KNOWLEDGE ORIENTED
SPECIES DISTRIBUTION MODELLING

Aidin Niamir

PhD dissertation committee:

Chair

Prof. dr. ir. A. Veldkamp	University of Twente

Promoters

Prof. dr. A. K. Skidmore	University of Twente
Prof. dr. R. Real Giménez	University of Malaga

Assistant promoter

Dr A. G. Toxopeus	University of Twente

Members

Prof. dr. ir. W.H. van der Putten	Wageningen University
Prof. dr. T. Hickler	Goethe-University Frankfurt
Prof. dr. ing. W. Verhoef	University of Twente
Prof. dr. V.G. Jetten	University of Twente
Dr A. R. Muñoz Gallego	University of Malaga

ITC dissertation number 250
ITC, P.O. Box 217, 7500 AE Enschede, the Netherlands

ISBN: 978-90-365-3687-5
DOI: 10.3990/1.9789036536875

Printed by ITC Printing Department
© Aidin Niamir, Enschede, the Netherlands, 2014

UNIVERSITY OF TWENTE.
ITC
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

KNOWLEDGE ORIENTED
SPECIES DISTRIBUTION MODELLING

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Friday 13 June 2014 at 12:45 hrs

by
Aidin Niamir
born on May 22nd, 1980
in Tehran, Iran.

This dissertation is approved by

Prof. dr. A. K. Skidmore, promoter
Prof. dr. R. Real Giménez, promoter
Dr A. G. Toxopeus , assistant promoter

Preface

The outcome of my master project was the beginning of my PhD research path. The results of the project had fully satisfied the statistical performance measures; however it was always subjected to doubts and questions by biologists' community. In that project my intention was to use available - though scarce - species observation records along with rigorous statistical modelling techniques to understand the habitat preference, and to map the distribution of a poorly-studied raptor in southern Spain. After presenting the final results to the local ornithologists, I realised that the produced maps were not consonant with their expectations, though the maps fully satisfied statistical measures. They could come up with examples which disapproved my findings, or could describe situations that would significantly contribute to the quality of the maps, if I had known them earlier. These encounterments were the defining moments that made me realize the gap between widely used inductive approaches and expert knowledge in habitat studies and species distribution modelling. I recognized that incorporation of knowledge into the novel distribution models has not been yet well-studied.

At this moment biodiversity conservation requires reliable information and comprehensive understanding regarding species interaction with their surrounding environment more than ever. Expert knowledge can be a vital and trustworthy source of information as an alternative for conservation management decisions, when inadequate data make the inductive models unreliable. There are many locations, for example my own country Iran, that are scars in systematic species observation data, but there are experts and local people who hold valuable knowledge in that regard, gathered from their personal experiences and observations. Harnessing this knowledge and put to use next to computational advances and statistical developments can be a solution for biodiversity conversation in these situations. My aim in this research was to develop a framework to

incorporate expert knowledge into species distribution modelling. I hope that the outcome of my study contributes in bridging the above mentioned gap and becomes a steppingstone in slowing down the biodiversity deterioration.

This dissertation is the result of my PhD project which was jointly based at the Department of Natural Resources, Faculty of Geo-Information Science and Earth Observation, University of Twente[*] (Enschede, the Netherlands) and the Department of Animal Biology, Faculty of Science, University of Malaga[†] (Malaga, Spain). The project was jointly supervised by Professor Dr Andrew Skidmore (Enschede) and Professor Dr Raimundo Real (Malaga). The project was partly funded by European Erasmus Mundus Program[‡], administered by the University of Twente, and performed at the two work places mentioned above. Foundation Migres[§] (Algecira, Spain) kindly hosted the fieldworks and surveys in 2009 and 2010.

Aidin Niamir

Enschede, March 2014

[*] http://www.itc.nl/nrs

[†] http://www.uma.es/departamentos/info/4438/departamento-biologia-animal/

[‡] http://eacea.ec.europa.eu/extcoop/call/index.htm

[§] http://www.fundacionmigres.org/

Acknowledgements

After five years I am finally seeing the end of my PhD road. An achievement only made possible with the help of family members, friends and colleagues all of whom I am deeply grateful to.

I would like to thank my supervisors at ITC, Andrew Skidmore and Bert Toxopeus who gave me the opportunity to pursue this PhD in the first place and offered their help and advice whenever I needed. Andrew, you taught me the principal of parsimony in science. You always advised me not to look for complex explanations when a simple one will do; 'thanks for everything I have learned during the past five years'. I really enjoyed the first half of every single meeting discussing politics, economy and culture, that widened my view to life. Bert, what can I say? Since I walked in to your room for the first time in 2005 to check those poor cross-eyed stuffed birds on the wall of your office, you were always more than a supervisor to me, you were my friend. You kept encouraging me and consistently insisting that everything will be fine while I was not very optimistic. Fortunately, you were right and I was wrong.

My supervisors in Malaga, Raimundo Real and Antonio Munoz deserve a big "muchas gracias". Throughout my PhD project, I spent more than 10 wonderful months in Andalucía. Thank you Raimundo, for kindly hosting me in the visits. You taught me the philosophy of science. My work greatly benefited from your conceptual ideas and your critical feedback, for which I express my sincere gratitude. I would like to thank Antonio for all his collaborations not only during my PhD, but ever since I started my Master project in 2008. You kindly connected me to a large number of ornithologists in Spain. More importantly, you taught me the priorities in life and how family comes first.

During my PhD, I was lucky to get the chance of meeting a big crowd of fantastic people in Enschede and Malaga. These people deserve a huge

"Thank You" because they made me feel home and comfortable where I was. First, I would like to thank Babak Naimi and Sam Khosravifard, with whom I had spent a lot of time in Enschede. Babak, I learned a lot from our endless scientific and in-depth discussions. I do appreciate your critical advices and valuable help. My friend Sam, I enjoyed every single discussion we had over lunch. You helped me to learn how to present science to the community. Thank you both for the great company. In the early days of my PhD, collaborations with David Rossiter and Thomas Groen helped me to find my way around geostatisitics. David, you directed me to open-source statistical computing, you actually taught me to fish. I am also grateful to Alexey Voinov for his honest and realistic comments when I was expressing my research ideas and concerns over our lunch meetings.

Life at ITC was made a lot easier; by Loes Colenbrander who took care of my graduate program from very beginning to the graduation ceremony, by Esther Hondebrink who dealt with all meeting arrangements, and departmental affairs, by Theresa van den Boogaard, Bettine Geerdink, and Marie Chantal Metz who dealt with all the bureaucracy that comes along with an international PhD student, and by Marion Pierik who did her best to put down all the financial anxieties. I am grateful to the staff of ITC library. I had a privilege of having Marga Koelen, Carla Gerritsen, and Desiree Snoek in the library making sure that I unrestrictedly access research materials, journal papers, books and even Dutch languages tests. I am thankful to the ITC helpdesk, and facility services that spoiled me with their great support. Special thanks to Aiko Mulder for preparing me spare laptops in a very short notice. I also would like to thank Roelof Schoppers for happily greeting me every morning in the past five years, and Benno Masselink and Job Duim for their support and dealing with my variety of requests.

I am grateful to the entire Biogeography, Diversity and Conservation Research Team of the University of Malaga. Special thanks to Jesús Olivero, Ana Luz Márquez, Pelayo Acevedo, David Romero, and Alberto

Jiménez Valverde for your great company. I felt welcome in the team from the very beginning. There is a working environment that one would wish to be part of it and this is due to the great atmosphere which is built and maintained by its people.

I would like to also express my appreciation and love to people who have always been there for me. My wonderful parents have always supported me in achieving my goals. Jaleh and Iraj, I always felt that I am standing on your strong shoulders and nothing can slow down my progress. You are the greatest mum and dad I could ever wish for. My sister Leila, you gave me pure love endlessly. Your "Sure you can!" motto always made me feel stronger.

The person who has shared her life with me since the early days of my PhD and certainly had to suffer the most through my ups and downs was Maral. My dear Maral, you have brought happiness to my life and with you I am the luckiest man in the world. I love you and I am looking forward to see what life will bring us next.

Cover caption:
Bonelli's Eagle (*Aquila fasciata*)
Photo: Juan Luis Muñoz

Table of Contents

Preface i
Acknowledgements iii
Chapter 1: GENERAL INTRODUCTION 1
1.1 Introduction 3
1.2 Species distribution modelling 5
1.3 Expert knowledge 10
1.4 Proposed approach 12
1.5 Research objectives 13
1.6 Outline of the dissertation 14
1.7 The study area 15
Chapter 2: FINESSING ATLAS DATA FOR SPECIES DISTRIBUTION MODELS 17
2.1 Abstract 19
2.2 Introduction 20
2.3 Methods 22
2.3.1 Study area 22
2.3.2 Species occurrence data 23
2.3.3 Target Species 24
2.3.4 Expert layers and rules 27
2.3.5 Explanatory variables 29
2.3.6 Bayesian expert system 29
2.3.7 Hybrid method 31
2.3.8 Evaluation 32
2.4 Results 33
2.5 Discussion 37
Acknowledgments 41
Chapter 3: USE OF TAXONOMY TO DELINEATE ENVIRONMENTAL EXTENT FOR SPECIES DISTRIBUTION MODELLING 43
3.1 Abstract 45
3.2 Introduction 46
3.3 Methods and Materials 48
3.3.1 Species data 48
3.3.2 Taxonomically delimited species occurrence datasets 49
3.3.3 Species distribution models 52

3.3.4 Model evaluation 53
3.4 Results 54
3.5 Discussion 60
Acknowledgments 63
Chapter 4: ACCOUNTING FOR KNOWLEDGE UNCERTAINTY IN SPECIES DISTRIBUTION MODELLING 65
4.1 Abstract 67
4.2 Introduction 68
4.3 Explanation of Dempster-Shafer theory of evidence 70
4.4 Experimental settings 75
4.4.1 Study area 75
4.4.2 Choice of species 75
4.4.3 Choice of experts 77
4.4.4 Environmental predictors 77
4.4.5 Conventional species distribution models 77
4.4.6 Performance measures 78
4.5 Using Dempster-Shafer evidence theory 79
4.6 Results 81
4.6.1 Assignment of belief functions 81
4.6.2 Combination of belief functions 85
4.6.3 Performance of the inductive models 90
4.6.4 Variability in knowledge domain 90
4.7 Discussion 92
Acknowledgements 96
Chapter 5: KNOWLEDGE ORIENTED SPECIES DISTRIBUTION MODELLING; SYNTHESIS 97
5.1 Introduction 99
5.2 Revisiting the research objectives 102
5.2.1 Finessing spatial quality of species occurrence data 102
5.2.2 Delimiting environmental extent 103
5.2.3 Accounting for self-acknowledged ignorance 105
5.3 Lesson learned 106
5.3.1 Broad application of expert knowledge 106
5.3.2 Method for expert knowledge elicitation 107
5.3.3 Expert knowledge has more to offer 108
5.4 Recommendations 109
5.4.1 Tools to support knowledge elicitation 109

5.4.2 Critically evaluate expert knowledge 110
5.4.3 Forwarding species distribution models in time 111
Appendix I: EXAMPLES OF SPECIES DISTRIBUTION MAPS OBTAINED USING ATLAS DATASET AND TAXONOMICAL EXTENTS 113
Appendix II: KNOWLEDGE ELICITATION ONLINE QUESTIONNAIRE 117
Bibliography 125
Summary 149
Samenvatting 151
Biosketch 157
Publications 158
PE&RC Training and Education Statement 160
ITC Dissertation List 162

Chapter 1:

GENERAL INTRODUCTION

species
Elith
particular
occurrences
Experts
ignorance
dissertation
variables
source
concepts
science
theoretical
presence
information
ecological
approaches
study
relevant
explanatory
records
Mediterranean
data
results
spatial
Soberón
complex
experts
different
absence
Peterson
modelling
environmental
uncertainty
expert
statistical
background
techniques
use
using
conservation
area
Species
sampling
Expert
provide
natural
interactions
may
used
available
Jiménez-Valverde
model
Spain
Knowledge
Austin
occurrence
locations
distribution
methods
approach
Geisan
research
models

1.1 Introduction

Species distribution models are defined as models that associate species occurrence at known locations with information on the environmental characteristics of those locations (Elith & Leathwick 2009). The spatial prediction of species distribution in unknown locations has been recognized as a fundamental component of conservation planning (Guisan & Zimmermann 2000; Austin 2002; Elith *et al.* 2002). These models have been widely employed to *understand* the relationship between a species and its biotic and abiotic environment; they are based on observations to test ecological or biogeographical hypotheses about species distribution and range. Species distribution models have also been used to *predict* the occurrence of a species, either for certain locations or at times where survey data were lacking or unobtainable (Franklin 2013). There is a variety of terms to address species distribution models with different emphases and implications: bioclimatic models (Heikkinen *et al.* 2006; Jeschke & Strayer 2008), ecological niche models (Peterson 2003; Soberón & Peterson 2005; Soberon & Nakamura 2009), habitat models (Luck 2002), and resource selection functions (Boyce & Mcdonald 1999).

Modern quantitative modelling of species distribution emerged when advances in physical geography, concepts from field-based ecological studies, and developments in statistics and information technology coincided. Every day huge amounts of distributional data including sampling and tracking, as well as museum and herbarium records, are made available in global and national biodiversity databases (Soberón & Peterson 2004). In response, a wide variety of modelling techniques has been specially designed and enhanced to accurately model species distribution using the available occurrence data (Jiménez-Valverde *et al.* 2008). Published examples indicate that species distribution models can characterize the distribution of species to a satisfactory level when well-surveyed species occurrence data and relevant explanatory environmental variables are analysed with a well specified algorithm. In contrast,

applications that extrapolate models in time or space, or are trained based on inadequate species observation data or explanatory environmental variables, are more challenging and the results are more equivocal (Elith & Leathwick 2009).

An adequate species dataset consists of records describing known occurrences (presence), as well as records where the species was not observed (absences). Radiotelemetry data collected in wildlife studies as well as museum records, which are now utilized based on the millions of records compiled in digital form from natural history collections (Graham *et al.* 2004a; Soberón & Peterson 2004), are lacking information about absences of species. Therefore, rigorous species distribution methods that require pseudo-absence data along with presence data as input are proposed to compensate for this shortfall. Contrary to popular belief, there are very few methods that only require presence data (Barbet-Massin *et al.* 2012). So-called presence-only methods such as MaxEnt (Phillips *et al.* 2006) or GARP (Stockwell 1999) actually do require the use of pseudo-absence or background data. Indeed, the value of such methods could be of particular importance, but there is a danger of developing species distribution models based on a variety of statistical assumptions without a conceptual ecological background (Jiménez-Valverde *et al.* 2008), as false absence data can have negative effects on species distribution models (Gu & Swihart 2004). An increasing number of recently published studies are questioning the theoretical aspects of distribution models and are discussing a reflection of the conceptual basis that underlies model practices (Soberón & Peterson 2005; Araujo & Guisan 2006; Real *et al.* 2006; Austin 2007; Jiménez-Valverde & Lobo 2007; Jiménez-Valverde *et al.* 2008; Lobo 2008; Kearney *et al.* 2010).

Expert knowledge is commonly used to generate and evaluate hypotheses, sample design, model development, and interpretation of results (Fazey *et al.* 2005). It forms an authentic source of information where empirical data are scarce or unaffordable. Expert knowledge can be the result of training, research, or personal experience and observation (Burgman *et al.* 2011).

Expert knowledge is also utilized in conservation science (Martin *et al.* 2012).The need to characterize dynamic and complex interactions between a species and its biotic surroundings, limited resources to collect new empirical data, and the urgency of conservation decisions (Sutherland 2006; Kuhnert *et al.* 2010) are the main reasons for incorporation of expert knowledge in conservation science, in particular species distribution modelling. In this dissertation we argue that the use of expert knowledge in species distribution modelling may converge theoretical concepts and advance statistical techniques. The outcome of knowledge oriented species distribution models versus conventional approaches as a proxy to reflect on the theoretical concepts that underlie species distribution modelling methodologies will be discussed. Alternative knowledge oriented approaches that outperform the most conventional agreed upon settings in species distribution modelling are also provided.

1.2 Species distribution modelling

Literature that provides comprehensive reviews on species distribution modelling include Guisan & Zimmermann (2002), Guisan & Thuiller (2005), Elith et al. (2006), Richards et al. (2007), Elith & Leathwick (2009), and Franklin (2009). Here we briefly present species distribution modelling using the framework adapted from Austin (2002). This framework consists of four parts: ecological, data, statistical, and graphical models (Figure 1.1).

- The ecological model encompasses ecological concepts or biogeographical hypotheses tested in the study.
- The data model "consists of the decision made regarding how the data are collected and how the data will be measured or estimated" (Austin 2002).
- The choice of statistical techniques, evaluation methods and experimental settings form the statistical model.

The graphical model includes visualization and graphical representation of the models

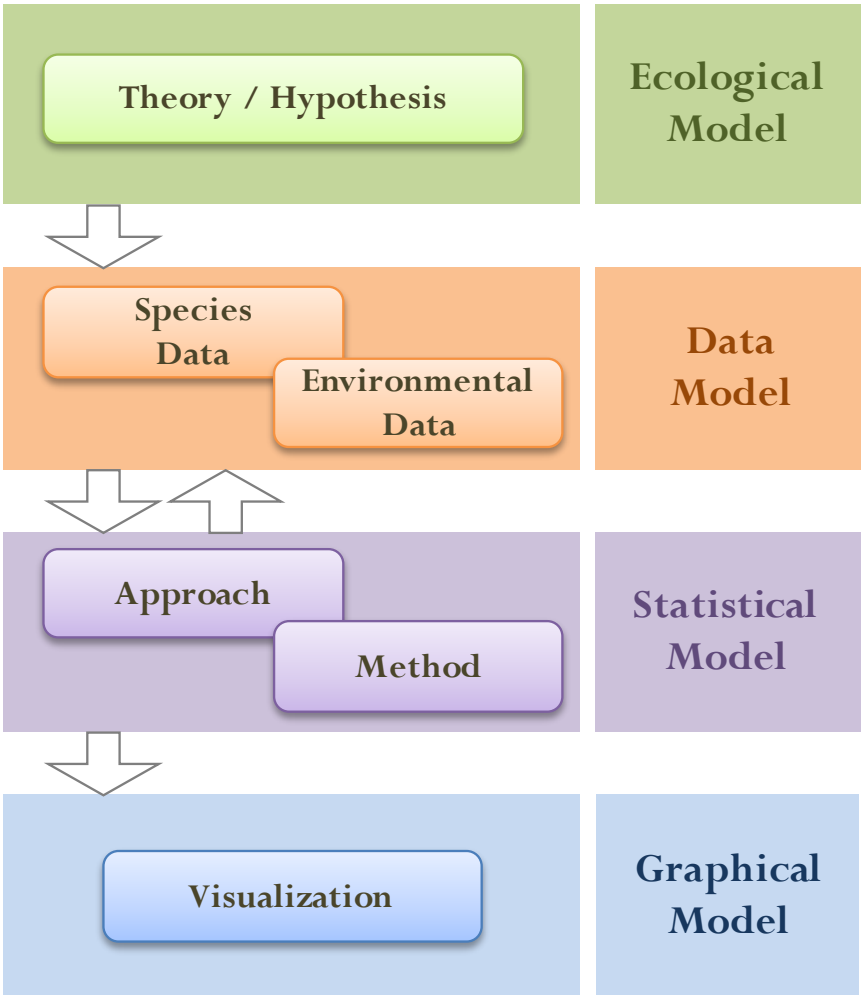


Figure 1.1 Classic components of species distribution modelling

Developing a species distribution model starts with a theoretical or conceptual model of the abiotic and biotic (De Araujo *et al.* 2014) factors affecting species occurrences in space and time. Hypotheses generation, using the physical environment preferences of species and species interactions with other species in competition or coexistence, form a primary step. Collecting adequate data on species occurrences and environmental explanatory variables comes next. Species occurrence data can be collected through a systematic sampling scheme, or obtained directly – where available – from biodiversity data bases. Species interactions with their environment are often complex and only some of the relevant explanatory variables can be translated into this spatial representation. Sometimes environmental preference and habitat favourability of species can be determined using expert knowledge. In the next step a function is derived using either the correlation between species occurrences and the environmental variables, or expert knowledge. The model will then be applied to the values of the environmental variables to predict the favourability of unobserved locations. A map of species occurrence can be produced using the predicted favourability.

