similar results with their simulated data reflecting plant and mammal distributions. However, the estimated minimum sampling coverage should probably not be considered as a rule of thumb. First, our results revealed considerable among-species variation in this minimum sample size. Second, the heterogeneity of the study area and the variables that are used to quantify the environmental conditions undoubtedly influence the number of sampling units needed to capture the link between species distribution and environmental conditions.

This application in southern Belgium illustrates that a substantial sampling coverage may be needed to derive accurate species distribution models from long-term monitoring data. A sampling coverage of 4-5% of the study area is actually much higher than the coverage implemented in most of existing monitoring programmes worldwide. It may then become logistically difficult to find a trade-off between the number of sampling units and the number of repeated surveys in order for the same monitoring project to integrate in its objectives both the estimation of changes in occupancy or abundance and the mapping of species distribution. Interestingly, Hooten et al. (2009, 2012) used an optimal hybrid sampling design to combine different objectives in a single long-term monitoring project. In line with such an approach, a fixed subset of sampling units may be repeatedly surveyed within and between seasons (static design) to estimate species occupancy and detection probability, while a roving subset of sampling units may be surveyed less frequently over time (dynamic design) to increase spatial knowledge for distribution mapping. Both static and dynamic designs have advantages and disadvantages (MacKenzie & Royle 2005; Wikle & Royle 2005), but an appropriate allocation of sampling effort between fixed and roving units may contribute to combining several monitoring and mapping objectives (Hooten *et al.* 2009). In this context, our methodological approach constitutes a pilot analysis able to provide an initial estimate of the total number of sampling units needed when monitoring data are not yet available and to help putting the monitoring effort into place in order to reach one of the objectives. It is now important to consider additional optimisation criteria and to further integrate such approaches into a more general analytical framework to evaluate whether this initial sampling design will be suited to document species distribution dynamics and to estimate changes in the selected state variables or, alternatively, how the design has to be modified to better reach the multiple (and sometimes conflicting) objectives. In this respect, adaptive sampling design (Hooten, Ross, & Wikle 2012) may prove a useful approach as it focuses on adjusting the sampling strategy on a regular basis as new information is gained in order to improve the cost-efficiency of the monitoring project.

ACKNOWLEDGEMENTS

We thank participants in the Breeding Bird Atlas of Wallonia project for fieldwork and Christophe Dehem, Marc De Sloover, Marc Fasol, Jean-Paul Jacob and Thierry Kinet for data or project management. Land use (COSW and IGN), land management (SIGEC) and soil (CNSW) maps were provided by the Direction Générale de l'Agriculture, des Ressources Naturelles et de l'Environnement (DGARNE) of the Service Public de Wallonie (SPW). The BBAW project was funded by the Service Public de Wallonie (SPW-DGO3). O.A. was funded by the National Research Fund, Luxembourg (FNR-AFR-PHD-08-63). N.T. and L.B. were funded through the EU BON project (contract no. 308454; FP7-ENV-2012, European Commission). This work has also been supported by the European infrastructure for biodiversity and ecosystem research (LifeWatch).

REFERENCES

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

Bean, W.T., Stafford, R. & Brashares, J.S. (2012) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, **35**, 250–258.

Boulangeat, I., Gravel, D. & Thuiller, W. (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, **15**, 584–593.

Braunisch, V. & Suchant, R. (2010) Predicting species distributions based on incomplete survey data: the trade-off between precision and scale. *Ecography*, **33**, 826–840.

Brotons, L., Herrando, S. & Pla, M. (2007) Updating bird species distribution at large spatial scales: Applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions*, **13**, 276–288.

Brotons, L., Sierdsema, H., Newson, S., Jiguet, F. & Gregory, R.D.N. (2006) Spatial modeling of large-scale bird monitoring data: Towards pan-European quantitative distribution maps. *Journal of Ornithology*, **147**, 29.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.

De Cáceres, M. & Brotons, L. (2012) Calibration of hybrid species distribution models: the value of general-purpose vs. targeted monitoring data. *Diversity and Distributions*, **18**, 977–989.

Cadena, C.D., Loiselle, B. a. & Daniel Cadena, C. (2007) Limits to elevational distributions in two species of emberizine finches: Disentangling the role of interspecific competition, autoecology, and geographic variation in the environment. *Ecography*, **30**, 491–504.

Carden, R., Aughney, T., Kelleher, C. & Roche, N. (2010) Irish Bat Monitoring Schmes: ATLAS Republic of Ireland. Report for 2008-2009. *Bat Conservation Ireland*.

Dunn, A.M. & Weston, M. A. (2008) A review of terrestrial bird atlases of the world and their application. *Emu*, **108**, 42–67.

Elith, J., Kearney, M.R. & Phillips, S.J.P. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Estrada, J., Pedrocchi, V., Brotons, L. & Herrando, S. (2004) Atles Dels Ocells Nidificants de Catalunya (1999-2002). Lynx Editions, Barcelona.

Feeley, K.J. & Silman, M.R. (2011) Keep collecting: accurate species distribution modelling requires more collections than previously thought. *Diversity and Distributions*, **17**, 1132–1140.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Gitzen, R.A., Millspaugh, J.J., Cooper, A.B. & Licht, D.S. (2012) Design and Analysis of Long-Term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.

Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution*, **3**, 860–869.

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.

Hirzel, A.H. & Le Lay, G. (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, **45**, 1372–1381.

Hoeting, J.A., Leecaster, M. & Bowden, D. (2000) An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological and Environmental Statistics*, **5**, 102–114.

Hooten, M.B., Ross, B.E. & Wikle, C.K. (2012) Optimal spatio-temporal monitoring designs for characterizing population trends. Design and analysis of long-term ecological monitoring studies (eds R.A. Gitzen, J.J. Millspaugh, A.B. Cooper & D.S. Licht), pp. 443–459. Cambridge University Press.

Hooten, M.B., Wikle, C.K., Sheriff, S.L. & Rushin, J.W. (2009) Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, **20**, 639–649.

Hortal, J. & Lobo, J.M. (2005) An ED-based Protocol for Optimal Sampling of Biodiversity. *Biodiversity and Conservation*, **14**, 2913–2947.

Jacob, J.-P., Dehem, C., Burnel, A., Dambiermort, J.L., Fasol, M., Kinet, T., van der Elst, D. & Paquet, J.-Y. (2010) Atlas Des Oiseaux Nicheurs de Wallonie. Publication d’Aves et du Département de l'Etude du Milieu Naturel et Agricole.

Jiménez-Valverde, A. (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, **21**, 498–507.

Jiménez-Valverde, A., Lobo, J.M. & Hortal, J. (2009) The effect of prevalence and its interaction with sample size on the reliability of species distribution models. *Community Ecology*, **10**, 196–205.

Julliard, R., Clavel, J., Devictor, V., Jiguet, F. & Couvet, D. (2006) Spatial segregation of specialists and generalists in bird communities. *Ecology Letters*, **9**, 1237–1244.

Kéry, M., Dorazio, R.M., Soldaat, L., van Strien, A.J., Zuiderwijk, A. & Royle, J.A.A. (2009) Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, **4**, 1163–1172.

Kéry, M., Gardner, B. & Monnerat, C. (2010) Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37**, 1851–1862.

Kéry, M., Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2013) Analysing and mapping species range dynamics using occupancy models (ed M Patten). *Journal of Biogeography*, **40**, 1463–1474.

Legendre, P. & Legendre, L. (2012) Numerical Ecology. Elsevier, Amsterdam.

Lima, S.L. (2009) Predators and the breeding bird: behavioral and reproductive flexibility under the risk of predation. *Biological Reviews*, **84**, 485–513.

Lindenmayer, D.B., Gibbons, P., Bourke, M., Burgman, M.A., Dickman, C.R., Ferrier, S., Fitzsimons, J., Freudenberger, D., Garnett, S.T., Groves, C., Hobbs, R.J., Kingsford, R.T., Krebs, C., Legge, S., Lowe, A.J., Mclean, R., Montambault, J., Possingham, H.P., Radford, J., Robinson, D., Smallbone, L., Thomas, D., Varcoe, T., Vardon, M., Wardle, G., Woinarski, J. & Zerger, A. (2012) Improving biodiversity monitoring. *Austral Ecology*, **37**, 285–294.

Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.

MacKenzie, D.I. (2012) Study design and analysis options for demographic and species occurrence dynamics. Design and analysis of long-term ecological monitoring studies (eds R.A. Gitzen, J.J. Millspaugh, A.B. Cooper & D.S. Licht), pp. 397–425. Cambridge University Press.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Hines, J.E. & Bailey, L.L. (2005) Occupancy Estimation and Modeling. Inferring Patterns and Dynamics of Species Occurrence. Elsevier, San Diego, USA.

MacKenzie, D.I. & Royle, J.A.A. (2005) Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

Maes, D., Vanreusel, W. & Van Dyck, H. (2013) Dagvlinders in Vlaanderen: Nieuwe Kennis Voor Betere Actie. Uitgeverij Lannoo nv., Tielt.

Maes, D., Vanreusel, W., Jacobs, I., Berwaerts, K. & Van Dyck, H. (2012) Applying IUCN Red List criteria at a small regional level: A test case with butterflies in Flanders (north Belgium). *Biological Conservation*, **145**, 258–266.

Martin, Y., Van Dyck, H., Dendoncker, N. & Titeux, N. (2013) Testing instead of assuming the importance of land use change scenarios to model species distributions under climate change. *Global Ecology and Biogeography*, **22**, 1204–1216.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied ecology*, **41**, 811–823.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.

Phillips, S.J.P., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Rhodes, J.R. & Jonzén, N. (2011) Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? *Ecography*, **34**, 1040–1048.

Rodhouse, T.J., Ormsbee, P.C., Irvine, K.M., Vierling, L.A., Szewczak, J.M. & Vierling, K.T. (2012) Assessing the status and trend of bat populations across broad geographic regions with dynamic distribution models. *Ecological applications : a publication of the Ecological Society of America*, **22**, 1098–113.

Rodríguez, J.P., Brotons, L., Bustamante, J., Seoane, J. & Rodriguez, J.P. (2007) The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, **13**, 243–251.

Rota, C.T., Fletcher, R.J., Evans, J.M. & Hutto, R.L. (2011) Does accounting for imperfect detection improve species distribution models? *Ecography*, **34**, 659–670.

Royle, J.A. & Nichols, J.D. (2003) Estimating Abundance from Repeated Presence-Absence Data or Point Counts. *Ecology*, **84**, 777–790.

Van Strien, A.J., van Swaay, C. a. M.M. & Termaat, T. (2013) Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, **50**, 1450–1458.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009) eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, **142**, 2282–2292.

Van Swaay, C.A.M., Nowicki, P., Settele, J., Van Strien, A.J., Swaay, C. a. M. & Strien, A.J. (2008) Butterfly monitoring in Europe: Methods, applications and perspectives. *Biodiversity and Conservation*, **17**, 3455–3469.

Thompson, S.K. (2012) Sampling, 3rd. ed. Wiley.

Timothy, J. & Sharrock, R. (1974) Minutes of the second meeting of the European Ornithological Committee. *Acta Ornithologica*, **14**, 404–411.

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. & Defourny, P. (2007) Fitness-related parameters improve presence-only distribution modelling for conservation practice: The case of the Red-backed shrike. *Biological Conservation*, **138**, 207–223.

Vorísek, P., Klvanová, A., Wotton, S. & Gregory, R.D. (2008) A Best Practice Guide for Wild Bird Monitoring Schemes. CSO/RSPB, Trebon, Czech Republic.

Wikle, C.K. & Royle, J.A.A. (2005) Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics*, **16**, 507–522.

Wintle, B. A & Bardos, D.C. (2006) Modeling species-habitat relationships with spatially autocorrelated observation data. *Ecological Applications*, **16**, 1945–1958.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J.P., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J.M., Williams, S.E. & Zimmermann, N.E. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.-A.J.-A.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.-C.M.-C.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D. a, Aastrup, P. & Svenning, J.-C.J.-C.C. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Article 3

Evaluating the reliability of species distribution models with an indirect measure of bird reproductive performance

The content of this chapter has been published*:

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Sardá-Palomera, F., Gargallo, G., Herrando, S., Brotons, L., Titeux, N. (2017). Evaluating the reliability of species distribution models with an indirect measure of bird reproductive performance. *Journal of Avian Biology, **48**,1575-1582*

*text in the thesis has been marginally been modified compared to the published

ABSTRACT

Measures of fitness such as reproductive performance are considered reliable indicators of habitat quality for a species. Such measures are, however, only available in a restricted number of sites, which prevents them from being used to quantify habitat quality across landscapes or regions. Alternatively, species presence records can be used along with environmental variables to build models that predict the distribution of species across larger spatial extents. Model predictions are often used for management purposes as they are assumed to describe the quality of the habitats to support a species. Yet, given that species are often present both in optimal and suboptimal areas, the use of data collected during the breeding season to build these models may potentially result in misleading predictions of habitat quality for the reproduction of the species, with potentially significant conservation consequences. In this study, we analysed the relationship between fitness parameters informing on habitat quality for reproduction and predictions of species distribution models at multiple spatial scales using two independent sets of data. For 19 passerine bird species, we compared an indirect measure of reproductive performance (ratio of juveniles-to-adults) – obtained from Constant Effort Sites (CES) mist-netting data in Catalonia – with the predictions of models based on bird presence records collected during the Catalan Breeding Bird Atlas (CBBA). A positive relationship between the predictions derived from species distribution models and the reproductive performance of the species was found for half of the species at one or more spatial scales. This result suggests that species distribution models may help to predict habitat quality for some species over some extents. However, caution is needed as this is not consistent for all species at all scales. Further work based on species- and scale-specific approaches is now required to understand in which situations species distribution models provide predictions that are in line with reproductive performance.

INTRODUCTION

The greatest threat to wild bird populations is the degradation and loss of habitat (Johnson 2007), i.e., “the resources and conditions present in an area that produce occupancy—including survival and reproduction—by a given organism” (Hall, Krausman, & Morrison 1997). To assess which habitats are more or less valuable for a particular species, especially for those of conservation concern, it is necessary to use robust measurements of habitat quality, that is, of “the ability of a given habitat to provide resources and conditions appropriate for the per capita contribution to population growth” (Johnson 2007). This implies that estimates of habitat quality require measures of demography. One of the most cited definitions of habitat quality is that of Van Horne (1983), who describes habitat quality in terms of fitness, as “the product of density, mean individual survival probability, and mean expectation of future offspring”. Gathering such demographic information to describe habitat quality can be costly and time consuming, and is usually needed over long periods, thus limiting its availability to a restricted number of sampling sites; therefore, the use of proxies to reflect how habitat quality varies across landscapes and regions is often necessary (Stephens *et al.* 2015).

Species distribution models statistically link species presence records with known environmental conditions to predict the likelihood of a species to occur at any site in a given area (Araújo & Guisan 2006; Elith, Kearney, & Phillips 2010). Model predictions are often based on species presence records available from atlas projects or museum inventories, and then used for management and conservation purposes (Brotons, Mañosa, & Estrada 2004; Guisan *et al.* 2013) under the assumption that they directly inform on the quality of the habitats for the species. Although the predictions of these models correlate well in general with species abundance (Weber *et al.* 2016), not much quantitative information is available regarding the link with population growth. Yet, outcomes of these models are often used as proxies for demographic parameters to differentiate between areas with high individual fitness or positive population growth and areas that are not suitable for viable populations (Pulliam 2000; Guisan & Thuiller 2005; Franklin 2009; VanDerWal *et al.* 2009; Pellissier *et al.* 2013). One of the most important assumptions of these models is that the species is present wherever the local environmental conditions are within the species demographic niche (populations at equilibrium, Peterson et al. 2011). However, these models frequently ignore possible deviations from this equilibrium, such as source-sink dynamics (Pulliam 2000; Thuiller *et al.* 2014) or ecological traps (Robertson & Hutto 2006; Hollander *et al.* 2011), where individuals occur under certain environmental conditions beyond the boundaries of their demographic niche. Previous studies examined the link between the predictions of species distribution models and a variety of fitness parameters informing on habitat quality (Titeux *et al.* 2007; Stephens *et al.* 2015). Although they found a positive correlation between model predictions and abundance for an endangered rodent species, Bean et al. (2014) reported an absence of correlation with two fitness parameters (i.e. survival and body condition). In contrast, Brambilla and Ficetola (2012) showed that their species distribution models predicted accurately the number of fledglings in the red-backed shrike. Pellissier et al. (2013) found a significant relationship between model predictions and nest success but only for one of the three wader species they studied. The contrasting results from studies examining the link with fitness parameters call into question the use of model predictions as proxies for habitat quality. This draws attention to the need for a better understanding of the link between model predictions and habitat quality (Falcucci *et al.* 2009).