Diverse factors operating dynamically with a variety of intensities at different scales (Pearson *et al.* 2002; Pearman *et al.* 2008; Gaston & Fuller 2009) determine the distributional area of a species, which is a complex expression of its ecology and evolutionary history (Brown 1995). Soberón and Peterson (2005) have defined three classes to assist in determining the location of a species; abiotic conditions, biotic factors, dispersal accessibility and evolutionary capacity. They further discuss the idea of using the **MAB** diagram (Figure 1.2). If a species is present (shown by a yellow star in the diagram) at a given location the following conditions have been met: abiotic conditions must be favourable (**A**), an appropriate suit of species is present and absent (**B**), and the area is within reach for the species from established distribution areas in ecological time (**M**). For further information please see Soberón & Peterson (2005) and Soberón & Nakamura (2009).

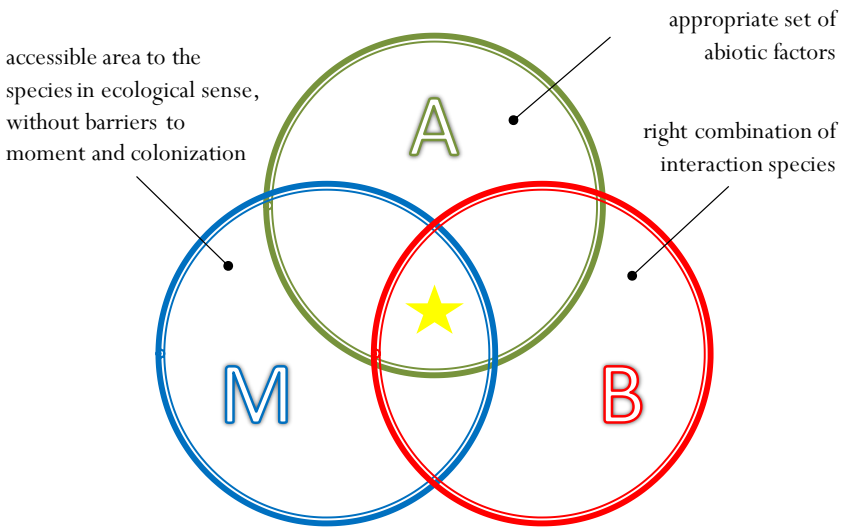


Figure 1.2 MAB diagram (adopted from Soberón & Peterson 2005)

Regarding the **M** factor some argue that absence records indicate whether the sampling site was unsuitable or suitable but unoccupied, perhaps because of inaccessibility (Jiménez-Valverde *et al.* 2009). However, other external factors (e.g. anthropogenic factors) may also result in the absence of a species from its suitable habitat or its presence in an unsuitable ones (Pulliam 2000). What we actually can derive from species observation records and statistical models is the realized, but also with some caution the potential, geographical distribution of species (Araujo & Guisan 2006; Peterson *et al.* 2007; Kearney *et al.* 2010). Potential distribution of a species refers to the places where a species could live, while realized distribution refers to the places where a species is actually living. Calibration and validation of a species distribution model to represent either of these two concepts need adequate species occurrence data (Jiménez-Valverde *et al.* 2008) on both presence and absence of species. Complicated statistical techniques have the potential to discriminate between species presence locations and unknown locations, hence edge

closer to a realized distribution of a species. Notwithstanding that their evaluation will be biased toward higher discrimination capacity (Elith *et al.* 2006; Tsoar *et al.* 2007).

The primacy of such complicated machine learning techniques over simpler linear approaches for species potential distributing modelling remained unproved. Almost all species distribution models actually do require data on both presence and absence of species. There is ongoing discussion on how to best model potential distribution of a species when reliable absence data are not available. There are several so-called presence-only species distribution models that use presence data against randomly sampled (i.e. pseudo-absence) or numerous (i.e. background) unknown locations.

Methods of detecting and dealing with absence locations can have an effect on results and the ecological interpretation of outputs (Lutolf *et al.* 2006; Niels Raes 2007; Ortega-Huerta & Peterson 2008; Phillips *et al.* 2009; Aranda & Lobo 2011; Beaugrand *et al.* 2011). Some acknowledge that the models would be ecologically meaningful and interpretative if presence-absence data were input to the modelling process (Kent & Carmel 2011). Another advantage of the availability of presence-absence data is that it enables analysis of bias, goodness-of-fit (Jiménez-Valverde 2014), and prevalence (Phillips *et al.* 2009; Lawson *et al.* 2014). However, this will have notable impact on the type of data that needs to be collected. Contrary to records describing known occurrences (presence), reliable absence data is rare and often unaffordable.

Often there are spatial correlations among and within environmental variables that violate the assumption of independency in most standard statistical procedures (Legendre 1993b). Spatial structure of environmental variables can be measured using the variance inflation factor as a measure of multicollinearity, and the Moran's *I* statistic as a measure of univariate spatial autocorrelation. However, decisions regarding eliminating correlating variables remain with the modeller. Selecting a relevant and ecologically informative set of variables is fundamental in species

distribution modelling, and interpretation of models may vary with different environmental explanatory variables.

Problems related to uncertainty in species distribution modelling are difficult to deal with and are often ignored (Elith & Graham 2009). This is largely because it is difficult to quantify errors, and the problem seems overwhelming. Although there have been a number of reviews on, as well as theoretical ways to treat, uncertainty sources in species distribution modelling (Elith *et al.* 2002; Rocchini *et al.* 2011), little has been done in practice to illuminate the actual effect of likely uncertainties on modelled predictions. Rocchini (2011) highlighted the need for a "map of ignorance" to provide an overview where the reliability of predictions is unknown. Uncertainty in species distribution models mostly results from the input data, i.e. inadequate species occurrence, bias in sampling, geographical position accuracy, and species misdetection (Heikkinen et al. 2006; Hortal et al. 2008; Heikkinen et al. 2012). This topic has received significant attention in recent years (Orton & Lark 2007; Refsgaard *et al.* 2007; Buisson *et al.* 2010; Luoto *et al.* 2010; Lukey *et al.* 2011; Rocchini *et al.* 2011; Synes & Osborne 2011; Mcbride *et al.* 2012a; Naimi *et al.* 2013). The use of pseudo-absence and background as a substitute for absence locations is considered to be another source of uncertainty. The geographical extent of background locations or pseudo-absence sampling, affects the output of the modelling practices; the results may vary between different sampling strategies (Acevedo *et al.* 2012).

1.3 Expert knowledge

An expert in our view is either a scientist, who researches and formally publishes their knowledge, or a practitioner, who applies scientific knowledge but does not necessarily conduct research or publish his knowledge. An expert may be specialist on a specific topic (for example, an authority in the migration sea birds), or a skilled individual (for example an expert on identifying raptors or tracking carnivores), or a sage practitioner

(for example an experienced Griffon Vulture biologist) (Perera *et al.* 2012). Experts are often well-known to their peers for their skill and knowledge (Choy *et al.* 2009), therefore peer-recognition may be a reliable initial filter for identifying experts.

Expert knowledge is implicit by nature. It is typically not documented and remains tacit until its expression is demanded. Experts express their knowledge in different forms, thus it may appear fragmented and properties such as uncertainty may remain unassessed (Forsyth 1984; Burgman *et al.* 2011; Perera *et al.* 2012). The value of expert knowledge now goes beyond traditional expert systems. They may be able to bypass complex systems and provide parsimonious solutions that focus on the key aspects of a situation (Perera *et al.* 2012). What experts know based on their observations, readings and discussions, can then deductively provide insights on complex patterns and processes. The growing use of expert knowledge in conservation science and in particular in species distribution models is driven by limited resources for collecting adequate empirical data on species occurrences and the imminent nature of many conservation decisions (Martin *et al.* 2012). Conservation scientists may find useful applications for expert knowledge in all aspects of research and development. A wide variety of experts are mentioned in the literature with reference to the use of their knowledge in conservation science. Indeed, expert knowledge is now commonly used in conservation science (Borsuk 2004; Comber *et al.* 2005; Fazey *et al.* 2005; Choy *et al.* 2009; Janssen *et al.* 2010; Kuhnert *et al.* 2010). Only a relatively small number of studies to date addressed the incorporation of expert knowledge into the species distribution modelling realm (Murray *et al.* 2009). Utilizing expert knowledge in distribution modelling was mainly practiced through the analytical hierarchy process and multi-criteria evaluation system, or fuzzy set theory (Zadeh 1965), where sampled data were unreliable or inadequate (Anselin et al. 1989; Skidmore 1989; Salski 1992; Store & Kangas 2001; Ozesmi & Ozesmi 2004; Rüger et al. 2005; Doswald et al. 2007; Estrada et al. 2008; Niamir et al. 2011; Olivero et al. 2013). These

approaches are very promising for producing deductive distribution models, and forming a basis for developing more quantitative models regarding the species-environment relationship. Expert knowledge would be an efficient source of information for the species distribution modeller (Murray *et al.* 2009; Newbold 2010). Experts provide valuable insight into species-environment interactions and they may offer predictions given specific environmental conditions (Kuhnert 2011). Furthermore, conservation managers might rely on experts as an alternative source of knowledge (Murray *et al.* 2009; Martin *et al.* 2012) and get them involved in the components (see Figure 1.1) of species distribution modelling.

There are several examples of attempts to employ expert knowledge for species distribution modelling (Dennis 1996; Pearce *et al.* 2001; Ferrier *et al.* 2002; Hobbs 2003; Choy *et al.* 2009; Bierman *et al.* 2010). In practice, however, incorporation of expert knowledge into the species distribution modelling procedure presents challenges (Ferrier *et al.* 2002) and has not been adequately utilized and promoted yet (Carpenter 2002; Lehmann *et al.* 2002; Franklin 2009).

1.4 Proposed approach

We believe that expert knowledge can potentially support species distribution modelling practices by proposing parsimonious solutions for complex natural processes. Expert knowledge is also an efficient source of information for filling the gaps where species data are scarce or unreliable. In this dissertation we propose two approaches to employ expert knowledge in species distribution modelling. One approach is to incorporate knowledge into the data model (see figure 1.1) to facilitate making informed decisions regarding species data preparation. This is a hybrid approach that makes use of the advantages of inductive and deductive approaches. Another approach is to incorporate knowledge directly into the statistical model (see figure 1.1) to directly set species preferences. We then evaluate the effect of incorporating expert

knowledge into components of species distribution modelling. We hypothesise that the use of expert knowledge improves the performance of distribution models particularly when species interactions with their surrogates are complex or data are unobtainable, biased or inadequate. This assumption is justified by the fact that inductive distribution modelling approaches are absolutely dependent on the species data. So, even if a method for distribution modelling provides an accurate relationship between a species and the habitat, the adequacy of the input would still significantly affect the outcome.

Expert knowledge about natural phenomena is subject to uncertainty. This uncertainty is partly due to lack of knowledge of the expert, which can be compensated by learning or adding more experts to the knowledge domain, and partly due to substantive uncertainty in natural phenomena themselves. Our proposed approach towards uncertainty management in expert knowledge is to consider both sources of uncertainty through a novel method that accommodates an expert's self-acknowledged uncertainty and level of ignorance. By ignorance here we refer to absence of fact, understanding, insight, or clarity about something unknown (Firestein 2012). In summary, our approach aims to utilize expert knowledge by taking the "unknown" pieces of information into account and pursuing ignorance in species distribution modelling.

1.5 Research objectives

The overall objective of this thesis is to contribute to understanding of how existing knowledge may enhance species distribution modelling practices. The four chapters of the thesis aim to explore and evaluate a variety of approaches to incorporate knowledge and ignorance into distribution modelling. In particular, we

- explore and evaluate the potential contribution of experts' knowledge for spatial quality enhancement of occurrence data,

where available data are too coarse for the desired application and finer resolution distribution maps are required;

- investigate the effect of considering species taxonomical information on the performance of the distribution models, by delimiting unknown localities using species ancestor's presence;
- propose a novel approach for accounting for self-acknowledge ignorance when modelling species distributions, using solely experts' knowledge, and compare the outcomes with conventional data-driven approaches;
- set a framework for a distribution modeller where expert knowledge is available and species occurrence data are scarce or unobtainable.

1.6 Outline of the dissertation

The dissertation consists of five chapters. Apart from the General Introduction and the Synthesis, the chapters are either under review or have been accepted in scientific peer reviewed journals. Each paper has been presented as a stand-alone chapter and deals with one specific research question. The structure and content of the manuscripts are largely retained for the purpose of this thesis.

- The 1st chapter, the General Introduction, gives an overview of the background and objectives of the dissertation.
- The 2nd chapter outlines the incorporation of existing knowledge into a conventional approach to predict the distribution of Bonelli's eagle at a resolution 100 times finer than available atlas data, in Malaga province, southern Spain.
- Chapter 3 demonstrates the use of species taxonomy to delineate environmental extent and evaluates how this affects the

performance of species distribution models for 356 terrestrial species in mainland Spain.

- The 4th chapter explores the application of the Dempter-Shafer Theory of Evidence in modelling the distribution of a well-known (Bonelli's eagle), and a poorly-studied (Short-toed Eagle) species in mainland Spain.
- Chapter 5, the Synthesis, summarizes and discusses the main findings, and outlines perspectives for future research.

1.7 The study area

Considering the research questions, the availability of materials and of a corporative network of experts, it was decided to conduct the study in Spain at two different scales: local scale (Malaga Province) and national scale (Spain). Peninsular Spain is situated in south-western Europe and covers an area of 493.518 km^2. Spain may be divided into three climatic areas: Atlantic, Mediterranean and Interior. The Atlantic area has mild winters and relatively cool summers, and abundant precipitation throughout the year. The Mediterranean part is characterized by hot summers and mild winters with low rainfall, occurring mainly during spring and autumn. Hot summers, mild winters, and scarce precipitation characterize the Interior (Capel Molina 1981). Spain also comprises important mountain ranges, situated mainly in the north (Pyrenees and Cantabrian mountains) and the southeast (Baetic mountains), with relevant mountain chains traversing central Spain from west to east (Iberian and Central systems). Ecologically, mainland Spain is a relatively homogenous unit with sufficient variability as well as the target species of this study being of particular interest in a European context. The province of Malaga (7267 km2) is a mountainous region situated in Andalusia, southern Spain, ranging in altitude from sea level along the Mediterranean shoreline to almost 2000 m. The climate is Mediterranean and the natural vegetation in

the valleys and lowlands has been transformed into olive groves, cereal crops, and coastal urbanization, interspersed with small fragments of Mediterranean scrubland.

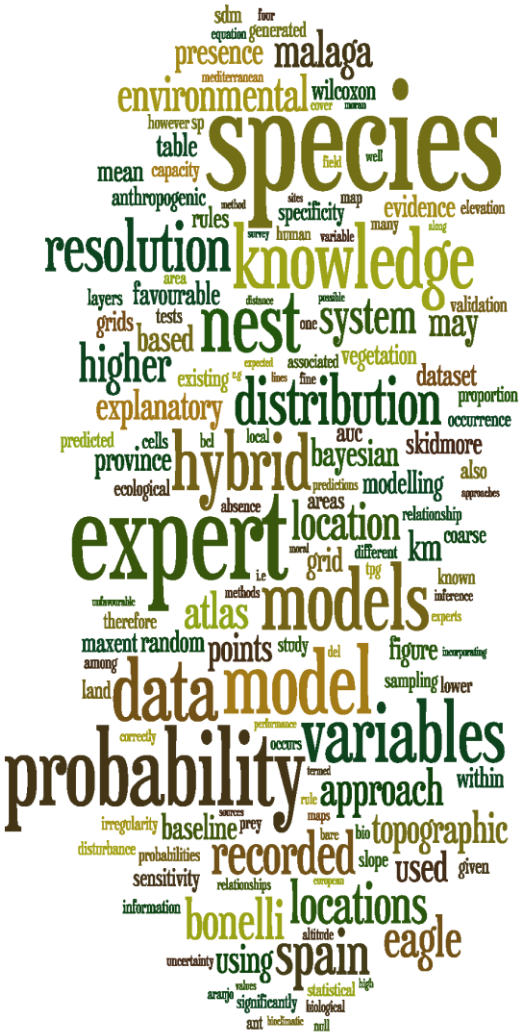
Chapter 2:

FINESSING ATLAS DATA FOR SPECIES DISTRIBUTION MODELS

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Munoz, A.R. & Real, R. *Diversity and Distributions* (2011) 17, 1173-1185

Parts of this chapter also appear in:

- Niamir, A. et al. "Downscaling species atlas data using expert-system", in the proceedings of the 5th International meeting of the International Biogeography Society, 2011, Iraklion, Greece
- Niamir, A. et al. "Enhancing coarse-resolution species atlas", in the proceedings of the Netherlands Ecological Annual Meeting 2011, Lunteren, the Netherlands
- Niamir, A. et al. "Generating finer resolution species distribution data using an expert system", in the proceedings of the Ecological society of Australia, annual conference 2010. Canberra, Australia.


sdm
four
equation
generated
presence
malaga
mediterranean
environmental
wilcoxon
cover
however
sp
species
table
field
well
mean
capacity
anthropogenic
method
sites
map
evidence
elevation
rules
specificity
many
along
survey
human
variable
resolution
knowledge
area
layers
favourable
distance
possible
validation
grids
tests
based
nest
one
system
may
higher
expected
associated
vegetation
existing
lines
fine
dataset
proportion
explanatory
distribution
occurrence
predicted
cells
bcl
local
auc
skidmore
province
hybrid
bayesian
also
ecological
predictions
modelling
approaches
absence
areas
relationship
expert
location
coarse
moral
different
km
grid
tpg
known
i.e
methods
inference
unfavourable
therefore
atlas
models
experts
maxent
random
points
study
figure
incorporating
among
del
land
data
model
sampling
lower
correctly
performance
occurs
probability
variables
termed
within
rule
approach
irregularity
sources
maps
baseline
prey
bio
topographic
disturbance
probabilities
recorded
bare
slope
used
given
sensitivity
relationships
european
information
bonelli
locations
altitude
eagle
uncertainty
using
spain
values
statistical
high
araujo
significantly
biological
ant
biseliatic
null

2.1 Abstract

The spatial resolution of species atlases and therefore resulting model predictions are often too coarse for local applications. Collecting distribution data at a finer resolution for large numbers of species requires a comprehensive sampling effort, making it impractical and expensive. This study outlines the incorporation of existing knowledge into a conventional approach to predict the distribution of Bonelli's eagle at a resolution 100 times finer than available atlas data in Malaga province, southern Spain. A Bayesian expert system was proposed to utilize the knowledge from distribution models to yield the probability of a species being recorded at a finer resolution (1 x 1 km) than the original atlas data (10 x 10 km). The recorded probability was then used as a weight vector to generate a sampling scheme from the species atlas to enhance the accuracy of the modelling procedure. The Maximum Entropy for species distribution modelling (MaxEnt) was used as the species distribution model. A comparison was made between the results of the MaxEnt using the enhanced and, the random sampling scheme, based on four groups of environmental variables: topographic, climatic, biological and anthropogenic. The models with the sampling scheme enhanced by an expert system had a higher discriminative capacity than the baseline models. The downscaled (i.e. finer scale) species distribution maps using a hybrid MaxEnt/expert system approach were more specific to the nest locations, and were more contrasted than those of the baseline model. The proposed method is a feasible substitute for comprehensive field work. The approach developed in this study is applicable for predicting the distribution of Bonelli's eagle at a local scale from a national level occurrence dataset; however, the usefulness of this approach may be limited to well-known species.

2.2 Introduction

Species distribution modelling (SDM) is being widely used to predict the occurrence of species at locations where survey data are lacking (Guisan & Thuiller 2005), which is the case for most of the earth's surface (Franklin 2009). To define the suitability of a location for a species, SDM links data about a species distribution data to the environmental characteristics of those locations and then extrapolates the relationship over space or time (Oindo *et al.* 2003; Guisan & Thuiller 2005). For the majority of species, data describing distributions are scarce and in the form of coarse resolution atlases (Newbold 2010). Species atlases provide a convenient display, using near-equal areas (grids) on maps (Araujo *et al.* 2005), of the extent-of-occurrence based on historical observations, museum records and often a complementary field survey. A species is marked as "present" if there is at least one record of that species at a location within a pre-defined grid (Bierman *et al.* 2010). Such grid maps are typically available in the form of systematic 10 × 10 km grid cells at a national level or in coarser resolutions at continental and global level, while application usually occurs on a local level of around one hundred hectares (Rouget 2003).

Atlas data plays an important role in conservation biogeography by providing species distribution data for developing new concepts, analytical approaches, and to address a range of conservation problems (Araujo *et al.* 2005). Therefore, to generate a reliable fine resolution distribution map, the gap between the available coarse resolution species occurrence datasets and the desired resolution has to be bridged (Hobbs 2003). Developing an empirical approach to deal with uncertainties in the downscaling process remains a challenge (Boitani *et al.* 1999; Barbosa *et al.* 2003; Araujo *et al.* 2005; Graham *et al.* 2008).

Projection to a finer resolution of statistical relationships calibrated at a coarse resolution is a conventional approach (Collingham *et al.* 2000; Araujo *et al.* 2005; Barbosa *et al.* 2009; Barbosa *et al.* 2010). However, the extrapolation of models built for one resolution to a different resolution

increases the uncertainty in the model predictions because they are extended beyond the model's original scale and error can vary among scales (Araujo *et al.* 2005). Another downscaling approach investigated point-sampling (Lloyd & Palmer 1998; Barbosa *et al.* 2003; Hartley *et al.* 2003; Mcpherson *et al.* 2006), where the species occurrence dataset consisted of a random point within each coarse resolution grid cell (50 x 50 or 10 x 10 km) along with fine resolution (1 x 1 km) environmental variables informing the SDM. A critical issue in this approach was the model's uncertainty, derived from prediction inconsistency over different random sampling iterations (Costa *et al.* 2010).

Furthermore, the expense of an intensive field survey is often high and cannot be afforded (Skidmore & Turner 1992; James *et al.* 2001). Since there are insufficient field data available, expert knowledge could be an efficient source of information (Murray *et al.* 2009). Translation of such knowledge into a rule-based script poses a challenge (Ferrier *et al.* 2002), and has not been widely promoted (Carpenter 2002). There have been concerns that prior knowledge (i.e. expert opinion) might drive the modelling process rather than guide it (Dennis 1996). There are several examples which attempted to incorporate expert knowledge into SDM procedures (Pearce *et al.* 2001; Ferrier *et al.* 2002; Hobbs 2003; Choy *et al.* 2009; Bierman *et al.* 2010), but this has not been adequately utilized in species data optimization and sampling strategies (Lehmann *et al.* 2002). A species is recorded as present in an atlas grid if the database holds at least one record of that species from a location within that grid. The term "recorded probability" is used to address the probability of a species being recorded at a location.

Bayesian inference is an accepted statistical tool among ecologists (Pereira & Itami 1991; Dennis 1996; Ellison 2004; Clark 2005; Mccarthy 2007). Bayes' Theorem provides a clear method for estimating parameters and expressing the degree of confidence and uncertainty in those estimates. Bayesian expert-systems (Lee *et al.* 1987) have been defined to handle complex, real world problems and attempt to solve problems by reasoning

like an expert (Forsyth 1984; Skidmore 1989; Heikkinen & Hogmander 1994). Expert-systems used to combine diverse data sources with remotely sensed images in order to map soils (Skidmore *et al.* 1996), vegetation (Skidmore 1989; Schmidt *et al.* 2004; Wang *et al.* 2009) and habitat characterization (Bierman *et al.* 2010; Wang *et al.* 2010).

This study outlines the incorporation of existing knowledge into a point-sampling approach to predict the distribution of Bonelli's eagle (*Aquila faciata*) in Malaga province, southern Spain, at a resolution 100 times finer than atlas (modelled) data (Marti & Del Moral 2003). An experiment was designed to evaluate the discriminative capacity of a Bayesian expert system model with an enhanced sampling scheme and four groups of environmental explanatory variables: topographic, climatic, biological, and anthropogenic. This model (termed "hybrid" model hereafter) consisted of a Bayesian expert system combined with a conventional SDM technique: Maximum Entropy species distribution modelling (MaxEnt)(Phillips *et al.* 2006). The aim was to assess whether an expert system could improve the accuracy and robustness of predictions for local conservation applications by utilizing existing ecological knowledge, where empirical data were missing or difficult to obtain.

2.3 Methods

2.3.1 Study area

The province of Malaga (7267 km²) is a mountainous region situated in Andalusia, southern Spain, ranging in altitude from sea level along the Mediterranean shoreline to almost 2000 meters. The climate is Mediterranean with mean annual rainfall ranging from 400 to 1200 mm and annual temperatures ranging from 12.6 to 19.2 °C (Font 2000). The natural vegetation has been transformed in the valleys and lowlands to olive

groves, cereal crops, and coastal urbanization interspersed with small fragments of Mediterranean scrublands.

2.3.2 Species occurrence data

Bonelli's eagle is a resident species in the Malaga province, with only juvenile birds dispersing. Therefore, it was assumed that the presence of active nests indicated the presence of the species throughout the year. Occurrence data for Bonelli's eagle in 10 × 10 km grids for Malaga province were obtained from the Atlas of Spanish breeding birds (Marti & Del Moral 2003) and updated with the Atlas of raptors of Malaga province (Jiménez & Muñoz 2008). Malaga province is covered by 104 atlas grids (10 x 10 km), with Bonelli's eagle marked "present" in 67 (70%) of the grids and "absent" in the others (Figure 2.2a), which indicates that the density for the species in Malaga province is the highest one known in Europe.

Three species sampling datasets were generated. The first dataset termed the 'random points' dataset consisted of 100 iterations of a random point in a coarse resolution grid (10 x 10 km) which was originally marked as 'present' in the atlas (n=67). The second dataset (termed the 'enhanced' dataset) was generated based on the same procedure, except that the recorded probability (as quantified by the expert system) was used as a weighting factor. This meant that a grid cell with a higher recorded probability had a higher probability of being sampled. In order to test whether the developed SDMs differ significantly from what would be expected by chance, 1000 random points were generated over the study area (Malaga province) and used as inputs for the null models (Raes & Ter Steege 2007).

An independent evaluation dataset was collected consisting of 80 nest locations (presences). Nest locations were provided by local ornithologists monitoring breeding territories from 2005 to 2009 in Malaga province.

The absence points (n=1000) were randomly generated using the atlas data (Marti & Del Moral 2003) within cells where the species was not observed.

2.3.3 Target Species

Bonelli's eagle is a medium to large sized raptor (Figure 2.1) with its western Palaearctic populations mainly distributed in the Mediterranean area (Del Hoyo 1994). The European population has suffered a marked decline and has been listed as an endangered European Species in recent decades (Rocamora 1994; Birdlife-International 2000, 2004, 2008, 2011). Spain supports 730 – 800 breeding pairs (Real 2003) or about 75% of the total European population. Malaga supports about 10% of the Spanish population of the species (Del Moral 2006), considered one of the last strongholds of the Bonelli's eagle in Europe (Balbontín 2005). Existing knowledge about favourable nest locations for Bonelli's eagle was elicited from literature and retrieved from the Web of Science® on 1 January 2010 through three key words: "Bonelli's eagle", "*Hieraaetus fasciatus*", and "*Aquila fasciata*".

Figure 2.1: Bonelli's Eagle (Photo: Juan Luis Muñoz)

Table 2.1 lists the studies that describe the variables that are considered to be important for determining the distribution of the species. Disagreement among different sources was discussed with local experts and their contribution has been evaluated subjectively. It was concluded that topographic attributes formed a critical factor in favourable nest locations for Bonelli's eagle. Experts consistently mentioned slope gradient and altitude as important topographic explanatory variables. Another critical factor frequently mentioned in the literature was formed by land cover attributes, e.g. distance to artificial surfaces and human disturbance.

Table 2.1 Details of studies which have been used for knowledge extraction on the habitat preferences of Bonelli's Eagle.

Data [1]	Region	Parameters
P	Cádiz, Spain (Balbontín 2005)	Topographic features, and Human disturbance
P	Catalonia, Spain (Bosch *et al.* 2010)	Human disturbance
A	Spain (Carrascal & Seoane 2009)	Climate, vegetation, and interspecific relationships
P	Murcia, Spain (Carrete *et al.* 2002)	Prey availability
P	Granada, Spain (Gil-Sánchez *et al.* 1996)	Distance to villages, topographic irregularity, and cultivation
P	Granada, Spain (Gil-Sanchez *et al.* 2004)	Human disturbance, interspecific relationships, and topographic variables
P	Castellón, Spain (Lopez-Lopez *et al.* 2006)	Cragy slopes, human disturbance, climate, and landuse
A	Valencia, Spain (Lopez-Lopez *et al.* 2007a)	Altitude, and slope

Table 2.1 (continued)

Data [1]	Region	Parameters
P	Castellón, Spain (Lopez-Lopez *et al.* 2007b)	Altitude
P	Alicante, Spain (Martinez *et al.* 2008a)	Landcover
P	Alicante and Murcia, Spain (Martinez *et al.* 2008b)	Intra- and interspecific relationships
SP, A	Western Europe (Moleon *et al.* 2009)	Prey availability, and local territorial features
A	Spain (Moreno-Rueda *et al.* 2009)	Anthropogenic disturbances, temperature, and prey diversity
A	Spain (Muñoz *et al.* 2005)	Slope, temperature, and precipitation
SP	Granada, Spain (Ontiveros & Pleguezuelos 2003a)	Landscape features, and prey accessibility
SP	Granada, Spain (Ontiveros & Pleguezuelos 2000)	Landcover, and Aspect
SP	Western Mediterranean (Ontiveros & Pleguezuelos 2003b)	Climatic constraints, and anthropogenic disturbances
P	Andalucia, Spain (Ontiveros *et al.* 2005)	Landcover and, prey availability
P	Portugal (Palma *et al.* 2006)	Prey accessibility, and landscape composition
P	Valencia, Spain (Rico *et al.* 1999)	Altitude
SP	Spain (Soutullo *et al.* 2008)	Anthropogenic disturbances

[1] Species data: SP, subpopulation; A, atlas/lattice; P, points

2.3.4 Expert layers and rules

Using the ASTER elevation model (Abrams *et al.* 2010) two topographic explanatory variables was compiled: elevation and slope irregularity. To generate the irregularity layer for slope (Skidmore 1989) and elevation, the maximum values were subtracted from the mean within a 1 x 1 km grid, resulting in higher values representing higher irregularity and lower values describing homogenous topographic landscape. The CORINE (Coordination of Information on the Environment of the European Environmental Agency (European Environment Agency 2012)) land cover 1995 and 2006 maps (Bossard *et al.* 2000) with a 250 x 250 meter resolution was also downloaded and aggregated to a 1 x 1 km resolution using the Raster package (Hijmans & Van Etten 2010). Land cover and topographic irregularity were selected as environmental layers in the expert model, reflecting the fact that this species is geographically widespread in Malaga province with its fragmented bare rocks and high human disturbance. These three environmental explanatory variables (expert layers) were then used in the expert system. Since spatial correlation among (Montgomery 1982) and within (Dormann *et al.* 2007; Naimi *et al.* 2011; Naimi *et al.* 2013) predictors violates the assumption of independence of most standard statistical procedures (Legendre 1993a), the spatial structure of the expert layers was investigated using the variance inflation factor (VIF) as a measure of multicollinearity, and the Moran's I statistic as a measure of univariate spatial autocorrelation. Formulation of expert rules was based on the integration of knowledge from several sources, including the reviewed literature, knowledge acquired through discussion with experienced local researchers, as well as personal field observation. Where disagreement occurred between the different sources, a subjective decision was made based on field knowledge. The expert rules and their associated probabilities are listed in Table 2.2.

Table 2.2 Variables and their corresponding rules included in the expert system to calculate the probability of recording a nest of Bonelli's eagle in Malaga.

Variables	Evidence	P(E|H)
Elevation Diff. (m)	less than 100 meters	0.3
Moran's *I* = 0.367	between 100 to 200 meters	0.5
VIF = 1.187	more than 200 meters	0.9
Slope gradient (%)	less than 30	0.2
[Max – Mean]	30 - 45	0.3
Moran's *I* = 0.483	45- 60	0.5
VIF = 1.271	more than 60	0.9
Land cover	Continuous urban fabric	0.1
Moran's *I* = 0.909	Discontinuous urban fabric	0.1
VIF = 1.468	Industrial or commercial units	0.1
	Road and rail networks and associated land	0.1
	Port areas	0.1
	Airports	0.2
	Mineral extraction sites	0.2
	Construction sites	0.2
	Sport and leisure facilities	0.2
	Non-irrigated arable land	0.3
	Permanently irrigated land	0.2
	Vineyards	0.3
	Fruit trees and berry plantations	0.3
	Olive groves	0.3
	Crops associated with permanent crops	0.3
	Complex cultivation patterns	0.3
	Agriculture and natural vegetation	0.3
	Agro-forestry areas	0.4
	Broad-leaved forest	0.3
	Coniferous forest	0.4
	Mixed forest	0.4
	Natural grasslands	0.5
	Sclerophyllous vegetation	0.7
	Transitional woodland-shrub	0.3
	Bare rocks	0.9
	Burnt areas	0.5
	Water courses	0.1
	Water bodies	0.1

2.3.5 Explanatory variables

The environmental explanatory variable sets have a direct interaction with species and were chosen based on ecological theory (Austin 2007). Usually there is a wide variety of potential predictor variables to choose from, but the effects of making a choice between variables are poorly known (Synes & Osborne 2011). In this case, the explanatory environmental variables were grouped into four categories (see Appendix S1 in Supporting Information); topographic, climatic, anthropogenic, and biological. The basis for grouping was to create consistent variable sets for input to, and comparison of the SDM predictions.

Topographic variables (TPG) consisted of elevation, slope, and aspect along a north-south axis derived from the ASTER elevation model (Abrams *et al.* 2010). For climatic variables (BCL), all 19 bioclimatic layers (1 x 1 km) (Hijmans *et al.* 2005) were used. The approximate distance to roads, the approximate distance to urban areas, and the land cover map comprised the anthropogenic variables (ANT). The most commonly used parameter for quantifying productivity and above-ground biomass of ecosystems is the Normalized Difference Vegetation Index (NDVI). For biological variables (BIO), atmospherically corrected SPOT4 and SPOT5 Vegetation Sensor were obtained from SPOT Vegetation Program*. The decade average (1998 -2008) of the 10-day composite NDVI-images (S10 product) at 1 km^2 resolution was then calculated (36 images). For more information on NDVI and faunal distribution see (Leyequien *et al.* 2007).

2.3.6 Bayesian expert system

Bayesian theory in SDM offers an alternative approach to statistical inference and differs from conventional frequentist inference in

* http://www.vgt.vito.be

fundamental ways. Frequentist inference estimates the probability of the evidence (E) given hypothesis (H), while Bayesian inference estimates the probability that a hypothesis is true given evidence and defines it as the degree of belief in the likelihood of the evidence (Wade 2000; Ellison 2004). Here a forward chaining expert system, originally developed by Skidmore (1989), was used to infer the posterior probability that a Bonelli's eagle nest occurs at a given cell based on the predictors and the expert rules. The forward chaining approach is essentially a data driven approach (Naylor 1984) and has been applied satisfactorily in remote sensing and image classification (Skidmore *et al.* 1996; Schmidt *et al.* 2004; Wang *et al.* 2010). Bayesian methods explicitly recognize and combine four components of knowledge: prior knowledge, data, model and posterior knowledge (Mccarthy 2007). In this study, prior knowledge from coarse resolution atlas data was combined with expert rules to estimate the recorded probability of a nest location at a finer resolution.

Let (N) be a nest occurring at location $(X_{i,j})$ and let (E_b) be an item of evidence (for $b=1,..,k$) known at location $(X_{i,j})$. Set a hypothesis (H) that a nest (N) occurs at location $(X_{i,j})$. A rule may be defined thus: given (E_b) then (H), that is: given a piece of evidence (E_b), then infer (H). However, there may be uncertainty associated with this rule (Forsyth 1984), and the probability of the rule may not be 0 (i.e. false) or 1 (i.e. true), but rather lie between 0 and 1, depending on how "sure" the experts are that the rule is true (Skidmore 1989).