In addition, most species are influenced by ecological processes and interactions acting beyond the local scale, e.g. the habitat patch (Kareiva & Wennergren 1995; Steffan-Dewenter *et al.* 2002). Hence, fitness parameters such as measures of reproductive performance collected in a specific site may not necessarily reflect only the habitat quality of this site, but may be also influenced by the quality of the neighbouring areas (Chalfoun & Martin 2007). For instance, the proportion of juveniles captured in a site may include those juveniles produced locally and also those produced in suitable nearby areas that had dispersed into the site (Greenwood & Harvey 1982). Hence, it is necessary to understand how the relationship between the predictions of the models based on presence records and the fitness parameters documenting habitat quality varies across multiple spatial scales. It is also worth testing this scale dependency using a sample of species, because factors such as juvenile dispersal are species-specific traits.

In this study we assessed, for 19 passerine bird species in a Mediterranean region, how the predictions derived from species distribution models are related to measures of reproductive performance obtained from data collected in Constant Effort Sites (CES) mist-netting stations. We used the species presence records collected in the frame of a breeding bird atlas project and we built species distribution models using the same procedure as the one applied in this atlas. The predictions of these models have been extensively used to guide conservation and management strategies aiming to protect or preserve suitable habitats for birds (e.g. Brotons et al. 2004, Bosch et al. 2010, Herrando et al. 2010, Sardà-Palomera et al. 2012). The proportion of juveniles captured in each CES across Catalonia was used as a measure of reproductive performance of the bird species. We tested at multiple spatial scales if the reproductive performance measured in the CES matched the predictions derived from the species distribution models.

METHODS

Study area

Catalonia is a 32,114 km^2 region located in the north-east of the Iberian Peninsula. This region is mostly dominated by Mediterranean climate but has a large altitudinal range (from 0 to 3143 m.a.s.l), which creates a highly heterogeneous landscape (Fig.1). Catalonia has a long tradition in ornithology that has allowed the development of the Catalan Breeding (CBBA) and Wintering (CWBA) Bird Atlases and several long-term bird monitoring programmes (Estrada *et al.* 2004; Brotons, Herrando, & Pla 2007; Herrando *et al.* 2011).

Presence records

Bird presence records during the breeding period were sourced from the CBBA. Bird data collection is described in details in Estrada et al. (2004) and an English version of the methodology chapter can be freely downloaded from the website of the Catalan Ornithological Institute (http://www.ornitologia.org/ca/quefem/monitoratge/atles/atles_nidificants/atles_nidificants_metodologia.pdf). During 1999-2002, volunteers and professionals surveyed the 10-km resolution squares (N=385) according to a grid system covering the whole of Catalonia. They recorded the presence, the breeding evidence and estimated the abundance of each species. In addition, a sample of 5-10 1-km resolution squares was selected within each 10-km resolution square, where observers recorded the presence of each species during two 1-hour timed surveys in early and late spring. These 1-km resolution squares were spatially distributed to cover the different habitat types following a stratified random sampling procedure. In our analyses, we only used information on bird presence from the 1-km resolution squares enclosed in 10-km resolution squares where breeding evidence was reported. The number of 1-km resolution squares with bird presence used to build the species distribution models for each species is reported in Table 1.

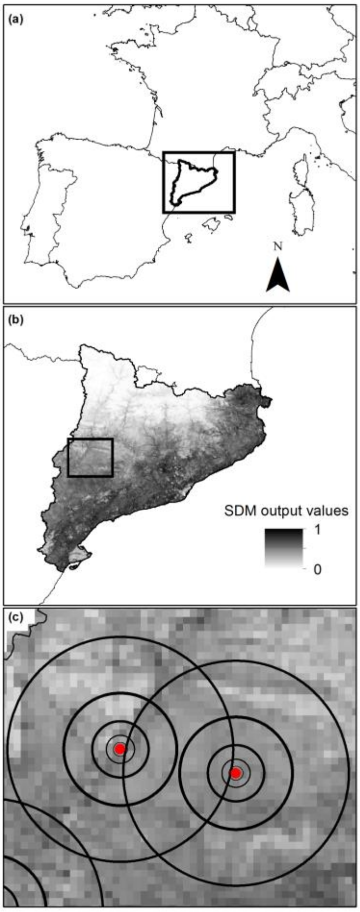


Figure 1. (a) Location of Catalonia in Southern-Europe. (b) Predictions of the species distribution model for *Sylvia melanocephala* (example species). (c) Subset of the study area with Constant Effort Sites (CES) marked with red dots and the buffers in black at 1, 2, 4, 8 and 16 km of distance around the CES.

Table 1. Detailed information for each bird species used in this study: number of presence records at 1-km resolution used in the modelling procedure, average modelling performance (AUC) ± standard deviation (SD) over the 10 evaluation datasets, number of Constant Effort Sites (CES) used in this study, total number of juveniles captured and total number of individuals captured during the study period in the CES.

Species	Acronym	Number of presence records	Model performance (AUC±SD)	Number of CES	Number of juveniles	Number of total captures
Acrocephalus arundinaceus	acraru	132	0.90±0.03	10	81	416
Acrocephalus scirpaceus	acrsci	102	0.93±0.01	24	542	2514
Aegithalos caudatus	aegcau	1274	0.73±0.01	43	594	996
Carduelis carduelis	carcar	1812	0.65±0.01	32	130	559
Certhia brachydactyla	cerbra	1451	0.69±0.01	44	245	561
Cettia cetti	cetcet	751	0.81±0.01	29	859	1490
Chloris chloris	chlchl	1637	0.67±0.01	35	190	739
Cyanistes caeruleus	cyacae	1590	0.70±0.01	43	461	957
Erithacus rubecula	erirub	1733	0.71±0.01	42	1388	2445
Hippolais polyglotta	hippol	842	0.74±0.01	44	178	1492
Luscinia megarhynchos	lusmeg	2038	0.67±0.01	49	664	2694
Parus major	parmaj	2435	0.64±0.01	50	1280	2046
Passer domesticus	pasdom	1994	0.68±0.01	36	443	2156
Serinus serinus	serser	2334	0.64±0.01	44	376	970
Sylvia atricapilla	sylatr	1701	0.68±0.01	46	1197	2720
Sylvia cantillans	sylcan	893	0.75±0.01	32	315	669
Sylvia melanocephala	sylala	1512	0.72±0.01	39	1574	2599
Troglodytes troglodytes	trotro	1565	0.70±0.01	36	240	524
Turdus merula	turmer	2613	0.61±0.01	50	1615	3721

Predictor variables

We decided to follow the same methodology to build the species distribution models as the one implemented in CBBA and CWBA, because the maps produced by these models in these atlases are broadly used in Catalonia by the administrations in charge of biodiversity conservation or land use planning. In addition, they are assumed to directly inform on the quality of the breeding or wintering habitats for the bird species and used as such in many scientific studies (Brotons, Mañosa, & Estrada 2004; Bosch *et al.* 2010; Herrando *et al.* 2010; Sardà-Palomera *et al.* 2012).

We used the same set of environmental variables as the ones that were used to build the models in the CWBA. These variables (n=55) were selected to incorporate the factors known *a priori* to determine the current distribution of birds at different spatial scales (Vaughan & Ormerod 2003). Our objective was to reproduce the atlas modelling approach and to generate predictive models that represented the distribution of the species as accurately as possible. As we did not project the outcomes of such models to other areas or time periods, we considered that potential overfitting due to the high number of variables was not a too serious issue for our study aims. Environmental variables were calculated within each 1-km resolution square to reflect:

1. Habitat and land-use: variables describing the different types of land use and land cover.

2. Climate: variables describing temperature and precipitation regimes over the course of the year.

3. Relief: variables describing altitude and slope.

4. Human influence: variables describing the potential impacts of human infrastructures (e.g. urban areas, roads).

5. Others: variables used to account for spatial patterns not directly explained by other environmental variables (e.g. mean latitudinal and longitudinal coordinates for each square, mean distance to the sea).

Detailed information about each predictor variable used in the CWBA, their sources and how they were generated can be found in Herrando et al. (2011). An English version of the methodology of the CWBA can be freely downloaded from the website of the Catalan Ornithological Institute (http://www.ornitologia.org/ca/quefem/monitoratge/atles/atles_hivern/atleshivern_metodologia.pdf).

Species distribution modelling

We applied the presence-only maximum entropy framework Maxent 3.3.1 (Phillips, Anderson, & Schapire 2006) because it is moderately sensitive to sample size and outperforms other methods when sample size is small (Hernandez *et al.* 2006; Wisz *et al.* 2008; Bean, Stafford, & Brashares 2011). Models were computed at a spatial resolution of 1 km using the Maxent default parameters, but limiting the response to environmental variables to linear and quadratic functions. For each species, the set of 1-km resolution squares with presence records was randomly split into a calibration dataset (70% of the presence records) and an evaluation dataset (remaining 30% of the records). This procedure was repeated 10 times using a subsampling approach. The predictive performance of the models was evaluated using the evaluation datasets and the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve. As bird presence records were collected during the breeding period, we used the average model outcomes across the 10 replicates to predict breeding habitat quality for each of the bird species in the 1-km resolution squares. The values of the model outcomes ranged from 0 to 1 and reflected an increasing habitat quality.

Bird mist-netting and fitness parameter

Bird mist-netting using a constant effort methodology and standard field procedures (Robinson, Julliard, & Saracco 2009) started in the early 1990s in Catalonia within the frame of the SYLVIA project and includes now over 60 Constant Effort Sites (CES) (Grup Català d'Anellament 2002). These CES are distributed all across Catalonia and are often located within protected areas, but they represent the whole diversity of habitat types in the region. SYLVIA focuses on the study of demographic parameters through captures of birds carried out from May 1st to August 6th. This allows covering the whole breeding season for all bird species across the whole climatic gradient in Catalonia, including the breeding period of late migrants (e.g. *Acrocephalus arundinaceus*, *Acrocephalus scirpaceus* or *Hippolais polyglotta*) and the second or replacement clutch of resident species (e.g. *Carduelis carduelis*, *Chloris chloris* or *Passer domesticus*). To match the time frame of the CBBA data used to build the models, we used the demographic data collected in CES active between 1996 and 2005.

To reflect breeding habitat quality for each species around the CES, we calculated for each year separately the number of juveniles over the total number of individuals captured (probability of capturing juveniles) (Peach, Buckland, & Baillie 1996), which is frequently used as a relative measure of reproductive performance in birds (DeSante *et al.* 1995; Grup Català d'Anellament 2002). This parameter was calculated for those species most commonly captured in the CES in Catalonia (N=19, Table 1) for which the amount of data was sufficiently large to allow a robust analysis, i.e. with over 400 captures during the study period. The number of CES used in subsequent analyses varied among the different species (N=10-50) because they were not all captured in each CES.

Habitat quality indices at multiple spatial scales

Based on the predictions of the species distribution models, we estimated the habitat quality for the 19 bird species at multiple scales to inform on the quality of the neighbouring areas around each CES. To do so, we drew buffers around the CES using five different radii (1, 2, 4, 8 and 16 km) and we calculated the area-weighted mean habitat quality (AWM-HQ) predicted by the models within these buffers (Fig.1). We acknowledge that there is an overlap between the buffers of neighbouring CES for the largest spatial scales and that the data are not fully independent from each other.

Statistical analyses

For each species and at each spatial scale, we used generalised linear mixed models (GLMMs) to test for a relationship between the measures of reproductive performance (i.e. the probability of capturing juveniles) and the breeding habitat quality estimated from the models, using a binomial error distribution and a logit link function model to avoid overdispersion. CES identities were included as random effects in all models to control for between-year variations of reproductive performance within the locations. As a high density of individuals may negatively affect reproductive performance through competition (Sardà-Palomera *et al.* 2011), we included the log-transformed total number of captured adults as a covariate. As some CES are close to each other in some parts of Catalonia, we performed an analysis of spline correlograms of the model residuals using the *ncf* package in R, and we did not detect any spatial autocorrelation (Bjornstad 2016).

RESULTS

Four of the studied species (*Acrocephalus scirpaceus*, *Chloris chloris*, *Sylvia melanocephala* and *Turdus merula*) showed a significant positive relationship between the probability of capturing a juvenile and the habitat quality derived from the species distribution models across all spatial scales (Fig.2 and Supplementary material Appendix 1, Fig.A1). *Carduelis carduelis* was the single species with a significant positive relationship between the probability of capturing a juvenile and the habitat quality only at the smallest spatial scales. In contrast, for *Erithacus rubecula*, *Parus major* and *Sylvia atricapilla*, this relationship was significantly positive only at the larger spatial scales. Nine of the studied bird species (*Aegithalos caudatus, Certhia brachydactyla*, *Cettia cetti*, *Cyanistes caeruleus*, *Hippolais polyglotta*, *Luscinia megarhynchos*, *Passer domesticus*, *Serinus serinus* and *Sylvia cantillans*) showed no significant relationship between the probability of capturing a juvenile and the predictions of the models at any of the spatial scales. For only two species (*Acrocephalus arundinaceus* and *Troglodytes troglodytes*) we found a significantly negative relationship between the probability of capturing a juvenile and the habitat quality estimated from the models, but only at the largest spatial scale for the former species and at the smallest spatial scales for the latter.

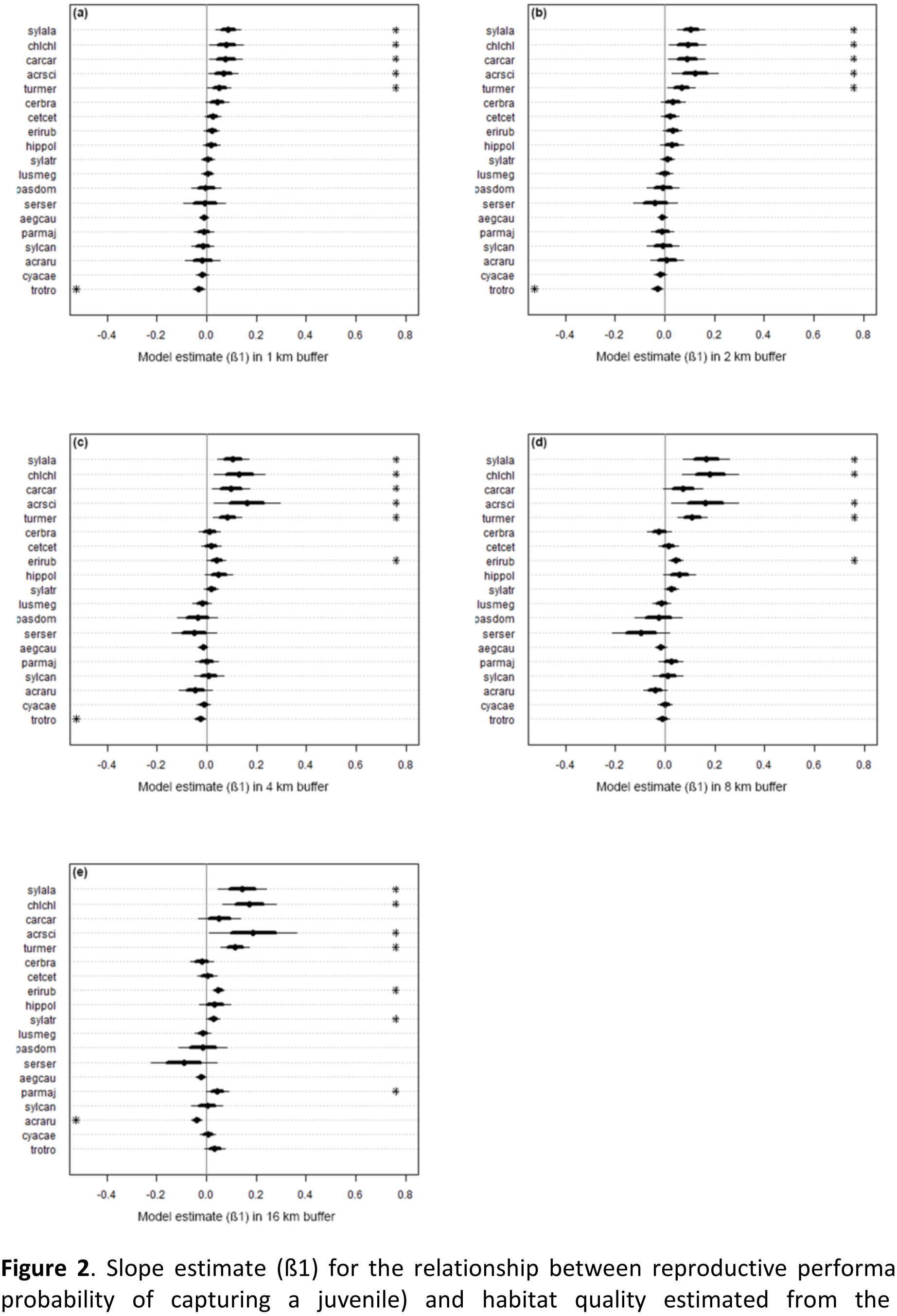


Figure 2. Slope estimate (ß1) for the relationship between reproductive performance (i.e. probability of capturing a juvenile) and habitat quality estimated from the species distribution models around the Constant Effort Sites at multiple spatial scales: (a) 1 km, (b) 2 km, (c) 4 km, (d) 8 km and (e) 16 km buffers. The different bird species are ordered in the same way in the different panels (see Table 1 for acronyms). Points indicate the slope estimate; bold line indicates the standard error and thinner line indicates the 95% confidence intervals. Asterisks indicate a significantly positive or negative relationship (P ≤ 0.05).