Bayes' Theorem was used to update the probability of the rule that the nest of Bonelli's eagle (H) occurs at $(X_{i,j})$ given an environmental explanatory variable:

$$P(H \mid E_b) = \frac{P(E_b \mid H) \times P(H)}{P(E_b)} \qquad \text{(Equation 1)}$$

where $P(E_b \mid H)$ is the probability that there is a piece of evidence (E_b) (e.g. bare rocks) given that a nest occurs at location $(X_{i,j})$, also known as conditional probability which is based on expert rules (Table 2.2); $P(H)$ is the probability for the hypothesis (H) that a nest occurs at location $(X_{i,j})$, initially obtained from the species atlas data (*a priori*). On iterating with further pieces of evidence, $P(E_b \mid H) : b = 1$ replaces $P(H)$ in equation 1. $P(E_b)$ is the probability of the evidence alone:

$$P(E_b) = \sum_{b=1}^{k} P(E_b \mid H) \times P(H) \qquad \text{(Equation 2)}$$

The evidence (E_b) must be spatially independent (Table 2.2), otherwise $P(E_b)$ would become larger or smaller, thereby decrementing or incrementing (H), causing the *posterior* probabilities to be incorrect.

2.3.7 Hybrid method

A hybrid method for species distribution was constructed, incorporating a Bayesian expert system and a machine learning model, to increase the discriminative capacity compared to methods without an expert system.

Machine learning methods in contrast to deductive and knowledge-driven approaches include various kinds of algorithms implemented to learn the classification rules directly from data (Breiman 2001a). The Maximum Entropy model for species distribution modelling (MaxEnt) (Phillips *et al.* 2006) has generated higher predictive accuracy than many other methods (Elith *et al.* 2006; Hernandez *et al.* 2006). MaxEnt also outperformed others where sampling was poor (Costa *et al.* 2010), data were collected with sampling bias (Phillips *et al.* 2009; Rebelo & Jones 2010) and across different sample sizes (Wisz *et al.* 2008). Therefore the MaxEnt version 3.3.1 was employed here as distribution modelling core. For more information about MaxEnt and its statistical explanation see (Elith *et al.* 2011).

2.3.8 Evaluation

The PresenceAbsence package (Freeman & Moisen 2008) in R 2.11.1 (R Development Core Team 2013) was used to assess the capacity of the downscaled models to discriminate between validation presence and absence points (validation dataset) by analysing their receiver operation characteristic (ROC) curves. The area under this curve (AUC) provided a threshold-independent measure across all possible classification thresholds for each model (Fielding & Bell 1997). Then the R's *wilcox.test* function was used to perform a Mann-Whitney-Wilcoxon rank-sum test (termed Wilcoxon test hereafter) to check whether the hybrid model had a higher AUC value than the MaxEnt model. AUC combines specificity and sensitivity (Fielding & Bell 1997) thus it is not possible to explore whether the improvement in model performance has to do with reduction of commission or omission, or both errors. AUC ignores the goodness-of-fit of the predictions (Lobo *et al.* 2008; Jiménez-Valverde 2012), therefore it is possible that a poorly fitted model retains good discrimination power (Hosmer & Lameshow 2000). To deal with this limitation sensitivity and specificity of models was also calculated.

2.4 Results

The probability of a nest being recorded (at the finest resolution of 1 x 1 km) was estimated using the Bayesian expert system. The mean of the posterior probabilities ($\mu = 0.24$) was low. Two-thirds (76%) of all fine resolution grids had a recorded probability of 0.5 or less. In line with expectation, the range of recorded probabilities was wide with a mean (95% CI) of 0.81 among presence coarse resolution grids, indicating that there are few cells with a high and many cells with a low recorded probability in each presence atlas grid (Figure 2.2b). Intersection of validation nest location with the recorded probability revealed that 30 (37.5%) of the nests were located in cells with a 0.75 or higher recorded probability, 47 (58.7%) were located in cells with a 0.5 or higher recorded probability, and 8 (10%) in cells with a 0.1 or lower recorded probability. The average recorded probability for nest locations (presences), was 0.57 (with standard deviation 0.27).

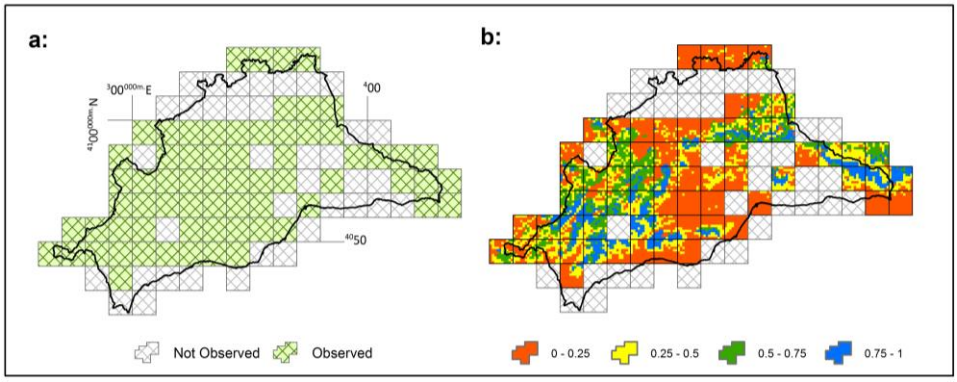


Figure 2.2: (a) Distribution of Bonelli's Eagle in Malaga, based on the Spanish Atlas of Breeding Birds and Atlas of raptors of Malaga province. (b) Probability of a nest being recorded over presence atlas grids based on expert knowledge.

The random, null-points, and enhanced species dataset along with four categories of environmental variables (TPG, BCL, ANT, and BIO) were input to MaxEnt. The baseline models (built using the random point dataset) were compared with the "hybrid" models (built using the enhanced dataset). The hybrid models had a higher mean AUC than the baseline models. However, not all models were equally improved by incorporating the expert system. The favourable nesting locations were accurately discriminated from unfavourable locations over the study area (Figure 2.2). The mean AUC score in models with ANT variables increased the most by replacing the baseline with hybrid models: from 0.71 to 0.83 (Wilcoxon tests, $p < 0.001$); and there was a 8% improvement in discrimination capacity of the models with TPG ($AUC_{Baseline} = 0.74$ to $AUC_{Hybrid} = 0.81$, Wilcoxon tests, $p < 0.001$) and BIO ($AUC_{Baseline} = 0.75$ to $AUC_{Hybrid} = 0.82$, Wilcoxon tests, $p < 0.01$), whereas the models with bioclimatic variables (BCL) only slightly increased the AUC score, from 0.75 to 0.79 (Wilcoxon tests, $p < 0.001$) (Table 2.3). The evaluation of the null models also revealed that all models are significantly different than what is expected by chance.

The mean of the probabilities (of favourable nest locations) was always significantly higher at 'presence' sites (i.e. n = 80 nest locations), than at 'absence' sites (Wilcoxon tests, for ANT, TPG and BIO: $p < 0.001$ and for BCL: $p < 0.01$). The proportion of presence validation points in areas with high probability (≥ 0.75) was always significantly higher than the proportion of localities available within those areas for both hybrid and baseline models. The proportion of presences in areas of low probability (≤ 0.25) was generally lower than expected by chance (Wilcoxon tests, for ANT, TPG and BIO: $p < 0.01$ and for BCL: $p < 0.05$). The mean probability of occurrence values were also consistently higher at presence validation points and lower at absence validation points than at any of the

random site samples (null-points) for hybrid models (Wilcoxon tests, for presence: $p < 0.01$ and for absence $p < 0.05$).

The hybrid approach had considerably higher specificity (pseudo-absence location correctly predicted), whereas the sensitivity (nest location correctly predicted) was decreased at lower thresholds. The threshold probability where specificity is equal to sensitivity was lower in the hybrid than the baseline models. However, the proportion of the correctly predicted nest locations was higher at the above mentioned threshold (Figure 2.3). This is reflected in the predicted distribution maps. The least favourable locations (≤ 0.25) were discriminated from less favourable locations ($0.25 \leq f \leq 0.5$) in the hybrid approach (Figure 2.3).

Uncertainties in the prediction (standard deviation of AUC) over 100 iterations were significantly smaller in hybrid models than in the baseline models (Table 2.3), revealing that the hybrid models were more robust in their discrimination.

Table 2.3: Evaluation of model performance, Mean and (Standard Deviation) of the AUC scores over 100 iterations for null model, baseline (random sampling) and hybrid (Bayesian expert system) models. Significant levels are associated with the Wilcoxon's test.

Model	Null model	Baseline	Hybrid
Topographic	0.60 (0.078)	0.74 (0.114)	0.81 (0.027) $^{++}$
Bioclimatic	0.56 (0.047)	0.75 (0.019)	0.79 (0.019)$^{++}$
Biological	0.55 (0.039)	0.75 (0.045)	0.82 (0.015)$^{+}$
Anthropogenic	0.57 (0.058)	0.71 (0.072)	0.83 (0.019)$^{++}$

$^{++}$ *P < 0.001,* $^{+}$ *P < 0.01,* ^{ns} *not significant*

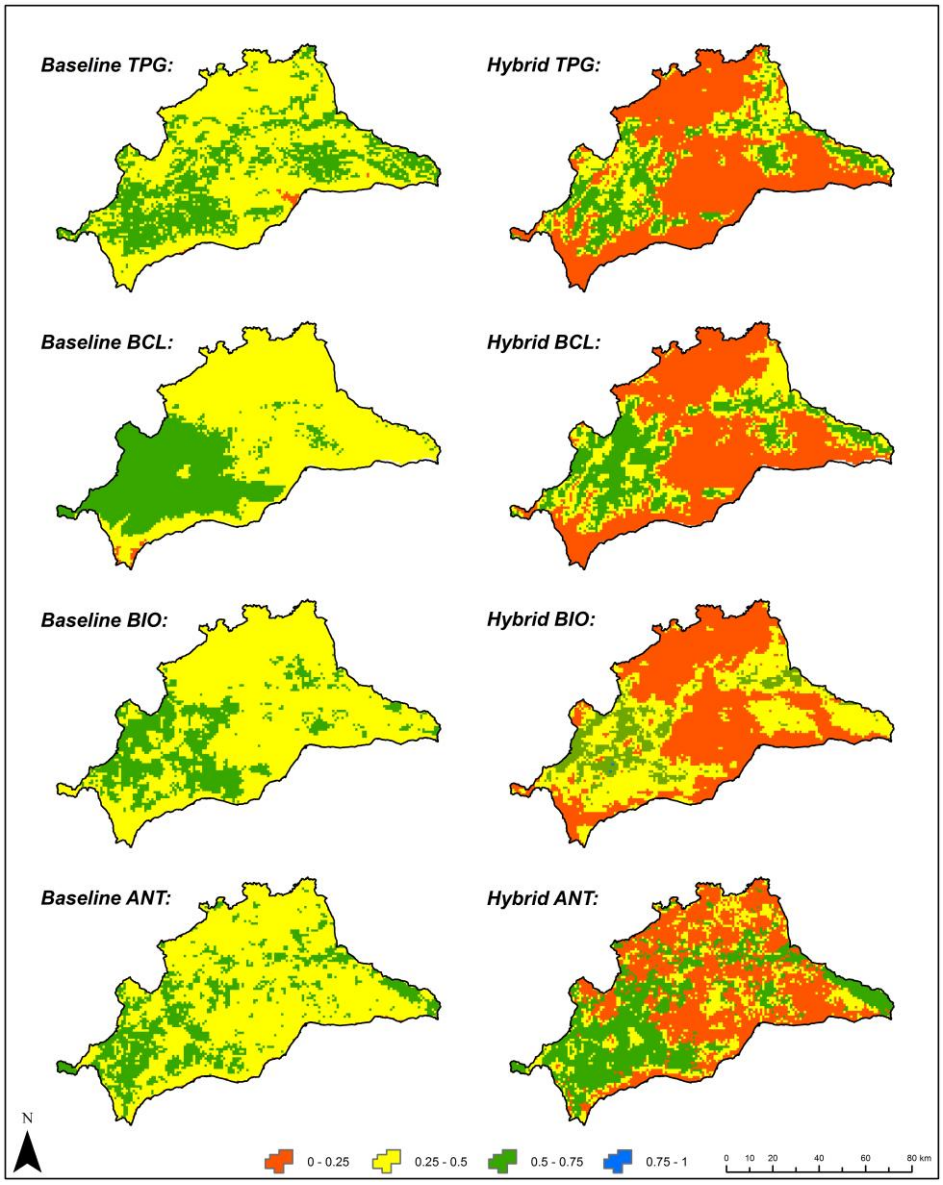


Figure 2.3 Comparison of fine resolution prediction maps generated by baseline (right column) and hybrid (left column) models using topographic (TPG), bioclimatic (BCL), biological (BIO), and anthropogenic (ANT) environmental explanatory variables for Bonelli's eagle in Malaga

2.5 Discussion

For all environmental explanatory variables, the downscaled predictions based on the hybrid approach were significantly related to the location of survey nests and improved the discriminative capacity of the distribution models. This suggests that such a hybrid approach may be a valid way to incorporate existing knowledge into the modelling procedure in regions that are poorly known or where fine resolution species information is too costly and time consuming to gather.

The main aim of this paper was to introduce the Bayesian expert system as a promising alternative, incorporating existing knowledge into the analysis of species atlas data. Since the proposed approach enhances the quality of input species data, other SDMs than MaxEnt may also be employed for the prediction of species distribution. In fact, any parametric model may be suitable to build up a hybrid approach using our expert system. No a priori assumption was made as to the form of the relationship between the proxy variable for the expert layers and the recorded probability. However, it is possible to take into account the importance of environmental explanatory variables based on the expert knowledge by assigning weights to (Equation 2), and consequently updating (Equation 1). The methodology presented in this paper extends the methodology by Skidmore (1989) by including a species distribution model in which the recorded probability of the species is modelled as a function of environmental variables.

Higher discrimination capacity of the hybrid approach is due to its higher specificity rather than sensitivity. When considering a 10 x 10 km atlas data grid, there are many favourable locations for a nest that might not be used for many reasons, e.g. the territorial behaviour of the species. Consequently, a model should be capable of discriminating unfavourable locations (even if they are biophysically favourable). A reasonable model would be expected to be not only sensitive, but also specific to the nest locations. Sensitivity is equal to specificity (sensitivity and specificity lines cross each other) at a higher proportion value in the hybrid model, and at a

threshold closer to the prevalence of nest presence compared to the baseline model (Figure 2.4).

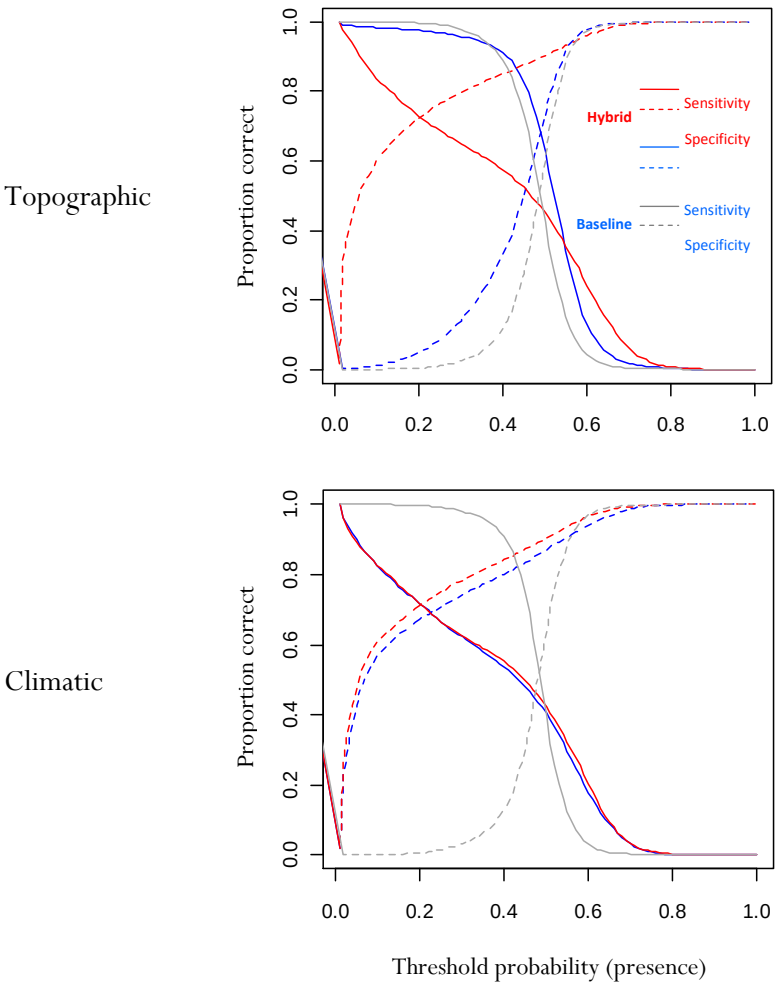
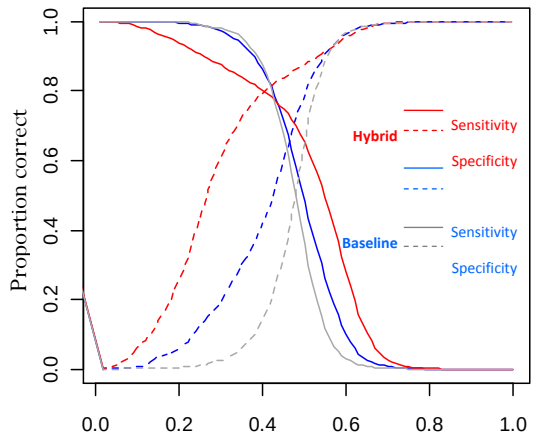


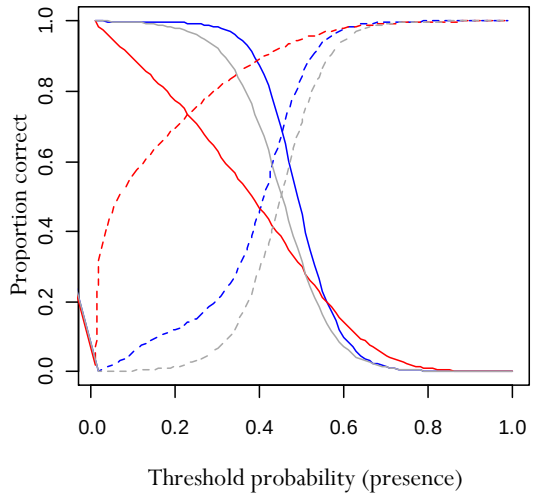
Figure 2.4 Variation of sensitivity (solid lines), the proportion of correctly predicted nest location versus specificity (dashed lines), the proportion of correctly predicted absences in baseline (blue) and hybrid (red) models, using four categories of environmental explanatory variables for Bonelli's eagle in Malaga. Grey lines represent null models.