DISCUSSION

We examined, for a sample of common passerine species over a large spatial extent, the relationship between the predictions obtained from species distribution models (which are typically considered as proxies for habitat quality) using atlas data and the production of juveniles estimated from mist-netting CES. For 21% of the species, we found a positive relationship between the predictions of the models and the estimated reproductive performance across all spatial scales. For almost half of the species, we found a positive relationship at least at one of the different spatial scales. On the other hand, there was no significant relationship between the modelling outcomes and the reproductive performance for the other half of species.

Our study was carried out at multiple spatial scales because the reproductive performance of a species at a site does not necessarily reflect only the local quality of the habitat, but also that of the neighbouring areas (Chalfoun & Martin 2007). Results showed that the relationship between the predictions of the models and the habitat quality estimated from the reproductive performance measure may considerably vary across spatial scales. *Erithacus rubecula*, *Sylvia atricapilla* and *Parus major* showed a significantly positive relationship only when the more distant areas around the mist-netting site were included in the analysis. The juveniles of these species most probably disperse more across the landscape than those of other species and their numbers may not necessarily reflect the quality of the local habitat conditions for breeding. It is known that passerine juveniles disperse from their natal home range after birth and that dispersal behaviour might vary across species depending on the landscape composition and the learning strategies (Skórka, Lenda, & Sutherland 2016). A quantitative analysis of the link between this scale dependency observed in our results and traits such as juvenile dispersal (DeSante *et al.* 1995) would provide insights into the mechanisms behind the observed pattern. However, such an analysis was beyond the scope of our study because those species-specific traits are very difficult to measure and depend on external factors such as weather conditions or geographic location. Thus, this level of detailed information is only available over small spatial extents and for a limited set of species that may not be representative of the bird species assemblages. As recently recommended by Pellissier et al. (2013), we opted for carrying out an analysis using as many species as possible over a large spatial extent to test the generality of the relationship between the predictions of species distribution models and fitness parameters informing on habitat quality. Our results suggest that species distribution models may help to predict habitat quality for some species over some extents, but also show that the generality of this relationship is not straightforward, as there is strong variability among species and across spatial scales. Further work is now warranted to further examine the link between species life histories and the ability of these models provide predictions that are in line with reproductive performance across varying spatial scales. Although we selected here the probability of capturing juveniles as a frequently used measure of reproductive performance (DeSante et al. 1995), it is also needed to test alternative measures to represent fitness more completely. It is worth mentioning that we also tested the relationship between the probability of capturing a female with brood patch and the habitat quality values obtained from the species distribution models (Appendix 2, Fig.A2). The results using this alternative reproductive performance measure showed a similar pattern to that obtained when using the probability of capturing juveniles.

In addition to the life history traits of the species, the relationship between the predictions of the models and habitat quality estimated from reproductive performance measures collected in the field may also be influenced by the modelling procedure itself. Even though the variables selected for the models were ecologically relevant predictors of bird distributions, they were not tailored to species-specific environmental requirements and were probably more relevant for some species than for others. The link between the observed distribution of the species and the quality of the habitat may also vary among species (Pulliam 2000). The predictions of the species distribution models such as those used in our study, however, inform on the potential of a species to be present in different locations from the statistical associations between the species presence records and the environmental conditions. Hence, these models might not necessarily capture information on the behavioural mechanisms that can lead animals to select poor or to avoid the most suitable habitats (Johnson 2007; Hollander *et al.* 2011; Robertson, Rehage, & Sih 2013) and the demographic consequences (e.g. reproduction performance) of such maladaptive selection (Kristan 2003; Lamb *et al.* 2016). Research efforts are underway to evaluate the extent to which these processes affect a large number of species (Hale & Swearer 2016; Robertson & Chalfoun 2016; Hollander *et al.* 2017).

In line with previous studies on this issue (Pellissier *et al.* 2013; Bean *et al.* 2014; Thuiller *et al.* 2014), we conclude that the predictions of species distribution models may prove useful but should be interpreted with extreme caution, especially when they are used to guide conservation or management actions that are expected to induce changes in habitat quality for the species. Depending on the type of data used to build the models and the link between the spatial distribution and the habitat requirements of the species (Pulliam

2000; Peterson *et al.* 2011), it might be risky to use the predictions of the models as direct proxies for habitat quality in the lack of careful species- and scale-specific assessments in the area of interest.

ACKNOWLEDGEMENTS

O.A. was funded by the National Research Fund, Luxembourg (AFR-PHD-08-63). S.H., L.B. and N.T. were funded through the EU BON project (contract no. 308464; FP7-ENV-2012 European Commission).

REFERENCES

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

Bean, W.T., Prugh, L.R.L.., Stafford, R., Butterfield, H.. S., Westphal, M. & Brashares, J.S. (2014) Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. *Journal of Applied Ecology*, **51**, 1116–1125.

Bean, W.T., Stafford, R. & Brashares, J.S. (2011) The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. *Ecography*, **35**, 250–258.

Bjornstad, O.N. (2016) *Ncf: Spatial Nonparametric Covariance Functions. R Package. Version 1.1-7.* https://CRAN.R-project.org/package=ncf.

Bosch, R., Real, J., Tintó, A., Zozaya, E.L. & Castell, C. (2010) Home-ranges and patterns of spatial use in territorial Bonelli's Eagles Aquila fasciata. *Ibis*, **152**, 105–117.

Brambilla, M. & Ficetola, G.F. (2012) Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species. *Journal of animal ecology*, **81**, 781–7.

Brotons, L., Herrando, S. & Pla, M. (2007) Updating bird species distribution at large spatial scales: Applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions*, **13**, 276–288.

Brotons, L., Mañosa, S. & Estrada, J. (2004) Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodiversity and Conservation*, **13**, 1039–1058.

Chalfoun, A.D. & Martin, T.E. (2007) Assessments of habitat preferences and quality depend on spatial scale and metrics of fitness. *Journal of Applied Ecology*, **44**, 983–992.

DeSante, D.F., Burton, K.M., Saracco, J.F. & Brett, W. (1995) Productivity indices and survival rate estimates from MAPS, a continent-wide programme of constant-effort mist-netting in North America. *Journal of Applied Statistics*, **22**, 935–947.

Elith, J., Kearney, M.R. & Phillips, S.J.P. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

Estrada, J., Pedrocchi, V., Brotons, L. & Herrando, S. (2004) *Atles Dels Ocells Nidificants de Catalunya (1999-2002)* (eds J Estrada, V Pedrocchi, L Brotons, and S Herrando). Institut Català d'Ornitologia (ICO) / Lynx Edicions, Barcelona.

Falcucci, A., Ciucci, P., Maiorano, L., Gentile, L. & Boitani, L. (2009) Assessing habitat quality for conservation using an integrated occurrence-mortality model. *Journal of Applied Ecology*, **46**, 600–609.

Franklin, J. (2009) *Mapping Species Distributions*. Cambridge University Press, Cambridge, UK.

Greenwood, P.J. & Harvey, P.H. (1982) The natal and breeding dispersal of birds. *Annual review of ecology and systematics*, **13**, 1–21.

Grup Català d'Anellament. (2002) The Sylvia Program: First annual report of the Catalan Constant Effort Site scheme. *Bird Populations*, **6**, 35–50.

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

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Hale, R. & Swearer, S.E. (2016) Ecological traps: current evidence and future directions. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 1–8.

Hall, L.S., Krausman, P.R. & Morrison, M.L. (1997) The habitat concept and a plea for standard terminology. *Wildlife Society Bulletin*, **25**, 173–182.

Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, **29**, 773–785.

Herrando, S., Brotons, L., Estrada, J., Guallar, S. & Anton, M. (2011) *Atles Dels Ocells de Catalunya a l'Hivern (2006-2009)*. Institut Català d'Ornitologia (ICO) / Lynx Edicions, Barcelona.

Herrando, S., Brotons, L., Guallar, S. & Quesada, J. (2010) Assessing regional variation in conservation value using fine-grained bird atlases. *Biodiversity and Conservation*, **19**, 867–881.

Hollander, F.A., Van Dyck, H., San Martin, G. & Titeux, N. (2011) Maladaptive habitat selection of a migratory passerine bird in a human-modified landscape. *PLoS one*, **6**, e25703.

Hollander, F.A., Titeux, N., Holveck, M.-J. & Van Dyck, H. (2017) Timing of breeding in an ecologically trapped bird. *The American Naturalist*, **in press**.

Van Horne, B. (1983) Density as a misleading indicator of habitat quality. *Journal of Wildlife Management*, **47**, 893–901.

Johnson, M.D. (2007) Measuring habitat quality: a review. *The Condor*, **109**, 489–504.

Kareiva, P. & Wennergren, U. (1995) Connecting landscape patterns to ecosystem and population processes. *Nature*, **373**, 299–302.

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

Lamb, C.T., Mowat, G., McLellan, B.N., Nielsen, S.E. & Boutin, S. (2016) Forbidden fruit:

human settlement and abundant fruit create an ecological trap an apex omnivore. *Journal of Animal Ecology*, doi: 10.1111/1365–2656.12589.

Peach, W.J.J., Buckland, S.T.T. & Baillie, S.R. (1996) The use of constant effort mist-netting to measure between-year changes in the abundance and productivity of common passerines. *Bird Study*, **43**, 142–156.

Pellissier, L., Meltofte, H., Hansen, J., Schmidt, N.M., Tamstorf, M.P., Maiorano, L., Aastrup, P., Olsen, J., Guisan, A. & Wisz, M.S. (2013) Suitability, success and sinks: how do predictions of nesting distributions relate to fitness parameters in high arctic waders? *Diversity and Distributions*, **19**, 1496–1505.

Peterson, A.T., Soberon, J., Pearson, R.G., Anderson, R.P. & Martinez-Meyer, E. (2011) *Ecological Niches and Geographic Distributions*, Princeton . Princeton University Press.

Phillips, S.J.P., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Robertson, B.A. & Chalfoun, A.D. (2016) Evolutionary traps as keys to understanding behavioral maladapation. *Current Opinion in Behavioral Sciences*, **12**, 12–17.

Robertson, B.A. & Hutto, R.L. (2006) A framework for understanding ecological traps and an evaluation of existing evidence. *Ecology*, **87**, 1075–1085.

Robertson, B.A., Rehage, J.S. & Sih, A. (2013) Ecological novelty and the emergence of evolutionary traps. *Trends in Ecology and Evolution*, **28**, 552–560.

Robinson, R.A., Julliard, R. & Saracco, J.F. (2009) Constant effort: Studying avian population processes using standardised ringing. *Ringing & Migration*, **24**, 199–204.

Sardà-Palomera, F., Brotons, L., Villero, D., Sierdsema, H., Newson, S.E. & Jiguet, F. (2012) Mapping from heterogeneous biodiversity monitoring data sources. *Biodiversity and Conservation*, **21**, 2927–2948.

Sardà-Palomera, F., Puigcerver, M., Vinyoles, D. & Rodríguez-Teijeiro, J.D. (2011) Exploring male and female preferences, male body condition, and pair bonds in the evolution of male sexual aggregation: The case of the common quail (Coturnix coturnix). *Canadian Journal of Zoology*, **89**, 325–333.

Skórka, P., Lenda, M. & Sutherland, W.J. (2016) Response of young and adult birds to the same environmental variables and different spatial scales during post breeding period. *Landscape Ecology*, **31**, 2063–2078.

Steffan-Dewenter, I., Münzenberg, U., Bürger, C., Thies, C. & Tscharntke, T. (2002) Scale-dependent effects of landscape context on three pollinator guilds. *Ecology*, **83**, 1421–1432.

Stephens, P.A., Pettorelli, N., Barlow, J., Whittingham, M.J. & Cadotte, M.W. (2015) Management by proxy? The use of indices in applied ecology. *Journal of Applied*

Ecology, **52**, 1–6.

Thuiller, W., Münkemüller, T., Schiffers, K.H., Georges, D., Dullinger, S., Eckhart, V.M., Edwards, T.C., Gravel, D., Kunstler, G., Merow, C., Moore, K., Piedallu, C., Vissault, S., Zimmermann, N.E., Zurell, D. & Schurr, F.M. (2014) Does probability of occurrence relate to population dynamics? *Ecography*, **37**, 1155–1166.

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H. & Defourny, P. (2007) Fitness-related parameters improve presence-only distribution modelling for conservation practice: The case of the Red-backed shrike. *Biological Conservation*, **138**, 207–223.

VanDerWal, J., Shoo, L.P., Johnson, C.N. & Williams, S.E. (2009) Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. *American Naturalist*, **174**, 282–291.

Vaughan, I.P. & 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.

Weber, M.M., Stevens, R.D., Diniz-Filho, J.A.F. & Grelle, C.E. V. (2016) Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography*, doi: 10.1111/ecog.02125.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J.P., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J.M., Williams, S.E. & Zimmermann, N.E. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.

SUPPORTING INFORMATION

Appendix 1

In addition to Fig.2 in the main text, we present for each species separately across all spatial scales the relationship between the probability of capturing a juvenile and the habitat quality derived from the species distribution models (Fig.A1).

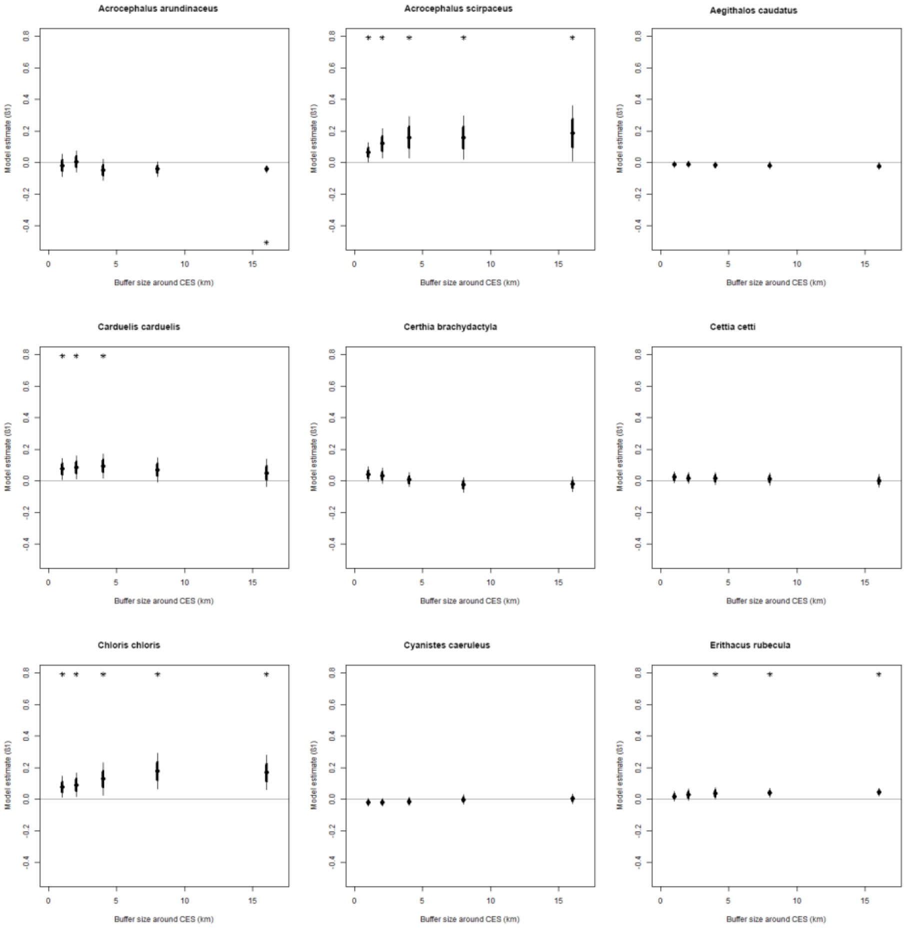


Figure A1. Slope estimate (ß$_1$) from the relationship between a reproductive performance parameter (the probability of capturing a juvenile) and the habitat quality estimated from the species distribution models around the Constant Effort Sites at multiple spatial scales (1, 2, 4, 8, and 16 km buffers) for each of the studied species separately. Points indicate the slope estimate, bold line indicates the standard error and thiner line indicates the 95% confidence intervals. Asteriks indicate a significant relationship (*P* ≤ 0.05).