Figure 2.4 (continued)

Anthropogenic



Biological



Expert systems offer many advantages and some disadvantages over conventional statistical approaches for species distribution modelling. The major advantage is that existing knowledge about the species-environment relationship can be encapsulated into the modelling process. In this study the inductive model, (MaxEnt) produced the suitability for nest occurrence, based on random points within coarse resolution grids. The

known ecological relationship between environmental variables and the location of a nest yielded the most likely nest location (recorded probability). The expert system handled the uncertainty in these relationships through the use of probability (e.g. it is fairly certain that the Bonelli's eagle will nest on bare rock, but this may not always be the case). This is in contrast with crisp models (Estrada *et al.* 2008), which try to describe the relationship in a binary format (i.e. does the nest of a Bonelli's eagle occur in sclerophyllous vegetation, yes or no?). Another advantage of expert systems is that expert judgment on the effect of scale can be quickly incorporated as the implicit relationship between data layers and the dependent variable being modelled becomes clear. For example many bare rock pixels were aggregated with other dominant pixels within 1 x 1 km grids. Therefore in this study, experts expressed their knowledge in the form of topographic irregularity to diminish the aggregation effects. This flexibility gives an expert the opportunity to build their own expert knowledge base, even manually drawing their ideas on a map using a graphical interface. Therefore, in such hybrid approaches the ecological realism and acceptability to the user community may improve as well as the predictive performance.

An obvious disadvantage of expert systems for species distribution modelling is that experts may not agree among themselves, causing inconsistencies in the existing knowledge on the ecological factors affecting distribution of the target species. The "robustness" of a recorded probability map generated by an expert system has to be gauged against the criteria defined by the expert or other source of knowledge, as well as against a validation data set. Another disadvantage of expert systems, as with all other modelling techniques, is that they do not respond well to incomplete knowledge or extrapolation into an area beyond the region of expertise (Murray *et al.* 2009). In such a situation, the probabilities associated with the rules may have to be adjusted to better reflect the gap in knowledge or the addition of explanatory variables may be required (Skidmore 1989). As the results revealed, existing knowledge was more

accurate in describing unfavourable locations for nests than favourable locations. Therefore it may be beneficial to model the unsuitability, using absence points instead of presence points, and then taking the inverse of the predicted distribution map (Lobo *et al.* 2010). In other words, the presented hybrid model may be employed to incorporate expert knowledge on either suitability or unsuitability for a certain species into distribution models.

A number of studies have investigated the performance of inductive (data-driven) SDM at predicting distribution and species ecological characteristics (Segurado & Araujo 2004; Elith *et al.* 2006; Guisan *et al.* 2007; Evangelista *et al.* 2008) and it appeared that generalist species with a wide geographical range yielded models with a lower discriminative capacity than species with strict geographical boundaries (Buisson *et al.* 2010). This is the stage that our hybrid approach will be most usefully applied. It is easier to discriminate favourable from unfavourable habitat for generalist species, by incorporating a deductive (knowledge-driven) method, thereby describing the ecological niche. The novel technique described here may not be successful for species that are not well known or minimally studied. However all SDMs suffer the same restrictions.

2.6 Acknowledgments

This research was funded by the European Union Erasmus Mundus program External Cooperation Windows 2008/2441/001 MUN ECW 15-7-2008, and by the Spanish Ministry of Education and Science and FEDER (project CGL2009-11316, BOS subprogram). The authors wish to thank the local ornithologists who kindly shared their knowledge and observations of the Bonelli's eagle in Andalucia, as well as the many volunteer fieldworkers who contributed to the Spanish atlas of breeding birds. Babak Naimi provided crucial tips on the use of R. The language editing was done by Eva Skidmore. Many thank to Mark Robertson and the

three anonymous reviewers for their comments on a previous version of this paper.

Chapter 3:

USE OF TAXONOMY TO DELINEATE ENVIRONMENTAL EXTENT FOR SPECIES DISTRIBUTION MODELLING

Niamir, A., Skidmore, A.K., Toxopeus, A.G. & Real, R.

Resubmitted after revision to *Global Ecology and Biogeography*

Parts of this chapter also appear in:

- Niamir, A. et al. “"Accounting for species taxonomy improves distribution models"”, in the proceedings of The International Statistical Ecology Conference 2014, Montpellier, France

data
presence
irrelevant
discrimination
However
area
improves
calibration
localities
capacity
atlas
relatively
Figure
shows
spatial
generalized
Lobo
modelling
dataset
absence
used
datasets
geographic
algorithms
trees
present
extent
family
conditions
recorded
latasti
results
order
generally
similar
based
grid
potential
geographical
environmental
assessed
probability
using
considered
appropriate
three
use
propose
including
orders
model
study
taxonomical
improved
may
target
absences
taxonomy
relevant
reliable
effect
Vipera
records
performance
inclusion
number
predict
common

3.1 Abstract

The use of species atlas data and correlational modelling approaches has become common practice in species distribution modelling. However, the effect of spatial extent on the performance of species distribution models remains unclear. We propose using species taxonomy to delineate environmental extent (and ipso facto study area), and assess how this affects the performance of species distribution models. Using atlas data, two taxonomically enhanced datasets were created for 356 terrestrial species from mainland Spain. As well as a species' presence, its absence was recorded. However, an absence was only recorded for localities where another species of the same family or order was known to be present. We then assessed the effects of environmental extent on the predictive performance of five species distribution models, namely on model calibration (Miller's statistic) and discrimination (area under the curve of the receiver operating characteristic plot), comparing the results to those obtained with the original atlas dataset. Using family and order datasets significantly improved the calibration of the species distribution models. While the discrimination capacity decreased for several species when using the family dataset, this drop was significant when the models were trained with the order dataset. The species distribution model algorithms differed only slightly in performance. Enhancement of atlas data by excluding irrelevant absences based on species taxonomy significantly improved the calibration of species distribution models and thus their capacity to predict environmental potential. We therefore recommend the inclusion of species taxonomical extent when atlas data are employed in distribution models.

3.2 Introduction

Using species atlas data in combination with a variety of modelling approaches has become a common practice in species distribution modelling. However, the effect of the spatial extent, being considered for a particular species, on the performance of the distribution models remains unclear. For the majority of species, atlases provide the most reliable and convenient sources depicting the area of distribution of a species (Araujo *et al.* 2005; Franklin 2009; Niamir *et al.* 2011). Species atlas data have been widely used to study species distribution patterns (Sillero *et al.* 2009; Barbosa *et al.* 2012) as well as for undertaking conservation planning at varying spatial and temporal scales (Hopkinson *et al.* 2000; Rouget 2003; Barbosa *et al.* 2010; Bombi *et al.* 2012). The extent of an atlas is often in accordance with geopolitical boundaries. Atlas data are usually a compilation of historical observations, museum records and, sometimes, complementary systematic field surveys. A species is marked 'present' if there is at least one record of that species within the extent of a pre-defined grid cell, and marked 'absent' for any grid cell without a record of the species. Customarily, the extent of the study area used for species distribution modelling is identical to the area reported in the atlas including all present and absent gird cells, regardless of the range of the species and without biological justification (Meyer & Thuiller 2006).

There have been attempts to investigate the effect of the extent of a study area on the performance of species distribution models (Jiménez-Valverde *et al.* 2008; Barve *et al.* 2011; Acevedo *et al.* 2012). The discrimination capacity of the models has shown a positive association with the extent of the study area (Lobo *et al.* 2010; Lobo & Tognelli 2011). Including more uninhabited and unsuitable localities – which are traditionally considered absences or background – improves the discrimination capacity of the models (Lobo *et al.* 2008). This is the case with specialist species that have a relatively narrow distribution caused by biological interaction or physical environment. By contrast, delimiting the study area to unoccupied, but

suitable, localities increases the calibration and informational contents of the distribution model (Anderson & Raza 2010; Acevedo *et al.* 2012), although it may reduce the discrimination capacity. Distribution models for generalist species with a broader environmental envelope (Carpenter *et al.* 1993) struggle to discriminate between suitable and unsuitable localities (Jiménez-Valverde *et al.* 2008).

The geographic range of a species is a basic unit of biogeography (Brown *et al.* 1996). Setting appropriate geographic range limits for a species is critical when creating reliable distribution models (Roy *et al.* 2009). However, the selection of spatial and environmental limits has been largely ignored by modellers (Hortal *et al.* 2012). These limits arise from complex interactions between biological interactions **B,** the physical environment **A,** and the dispersal ability **M** of a species (see BAM diagram Soberon and Peterson 2005). **M** is the geographic region that has been accessible to the species over relevant periods of time. The extent of this region is determined by two factors operating at different intensity and scale: the region the species has had access to, dispersing from some original area, and the evolutionary capacity the species has shown in adapting to new conditions (Soberón & Peterson 2005; Soberon & Nakamura 2009). Acevedo et al. (2012) assessed the use of geographic criteria, such as trend surface analysis, to restrict the inclusion of absence records from beyond the geographic domain of the species in distribution models. However, they recognized that a spatial pattern generated with trend surface analysis cannot be considered a geographical representation of **M**, as the concepts underlying each approximation do not necessarily converge on the same geographical space. Brave et al. (2011) proposed the use of information related to the dispersal capacity of a species to define an appropriate extent for modelling purposes. The ideal extent for model development should comprise a set of localities that has been accessible to the species of interest over relevant time periods.

An appropriate extent for studying the distribution of a target species includes 'reliable' presences as well as reliable absences. Recording reliable

absences and therefore excluding unknown/not-sampled localities, requires more effort than recording presences (Mackenzie 2005). Therefore, inclusion of 'not', or 'inadequately', surveyed localities in the modelling procedure affects the performance of species distribution models. In this study, we propose a species-specific, practicable procedure to refine the extent of a study area, and consequently, the extent of certain environmental conditions, based on species taxonomy. We argue that to develop a reliable species distribution model, the extent of the study area should be appropriately defined. This may be done by assuming that in the atlas data "presence" localities relating to other species of the same family or the same order may be used to qualify the absence localities of the target species as true and relevant. Considering a species' ancestors' distribution would minimize inclusion of irrelevant absences. Since atlas data are basically a compilation of observation reports, the probability of unnoticed presence of the target species at localities where other species of the same order are recorded, is relatively lower than at localities without such recordings. By irrelevant absences we mean those localities that either due to their environmental remoteness are not informative for an ecological model, or were not sampled specifically for the target species. In this context, our main objective is to propose an approach which incorporates the taxonomy of a species and therefore delineate the extent of study for species distribution model. We assess the effect of this enhancement on the calibration and discrimination capacity of 5 conventional modelling algorithms over 356 terrestrial species in mainland Spain.

3.3 Methods and Materials

3.3.1 Species data

The target species were terrestrial vertebrates of mainland Spain including mammals, birds, reptiles, and amphibians. We obtained the species distribution, consisting of presence and absence data, from the Spanish

Terrestrial Species Inventory* using Universal Transverse Mercator (UTM) 10 × 10 km grid cells (n = 4684). A species was recorded as present in a grid cell if the Spanish Terrestrial Species Inventory database holds at least one reliable record of that species from a location within that grid. The species list was first screened counting the number of UTM grid cells where a species was marked as present, removing species with less than 20 'present' or 'absent' records in the study area to assure an appropriate number of species occurrences to train models. In total 356 species were selected; 81 mammal species from 14 families in 7 orders, 213 bird species from 39 families in 15 orders, 36 reptile species from 7 families in 3 orders, and 26 amphibian species from 6 families in 2 orders.

3.3.2 Taxonomically delimited species occurrence datasets

We generated three species occurrence datasets. In all three datasets the presence of a species was exactly the same as those in the Spanish Terrestrial Species Inventory, while absences differed. The absences of species in the first dataset (called the atlas dataset hereafter) were identical to the absences of species in the national inventory. To generate the second dataset (the order dataset hereafter) the absence of a species was only recorded where at least one other species from the same order was recorded as present.

We extended this restriction to the family level for our third dataset (the family dataset), with the absence of a species being recorded only in the presence of at least one species from the same family. Figure 3.1 illustrates schematically the procedure used to generate the three described datasets.

* Ministry of Agriculture, Food and Environment [Access May 2013]

http://www.magrama.gob.es/es/biodiversidad/temas/inventarios-nacionales/inventario-especies-terrestres/default.aspx

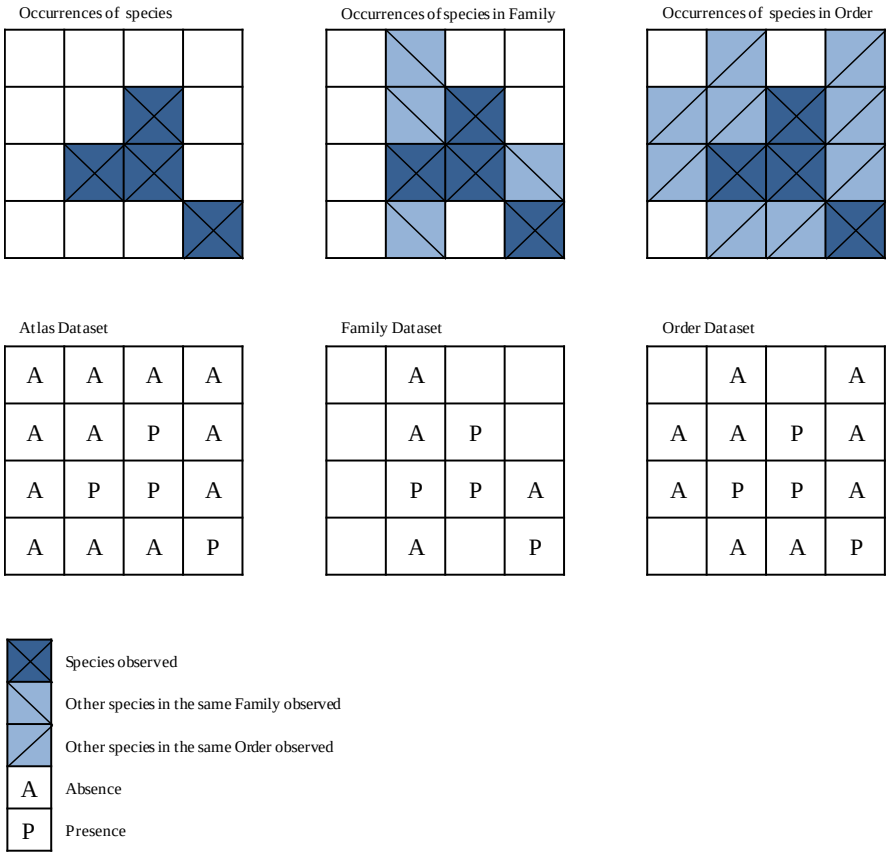


Figure 3.1 Schematic explanation of atlas, family and order datasets

Table 3.1 Environmental predictors used to model the distribution of species and their range of values in the study area (mainland Spain)

Predictor	Code	Range
Mean altitude (m)[a]	*Alti*	0 to 2630
Mean slope (degrees) (calculated from *Alti*)	*Slop*	0 to 19.6
Mean annual precipitation (mm)[b]	*Prec*	154 to 2621
Mean relative air humidity in January at 07:00 h (%)[b]	*HJan*	60.7 to 95.0
Mean relative air humidity in July at 07:00 h (%)[b]	*HJul*	50.6 to 92.4
Mean annual insolation (h/year)[b]	*Inso*	1603 to 3172
Mean annual solar radiation (kwh/m2/day)[b]	*SRad*	304 to 507
Mean temperature in January (◦C)[b]	*TJan*	3.7 to 13.2
Mean temperature in July (◦C)[b]	*TJul*	13.5 to 28.8
Mean annual temperature (◦C)[b]	*Temp*	3.9 to 19.4
Mean annual number of frost days (minimum temperature ≤ 0 ◦C)[b]	*DFro*	0 to 117.7
Mean annual potential evapotranspiration (mm)[b]	*PET*	599 to 1036
Mean annual actual evapotranspiration (mm) (=minimum [*PET*, *Prec*])	*AET*	154.6 to 923.9
Maximum precipitation in 24 h (mm)[b]	*MP24*	52.1 to 474.5
Relative maximum precipitation (=*MP24*/*Prec*)	*RMP*	0.1 to 1.0
Mean annual number of days with precipitation ≥ 0.1 mm[b]	*DPre*	16.1 to 193.2
Annual temperature range (◦C) (=*TJul*−*TJan*)	*TRan*	8.2 to 21.5
Annual relative air humidity range (%) (=|*HJan*−*HJul*|)	*HRan*	0.1 to 41.3
Distance to the nearest highway (km)[c]	*DHi*	0.5 to 105.2
Distance to the nearest town with more than 100,000 inhabitants (km)[c]	*U100*	0.1 to 154.0
Distance to the nearest town with more than 500,000 inhabitants (km)[c]	*U500*	0.1 to 301.9
Latitude (◦N)[c]	*Lati*	36.1 to 43.7
Longitude (◦E)[c]	*Long*	-9.3 to 3.3

[a] U.S. Geological Survey (1996);
[b] Font (1983, 2000);
[c] I.G.N. (1999); data on the number of inhabitants of urban centres taken from the Instituto Nacional de Estadística (http://www.ine.es) for Spain.

3.3.3 Species distribution models

Five distribution models, specifically the generalized linear model (Mccullagh 1989), generalized additive model (Hastie & Tibshirani 1990), boosted regression trees (Friedman 2001), random forests (Breiman 2001b), and support vector machine (Drake et al. 2006), were tested for each species using the three datasets and 23 environmental predictors (Barbosa et al. 2009) related to topography, climatology, and human activity (Table 3.1). The environmental predictors were chosen on the basis of availability and potential predictive power for terrestrial species in Spain (Barbosa et al. 2003; Naimi et al. 2011).

The generalized linear model (GLM) uses a parametric function to link the response variable to a linear, quadratic or cubic combination of explanatory variables. We used a GLM ordinary polynomial with an automatic stepwise model selection based on the Akaike Information Criterion. The generalized additive model (GAM) uses nonparametric functions to fit nonlinear functions. In this study, a GAM with a cubic spline smoother and an automated stepwise process was employed. The boosted regression trees (BRT) algorithm fits complex nonlinear relationships by combining two algorithms of regression trees and boosting; the former to relate a response to their predictors by recursive binary splits, and the latter to combine many single models to improve the performance. We followed the recommended default settings: viz. a maximum number of trees equal to 3000 and a learning rate of 0.001 (Elith et al. 2008). The random forests (RF) algorithm selects many bootstrap samples from the data and fits a large number of regression trees to each of these subsamples. Each tree is then used to predict those that were not selected as bootstrap samples. The classification is given by considering each tree as a 'vote', and the predicted class of an observation is determined by the majority vote among all trees (Cutler et al. 2007). The models presented here used 500 trees. The support vector machine (SVM) is a machine-learning generalised linear classifier that estimates the potential distribution of a species subject to the

environmental conditions, by separating the ecological space by hyper-planes into suitable and unsuitable environmental conditions (Reiss et al. 2011). The optimality criterion used to find the separating hyper-plane is the maximised distance to the training data points. All models were executed in the R environment v. 3.0.1 (R Development Core Team 2013).

3.3.4 Model evaluation

The area under the curve of the receiver operating characteristic (ROC) plot was computed to assess the discrimination capacity of the models for each data set. A ROC curve plots sensitivity values (true positive fraction) on the y-axis against 1– specificity values (false positive fraction) for all thresholds on the x-axis. The area under the curve (AUC) is a single measure of model discrimination capacity, which provides a threshold-independent measure across all possible classification thresholds for each model (Fielding & Bell 1997). We randomly split the data using 70% to train the models and the remaining 30% to measure the discrimination capacity of the models.