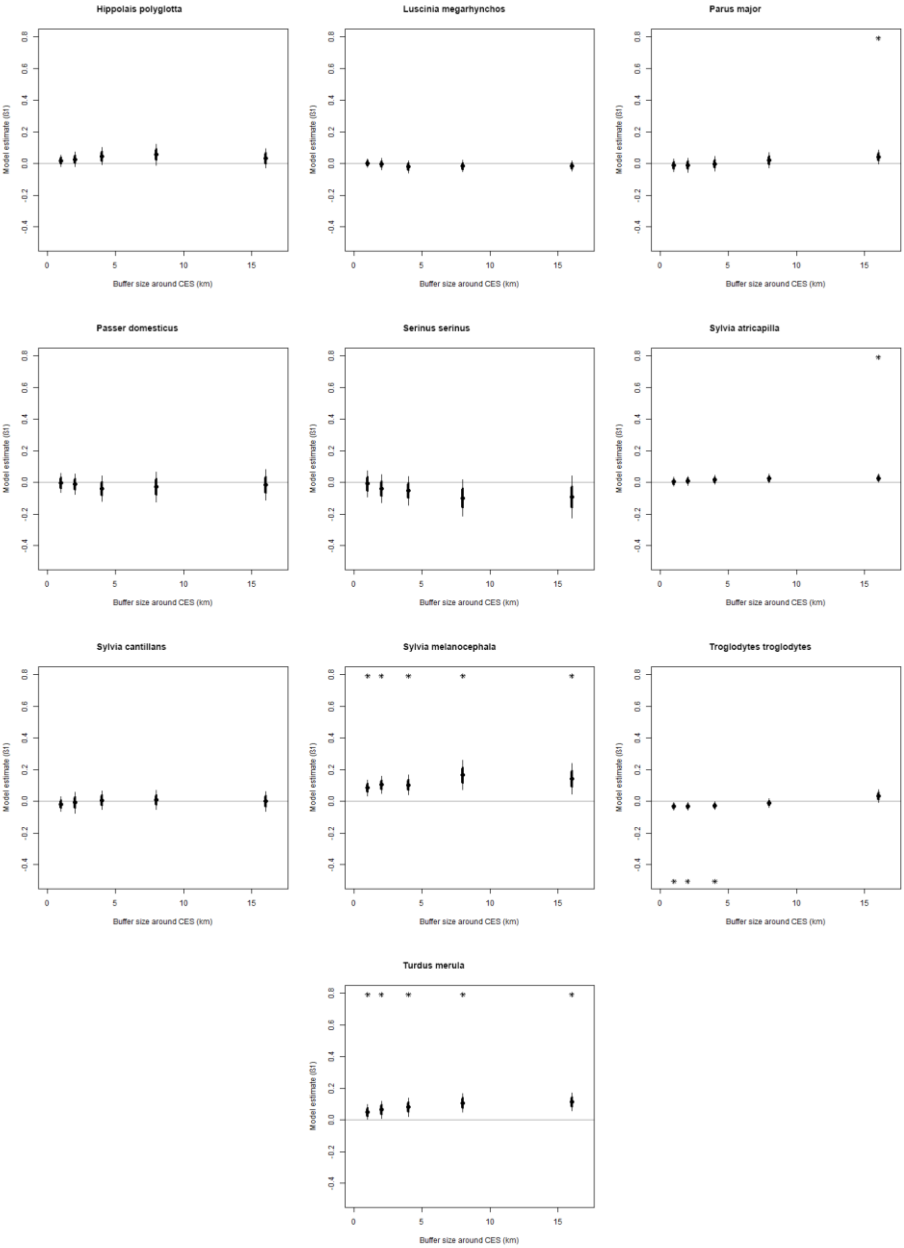


Figure A1 (continued). Slope estimate (ß$_1$) from the relationship between a reproductive performance parameter (the probability of capturing a juvenile) and the habitat quality estimated from the species distribution models around the Constant Effort Sites at multiple spatial scales (1, 2, 4, 8, and 16 km buffers) for each of the studied species separately. Points indicate the slope estimate, bold line indicates the standard error and thiner line indicates the 95% confidence intervals. Asteriks indicate a significant relationship ($P \leq 0.05$).

Appendix 2

As mentioned in the main text, we also tested the relationship between the probability of capturing a female with brood patch in the CES and the habitat quality values obtained from the species distribution models. We assumed that a high proportion of females with brood patch indicates high breeding habitat quality for the species around the CES. We applied the same statistical analysis as for the probability of capturing juveniles and the results are presented in Fig.A2.

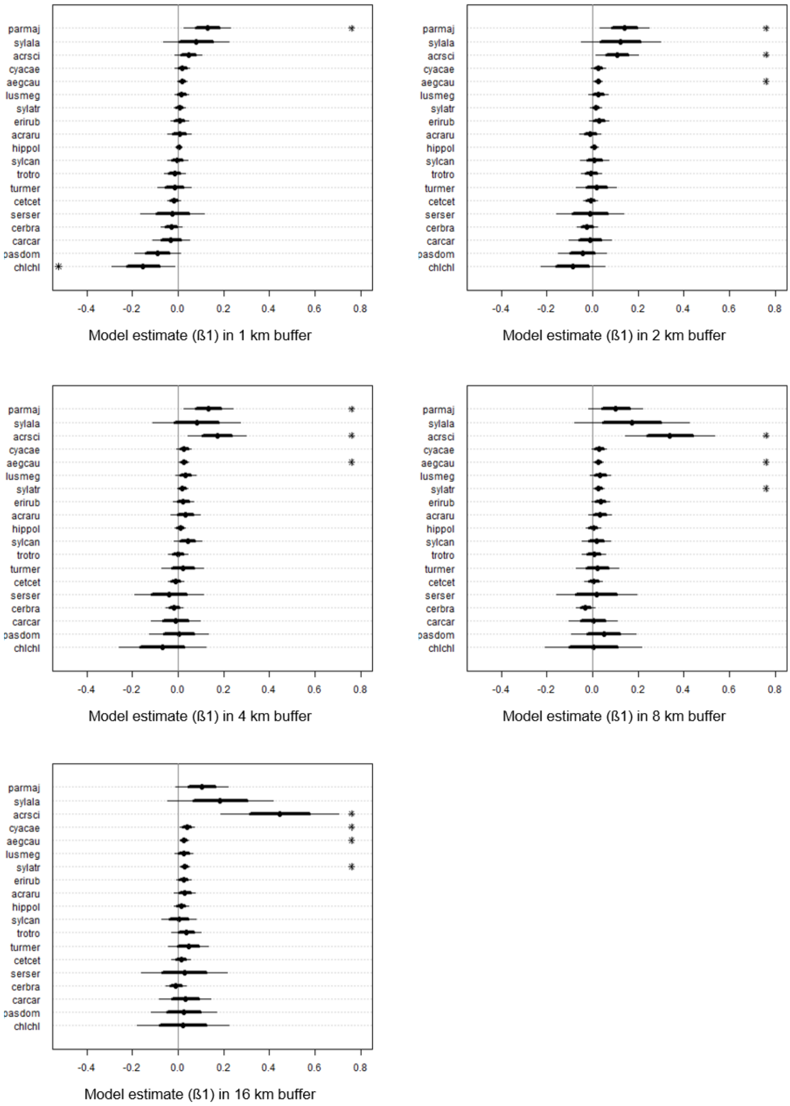


Figure A2. Slope estimate (ß1) for the relationship between reproductive performance (i.e. probability of capturing a female with brood patch) and habitat quality estimated from the species distribution models around the Constant Effort Sites at multiple spatial scales: (a) 1 km, (b) 2 km, (c) 4 km, (d) 8 km and (e) 16 km buffers. The different bird species are ordered in the same way in the different panels (see Table 1 in the main text for acronyms). Points indicate the slope estimate; bold line indicates the standard error and thinner line indicates the 95% confidence intervals. Asterisks indicate a significantly positive or negative relationship (P ≤ 0.05).

Article 4

Ecological traps and species distribution models: a challenge for prioritizing areas of conservation importance

The content of this chapter has been submitted:

Titeux, N., Aizpurua, O., Hollander, F., Sardà-Palomera, F., Hermoso, V., Paquet, J.Y., Mestdagh, X., Settele, J., Brotons, L., Van Dyck, H. Ecological traps and species distribution models: a challenge for prioritizing areas of conservation importance. *Biological Conservation* **under review.**

ABSTRACT

Species distribution models establish a statistical link between distribution records and environmental variables to predict spatial variations in the probability of species occurrence. Model outcomes are often used to prioritize areas for species conservation or management. Yet, organisms may prefer using low-quality habitats created by novel human activities. This ‘ecological trap’ phenomenon may challenge the usefulness of model outcomes to weigh the conservation importance of different areas. Previous work in Southern Belgium showed that Red-backed shrikes (*Lanius collurio*) are more attracted to clear-cut areas in plantation forests than to the traditional farmland habitat where their reproductive performance is higher. We built species distribution models with distribution data from a breeding bird atlas project and estimated how the probability of shrike occurrence varies between farmland and clear-cut areas in forest. These probabilities were then compared to measures of reproductive performance that proximately informed on breeding habitat quality for shrikes. Models predicted the probability of shrike occurrence with good accuracy, a criterion frequently used to support the use of model outcomes for conservation or management applications. However, models also predicted higher probability of occurrence in the clear-cut areas that are of lower quality than farmland for the reproduction of shrikes. With human-induced rapid environmental changes, species of conservation concern may use novel, but low-quality, habitats at high densities. When species switch or expand to a novel habitat, fitness consequences should be evaluated before using model outcomes in standard conservation-planning tools for guiding the spatial prioritization of management or conservation options.

INTRODUCTION

Species distribution models (SDMs) are widely used to predict the probability of species occurrence from the statistical relationship between distribution data and environmental variables (Elith and Leathwick, 2009; Franklin, 2013; Guisan and Zimmermann, 2000). Most techniques analyse the environmental conditions that are used by the species relative to those that are not used (or available) across the study area (Barbet-Massin et al., 2012; Elith et al., 2006). Probabilities of occurrence derived from SDMs are proposed to guide conservation practices such as weighing the importance of different areas for the management of threatened species (Guisan et al., 2013; Lavers et al., 2014; Pawar et al., 2007). Organisms may, however, settle in habitat types that are not necessarily the best options for their reproduction, survival and, therefore, persistence. This pattern may arise from a range of processes (Pulliam, 2000), such as the limited availability of optimal conditions (Braunisch et al., 2008; Titeux et al., 2007), source-sink dynamics (e.g. Pulliam & Danielson, 1991), dispersal limitations (e.g. Pinto and MacDougall, 2010), or behavioural maladaptation during habitat selection (Hale and Swearer, 2016; Robertson and Chalfoun, 2016; Robertson and Hutto, 2006). Probabilities of occurrence obtained from SDMs may then be of limited value to prioritize areas for the conservation of the species (Bean et al., 2014; Hirzel and Le Lay, 2008; Stephens et al., 2015).

Mechanistic approaches that incorporate demographic, physiological, evolutionary or behavioural processes in the modelling framework have been proposed as valuable alternative to SDMs (Buckley et al., 2010; Kearney and Porter, 2009; Sánchez-Clavijo et al., 2016). The development of models integrating the underlying mechanisms remains, however, strongly constrained by the availability of appropriate input data at relevant spatial scales (Thuiller et al., 2013; Urban et al., 2016). In addition, conservation practitioners are often not inclined to implement sophisticated approaches due to the need for a considerable level of conceptual and technical expertise that is not easily available to them (Addison et al., 2013). For these reasons, SDMs are expected to be further used for conservation and management applications (Dormann et al., 2012; Guisan et al., 2013). Therefore, there is a strong need to identify those situations under which SDMs are likely to produce outcomes that may not be useful for such practices (Bean et al., 2014; Stephens et al., 2015). One option to address this question is to interpret the probabilities of occurrence derived from SDMs relative to proxies of habitat quality that are relevant to the study species (Chalfoun and Martin, 2007; Johnson, 2007), i.e. derived from data on reproduction and survival (Jarnevich et al., 2015; Pellissier et al., 2013; Thuiller et al., 2014). Using these measurements to challenge the SDM outcomes may help evaluate whether probabilities of occurrence are useful to prioritize important areas for the conservation of the species.

Human-induced rapid environmental changes may create novel conditions (Robertson et al., 2013) that are more attractive to the organisms than alternative options of higher quality (Hale and Swearer, 2016; Robertson and Hutto, 2006; Schlaepfer et al., 2002). These changes are sudden in evolutionary time and organisms may, therefore, get trapped in habitats of low quality due to their behavioural response during habitat selection (Robertson and Chalfoun, 2016). Such ecological traps have the potential to lead to population decline and extinction (Kristan, 2003; Lamb et al., 2017) and represent an important conservation issue worldwide (Fletcher et al., 2012; Hale et al., 2015; Hale and Swearer, 2016; Rotem et al., 2013). Depending on the prevalence of low- and high-quality habitats across the landscape and on their relative attractiveness, individuals may occupy low-quality habitats at higher densities than high-quality habitats (Delibes et al., 2001). Since most SDM techniques analyse the relative use of different types of habitats by the species, they are likely to identify the densely occupied but low-quality habitats as the most important conservation areas for the species. Density has long been shown to represent a potentially misleading conservation metric (Skagen and Yackel Adams, 2011; Van Horne, 1983) and the importance of recognising ecological traps in conservation planning is a long-standing concern (Battin, 2004), but the relevance of using SDMs to identify suitable areas for the conservation of ecologically trapped organisms has yet received little attention (Sánchez-Clavijo et al., 2016; Sánchez-Mercado et al., 2014).

Hollander *et al.* (2011) have previously shown the existence of an ecological trap in the Red-backed shrike (*Lanius collurio*), a territorial, insectivorous, migratory bird species. Shrikes have long been using meadows and pastures in farmland areas to breed, but forest management since the last few decades has created large clear-cut areas with regrown vegetation in spruce plantation forests that offer a novel habitat to the species. Dominant males arrive in the clear-cut areas earlier than in the farmland habitat (Hollander et al., 2011) where they also defend their territories less vigorously (Hollander et al., 2012). However, reproductive performance (i.e. nest success, brood size and brood quality) is markedly lower in clear-cut areas than in the traditional farmland habitat because of between-habitat differences in food availability (Hollander et al., 2017, 2013) and nest predation (Hollander et al., 2015).

In this study system, we examined the relevance of using the outcomes of SDMs built with shrike distribution records to weigh the importance of clear-cut and farmland areas for the conservation of this ecologically trapped species. We used an independent breeding bird atlas dataset in the same region as

Hollander *et al.* (2011) to develop the SDMs and to estimate the probability of species occurrence in the two breeding habitat types. We tested whether the between-habitat differences in these probabilities are in line with the differences in shrike reproductive performance. Our results provide insights into the relevance of using the probabilities of occurrence derived from SDMs to identify and prioritize conservation areas when species switch or expand to novel habitats as a result of maladaptive habitat selection.

METHODS

Study areas

The study was conducted in the southeast of Belgium where agriculture and forestry created a mixed landscape of farmland and forest areas. We selected the same two study areas of 400 km^2 each as in previous work on shrike habitat selection (Hollander et al., 2011) (Fig. 1). Farmland includes cultivated areas, meadows and pastures for livestock rearing, whereas forests are mostly composed of Norway spruce (*Picea abies*) plantations for timber production (Fig. 2). These forests are intensively managed through large-sized clear-cutting in even-aged plantations.

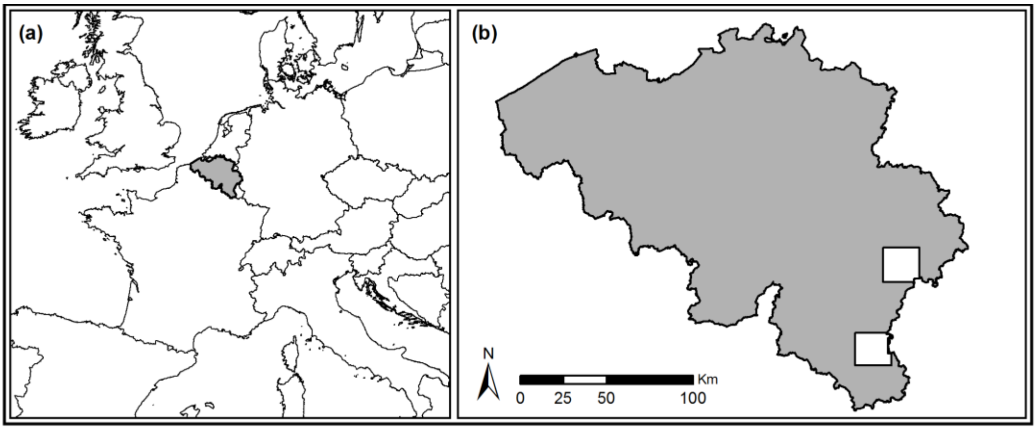


Figure 1. (a) Location of Belgium in NW Europe. (b) Location of the two study areas in SE Belgium.