We also assessed the calibration of the models using Miller's calibration statistic (Miller *et al.* 1991; Pearce & Ferrier 2000) based on whole datasets for each species. Miller's calibration statistic evaluates the ability of a distribution model to correctly predict the proportion of species occurrences with a given environmental profile. It is based on the hypothesis that the calibration line – perfect calibration – has an intercept of zero and a slope of one. The calibration plot shows the model's estimated probability (x-axis) against the mean observed proportion of positive cases (y-axis) for 10 equally sized probability intervals. To be able to compare the models' performance we calculated the Root Mean Square Error (RMSE) of the calibration graph for each model (Armstrong & Collopy 1992).

The effect of taxonomic extent – varying in three datasets – on the model performance measures for each species was assessed using all five species distribution models. We assessed the variation in discrimination capacity (AUC), and calibration (RMSE) of the models based on the atlas, order, and family datasets using the Mann–Whitney U test (Hollander & Wolfe 1973).

3.4 Results

Delimiting absences using species taxonomy (Table 3.2) reveals that the inclusion of family and order datasets significantly improved the calibration of the models. While discrimination capacity dropped for several group of species and modelling techniques when using the family dataset, this drop became more significant when the models were trained with the order dataset. Overall, the calibration of models improved for 83% of species for both the family and order dataset, while the discrimination capacity decreased in 51% and 43% of species respectively.

Amphibians benefited most: 90% for calibration and 74% for discrimination, when using the family dataset. Calibration improved for 75% of the bird species with both the family and the order dataset (Figure 3.2). For some species the drop in discrimination capacity was significant. Closer inspection of these species revealed that either they were geographically widely spread, or they belonged to narrowly spread families/orders, resulting in few absence occurrences remaining after correction with the atlas data. Those species for which delimiting their extent sharply downgraded the calibration of the models were mainly specialist species with geographically limited occurrences. Figure 3.3 illustrates the distribution maps obtained using atlas, family and order datasets and five species models for *Vipera latasti* as an example (See Appendix I for more examples).

Table 3.2 Summary of results of the SDM algorithms used to assess the effect taxonomical extent has on the performance of five models. Numbers represent the proportion of species models with improvement in their performance measures.

		SDM algorithms				
		GLM	GAM	RF	BRT	SVM
Calibration (RMSE)		**Family dataset**				
	Mammals	0.91 (+ **)	0.90 (+ **)	0.82 (+ ns)	0.91 (+ **)	0.75 (+ **)
	Birds	0.88 (+ **)	0.69 (+ **)	0.82 (+ *)	0.74 (+ **)	0.60 (+ **)
	Reptiles	0.80 (+ **)	0.77 (+ **)	0.89 (+ *)	0.75 (+ **)	0.91 (+ **)
	Amphibians	0.92 (+ **)	0.88 (+ **)	0.88 (+ *)	0.96 (+ **)	0.88 (+ **)
		Order dataset				
	Mammals	0.91 (+ *)	0.87 (+ **)	0.85 (+ ns)	0.86 (+ *)	0.82 (+ **)
	Birds	0.76 (+ **)	0.74 (+ **)	0.84 (+ *)	0.76 (+ **)	0.67 (+ *)
	Reptiles	0.77 (+ *)	0.80 (+ **)	0.88 (+ ns)	0.72 (+ **)	0.91 (+ **)
	Amphibians	0.84 (+ **)	0.84 (+ *)	0.88 (+ ns)	0.92 (+ *)	0.96 (+ **)
Discrimination (AUC)		**Family dataset**				
	Mammals	0.59 (+ ns)	0.57 (+ *)	0.45 (- *)	0.64 (+ *)	0.68 (+ ns)
	Birds	0.48 (- *)	0.38 (- **)	0.41 (- *)	0.40 (- *)	0.47 (- *)
	Reptiles	0.39 (- ns)	0.38 (- ns)	0.36 (- ns)	0.55 (+ ns)	0.56 (+ *)
	Amphibians	0.73 (+ *)	0.77 (+ ns)	0.69 (+ *)	0.80 (+ ns)	0.73 (+ ns)
		Order dataset				
	Mammals	0.69 (+ *)	0.54 (+ *)	0.58 (+ ns)	0.64 (+ *)	0.68 (+ **)
	Birds	0.47 (- **)	0.46 (- **)	0.47 (- **)	0.47 (- *)	0.48 (- **)
	Reptiles	0.41 (- ns)	0.44 (- *)	0.31 (- *)	0.56 (+ ns)	0.44 (- ns)
	Amphibians	0.42 (- ns)	0.34 (- *)	0.27 (- *)	0.61 (+ ns)	0.53 (+ ns)

+ : General improvement in calibration or discrimination capacity; − : General decrease in calibration or discrimination capacity; **: P<0.05; *: P<0.1; ns: non-significant

The species distribution model algorithms differed slightly in performance when trained using the family and order dataset. Variation of changes using the RF algorithm was relatively small compared to the other algorithms. GLM and GAM were affected most by both the calibration and discrimination measures. Discrimination capacity remained stable in the SVM algorithm while calibration improved for 79% and 84% of the species using family and order datasets, respectively. Figure 3.4 illustrates the variation in calibration and discrimination capacity of the models for all species.

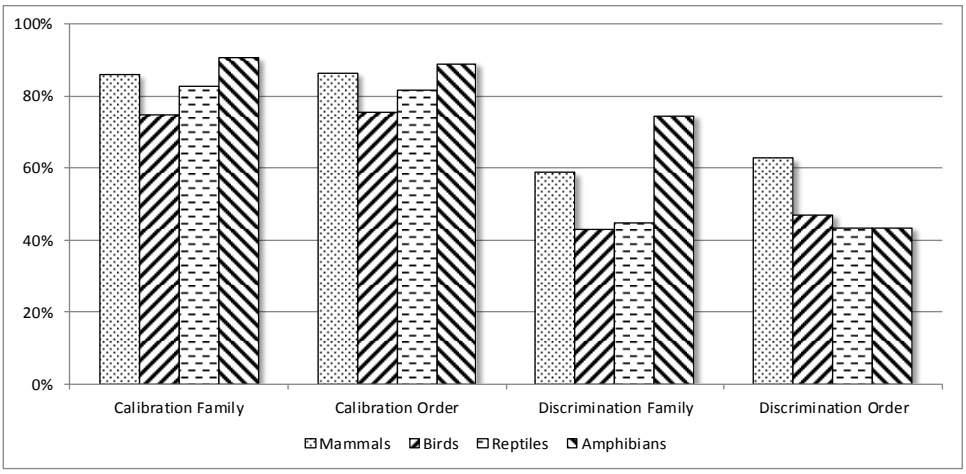


Figure 3.2 Proportion of species that delineating their extent – using family and order dataset – improved calibration and discrimination capacity of the distribution models

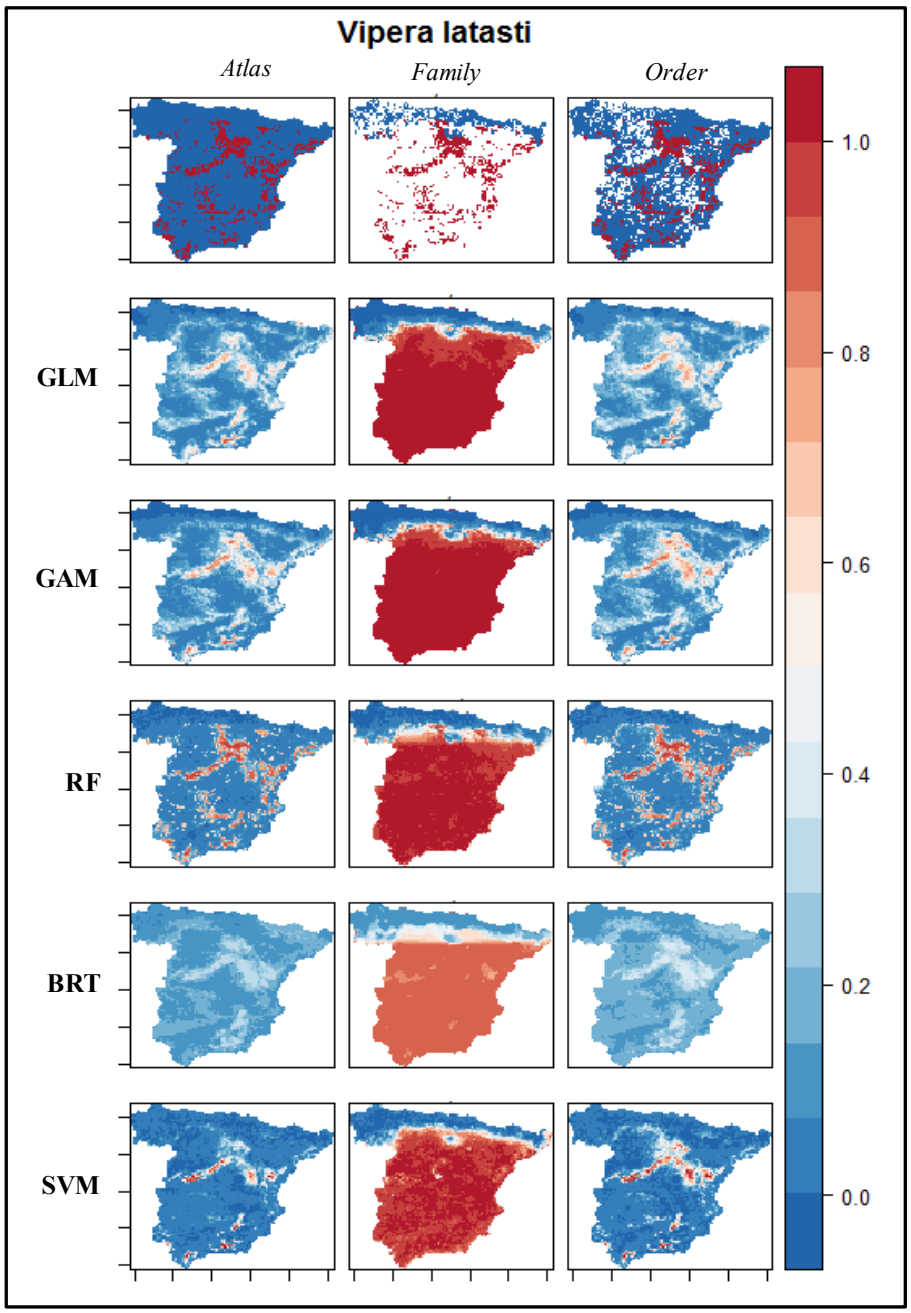


Figure 3.3 Species distribution maps obtained using atlas dataset as well as taxonomical extents (family and order) to refine five species models for *Vipera latasti* (an example). In the top row the red dots indicate species presences and the blue dots indicate species absences per dataset.

Figure 3.4 Variation of changes in calibration and discrimination capacity per model when the extent of the study area is enhanced by the extent of the family (left column) and the order (right column) of the species. Calibration of distribution models for species in right quadrants, as well as discrimination capacity of those in upper quadrants improved when the absence localities were enhanced by taxonomical information. Blue rectangles represent mammals, light green circles represent birds, red up-pointing triangles represents reptiles, and purple down-pointing triangles represents amphibians.

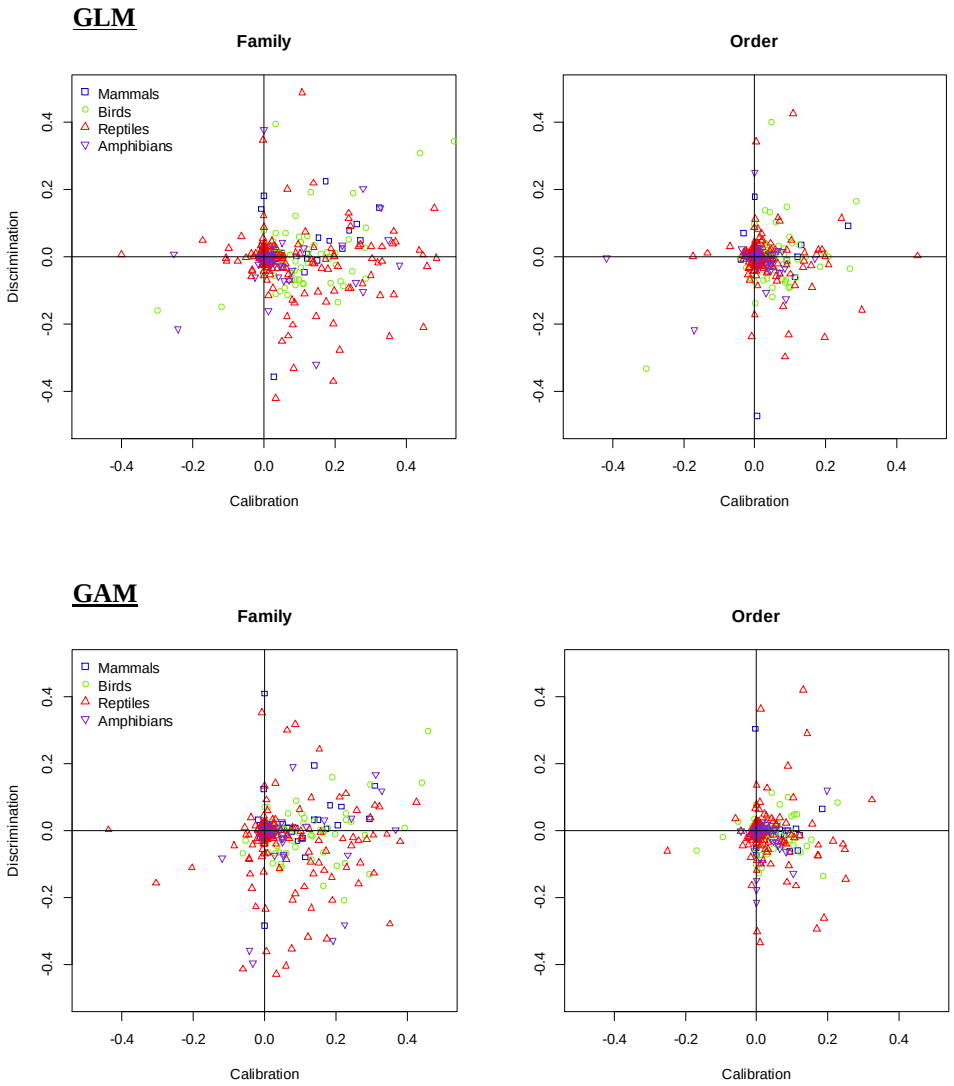


Figure 4.4 (continued)

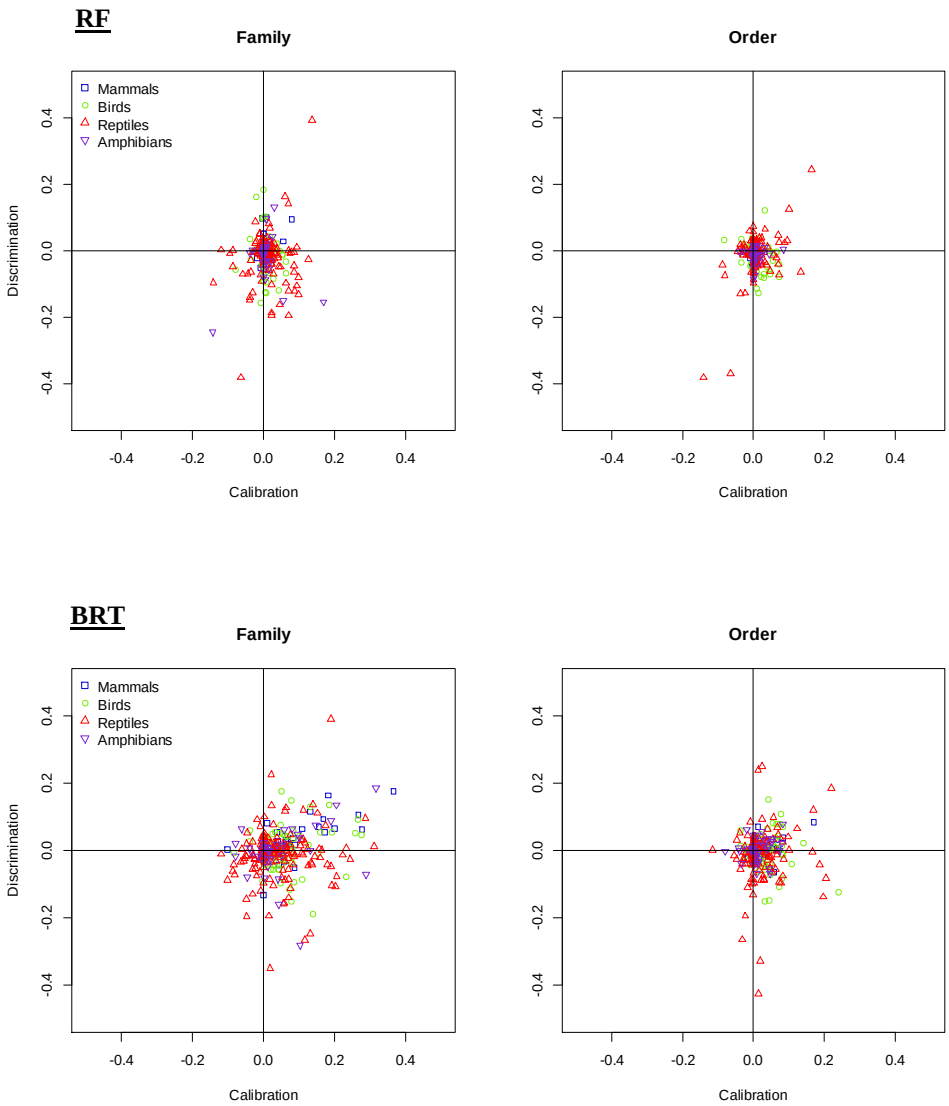
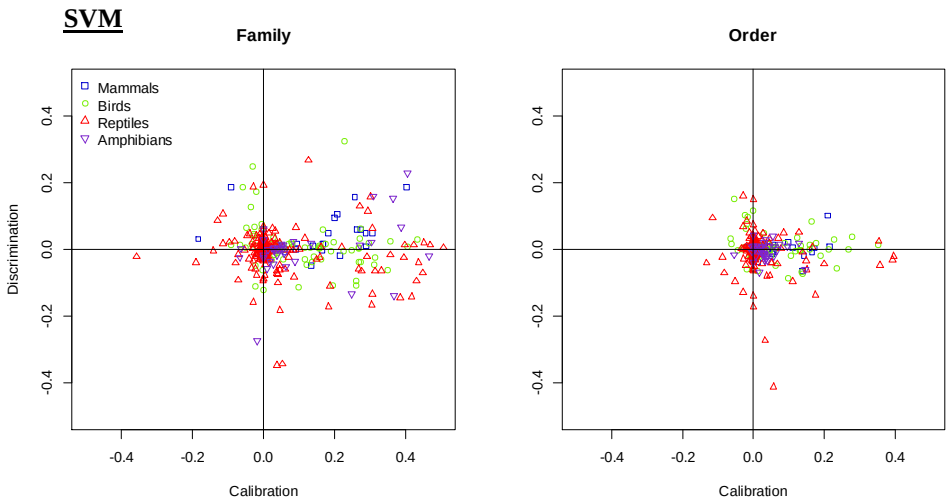


Figure 4.4 (continued)



3.5 Discussion

Enhancement of the atlas data, through the exclusion of irrelevant absences based on species taxonomy generally improves the calibration of species distribution models and thus their capacity to predict environmental potential for a species. Our proposed approach demonstrates a species-specific and practical way to use the atlas data to define an appropriate extent for species distribution models within a specified study area.