Study species

The Red-backed shrike (*L. collurio*) is a passerine bird that has a wide breeding range across the Western Palearctic (Lefranc and Worfolk, 1997). When arriving from overwintering sites in southern Africa, shrikes establish breeding territories in open habitats. Meadows and pastures with scattered and thorny hedges and bushes are the traditionally used breeding habitat (Titeux et al., 2007), but regrown vegetation in forest clear-cut patches is a novel habitat for the species (Hollander et al., 2011).

During three consecutive years (2008-2010), we studied the reproductive performance of the Red-backed shrikes in 118 breeding sites in meadows and pastures (hereafter ‘farmland’: N=58) and in clear-cut patches (hereafter ‘forest’: N=60) across the two study areas (see details in Hollander et al., 2011). We searched for nests on a daily basis from mid-May to late July and nests were revisited throughout the season to determine (1) nest success (i.e. production of at least one fledgling), (2) brood size (i.e. number of nestlings older than 12 days) and (3) brood quality (i.e. average nestling body conditions). We estimated brood quality from tarsus length, wing length (± 0.01 mm, digital callipers) and body mass (± 0.1 g, KERN laboratory balance) measured on 12-day old nestlings (range: 11-15 days). These three measures were combined into the first axis of a Principal Component Analysis (PC1: explained variance = 89%, eigenvalue = 2.60, loadings: tarsus length = 0.58, body mass = 0.58, wing length = 0.57), which we averaged per nest (Hollander et al., 2017). Hollander *et al.* (2011) provide evidence that these three measures of reproductive performance are lower in forest than in farmland (Fig. 3).

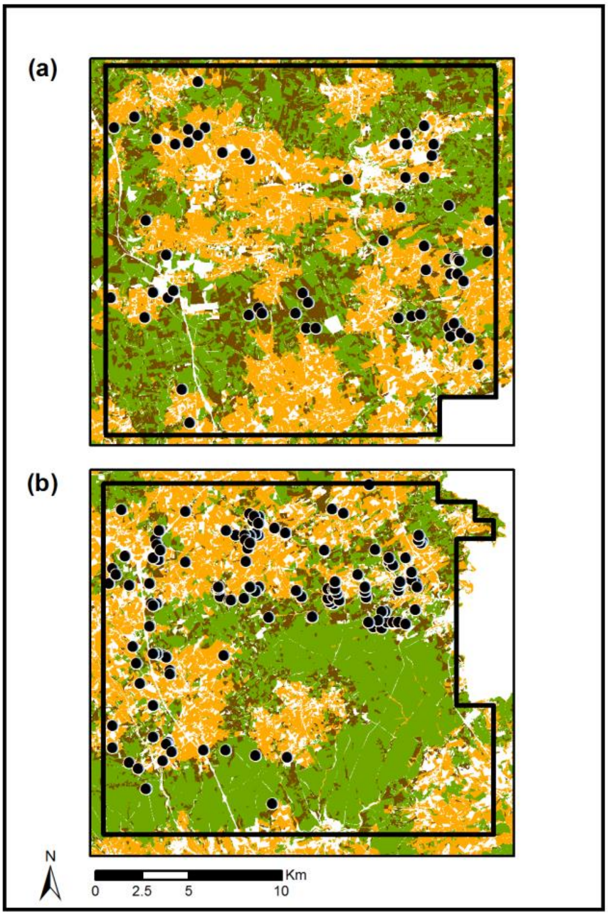


Figure 2. Red-backed shrike distribution records (black dots) from the Breeding Bird Atlas of Wallonia (BBAW) in the northern (a) and southern (b) study areas with the main land cover types in the background: meadows and pastures (orange), mature plantation forest (green) and clear-cut patches (brown).

Species data

We used Red-backed shrike distribution data collected during the Breeding Bird Atlas of Wallonia (BBAW) project (Jacob et al., 2010) to build the SDMs. For species such as the Red-backed shrike, atlas project fieldworkers surveyed potentially suitable habitats within 40 km^2 sampling units to localize breeding individuals with as much exhaustiveness and spatial precision as possible. As fieldworkers were asked not to neglect any habitat type, we assumed that the sampling bias was negligible and that the atlas data reflected the relative use of different habitat types by the species. Only the presences with a spatial precision smaller than 100 meters were used in subsequent analyses because this species establishes breeding territories covering 1-3 ha (Titeux et al., 2007). Although atlas data were collected between 2001 and 2008, we only used the presences from the period 2004-2008 (Fig. 2) to ensure as much temporal match as possible with the data used to quantify reproductive performance.

Environmental data

We selected 15 environmental variables (Table 1) considered to characterize the most important conditions for the reproduction of the study species in the region (Aizpurua et al., 2015; Titeux et al., 2007). These variables were constructed using readily available GIS layers and aerial photographs. All environmental variables were calculated within 200-m resolution grid cells covering the two study areas in order to approximately match the mean size of the shrike breeding territories. These variables were intended to describe the main vegetation types and landscape features relevant for shrikes in farmland and forest. All selected variables were only weakly correlated with each other (Pearson correlation coefficients < 0.6).

Table 1. Environmental variables calculated within the 200-m resolution grid cells covering the two study areas and used in the modelling procedures.

Variable	Source	Year	Units
Average orientation	DEM		°
Average slope	DEM		%
Topographic moisture index †	DEM		
Distance to closest urban area	COSW	2008	m
Surface of cultivated areas	COSW	2008	ha
Surface of meadows and pastures	COSW	2008	ha
Surface of semi-natural grasslands	COSW	2008	ha
Surface of broadleaved forest	MRW	1993	ha
Surface of coniferous forest	MRW	1993	ha
Surface of mixed forest	MRW	1993	ha
Surface of wetlands	COSW	2008	ha
Number of isolated trees and bushes	IGN	2009	
Total length of tree lines and hedgerows	IGN	2009	m
Surface of spruce plantation clear-cuts	Aerial pictures	2006-2010	ha
Dominant soil type ††	CNSW	2007	

DEM: Digital Elevation Model
CNSW: ‘*Cartographie Numérique des Sols de Wallonie*’ – Soil map of Wallonia (scale: 1:10000)
COSW: ‘*Cartographie de l’Occupation du Sol en Wallonie*’ – Land use map of Wallonia (scale: 1:10000)
IGN: ‘*Institut Géographique National*’ – Land use map of Belgium (scale: 1:10000)
MRW: ‘Ministère de la Région Wallonne’ – Land cover map of Wallonia (pixel resolution: 20 meters)
Aerial pictures: ‘*Service Public de Wallonie* (http://geoportail.wallonie.be)’ – Orthophotos 2006-2007 (pixel resolution: 50 cm) and 2009-2010 (pixel resolution: 25 cm)
† Topographic moisture index was calculated following Beven and Kirkby (1979).
†† Categorical variable

Habitat use in farmland and forest

To estimate how Red-backed shrikes use both types of habitats, we first calculated the area that is potentially available for them in farmland (i.e. surface of meadows and pastures) and in forest (i.e. surface of clear-cut patches) across the study areas (Fig. 2). Second, we estimated the extent to which these potentially suitable areas are actually occupied by the shrikes. To do so, we overlaid the shrike records from the atlas data with aerial photographs (see Table 1), we counted the number of records in farmland (meadows and pastures) and in forest (clear-cut patches), and we calculated for each habitat type separately the ratio between the number of records and the area potentially available for the shrikes.

Species distribution models

As atlas project fieldworkers covered the sampling units with as much exhaustiveness as possible, we built SDMs with presence-absence data and modelling algorithms using the BIOMOD2 package (Thuiller et al., 2016) implemented in R (R Core Team, 2015): artificial neural networks (ANN), flexible discriminant analysis (FDA), generalized boosting models (GBM), generalized linear models (GLM) and multivariate adaptive regression splines (MARS). Grid cells completely covered with mature forests or urban areas were excluded and SDMs were developed to document variation in the probability of shrike occurrence across the remaining areas where environmental conditions were not incompatible with the reproduction of the species (Titeux et al., 2007). We used the 200-m resolution grid cells where the species was recorded during the atlas fieldwork in 2004-2008 as presence data (n=184). We randomly selected grid cells where shrikes were not recorded during that period to cover 20% of the study area with absence data (n=2927) and we used equal weighting for presences and absences so that the whole set of presence data had the same weight as the set of absence data (Barbet-Massin et al., 2012).

For each algorithm, the models were calibrated 100 times using a random sample of 80% of the presence and absence data (calibration data) and were then evaluated against the remaining 20% of the data (evaluation data) using the area under the curve (AUC) of a receiver-operating characteristic (ROC) plot. The probabilities of occurrence from the different modelling algorithms were combined using the ensemble forecasting procedure implemented in the BIOMOD2 package using only individual models with AUC values above 0.6. This procedure resulted in an average probability of shrike occurrence in each grid cell that ranged between 0 and 1.

We delineated 100-m radius circles around the 299 nest locations found during the period 2008-2010 in the 118 breeding sites where shrike reproductive performance was measured (Hollander et al., 2011). This radius was set to represent the geographical area covered by the territories of the breeding shrikes around their nests. Within each circle, we calculated the area-weighted mean probability of shrike occurrence derived from the SDMs using the fraction of the circle intersecting each grid cell as a weighting factor. We analysed the difference between farmland and forest in the area-weighted mean probability of shrike occurrence around the nest locations.

RESULTS

About 15% of the grid cells across the study areas were covered with mature forests or urban areas. As these cells were considered as incompatible with the reproduction of the shrikes, they were not used in the analyses. The remaining grid cells were at least partly covered with potentially suitable vegetation for shrikes in farmland (meadows and pastures: 25,128 ha) or in forest (clear-cut patches: 6,106 ha). From the atlas data used to build the models, 72 shrike records were located in meadows or pastures and 110 were located in forest clear-cut patches. Ten presences were not used in the calculations because their allocation to one of the two habitat types with the aerial photographs was uncertain. On this basis, we estimated that shrikes occupied 0.3% of available areas in farmland and 2% in forest clear-cuts.

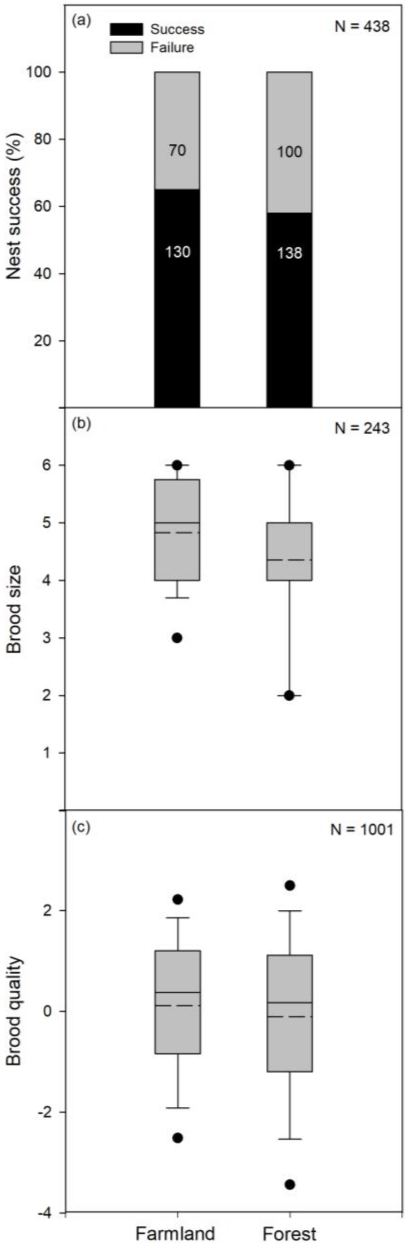


Figure 3. Reproductive performance of Red-backed shrikes in the farmland (pastures and meadows) and the forest (clear-cut areas) habitats (modified from Hollander *et al.* 2011). (a) Proportion of breeding attempts associated with nest success (black) and failure (grey). (b-c) Box-and-whisker plots and quartile distribution (⊥ and ⊤: 5th and 95th percentiles, •: outlying values, ---: mean value, −: median value) for brood size (i.e. number of nestlings older than 12 days per nest) and brood quality (i.e. average nestling body conditions per nest). The number of nests (for nest success and brood size) or nestlings (for brood quality) is indicated in the panels.

The AUC value obtained from the ensemble forecasting procedure was 0.804. This indicates that there is an 80% chance that a grid cell with shrike presence in the atlas data receives a probability of shrike occurrence higher than that of a randomly selected grid cell with absence data. The area-weighted mean probability of shrike occurrence derived from SDMs within the 100-m radius circles around the nest locations was significantly lower in farmland (mean value: 0.487; standard deviation: 0.080) than in forest clear-cuts (mean value: 0.537; standard deviation: 0.108) (Wilcoxon rank sum test, W=7746.5, P < 0.0001, Fig. 4).

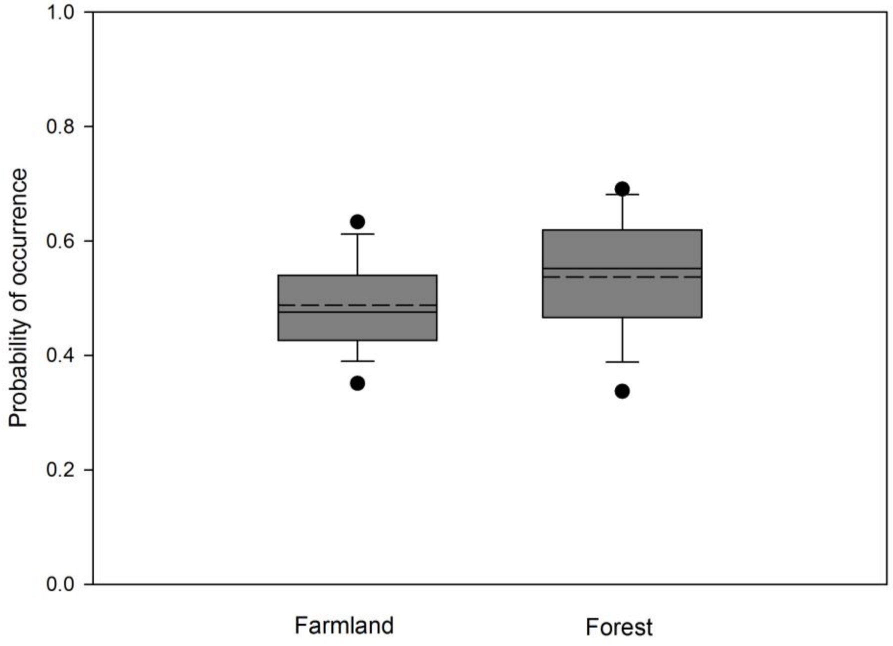


Figure 4. Box-and-whisker plots and quartile distribution (┴ and ┬: 5th and 95th percentiles, •: outlying values, ---: mean value, −: median value) for the probabilities of occurrence derived from the species distribution models within the 100-meters distance around the Red-backed shrike nest sites in farmland (meadows and pastures: N=144) and in forest (clear-cut areas: N=155).

DISCUSSION

We built SDMs with distribution data from a breeding bird atlas project to estimate how the probability of shrike occurrence varies across a landscape where the species occupies two different habitat types: i) meadows and pastures (farmland) and ii) clear-cut patches (forest). Although shrikes were known to prefer using the less suitable forest clear-cut habitat for breeding (e.g. Hollander *et al.* 2011), we intentionally ignored this behavioural maladaptation when building the SDMs with presence and absence data. If the outcomes of such SDMs were used to prioritize areas across the landscape for the conservation of the species (e.g. Lavers et al., 2014; Pawar et al., 2007), clear-cut patches in forest would be identified as more important for shrikes than meadows and pastures in farmland. This finding would be opposite to the observed difference in shrike reproductive performance between farmland and forest clear-cut habitats (Hollander et al., 2011).

Shrikes occupy clear-cut patches in plantation forests in many parts of their breeding range (e.g. Lefranc & Worfolk 1997; Karlsson 2004; Lislevand 2012). This novel ecosystem is also used by several other species worldwide, including other birds (Stjernman et al., 2013), butterflies (Viljur and Teder, 2016) and mammals (Bogdziewicz and Zwolak, 2014), amongst others. Although clear-cut patches in plantation forests potentially constitute real habitat opportunities for some species linked to early successional vegetation (Paz Acuña and Estades, 2011; Swanson et al., 2011; Żmihorski et al., 2016), they may also act as an ecological trap for a number of other organisms (Hollander et al., 2017). Beyond this specific ecosystem, an increasing amount of studies have documented the existence of ecological traps in a variety of taxonomical groups (e.g., insects, amphibians, reptiles, birds, arthropods, mammals) and human-modified environments such as cities, farmland, forest, or aquatic ecosystems (Hale et al., 2015; Hale and Swearer, 2016; Robertson and Chalfoun, 2016; Rotem et al., 2013). It has been recently suggested that traps may also arise as an unintended consequence of habitat restoration (Hale and Swearer, 2017) or invasive plant management (Carter et al., 2017). Hence, the issue we raise here and the recommendations we make below go largely beyond the specific case of shrikes in plantation forests.

We showed that the individual-level difference in habitat preference between the two types of habitats (Hollander et al., 2012, 2011) results in a higher shrike density in the preferred forest clear-cut habitat. In contrast, available areas in meadows and pastures are less preferred and occupied at a lower density. As most SDM techniques – such as the ones we used here – discriminate between the environmental conditions that are used by the species and those that are not used (Barbet-Massin et al., 2012; Elith et al., 2006; Guisan and Zimmermann, 2000), the probabilities of shrike occurrence estimated by the models were higher around the nests located in forest clear-cut areas than around those established in farmland. Interestingly, AUC values in the same order of magnitude as the one we obtained here usually provide support for the use of SDM outcomes with some degree of reliability because these values are considered to reflect a good predictive accuracy of the models (Swets, 1988). With our case study on an ecologically trapped organism, we showed that, even when delivering accurate estimates of occurrence probabilities, SDMs might misdirect conservation strategies if their outcomes were used to weigh the importance of different areas for the reproduction of the species. Guillera-Arroita *et al.* (2015) stressed the importance of using properly collected presence and absence data in SDMs to derive estimates of occurrence probabilities that suit certain types of management or conservation applications. Our results show that even occurrence probabilities estimated from presence and absence data collected according to a well-designed sampling scheme are insufficient to serve this purpose when behavioural maladaptation during habitat selection modifies the link between habitat quality and habitat use.

When a species is known to occupy novel conditions that have been emerging in the landscape as a result of human-induced rapid environmental changes (Robertson et al., 2013), we recommend to avoid relying blindly upon the outcomes of statistical SDM approaches except for documenting spatial variation in the probabilities of occurrence of the species. Before SDM outcomes could be used to guide the prioritization of management options amongst areas that apparently vary in their importance for the conservation of the species, there is a need to evaluate to which degree these newly occupied conditions contribute to population persistence relative to the traditionally used and still available habitat types. Fitness-related factors need to be estimated as to provide a proximate underpinning of the relative quality of the different habitat types occupied by the species (Chalfoun and Martin, 2007; Johnson, 2007; Pellissier et al., 2013; Thuiller et al., 2014). If this evaluation indicates that part of the population has been moving to a novel ecosystem where habitat quality is lower than in previously occupied sites, SDM outcomes should not be used to prioritize areas for conservation or management applications because the pattern of habitat use is not likely to reflect variation in habitat quality underlying reproduction or survival. Both theoretical and empirical studies have shown that the attractiveness of a trap habitat may induce source-sink dynamics with detrimental consequences for the populations (Kristan, 2003; Lamb et al., 2017).

The severity of this conservation issue will mostly depend on i) the attractiveness of the novel ecosystem compared to the traditionally used habitat(s) and ii) the relative availability of the different habitat types across the landscape (Sánchez-Clavijo et al., 2016). Quantifying the attractiveness of different habitat types occupied by the species demands a great deal of field and experimental work (Hollander et al., 2012; Pärt et al., 2007; Robertson and Hutto, 2006), but this effort will be needed to cope with the potential mismatch between habitat attractiveness and habitat quality in the modelling process itself. A variety of methods have been proposed to assess the consequences of maladaptive habitat selection on population dynamics through a separate modelling of habitat attractiveness and habitat quality (Delibes et al., 2001; Donovan and Thompson III, 2001; Fletcher et al., 2012). However, most of the proposed approaches are theoretical models and are not spatially explicit (Sánchez-Clavijo et al., 2016), which limits their use to identify spatial priorities for management or other conservation actions.

It is now warranted to evaluate whether mechanistic modelling approaches that explicitly relate spatial distribution data to demographic rates of the populations or individual behavioural traits (e.g. DeAngelis & Mooij 2005; Kearney & Porter 2009; Schurr *et al.* 2012) offer promising alternatives to prioritize areas for the conservation of ecologically trapped organisms. Among the diversity of spatially explicit modelling techniques that integrate mechanisms, individual-based approaches simulate how population-level patterns emerge from the interactions between individuals and from the variations in individual actions and behaviour (DeAngelis and Mooij, 2005). These techniques have the potential to deal with the behavioural motivation of the individuals to select among several options varying in attractiveness. To our knowledge, individual-based models have been used to evaluate the effect of maladaptive habitat selection on the spatial distribution of virtual species in simulated landscapes (e.g. Sánchez-Clavijo, Hearns & Quintana-Ascencio 2016) but have never been applied to real-world data.

In the last few decades, a number of species in the terrestrial, marine and freshwater realms have been shown to track changing climate conditions and shift their distribution along latitude, altitude and depth gradients (Scheffers et al., 2016). As organisms are moving towards more favourable climate conditions, these distributional changes are often assumed to reflect adaptive behavioural responses to a changing climate. However, the fitness consequences of such behavioural flexibility have received little attention and recent studies have suggested that individuals may encounter traps when moving and experiencing novel conditions (Hale et al., 2016). Range shifts are the results of moving individuals (i.e. the 'disperser' phenotypes, e.g. Debeffe *et al.* 2014) that may be more likely to face traps than other phenotypes. Hence, this process may affect or remove a biased sample of individuals and induce changes in the personality or phenotype structure of the populations, with potentially significant conservation consequences. Although SDMs are increasingly used to predict range shifts (e.g. Elith, Kearney & Phillips 2010) and to identify future spatial conservation priorities (e.g. Faleiro et al., 2013) under climate change, most modelling approaches and studies largely ignore potential maladaptive responses of the organisms when they track suitable climate conditions. Approaches are being developed to integrate ecological and evolutionary processes into the modelling framework so as to improve our ability to predict species range dynamics (Normand et al., 2014; Thuiller et al., 2013; Urban et al., 2016). Yet, to the best of our knowledge, these tools have not explicitly considered the potential interaction between range shift and ecological traps under changing climate. We encourage novel research at the crossroad between species distribution modelling, ecological traps, and environmental change impact assessment.

ACKNOWLEDGEMENTS

We thank the participants and the management team of the BBAW project for fieldwork and atlas data provision. GIS layers and aerial pictures were provided by the Direction Générale de l’Agriculture, des Ressources Naturelles et de l’Environnement (DGARNE) of the Service Public de Wallonie (SPW). The IRScNB-KBIN provided us with a national license for capturing shrikes. This is publication BRC XXX of the Biodiversity Research Centre (Université catholique de Louvain).

REFERENCES

Addison, P.F.E., Rumpff, L., Bau, S.S., Carey, J.M., Chee, Y.E., Jarrad, F.C., McBride, M.F., Burgman, M.A., 2013. Practical solutions for making models indispensable in conservation decision-making. Divers. Distrib. 19, 490–502. doi:10.1111/ddi.12054

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Biver, G., Brotons, L., Titeux, N., 2015. Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species. J. Appl. Ecol. 52, 1608–1616. doi:10.1111/1365-2664.12515

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338. doi:10.1111/j.2041-210X.2011.00172.x

Battin, J., 2004. When good animals love bad habitats: ecological traps and the conservation of animal populations. Conserv. Biol. 18, 1482–1491. doi:10.1111/j.1523-1739.2004.00417.x

Bean, W.T., Prugh, L.R., Stafford, R., Butterfield, H.S., Westphal, M., Brashares, J.S., 2014. Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. J. Appl. Ecol. 51, 1116–1125. doi:10.1111/1365-2664.12281

Bogdziewicz, M., Zwolak, R., 2014. Responses of small mammals to clear-cutting in temperate and boreal forests of Europe: a meta-analysis and review. Eur. J. For. Res. 133, 1–11. doi:10.1007/s10342-013-0726-x

Braunisch, V., Graf, R.F., Bollmann, K., Hirzel, A.H., 2008. Living on the edge - Modelling habitat suitability for species at the edge of their fundamental niche. Ecol. Modell. 214, 153–167.

Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J., Sears, M.W., 2010. Can mechanism inform species' distribution models? Ecol. Lett. 13, 1041–1054. doi:10.1111/j.1461-0248.2010.01479.x

Carter, E.T., Ravesi, M.J., Eads, B.C., Kingsbury, B.A., 2017. Invasive plant management creates ecological traps for snakes. Biol. Invasions 19, 443–453. doi:10.1007/s10530-016-1289-9

Chalfoun, A.D., Martin, T.E., 2007. Assessments of habitat preferences and quality depend on spatial scale and metrics of fitness. J. Appl. Ecol. 44, 983–992. doi:10.1111/j.1365-2664.2007.01352.x

DeAngelis, D.L., Mooij, W.M., 2005. Individual-based modeling of ecological and evolutionary processes. Annu. Rev. Ecol. Evol. Syst. 36, 147–168. doi:10.1146/annurev.ecolsys.36.102003.152644

Debeffe, L., Morellet, N., Bonnot, N., Gaillard, J.M., Cargnelutti, B., Verheyden-Tixier, H., Vanpé, C., Coulon, A., Clobert, J., Bon, R., Hewison, A.J.M., 2014. The link between behavioural type and natal dispersal propensity reveals a dispersal syndrome in a large herbivore. Proc. R. Soc. B Biol. Sci. 281, 20140873. doi:10.1098/rspb.2014.0873

Delibes, M., Gaona, P., Ferreras, P., 2001. Effects of an attractive sink leading into maladaptive habitat selection. Am. Nat. 158, 277–285. doi:10.1086/321319

Donovan, T.M., Thompson III, F.R., 2001. Modeling the ecological trap hypothesis: a habitat and demographic analysis for migrant songbirds. Ecol. Appl. 11, 871–882. doi:10.1890/1051-0761(2001)011[0871:MTETHA]2.0.CO;2

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M.R., Morin, X., Römermann, C., Schröder, B., Singer, A., 2012. Correlation and process in species distribution models: bridging a dichotomy. J. Biogeogr. 39, 2119–2131. doi:10.1111/j.1365-2699.2011.02659.x

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.M., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop.). 29, 129–151. doi:10.1111/j.2006.0906-7590.04596.x

Elith, J., Kearney, M.R., Phillips, S., 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1, 330–342. doi:10.1111/j.2041-210X.2010.00036.x

Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. doi:10.1146/annurev.ecolsys.110308.120159

Faleiro, F. V., Machado, R.B., Loyola, R.D., 2013. Defining spatial conservation priorities in the face of land-use and climate change. Biol. Conserv. 158, 248–257. doi:10.1016/j.biocon.2012.09.020

Fletcher, R.J., Orrock, J.L., Robertson, B.A., 2012. How the type of anthropogenic change alters the consequences of ecological traps. Proc. R. Soc. London Ser. B-Biological Sci. 279, 2546–2552. doi:10.1098/rspb.2012.0139

Franklin, J., 2013. Species distribution models in conservation biogeography: developments and challenges. Divers. Distrib. 19, 1217–1223. doi:10.1111/ddi.12125

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., McCarthy, M.A., Tingley, R., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24, 276–292. doi:10.1111/geb.12268

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C.S., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M., 2013. Predicting species distributions for conservation decisions. Ecol. Lett. 16, 1424–1435. doi:10.1111/ele.12189

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135, 147–186. doi:10.1016/S0304-3800(00)00354-9

Hale, R., Coleman, R., Pettigrove, V., Swearer, S.E., 2015. Identifying, preventing and mitigating ecological traps to improve the management of urban aquatic ecosystems. J. Appl. Ecol. 52, 928–939. doi:10.1111/1365-2664.12458

Hale, R., Morrongiello, J.R., Swearer, S.E., 2016. Evolutionary traps and range shifts in a rapidly changing world. Biol. Lett. 12, 20160003. doi:10.1098/rsbl.2016.0003

Hale, R., Swearer, S.E., 2017. When good animals love bad restored habitats: how maladaptive habitat selection can constrain restoration. J. Appl. Ecol. 54, 1478–1486. doi:10.1111/1365-2664.12829

Hale, R., Swearer, S.E., 2016. Ecological traps: current evidence and future directions. Proc. R. Soc. London Ser. B-Biological Sci. 283, 20152647. doi:10.1098/rspb.2015.2647

Hirzel, A.H., Le Lay, G., 2008. Habitat suitability modelling and niche theory. J. Appl. Ecol. 45, 1372–1381. doi:10.1111/j.1365-2664.2008.01524.x

Hollander, F.A., Titeux, N., Holveck, M.-J., Van Dyck, H., 2017. Timing of breeding in an ecologically trapped bird. Am. Nat. 189, 515–525. doi:10.1086/691329

Hollander, F.A., Titeux, N., Van Dyck, H., 2013. Habitat-dependent prey availability and offspring provisioning explain an ecological trap in a migratory bird. Funct. Ecol. 27, 702–709. doi:10.1111/1365-2435.12077

Hollander, F.A., Titeux, N., Van Dyck, H., 2012. Territorial resource valuation deviates from habitat quality in an ecologically trapped, long-distance migratory bird. Behav. Ecol. Sociobiol. 66, 777–783. doi:10.1007/s00265-012-1325-4

Hollander, F.A., Van Dyck, H., San Martin, G., Titeux, N., 2015. Nest predation deviates from nest predator abundance in an ecologically trapped bird. PLoS One 10, e0144098. doi:10.1371/journal.pone.0144098

Hollander, F.A., Van Dyck, H., San Martin, G., Titeux, N., 2011. Maladaptive habitat selection of a migratory passerine bird in a human-modified landscape. PLoS One 6, e25703. doi:10.1371/journal.pone.0025703

Jacob, J.-P., Dehem, C., Burnel, A., Dambiermort, J.L., Fasol, M., Kinet, T., van der Elst, D., Paquet, J.-Y., 2010. Atlas des oiseaux nicheurs de Wallonie. Aves & Région Wallonne, Gembloux (Belgium).

Jarnevich, C.S., Stohlgren, T.J., Kumar, S., Morisette, J.T., Holcombe, T.R., 2015. Caveats for correlative species distribution modeling. Ecol. Inform. 29, 6–15. doi:10.1016/j.ecoinf.2015.06.007

Johnson, M.D., 2007. Measuring habitat quality: a review. Condor 109, 489–504. doi:10.1650/8347.1

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

Kearney, M.R., Porter, W., 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecol. Lett. 12, 334–350. doi:10.1111/j.1461-0248.2008.01277.x

Kristan, W.B., 2003. The role of habitat selection behavior in population dynamics: source-sink systems and ecological traps. Oikos 103, 457–468. doi:10.1034/j.1600-0706.2003.12192.x

Lamb, C.T., Mowat, G., McLellan, B.N., Nielsen, S.E., Boutin, S., 2017. Forbidden fruit: human settlement and abundant fruit create an ecological trap for an apex omnivore. J. Anim. Ecol. 86, 55–65. doi:10.1111/1365-2656.12589

Lavers, J.L., Miller, M.G.R., Carter, M.J., Swann, G., Clarke, R.H., 2014. Predicting the spatial distribution of a seabird community to identify priority conservation areas in the Timor Sea. Conserv. Biol. 28, 1699–1709. doi:10.1111/cobi.12324

Lefranc, N., Worfolk, T., 1997. Shrikes: a guide to the shrikes of the World. Pica Press, Sussex, England.

Lislevand, T., 2012. Habitat and nest placement of Red-backed Shrikes Lanius collurio breeding in clear-cuts in southern Norway. Ornis Nor. 35, 28–36.