Our results confirm that different species datasets – each differing in the way absences are included – condition the results of species distribution models, as has been earlier suggested (Engler *et al.* 2004; Pearce & Boyce 2006; Elith & Leathwick 2007; Jiménez-Valverde *et al.* 2008; Lobo *et al.* 2010). In addition, our analysis shows the effect in terms of calibration and discrimination variation, of refining models by using relevant absences. Increasing the extent of the study area by including unobserved localities as absences improves discrimination, as the additional absence localities tend to differ environmentally from the presence localities and be discriminated

more easily (Lobo *et al.* 2008). When transferring such models in space and time, calibration provide more information about the potential performance of the models than the discrimination capacity, as discrimination is context specific (Jiménez-Valverde *et al.* 2013).

Our restriction of the study extent sometimes results in less discrimination, as presences and remaining absences were environmentally closer and, consequently, more difficult to discriminate. The majority of models that show a decline in their discrimination capacity belong to species whose order or family occupy relatively small areas or to highly prevalent species within the area occupied by their family or order. The opposite is the case for species with a wider taxonomical extent. Since the extent of a species' family is always equal to or smaller than the extent of the order it belongs to, fewer absences were considered irrelevant and excluded when considering the order than when considering the family. This explains the higher variation in discrimination capacity of models using a family dataset (std. = 0.0855) compared with models using an order dataset (std. = 0.0493).

Calibration generally improves when models use enhanced species datasets. If absences from beyond the taxonomical extent of the species are excluded, then the model will generate a more reliable probability of presence. Extreme improvements can be seen with generalist species distributed over relatively large extents. Specialist species, especially those belonging to widely distributed taxa, do not benefit much in terms of calibration when their taxonomical extent is considered, mainly because the calibration of their models was already accurate. In summary, generally, enhancement of species atlas data based on taxonomical extent produces more informative models.

To our knowledge few studies have presented a framework for including the extent of a study area in species distribution modelling. Our finding supports discussions by Acevedo et al. (2012) on geographical criteria, and Brave et al. (2011) on dispersal history in species distribution models. They

both suggest interesting approaches to delimit **M** from the point of view of geographical barrier and species dispersal capacity. In this study we propose a practical way to set the extent of a study area based on species taxonomical extent, and demonstrate that this approach enhances model performance in terms of calibration and discrimination. There may be two reasons for this enhancement. Firstly, absences become more reliable when taxonomical extent has being considered. Since atlas data are basically compilations of historical reports, the probability of an undetected presence of the target species at localities that do have a record of other species of the same taxon is lower than at localities without such a record. If neither the target species nor any other species of the same order or family is present in a location, the environmental conditions in that location may be hostile to the whole group of species. In addition, species from the same taxon tend to have similar biological interactions and abiotic preferences, and reports of absences from certain locations guarantee that habitats similar to those of the target species were sampled, so decreasing the probability of undetected presence. This probability decreases further by considering other species of the same family. Secondly, absences have also become more relevant. Restricting the extent of the study area by using taxonomy helps restrict the environmental domain of the training dataset to a historically relevant domain for the species, as taxonomy at order and family level is based on phylogeny, which implies a common ancestry and an evolutionary history in which the different lineages had to respond to common or similar environments in differing ways (Brown 1995; Brown *et al.* 1996). Figure 3.3 shows another example of irrelevant absences, and also of the situations that make a true absence relevant or not. In this case the order dataset produced a model similar to that of the atlas dataset, while the family dataset produced a different model. Models for *Vipera latasti* considered the species as belonging to Mediterranean mountains when using all absences or absences occupied by other snake species. However, when absences were restricted to those occupied by other viper species, then *Vipera latasti* was considered as a Mediterranean

species, only absent from the Euro-Siberian part of the Iberian Peninsula. This is so because no viper species is present in flat areas due to human persecution and, thus, absence from these areas are not relevant to distinguish the distribution of *Vipera latasti* from those of other vipers. Consequently, if the aim is to infer where *Vipera latasti* is more likely to be found nowadays, then atlas and order datasets are relevant, but if the aim is to characterize *Vipera latasti* according to natural constraints, then the family dataset is more relevant.

As the use of species atlas data is currently growing exponentially (Franklin 2013), we want to draw attention to potential misuse when applied in an uncritical manner. This study shows that family and order data improved model performance and should be incorporated where atlas data are used to model species distribution. It is increasingly recognized that, ideally, the extent of the study area for species distribution modelling should consist of reliable presence records as well as reliable and relevant absence records in order to enhance understanding and prediction (Lobo *et al.* 2010; Hortal *et al.* 2012). The point emphasized here is that when the absence records are basically the undetected records or due to the factors that are irrelevant for our modelling objectives, then excluding the localities of other family or order members improves the quality of a species' absence data. We have shown that calibration of species distribution models generally improves when the species data are enhanced by taxonomical information.

3.6 Acknowledgments

The study was jointly supported by the University of Twente and the European Union Erasmus Mundus program External Cooperation Windows, through the agreement 2008/2441/001 MUN ECW 15-7-2008 and by the project CGL2009-11316/BOS from the Spanish Government and FEDER. The foundation of this study was developed during the short-term scientific mission of AN to the University of Malaga in April 2013 founded by the European Cooperation in Science and Technology

programme on Harmonizing Global Biodiversity Modelling (HarmBio) COST-STSM-ECOST-STSM-ES1101-240313-025740. The language editing was done by Eva Skidmore. Many thank to Antoine Guisan the three anonymous reviewers for their constructive comments on a previous version of this paper.

Chapter 4:

ACCOUNTING FOR KNOWLEDGE UNCERTAINTY IN SPECIES DISTRIBUTION MODELLING

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Munoz, A.R. & Real, R.

Resubmitted after revision to *Journal of Biogeography*

Parts of this chapter also appear in:

- Niamir, A. et al. "Using Dempster-Shafer's evidence theory for species distribution modelling ", in the proceedings of The International Statistical Ecology Conference 2014, Montpellier, France
- Niamir, A. et al. "Dempster-Shafer Evidence Theory for improved species distribution mapping", in the proceedings of Spatial statistics 2011: mapping global change, 2011, Enschede, The Netherlands

Dempster-Shafer
cover
function
However
experts
inductive
theory
given
variability
either
one
model
case
variables
knowledge
provides
obtained
sources
higher
assess
occurrence
rule
BRT
Figure
using
maps
independent
asked
environmental
models
GLM
DST
two
equal
range
performance
land
respectively
European
relatively
methods
scale
Sup
also
always
belief
level
information
confidence
compared
habitat
applied
similar
requires
questionnaires
confident
variable
discrimination
AUC
poorly-studied
therefore
general
conventional
mainly
slope
trees
measures
areas
plausibility
eagle
events
proposed
favourable
ignorance
express
based
Experts
values
elevation
procedure
domain
i.e
particular
temperature
short-toed
framework
combined
well-known
map
well
well-calibrated
performed
sufficient
aspect
support
evidential
calibration
data
probabilities
Spain
distribution
study
process
modelling
predictors
target
capacity
degree
functions
approach
classes
results
area
evidence
Bonelli
expert
used
panel
Dempster
lower
breeding
set
low
online
uncertainty
selected
assigned
Unc
combination
probability
may
conservation
power
defined
bioclimatic

4.1 Abstract

Conservation managers rely on the knowledge of experts where information describing species distribution is scarce or unobtainable. However, experts' knowledge is always subject to uncertainty and accounting for this uncertainty into a modelling procedure poses a challenge. We propose using the Dempster-Shafer Theory of Evidence (DST) to accommodate knowledge uncertainty, and assess how this affects the performance of species distribution models. We applied DST to model the distribution of a well-known, and a poorly-studied species in mainland Spain. We invited experts to form a knowledge domain, and asked them through online questionnaires to express their knowledge on the habitat of the target species by assigning a probability value for a given environmental variable. Experts were also asked to acknowledge their confidence level on the assigned probability values. Then we calculated evidential belief functions and combined them using Dempster's rules of combination. We evaluated and compared the calibration and discrimination capacity of the DST models with two other conventional inductive models. The DST models yields similar results to conventional inductive methods for both species. Although the habitat of the well-known species was well-discriminated, DST models were generally underestimating the habitat favourability. In contrast, the models for the poorly-studied species were well-calibrated though the discrimination capacity remained low. Our proposed approach offers a practical alternative where knowledge of a species' geographic distribution is needed, and the distribution data is not sufficient. The particular strength of the developed approach is that it explicitly accommodates knowledge uncertainty. It also provides a framework to propagate and aggregate uncertainty, and it capitalizes on the range of data sources usually considered by an expert.

4.2 Introduction

Biodiversity conservation requires a reliable description of species distribution over time and space. Since information describing distributions is scarce or unobtainable for a majority of species (Newbold 2010), conservation managers can rely on experts as an alternative source of knowledge (Murray *et al.* 2009; Martin *et al.* 2012). Experts provide valuable insight into species-environment interactions and may come up with "beliefs" concerning the occurrence of species or offer predictions given specific evidence (Kuhnert 2011). In practice, however, incorporation of expert knowledge into a species distribution modelling procedure presents challenges (Ferrier *et al.* 2002), and has not yet been widely promoted (Carpenter 2002; Franklin 2009).

"Beliefs" are the simplest form of mental representation in which experts express their knowledge. Expert knowledge on species distribution, either gained through direct observation or deducted from indirect sources of information, is always subject to uncertainty. Much literature exists discussing uncertainty in expert knowledge and describing analysis frameworks (Reddy 1996; Refsgaard *et al.* 2007; Janssen *et al.* 2010). Walker et al. (2003) categorized the nature of uncertainty into epistemic uncertainty and stochastic uncertainty. Stochastic uncertainty is due to inherent, non-deterministic variability, e.g. climate variability. Such uncertainty is unavoidable and will always be present when dealing with nature. In contrast, the epistemic uncertainty – or recognised ignorance (Brown 2004) – is due to imperfect knowledge that may be reduced with study and research. Uncertainty in expert knowledge stems from both incompletely understanding of a phenomenon, an incomplete ability to accurately generalize knowledge beyond the scope of observations. Therefore, accounting for recognized ignorance in the process of knowledge acquisition may improve the quality of elicited knowledge and helps to distinguish this epistemic uncertainty from the innate uncertainty of natural phenomena.

Various strategies have been developed for knowledge elicitation and incorporation into species distribution modelling procedures. Multi-criteria decision methods such as the analytical hierarchy process (Anselin *et al.* 1989; Store & Kangas 2001; Doswald *et al.* 2007), expert system formalisms implemented in GIS (Yang *et al.* 2006), and fuzzy sets theory (Rüger *et al.* 2005) have been widely applied in habitat suitability modelling. However, empirical models tend to outperform mentioned models when sufficient species location data are available. Another approach is to conduct expert knowledge using Bayesian models (Choy *et al.* 2009; Kuhnert *et al.* 2010; Niamir *et al.* 2011), which accommodate the experts' knowledge in the form of probabilities. This probabilistic analysis requires that an expert has information on the probability of either of two events (presence or absence), which is unlikely in the case of species distribution modelling.

The Dempster-Shafer Theory of Evidence (DST) (Dempster & Weisberg 1968; Shafer 1976) provides a flexible framework for approaching the representation of knowledge and therefore the recognized ignorance. This theory draws inferences from incomplete and uncertain knowledge provided by various independent knowledge sources. An advantage of this theory is its ability to deal with ignorance. In particular, it provides explicit estimation of the extent of imprecision and conflict resulting from knowledge provided by various experts, and can deal with any union of hypotheses (Le Hegarat-Mascle *et al.* 1997).

The DST handles ambiguous and incomplete information using evidential functions. The evidential functions can then be used to represent incomplete knowledge, and to distinguish between the lack of information and obverse information (An *et al.* 1994). Evidential functions can be derived from probabilities (Lee *et al.* 1987; Rombaut & Min Zhu 2002), from the distance to cluster centres (Bloch 1996), or from fuzzy membership functions (Binaghi *et al.* 2000; Boudraa *et al.* 2004). The DST has been successfully applied in the geo-information domain (Malpica *et al.* 2007), particularly in remote sensing and image processing (Mertikas &

Zervakis 2001), decision making (Beynon *et al.* 2001), and uncertainty management (Comber *et al.* 2004; Clements *et al.* 2006; Baraldi & Zio 2010).

The purpose of this paper is to demonstrate the application of the DST to species distribution modelling as well as the value of applying evidential functions using experts' knowledge solely. To assess the performance of the proposed approach which utilize the quality of expert's knowledge, we modelled the distribution of a well-known species, the Bonelli's eagle (*Aquila fasciata*), as well as a poorly-studied species, the short-toed eagle (*Circaetus gallicus*) in Spain. We formed an independent expert panel and derived the evidential functions from an online questionnaire to test the feasibility of applying such an approach for extracting and incorporating experts' knowledge uncertainty in species distribution modelling. This approach facilitates the aggregation of knowledge from multiple experts, and is applicable in a wide variety of conservation purposes.

4.3 Explanation of Dempster-Shafer theory of evidence

The DST is a generalisation of traditional probability that allows uncertainty to be quantified (Dempster & Weisberg 1968; Shafer 1982). The DST has conceptual differences from Bayesian Theory. Bayesian theory requires an explicit formulation of conditioning and prior probabilities of events. DST embeds conditioning information into its belief function, making it appropriate for situations where it is difficult to either collect or posit such probabilities, or isolate their contribution (Lee *et al.* 1987; Hoffman & Murphy 1993; Comber *et al.* 2004).

This theory first supposes the definition of a set of hypotheses Θ called the frame of discernment, defined as follows:

$$\Theta = \{ H_1 , H_2 , \ldots , H_N \}.$$

It is composed of N hypotheses. Let's denote $P(\Theta)$ as the power set composed with the 2^N propositions A of Θ, where $\varnothing$ denotes the empty set:

$$P(\Theta) = \{ \varnothing, \{H_1\}, \{H_2\}, \ldots, \{H_N\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \ldots, \Theta\}.$$

A key point of evidence theory is the Basic Probability Assignment (BPA) which is a function from $P(\Theta)$ to [0,1] defined by:

$$m: \quad P(\Theta) \rightarrow [0,1]$$
$$A \rightarrow m(A)$$

and satisfies the following conditions:

$$\sum_{A \in P(\Theta)} m(A) = 1; m(\varnothing) = 0$$

From a general point of view, contrary to probability theory, which assigns the probability mass to single elementary events, the DST makes basic probability assignments *m(A)* on sets of outcomes. The basic belief assignment *m(A)* expresses the degree of belief that a specific element *x* belongs to the set *A* only, and not to any subset of *A* (Baraldi & Zio 2010).

Unlike probability theory that imposes restrictive conditions on the specification of the likelihood of events as a result of the requirement that the probabilities of the occurrence event must sum to one, there are two measures of likelihood in the DST called evidential belief functions. The evidential belief functions are the degree of support (*Sup*) and the degree of

plausibility (*Pls*), which represent the upper and lower probability that the available evidence supports a particular proposition (Dempster 1967), as well as the degree of uncertainty (*Unc*) and the degree of disbelief (*Dis*) (Figure 4.1). Thus, plausibility is always equal or larger than support, and subtracting support from plausibility reveals uncertainty representing the ignorance in one's belief. If there is no doubt, support is equal to plausibility. However, if support and disbelief are set to zero, then there can only remain uncertainty representing full ignorance, given the evidence.

The Dempster's (1968) rule of combination is the first one within the framework of evidence theory which can combine two BPAs m_1 and m_2 to yield a new BPA:

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-k},$$

with

$$k = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$$

where k measures the degree of conflict between m_1 and m_2, $k = 0$ corresponds to the absence of conflict, whereas $k = 1$ implies complete contradiction between m_1 and m_2. The belief function resulting from the combination of J information sources S_J is defined as

$$\mathrm{m} = \mathrm{m}_1 \oplus \mathrm{m}_2 \ldots \oplus \mathrm{m}_j \ldots\ldots \oplus \mathrm{m}_J$$

This provides a framework for the estimation of evidential belief functions (Dempster 1967; Shafer 1976), which are integrated according to Dempster's rule of combination. For each spatial variable used as an

evidence, two independent functions must be estimated, either support and disbelief, or support and plausibility (Chung & Fabbri 1993; An et al. 1994). Given evidences X1 and X2, each with Sup and Dis functions, the combined Sup, Dis, and Unc are shown in the following equations (Wright & Bonham-Carter 1996; Carranza & Hale 2003):

$$Sup_{X_1X_2} = \frac{Sup_{X_1}Sup_{X_2} + Sub_{X_1}Unc_{X_2} + Sub_{X_2}Unc_{X_1}}{\beta} ,$$

$$Dis_{X_1X_2} = \frac{Dis_{X_1}Dis_{X_2} + Dis_{X_1}Unc_{X_2} + Dis_{X_2}Unc_{X_1}}{\beta} ,$$

$$Unc_{X_1X_2} = \frac{Unc_{X_1}Unc_{X2}}{\beta} ,$$

$$\beta = 1 - Sup_{X_1}Dis_{X_2} - Dis_{X_1}Sup_{X_2} ,$$

where β ensures that the sum of *Bel*, *Dis*, and *Unc* is equal to 1.

The orthogonal sum thus allows two functions to be combined into a third function, which effectively pools pieces of evidence to support propositions of interest and multi-source information can be easily fused in the framework of evidence theory. Derivation of the evidential functions is the most crucial step since it represents the knowledge as well as the uncertainty surrounding selected evidence. This step requires prudence when applying DST to species distribution modelling. It can be performed by statistics (inductive methods), if a study area is well sampled (An *et al.* 1994; Carranza *et al.* 2005) and there are sufficient occurrences of the target species. In situations where the occurrence dataset is insufficient and the study area is not well-studied, the assignment of function values must

rely deductively on knowledge and expertise (Moon 1989; Tangestani & Moore 2002). In this study we used a panel of experts to assign the function values for the target species.