Normand, S., Zimmermann, N.E., Schurr, F.M., Lischke, H., 2014. Demography as the basis for understanding and predicting range dynamics. Ecography (Cop.). 37, 1149–1154. doi:10.1111/ecog.01490

Pärt, T., Arlt, D., Villard, M.-A., 2007. Empirical evidence for ecological traps: a two-step model focusing on individual decisions. J. Ornithol. 148, 327–332. doi:10.1007/s10336-007-0226-1

Pawar, S., Koo, M.S., Kelley, C., Ahmed, M.F., Chaudhuri, S., Sarkar, S., 2007. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. Biol. Conserv. 136, 346–361. doi:10.1016/j.biocon.2006.12.012

Paz Acuña, M., Estades, C.F., 2011. Plantation clearcut size and the persistence of early-successional wildlife populations. Biol. Conserv. 144, 1577–1584. doi:10.1016/j.biocon.2011.02.003

Pellissier, L., Meltofte, H., Hansen, J., Schmidt, N.M., Tamstorf, M.P., Maiorano, L., Aastrup, P., Olsen, J., Guisan, A., Wisz, M.S., 2013. Suitability, success and sinks: how do predictions of nesting distributions relate to fitness parameters in high arctic waders? Divers. Distrib. 19, 1496–1505. doi:10.1111/ddi.12109

Pinto, S.M., MacDougall, A.S., 2010. Dispersal limitation and environmental structure interact to restrict the occupation of optimal habitat. Am. Nat. 175, 675–686. doi:10.1086/652467

Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349–361. doi:10.1046/j.1461-0248.2000.00143.x

Pulliam, H.R., Danielson, B.J., 1991. Sources, Sinks, and Habitat Selection: A Landscape Perspective on Population-Dynamics. Am. Nat. 137, S50–S66. doi:10.1086/285139

R Core Team, 2015. R: A language and environment for statistical computing.

Robertson, B.A., Chalfoun, A.D., 2016. Evolutionary traps as keys to understanding behavioral maladapation. Curr. Opin. Behav. Sci. 12, 12–17. doi:10.1016/j.cobeha.2016.08.007

Robertson, B.A., Hutto, R.L., 2006. A framework for understanding ecological traps and an evaluation of existing evidence. Ecology 87, 1075–1085. doi:10.1890/0012-9658(2006)87[1075:AFFUET]2.0.CO;2

Robertson, B.A., Rehage, J.S., Sih, A., 2013. Ecological novelty and the emergence of evolutionary traps. Trends Ecol. Evol. 28, 552–560. doi:10.1016/j.tree.2013.04.004

Rotem, G., Ziv, Y., Giladi, I., Bouskila, A., 2013. Wheat fields as an ecological trap for reptiles in a semiarid agroecosystem. Biol. Conserv. 167, 349–353. doi:10.1016/j.biocon.2013.08.028

Sánchez-Clavijo, L.M., Hearns, J., Quintana-Ascencio, P.F., 2016. Modeling the effect of habitat selection mechanisms on population responses to landscape structure. Ecol. Modell. 328, 99–107. doi:10.1016/j.ecolmodel.2016.03.004

Sánchez-Mercado, A., Ferrer-Paris, J.R., García-Rangel, S., Yerena, E., Robertson, B.A., Rodríguez-Clark, K.M., 2014. Combining threat and occurrence models to predict potential ecological traps for Andean bears in the Cordillera de Mérida, Venezuela. Anim. Conserv. 17, 388–398. doi:10.1111/acv.12106

Scheffers, B.R., De Meester, L., Bridge, T.C.L., Hoffmann, A.A., Pandolfi, J.M., Corlett, R.T., Butchart, S.H.M., Pearce-Kelly, P., Kovacs, K.M., Dudgeon, D., Pacifici, M., Rondinini, C., Foden, W.B., Martin, T.G., Mora, C., Bickford, D., Watson, J.E.M., 2016. The broad footprint of climate change from genes to biomes to people. Science 354, aaf7671. doi:10.1126/science.aaf7671

Schlaepfer, M.A., Runge, M.C., Sherman, P.W., 2002. Ecological and evolutionary traps. Trends Ecol. Evol. 17, 474–480. doi:10.1016/S0169-5347(02)02580-6

Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F., Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A., Zimmermann, N.E., 2012. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. J. Biogeogr. 39, 2146–2162. doi:10.1111/j.1365-2699.2012.02737.x

Skagen, S.K., Yackel Adams, A.A., 2011. Potential misuse of avian density as a conservation metric. Conserv. Biol. 25, 48–55. doi:10.1111/j.1523-1739.2010.01571.x

Stephens, P.A., Pettorelli, N., Barlow, J., Whittingham, M.J., Cadotte, M.W., 2015. Management by proxy? The use of indices in applied ecology. J. Appl. Ecol. 52, 1–6. doi:10.1111/1365-2664.12383

Stjernman, M., Green, M., Lindström, Å., Olsson, O., Ottvall, R., Smith, H.G., 2013.

Habitat-specific bird trends and their effect on the Farmland Bird Index. Ecol. Indic. 24, 382–391. doi:10.1016/j.ecolind.2012.07.016

Swanson, M.E., Franklin, J.F., Beschta, R.L., Crisafulli, C.M., DellaSala, D. a, Hutto, R.L., Lindenmayer, D.B., Swanson, F.J., 2011. The forgotten stage of forest succession: early-successional ecosystems on forest sites. Front. Ecol. Environ. 9, 117–125. doi:10.1890/090157

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 204, 1285–1293.

Thuiller, W., Georges, D., Engler, R., Breiner, F., 2016. biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7.

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K., Gravel, D., 2013. A road map for integrating eco-evolutionary processes into biodiversity models. Ecol. Lett. 16, 94–105. doi:10.1111/ele.12104

Thuiller, W., Münkemüller, T., Schiffers, K.H., Georges, D., Dullinger, S., Eckhart, V.M., Edwards, T.C., Gravel, D., Kunstler, G., Merow, C., Moore, K., Piedallu, C., Vissault, S., Zimmermann, N.E., Zurell, D., Schurr, F.M., 2014. Does probability of occurrence relate to population dynamics? Ecography (Cop.). 37, 1155–1166. doi:10.1111/ecog.00836

Titeux, N., Dufrêne, M., Radoux, J., Hirzel, A.H., Defourny, P., 2007. Fitness-related parameters improve presence-only distribution modelling for conservation practice: The case of the red-backed shrike. Biol. Conserv. 138, 207–223. doi:10.1016/j.biocon.2007.04.019

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.-B., Pe’er, G., Singer, A., Bridle, J.R., Crozier, L.G., De Meester, L., Godsoe, W., Gonzalez, A., Hellmann, J.J., Holt, R.D., Huth, A., Johst, K., Krug, C.B., Leadley, P., Palmer, S.C.F., Pantel, J.H., Schmitz, A., Zollner, P.A., Travis, J.M.J., 2016. Improving the forecast for biodiversity under climate change. Science 353, aad8466. doi:10.1126/science.aad8466

Van Horne, B., 1983. Density as misleading indicator of habitat quality. J. Wildl. Manage. 47, 893–901.

Viljur, M.-L., Teder, T., 2016. Butterflies take advantage of contemporary forestry: Clear-cuts as temporary grasslands. For. Ecol. Manage. 376, 118–125. doi:10.1016/j.foreco.2016.06.002

Żmihorski, M., Berg, Å., Pärt, T., 2016. Forest clear-cuts as additional habitat for breeding farmland birds in crisis. Agric. Ecosyst. Environ. 233, 291–297. doi:10.1016/j.agee.2016.09.023

Discussion

The overall objective of this PhD-thesis is to provide evidence of the usefulness and applicability of species distribution models (SDMs) in conservation projects, understanding in which situations these models can be confidently used as tools to support decision-making (Guisan *et al.* 2013). For this aim, we conducted two studies (articles 1 and 2) analysing the usefulness of species distribution models (SDMs) to predict the distribution of the species and two other studies (articles 3 and 4) linking model outcomes with demographic parameters, specifically reproductive performance.

Long-term monitoring projects aim to deliver information on the changing status of biodiversity, but they also provide an appropriate source of data to document the distribution of species in space through species distribution modelling. In article 2, we have provided an analytical framework to help in the initial design of monitoring projects able to generate appropriate data to produce accurate SDMs and maps. Our analytical framework using data from large-scale last-generation atlases examined the influence of sample size (number of monitoring plots sampled) on modelling performance, to identify how large the number of sampled plots should be in a monitoring project to derive accurate species distribution maps based on models. This approach may be applied before the start of field data collection when decisions about sampling design are made. Our result of most appropriate number of monitoring plots (4-5% of surface coverage) was nevertheless higher than the coverage implemented in most of existing monitoring programmes worldwide. We proposed a hybrid sampling design were a fixed subset of sampling plots may be repeatedly surveyed within and between years and a roving subset of sampling plots surveyed less frequently over time (Hooten *et al.* 2009; Hooten, Ross, & Wikle 2012). With this hybrid sampling design, different objectives are combined in a single long-term monitoring project: the estimation of species occupancy and detection probability and/or the spatial knowledge for distribution mapping. It is also important to consider whether the initial sampling is suited to document species distribution dynamics and to estimate changes in abundance and occurrence of the species.

An example of this hybrid sampling design has been implemented in Luxembourg for the common bird monitoring project (Titeux *et al.* 2009), consisting in a subset of fixed yearly sampled monitoring plots (representing 1.2% of country's surface coverage) and three subsets of roving monitoring plots (each representing 2.4% of the country's surface coverage) which are sampled every three years. With this design, every three years 8.4% of the surface coverage of Luxembourg is monitored for common breeding birds, giving a great potential for species distribution modelling in order to guide appropriate conservation decision-making. In article 1 we included the yearly sampled, fixed subset of monitoring plots of this sampling design (which we referred to as "general-purpose stratified random sampling") in our comparison of the performance of three sampling strategies to identify important areas for the Red-backed Shrike in Luxembourg. Of the three strategies used, species distribution modelling was the most efficient tool to guide field data collection. Being aware of the existence of some scepticism between the practitioners towards the use of species distribution modelling, we hope to facilitate the involvement of practitioners in the development of SDMs to maximize the robustness of modelling applications in conservation practice and management decision-making. We highlight the idea exposed by Guisan *et al.* (2013) of the need of individuals or institutions as "translators" between modellers and decision makers.

The discussion regarding the ecological interpretation of modelling outcomes in terms of demographic parameters is the subject of many articles (Brambilla &

Ficetola 2012; Pellissier *et al.* 2013; Bean *et al.* 2014), and it is directly linked with the type of data used in building the model itself (species occurrence and environmental variables data) and the basic assumptions linked in the modelling process.

In article 3, we have analysed the degree in which SDM outcomes can be used as appropriate predictors of habitat quality for the reproduction of bird species. Our results showed that SDMs may help to predict habitat quality for the reproduction of some species, although results were not consistent for all species at all scales. We highlighted that caution is needed when interpreting modelling outcomes, especially if they are used to guide nature conservation and management actions.

Based on the findings of Hollander *et al.* (2011), who have shown the existence of an ecological trap in the Red-backed Shrike in southern Belgium, in article 4 we developed further on the link among habitat use, habitat quality and reproduction performance. The results of our article show that, in this context, the models predicted a higher habitat quality for the shrikes in the trap habitat, where individuals had a reduced reproductive performance, highlighting a mismatch between modelling outcomes and shrike reproductive performance. Therefore, we emphasize again that caution is needed when interpreting SDMs outcomes to guide spatial prioritization of management or conservation.

It is now warranted to evaluate if mechanistic modelling approaches that explicitly link spatial distribution data to demographic rates of the populations or behavioural traits of the individuals (e.g. DeAngelis & Mooij 2005; Kearney & Porter 2009; Schurr et al. 2012) offer a promising alternative to correlative modelling in order to prioritise areas across the landscape for the conservation of ecologically trapped organisms. Among the diversity of spatially explicit modelling techniques based on mechanisms, individual-based approaches simulate how population-level patterns emerge from the interactions between individuals and from the variations in individual actions and behaviour (DeAngelis & Mooij 2005). These techniques have the potential to deal with the behavioural motivation of the individuals to select among several options varying in attractiveness of the different habitats.

Regarding our general results, we want to emphasize the importance of the appropriate use of SDM outcomes. On one hand, in our study conducted in Luxembourg with the Red-backed Shrike (article 1), we have concluded that SDM outcomes were a useful tool to guide and identify areas harbouring the species, therefore appropriate for inferring about basic knowledge on species distribution. On the other hand, in the study conducted with the same species a few kilometres farther (article 4), we have concluded that SDM outcomes are not always appropriate to guide on conservation actions, specifically when the species is present in different habitat types. In Luxembourg, the species can mainly be found in meadows and pastures as there are no large spruce forest plantations and therefore there are no large clear-cut areas were the Red-backed Shrike could breed. Our results highlight the relevance of studying each case in detail, clearly setting SDMs objectives and avoiding blind interpretation of SDM outcomes, for the most outstanding results and applications.

SDMs have become a fundamental tool in Ecology and Biogeography and play a key role in supporting decision-making for the conservation biodiversity (Guisan & Thuiller 2005; Franklin 2013; Guisan *et al.* 2013). The increasing number of peer-review scientific papers in ISI Web of Science with keywords such as “species distribution model”, “habitat suitability model” and “ecological niche model” shows that the interest on the subject is still a major issue. However, Guillera-Arroita *et al.* (2015) published a review considering 100 journal articles within the subject and concluded that matching SDM to the application needs is crucial to avoid inappropriate management decisions. In addition, the type of incoming species distribution and environmental data (now improving with remotely sensed data) and their quality fundamentally affect the modelling outcomes.

The ensemble of articles presented in this thesis provide more evidence of species distribution modelling as useful tools related to bird conservation projects. We encourage local conservation-managers to take part in the modelling procedures and to modellers to consider in detail the importance of studying the species. We have shown the potential of SDMs for punctual detection of species and as tools to guide on the design of long-term monitoring projects but also the relevance of paying special attention to the interpretations of modelling outcomes. We believe that our results have opened new perspectives in the understanding of the uses of species distribution models.

REFERENCES

Bean, W.T., Prugh, L.R., Stafford, R., Butterfield, H.S., Westphal, M. & Brashares, J.S. (2014) Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. *Journal of Applied Ecology*, **51**, 1116–1125.

Brambilla, M. & Ficetola, G.F. (2012) Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species. *Journal of animal ecology*, **81**, 781–7.

DeAngelis, D.L. & Mooij, W.M. (2005) Individual-Based Modeling of Ecological and Evolutionary Processes. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 147–168.

Franklin, J. (2013) Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions*, **19**, 1217–1223.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015) Is my species distribution model fit for purpose ? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.

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

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P. & Buckley, Y.M. (2013) Predicting species distributions for conservation decisions. *Ecology Letters*, **16**, 1424–1435.

Hollander, F.A., Van Dyck, H., San Martin, G. & Titeux, N. (2011) Maladaptive habitat selection of a migratory passerine bird in a human-modified landscape. *PLoS one*, **6**, e25703.

Hooten, M.B., Ross, B.E. & Wikle, C.K. (2012) Optimal spatio-temporal monitoring designs for characterizing population trends. *Design and analysis of long-term ecological monitoring studies* (eds R.A. Gitzen, J.J. Millspaugh, A.B. Cooper, & D.S. Licht), pp. 443–459. Cambridge University Press.

Hooten, M.B., Wikle, C.K., Sheriff, S.L. & Rushin, J.W. (2009) Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, **20**, 639–649.

Kearney, M.R. & Porter, W.P. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, **12**, 334–50.

Pellissier, L., Meltofte, H., Hansen, J., Schmidt, N.M., Tamstorf, M.P., Maiorano, L., Aastrup, P., Olsen, J., Guisan, A. & Wisz, M.S. (2013) Suitability, success and sinks: how do predictions of nesting distributions relate to fitness parameters in high arctic waders? *Diversity and Distributions*, **19**, 1496–1505.

Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F., Kissling, W.D., Linder, H.P., Midgley, G.F., Schröder, B., Singer, A. & Zimmermann, N.E. (2012) How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, **39**, 2146–2162.

Titeux, N., Biver, G., Lorgé, P. & Hoffmann, L. (2009) Implementation of the Common Bird Monitoring scheme in Luxembourg. *Bird Census News*, **22**, 42–50.

Conclusion

In this section we answer the questions posed at the beginning of this dissertation and summarize the main conclusions (Fig.1):

Question 1: Are SDMs as useful as experts to improve our knowledge on the distribution of a species of conservation concern?

Yes. Our study explicitly shows that species distribution models can efficiently guide the collection of field data towards important areas for a bird species of conservation concern. Our results may facilitate the involvement of practitioners in the development of species distribution models with the objective of maximising the robustness of modelling applications in conservation practice and management decision-making.

Question 2: How to optimise sampling design in long-term bird monitoring projects so that the data collected in the field could be used to map the distribution of the species?

In article 2, we showed that modelling performance was sensitive to particularly small number of bird monitoring plots and the minimum number of plots required to ensure an acceptable level of modelling performance was strongly related to the prevalence of the species in the sampling plots. On average, the minimum bird monitoring plots was larger in low-prevalence species than in high-prevalence species. Although we argue that monitoring data may be cost-effectively collected and used in species distribution models to document the spatial distribution of the species, the minimum sampling coverage obtained in our study in southern Belgium is higher than the coverage implemented in most of the existing monitoring programmes worldwide. Therefore, we propose a hybrid sampling design to combine objectives: a fixed subset of monitoring plots repeatedly surveyed within and between seasons to estimate species occupancy and detection probability and a roving subset of monitoring plots surveyed less frequently over time to increase spatial knowledge for distribution mapping.

Question 3: Are SDMs reliable to inform on habitat quality for breeding birds at the regional scale?

Not always. In article 3, we have shown that for almost half of the studied bird species there was a positive relationship between predictions of the models and the estimated reproductive performance. Depending on the type of data used to build the models and the link between the spatial distribution and the habitat requirements of the species, it might be risky to use the predictions of the models as direct proxies for habitat quality in the lack of careful species- and scale-specific assessments in the area of interest. The predictions of species distribution models may prove useful but should be interpreted with extreme caution, especially when they are used to guide conservation or management actions that are expected to induce changes in the habitat quality for the species.

Question 4: Are SDMs reliable to inform on habitat quality for an ecologically trapped bird?

No. In our study on an ecologically trapped organism (Red-backed Shrike), we showed a mismatch between the outcomes of a species distribution model based on atlas distribution data, and the reproductive performance of the species. If the outcomes of such species distribution models were used to prioritize areas across the landscape for the conservation of the species, forest would be identified as more important for shrikes than meadows and pastures in farmland. This finding would be opposite to the observed difference in shrike reproductive performance between farmland and forest. Therefore, carefulness is needed before blindly applying species distribution models in an area where a species occupies multiple habitat types and when novel conditions are emerging in the landscape.

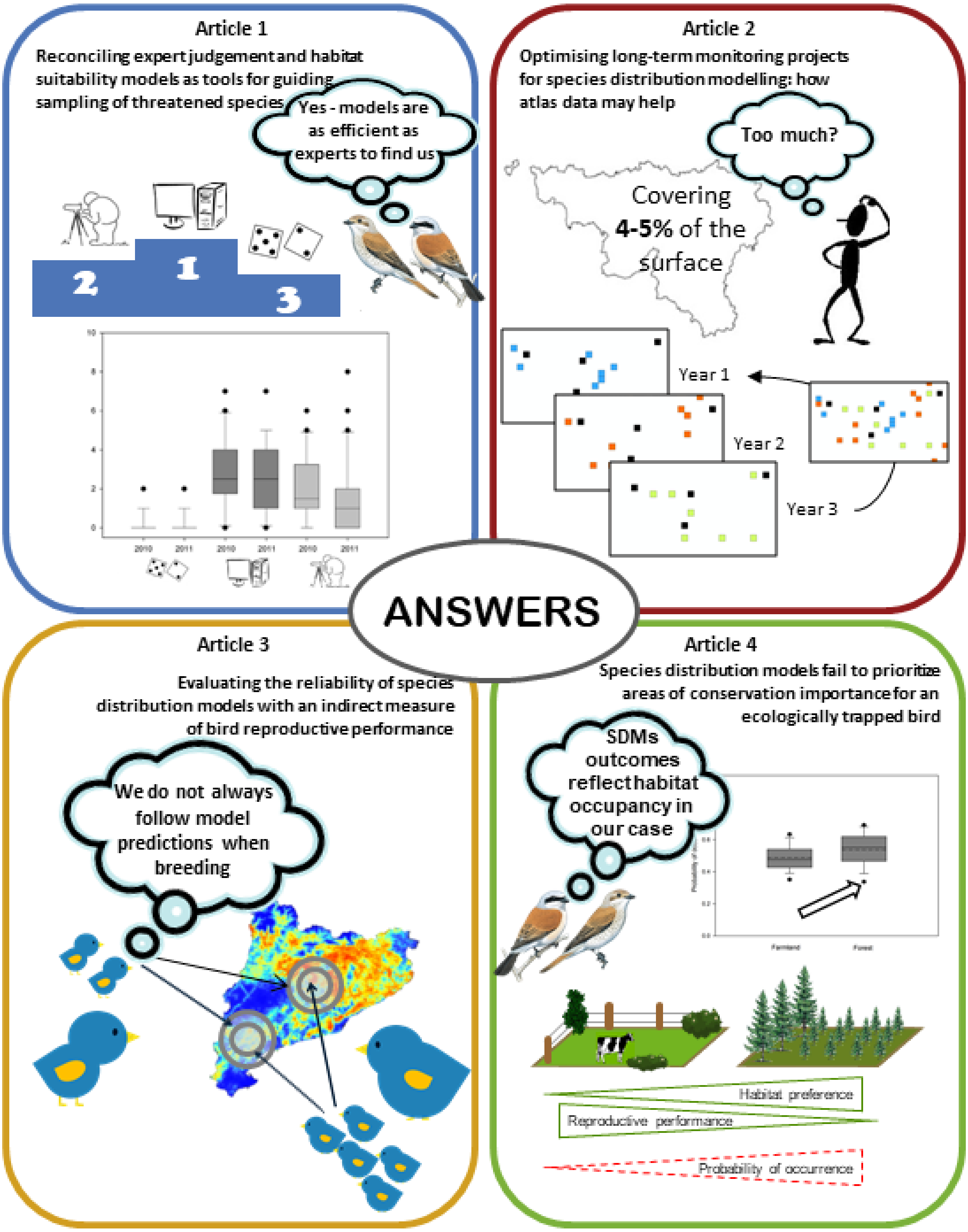


Figure 1. Graphic summary of the main results and conclusions withdrawn from each of the articles presented in this thesis.

Appendix

About the author

Olatz Aizpurua San Roman was born in Pamplona in 1980. During her first 10 years of life she moved with her family through seven different cities and studied in four different languages, which has probably determined her love for languages and travelling. The family settled in Navarra in 1990, where she studied and finished her science-based high school diploma in June 1998.

She studied Biology, specialized in Environment and Agriculture at the University of Navarra in Pamplona. After finishing her second-degree year, she had a great experience as an intern at Odiel marshes (Huelva). There she was first introduced to ornithology and had her first bird ringing and bird tagging experiences. She finished her Biology degree at the University of Southern Denmark thanks to the European Erasmus exchange program in June 2003.

In September 2007, after her Master in Terrestrial Ecosystems in Bordeaux, she moved to Luxembourg. A year later, the Public Research Center Gabriel Lippman, now called LIST (Luxembourg Institute of Science and Technology) hired Dr Nicolas Titeux and opened a Nature conservation and Spatial Ecology team. She had the pleasure to join his team and to start her PhD in April 2009. The PhD thesis was in tight collaboration with the Forest Sciences Centre of Catalonia and the University of Lleida. During her PhD thesis years, she was able to participate in many conferences and training courses.

After finishing her PhD contract with LIST (2014) she moved with her family to Panamá where they lived until 2017. During her tropical experience, she collaborated as an intern in the Smithsonian Tropical Research Institution and worked in the publication of the articles of her thesis.

Nowadays, she is back and settled at home, in the Basque Country, where she tries to find a new equilibrium, enjoying her family and engaging in a new professional career.

Acknowledgements

This adventure started too long ago… so my relief writing this and finishing this is… huge! But, I should start from the beginning to make the acknowledgements properly. I was in Luxembourg, searching a research subject, a research team, an institution and funding. I have to thank in the first place **Nicolas Titeux**. I first contacted him in 2007, asking him for a research project. It took long and we worked hard already before really starting the PhD but we finally made it! In April 2009, I joined his new Nature Conservation team in the Luxembourg Institute of Science and Technology and started the MOBIMO project. I was the first PhD student in the team… and I am also the one "leaving" on last. I have been closely supervised by Nicolas, who took always the time to listen, discuss and help me with my on-going work. The years passed within the team were very enriching scientifically and personally. The years passed "out of the team", have been very different, my life has gone far from research, but Nicolas has always been there to push me and encourage me to keep on going and finishing. Merci beaucoup.

My co-supervisor, **Lluís Brotons** was also there from the very beginning of the story. He gave me Nicolas's contact. Moltes gràcies Lluís for that first hint and for always having the doors open to me at the Centre Tecnològic Forestal de Catalunya. To join the Biodiversity and Landscape Ecology lab was always a good training, enriching experience and fun. Solsona mola!

I also need to thank **Jose Antonio Bonet**, my academic tutor and guide in all the administrative changes related with the University of Lleida, especially when the PhD plan changed, and I needed someone to personally go and ask for me, he was there! Gracias!

All three of them, have been available for any scientific or administrative question timely by mail. The communication has been always easy despite the geographical distance. Being part of these two-research groups has been a great opportunity to learn and develop plenty of skills.

I am also grateful to the **Fonds National de Recherche du Luxembourg** (FNR-AFR-PHD 08-063) for their funding within the Aides à la Formation Recherche program. The working contract and the conditions were always favorable to me. I also have to thank the **Luxembourg Institute of Science and Technology** and

Lucien Hoffmann for making possible the development of my work in his department in LIST.

I really enjoyed being part of the Nature Conservation team, we were not many and we had a very friendly atmosphere: **Adeline Gillet**, **Hubert Baltus**, **Alain Dohet**, **Lionel L'Hoste** and **Gérard Schmidt** thanks for those lunch periods and dinners. I am a team-person, that is for sure! From the team, I most closely worked with **Youri Martin**, **Xavier Mestdagh** and **Lisette Cantú-Salazar**. Youri and I stood back-to-back during many years, we discussed many R questions together and he answered plenty of times "I don't drink coffee" to my invitations for a break Thank you for making those years great! Lisette, many thanks for reading and commenting all these "almost" finished versions of the thesis.

Many people have helped me during my fieldwork in Luxembourg, during 2009, 2010 and 2011. **Franck Hollander** shared the best tricks to detect the shrike in the field. In 2009, **Christophe Bouillon** and **Dido Gosse.** Jefa! Dido, the first inter of the team. We spent many hours in the field, driving, following red-baked shrikes and became very good friends. I guess that the friendship was the best result from that first field work season! (for sure it was not the data). In 2010 and 2011, many more participated in the field data collection: **Lionel L'Hoste, Alain Dohet, Doriane Müller, Huber Baltus, Joseph Dunlop, Jean-Phillippe Rolin, Jenny Renault, Xavier Mestdagh** and **Nicolas Titeux.** Thank you very much to all of you! I would have not been able to collect all the data without your help!

Many other people from the Luxembourg have made this thesis possible, especially those that got me into the Letzebuerguesch vull -a emwelt: **Jim Schmitz**, **Guy Mirgain**, **Ed Melchior**, from whom I learnt a lot about bird rigging. **Gilles Biver** and **Patric Lorgé** who were always collaborative and helpful.

Those years in Luxembourg, where also great thanks to my office mates, the first filling of the office: **Anouk Kaulmann**, **Eric Biehler**, **Moustafa Selim**, **Maria Helena Novais**, **Carole Molito**r, later other new students came, thanks to all! A special mention to **Mauricio Montaño**, that through our "window" helped me in many aspects of my thesis and life.

Luxembourg has been my home during many years, the first few years were a bit hard, but joining LIST made me meet a really good and fun group of friends.

Too many to mention all of you: **Dasa, Mara, Mayerling, Catalina, Núria, Laurent, Javi, Maria, Dave, Natasa, Fabian**, … all of you, villmools merci!

In my other team, the one in Solsona, I also should thank many that either helped me directly with my work and those that made my stays nice and fun: **Dani Villero**, **Magda Pla**, **David Giralt**, **Francesc Sardà-Palomera**, **Miquel de Cáceres**, **Assu Gil-Tenna**, **Núria Aquilué**, **Elena Lopez-Zozaya**, **Toni Llongarriú** and **Laia Espasa.**

Later, during the years post-Luxembourg, in the other side of the world, the work on my thesis slowed down, but I still need to thank people that someway influenced and made it possible for me to work on it and finish it. **Julia** and **Jonathan** for getting me into STRI community, this was great working ambience! I also need to thank **Ekhiñe** and **Paula** that helped me getting time to dive back into my thesis.

Special thanks to all of those that are not mentioned, to those that are around me always, no matter in which continent I am, those that deserve to be in capital letters. This is not needed as you are all always in my heart. :P Special mention to **Adriano**, that saved me from a big R analysis during a tropical Christmas visit. Thanks also to all of you that have not asked if I had finished my thesis and to those that have actually dared to ask when nobody did ;)

This work is dedicated to my dearest, to my mother, who left us during these years. To my dad and sister who helped me when I asked. To my children, that were born during these years, Eneko in 2012 and Oihane in 2016, and specially to Jean-Marc, who has been with me every day of all this process, supporting and encouraging me in every good and bad moment. This thesis would not have been possible without all your support and love. Eskerrik asko jota y me!

Publications

Aizpurua, O., Paquet, J.-Y., Brotons, L. & Titeux, N. (2015) Optimising long-term monitoring projects for species distribution modelling: how atlas data may help.

Ecography, **38**, 29–40. Impact factor (2015): 5.355

Aizpurua, O., Cantú-Salazar, L., San Martin, G., Biver, G., Brotons, L. & Titeux, N. (2015) Reconciling expert judgement and habitat suitability models as tools for guiding sampling of threatened species.

Journal of Applied Ecology, **52**, 1608–1616. Impact factor (2015): 6.02

Aizpurua, O., Cantú, L., San Martin, G., Sardà-Palomera, F., Gargallo, G., Herrando, S., Brotons, L. & Titeux, N. (2016) Evaluating the reliability of species distribution models with an indirect measure of bird reproductive performance.

Journal of Avian Biology, **48**, 1575-1582. Impact factor (2015): 2.192

Titeux, N., **Aizpurua, O.**, Hollander, F., Sardà-Palomera, F., Hermoso, V., Paquet, J.-Y., Mestdagh, X., Settele, J., Brotons, L. & Van Dyck, H. (*under review* in Biological Conservation) Ecological traps and species distribution models: a challenge for prioritizing areas of conservation importance.

Biological Conservation. Impact factor (2014): 3.762

Training during PhD thesis

2009
Curs practic d'introducció als sistemes d'informació geogràfica
September 24th – October 15th Solsona (Spain)
Universidad de Lleida - Centre Tecnològic Forestal de Catalunya

2010
Curso avanzado de herramientas informáticas para la gestión de información ambiental: sistemas de información geográfica Arc Gis 9.1
August 9th – 13th Pamplona (Spain)
Universidad de Navarra – Delegación de La Rioja – Navarra del Colegio Oficial de Biólogos

Introduction to R
November 22nd Luxembourg (Luxembourg)
University of Luxembourg – Lucilinx

2011
Introducció a la modelització especial de la biodiversitat
February 17th – 18th Solsona (Spain)
Institució Catalana d'Història Natural – Centre Tecnològic Forestal de Catalunya

Project Management for Research – a one-day programme
April 11th Luxembourg (Luxembourg)
Fast Training – Fonds National de la Recherche Luxembourg

Scientific presentation skills training
April 15th – 16th Merzig (Germany)
Centre de Recherche Public Gabriel Lippmann – CEB Merzig

Análisis de datos ecológicos en R
September 19th – 23rd Granada (Spain)
CEAMA – Fundación General UGR-Empresa

2014
Bayesian population analysis using BUGS and JAGS
January 13th -17th Zürich (Switzerland)
University of Zürich

Conferences during PhD thesis

2010
Bird Numbers 2010 – 18th Conference of the European Bird Census Council
March 22nd – 26th, Cáceres (Extremadura, Spain)
Poster presentation: Application of niche-based modelling techniques to long-term common bird monitoring data

2011
Luxembourg Life Science PhD Day 2011
September 14th – Luxembourg (Luxembourg)
Oral presentation: Application of niche-based modelling techniques to long-term common bird monitoring data

2013
Luxembourg Life Science PhD Day 2013
September 10th – Luxembourg (Luxembourg)
Poster presentation: Using expert opinion to evaluate usefulness of habitat suitability models to find new presence locations of a nearly threatened species

Bird Numbers 2013 - 19th Conference of the European Bird Census Council
September 17th – 21st, Cluj-Napoca (Rumania)
Poster presentation: Using expert opinion to evaluate usefulness of habitat suitability models to find new presence locations of a nearly threatened species
Oral presentation: Optimizing long-term monitoring projects for species distribution modelling

19th Biennial International Ecological Modelling Conference
October 28th – 31st, Toulouse (France)
Oral presentation: Evaluating reliability of habitat suitability models with indirect measures of bird survival and reproductive performance data

2016
Bird Numbers 2016 - 20th Conference of the European Bird Census Council
September 5th – 9th, Halle (Germany)
Oral presentation: Species distribution models misinform on habitat suitability in an ecologically trapped bird

Fonds National de la
Recherche Luxembourg

Supported by the Fonds National de la Recherche
Luxembourg (PHD-08-063)