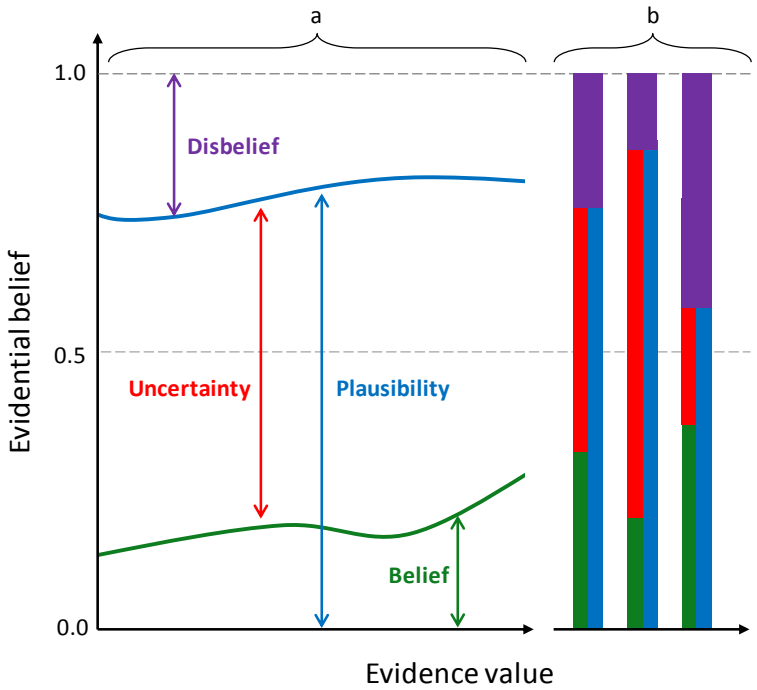


Figure 4.1 Schematic relationship of evidential belief functions for (a) a continuous, and (b) for a categorical environmental domain.

4.4 Experimental settings

4.4.1 Study area

The study area is mainland Spain, covering 493,518 km^2. Spain may be divided into three climatic areas: Atlantic, Mediterranean and Interior. The Atlantic area has mild winters and relatively cool summers, and abundant precipitation throughout the year. The Mediterranean part is characterized by hot summers and mild winters with low rainfall occurring mainly during spring and autumn. Hot summers, mild winters, and scarce precipitation characterize the Interior (Capel Molina 1981). Forests are widely distributed in Spain and cover approximately half of the country (Costa *et al.* 1997). Spain also comprises important mountain ranges, situated mainly in the north (Pyrenees and Cantabrian mountains) and the southeast (Baetic mountains), with relevant mountain chains traversing central Spain from west to east (Iberian and Central systems). Ecologically, mainland Spain is a relatively homogenous unit with sufficient variability as well as being of particular interest for the target species of this study in a European context.

4.4.2 Choice of species

We selected two raptors for this study; the Bonelli's eagle, a well-known species, and the poorly-studied short-toed eagle (Figure 4.2). The Bonelli's eagle is a medium to large, cliff-nesting bird of prey with its western Palaearctic populations mainly located in the Mediterranean area (Del Hoyo 1994). The European population has suffered a significant decline over the last decade (Birdlife-International 2004), and the Bonelli's eagle has consequently been listed as an endangered European Species (Birdlife-International 2011). Spain supports 730 – 800 breeding pairs (Real 2003) or about 80% of the total European population. Since the Bonelli's eagle is included as a priority target species for special conservation measures in European and Spanish legislations, high-priority conservation has been urged, and consequently it is a fairly well-studied species (Muñoz *et al.*

2013; Muñoz & Real 2013). The short-toed eagle is a medium-sized bird of prey. The European population migrates mainly to sub-Saharan Africa (Del Hoyo 1994), leaving in September-October and returning in March-April. This species builds its nests on trees, mainly pines, in forests with relatively little human disturbance (López-Iborra *et al.* 2010). Nowadays, the species seems to be more abundant than in the past decades, with estimates of approximately 10,000 breeding pairs (Palomino *et al.* 2011). Currently, there are several reports on their wintering in southern Spain (Jiménez & Muñoz 2008). Due to its shy behaviour, commonness, and low-priority conservation status, the short-toed eagle remains one of the least-studied raptor species.

Figure 4.2: A juvenile Short-toed Eagle (top), A typical nest of Bonelli's Eagle occupied with two nestlings (Bottom)

(Photos: Juan Luis Muñoz)

4.4.3 Choice of experts

Independent experts were identified through their membership of a scientific society and by their relevant publications. All selected experts had performed long term extensive studies on the selected species in Spain. They were invited to participate and reply to the online questionnaire independently. Nine experts accepted and formed the knowledge domain for this study. Seven submitted online questionnaires for both species, while two did so for one species only. Therefore the knowledge domain for each species consisted of eight experts. Experts did not exchange knowledge or fill in the questionnaires collaboratively.

4.4.4 Environmental predictors

We studied the literature on species habitat preferences for both the Bonelli's eagle and the short-toed eagle in order to collect a list of relevant environmental predictors. The predictors were chosen on the basis of availability and potential predictive power for the target species in Spain. We then presented the expert panel with the land cover classes (European Environment Agency 2012), topographical attributes (i.e. elevation, slope, and aspect) obtained from SRTM data (USGS, 2004), as well as two bioclimatic variables (i.e. maximum temperature in the warmest quarter, and minimum temperature in the coldest quarter) taken from WorldClim (Hijmans *et al.* 2005).

4.4.5 Conventional species distribution models

To compare the performance of our proposed approach with conventional species distribution models, two inductive distribution models; generalized linear model (Mccullagh 1989), and boosted regression trees (Friedman 2001), were performed for each species over the independent species datasets – obtained from the atlas of Spanish breeding birds (Marti & Del Moral 2003) - with the same environmental predictors. The generalized

linear model (GLM) uses parametric functions to link the response variable to a linear, quadratic and/or cubic combination of environmental predictors. We used a GLM ordinary polynomial with an automatic stepwise model selection based on the Akaike Information Criterion. The boosted regression trees (BRT) fits complex nonlinear relationships by combining two algorithms of regression trees and boosting; the former to relate a response to their predictors by recursive binary splits, and the latter to combine many single models to improve the performance. We followed the recommended default settings; maximum number of trees equal to 3000 and learning rate equal to 0.001 (Elith *et al.* 2008). For each species and in 100 realizations, we randomly sampled 50% of species datasets. These realizations were used to train the models. The idea was to compute the result of the model repeatedly using varied input values and then to assess the accuracy of each (Heuvelink 1998). This so-called Monte Carlo simulation allowed us to assess uncertainty in the inductive models.

4.4.6 Performance measures

We obtained observation data for a 10 x 10 km grid cell (based on the UTM projection) from the atlas of Spanish breeding birds (Marti & Del Moral 2003) for both species. These datasets were used to evaluate the discrimination capacity, and the calibration of the distribution models. We evaluated the discrimination capacity of species distribution models by analysing their receiver operation characteristic (ROC) curves. The area under such a curve (AUC) provides a threshold-independent measure across all possible classification thresholds for each model (Fielding & Bell 1997). To assess the goodness-of-fit of models we analysed the calibration plots of models. The calibration plot shows the model's estimated probability (x-axis) against the mean observed proportion of positive cases (y-axis) for equally sized probability intervals (Lemeshow & Hosmer 1982). The perfectly calibrated predictions follow the diagonal line in the plot. For more information see Jiménez-Valverde et al. (2013).

4.5 Using Dempster-Shafer evidence theory

We used a structured procedure to interrogate the experts. Experts were provided with online questionnaires consisting of factual questions (See Appendix II). The procedure to assign evidential belief functions and to combine evidential maps was as follows:

1. In each question – corresponding to a piece of evidence – the expert is asked to self-evaluate the level of confidence in her/his knowledge in order to express the favourable habitat of the target species using the given evidence according to a five-point scale ranging from "*not confident at all*" to "*quite confident*". We then converted these rankings, using Figure 4.3, into belief equivalents to obtain the degree of uncertainty to be applied to that specific piece of evidence.
2. The expert assigns her/his belief in form of likelihood (P) given the evidence. We set four linguistic probabilities (Xu et al. 2003; Bardossy & Fodor 2004) "very unlikely", "unlikely", "likely", and "very likely" in order to subsequently derive a numerical value to represent belief (Figure 4.3). For continuous evidence layers we implemented this process in a fuzzy inference system (Zadeh 1965), into which the belief functions were imported as membership functions for the given evidence.
3. Given the probability of occurrence and the degree of uncertainty for the given evidence, the degree of support (Sup), and the degree of plausibility (Pls) were calculated as follows:

$$Sup = P - (Unc \times P)$$

$$Pls = P + (Unc \times (1-P))$$

This approach insured asymmetric weights of uncertainty for the upper (Pls) and the lower (Sup) probabilities.

4. We applied the evidential belief functions to their corresponding piece of evidence in the original spatial scale (i.e. land cover and topographical attributes in UTM 0.1×0.1 km, and bioclimatic variables in UTM 1×1 km). Then, the evidence maps have been aggregated to UTM 10 × 10 km.
5. The aggregated maps then were combined using Dempster's (1968) rules of combination to produce the support map and the uncertainty map for each of the evidences.
6. Finally, the degree of support and the degree of uncertainty of all the evidence were combined using the same rules to obtain the final support map of the target species along with the degree of uncertainty.

We asked experts to provide us with comments at the end of each section. We modified the questionnaires in order to avoid any pairwise or multiple comparisons. Experts were also asked to express their main sources of knowledge.

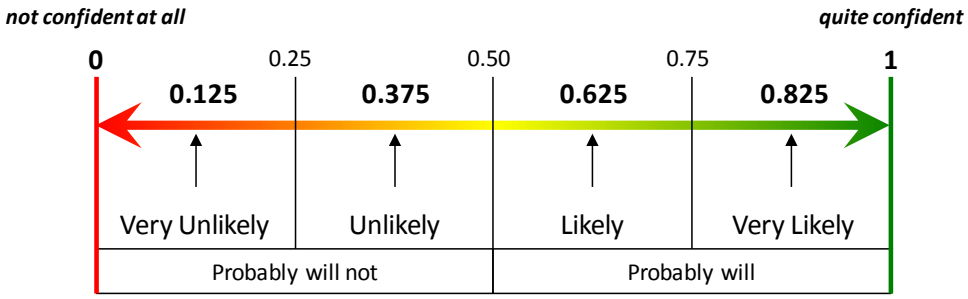


Figure 4.3 Word-to-probability relationship adopted from (Xu *et al.* 2003)

4.6 Results

4.6.1 Assignment of belief functions

We produced the support and the plausibility maps of each evidence based on the knowledge domain using combination rules for both species (Figure 4.4). We then analysed the discrimination capacity of the support and plausibility maps (Table 4.1). For Bonelli's Eagle slope gradient was the most discriminative evidential layer, with support and plausibility functions having AUC values of 0.69 and 0.68, respectively. The discrimination capacity for aspect and bioclimatic variables was quite low with AUC values of around 0.57. A few experts (2 out of 8) did not assign probability values based on aspect for the Bonelli's eagle. They commented that this variable is not a discriminative parameter for this species. For the short-toed eagle, both support and plausibility maps for slope produced the highest discrimination capacity with AUC values of 0.67 and 0.68, respectively. The support function was more discriminative (AUC = 0.68) than the plausibility function (AUC = 0.61), suggesting that experts overestimated their uncertainty when using land cover classes. Three of the experts did not use minimum temperature in the coldest quarter on the basis that the short-toed eagle is a migratory species. However, the other experts acknowledged that those few birds wintering in southern Spain would avoid freezing temperatures and therefore assigned their belief functions using this variable.

Table 4.1 Comparison of discrimination capacity (AUC) of inductive models (i.e. GLM and BRT), and DST approach per evidential and combined maps for Bonelli's Eagle and the short-toed eagle in Spain. *Sup*: Support function, *Pls*: Plausibility function.

		Bonelli's		**Short-toed**	
	GLM	0.74		0.66	
	BRT	**0.78**		**0.72**	
		Sup	*Pls*	*Sup*	*Pls*
Evidential Maps	Land cover (LC)	**0.70**	0.68	0.68	0.64
	Elevation (EL)	0.57	0.56	0.58	0.57
	Slope (SL)	0.69	**0.69**	**0.69**	**0.68**
	Aspect (AS)	0.55	0.55	0.55	0.54
	Min temp in the coldest quarter (TM)	0.58	0.58	0.51	0.50
	Max temp in the warmest quarter (TX)	0.56	0.57	0.58	0.60
Combined Maps	C1 = LC ⊕ EL	0.71	0.70	0.62	0.61
	C2 = LC ⊕ EL ⊕ SL	0.73	0.75	**0.69**	**0.67**
	C3 = LC ⊕ EL ⊕ SL ⊕ AS	0.74	0.74	0.60	0.61
	C4 = LC ⊕ EL ⊕ SL ⊕ AS ⊕ TM	**0.78**	**0.77**	0.68	0.60
	C5 = LC ⊕ EL ⊕ SL ⊕ AS ⊕ TM ⊕ TX	0.75	0.74	0.60	0.57

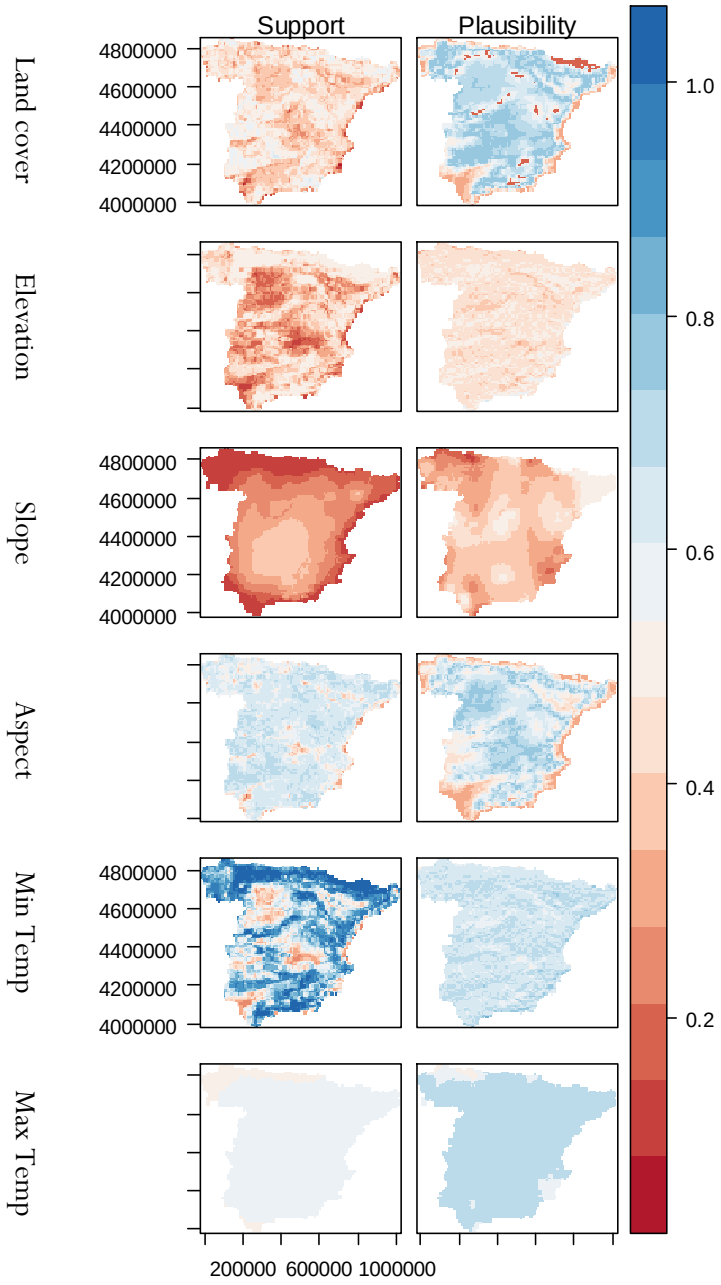


Figure 4.4a Combination of evidential functions based on experts' judgments for the Bonelli's eagle. The degree of support and the degree of plausibility (columns) for evidential maps (rows). Maps are presented on a graduated blue-red scale; blue indicating favourable and red indicating unfavourable locations.

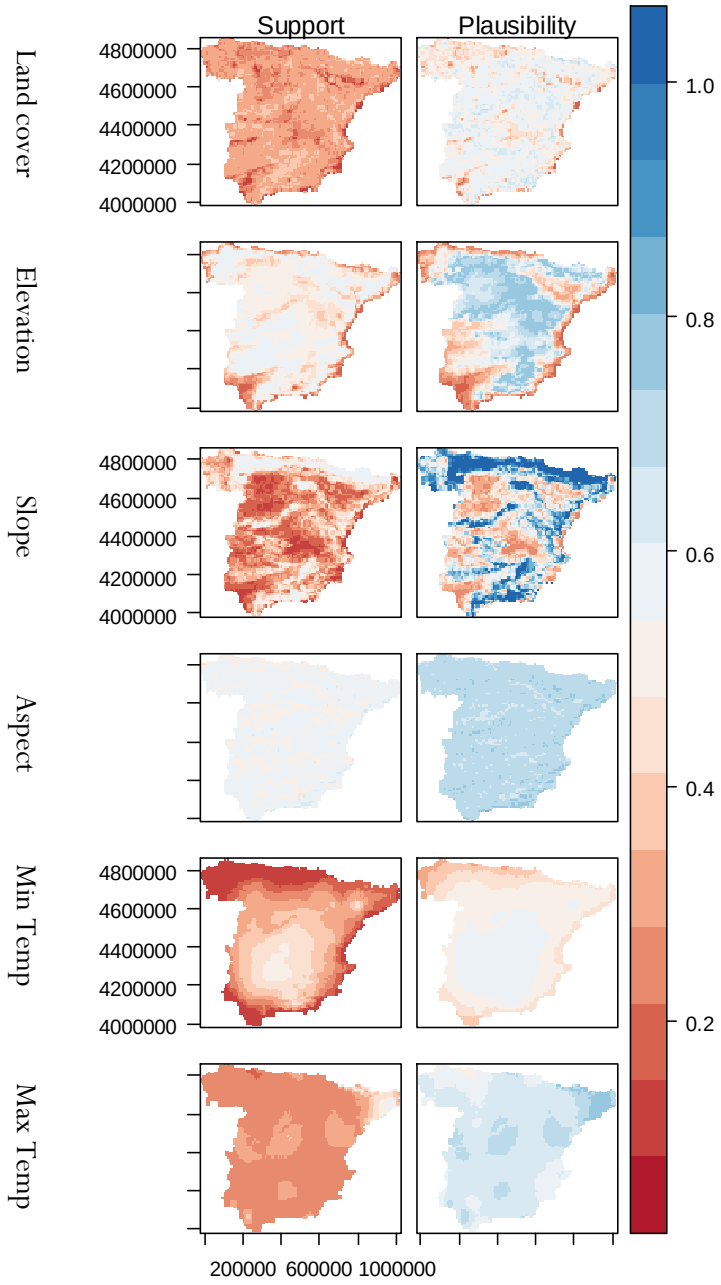


Figure 4.4b Combination of evidential functions based on experts' judgments for the short-toed eagle. The degree of support and the degree of plausibility (columns) for evidential maps (rows). Maps are presented on a graduated blue-red scale; blue indicating favourable and red indicating unfavourable location.

4.6.2 Combination of belief functions

We applied Dempster's combination rule recursively, so that the above mentioned evidence maps were combined in a pairwise manner (Figure 4.5). When we combined all the evidential maps for the short-toed eagle the AUC for the support and plausibility maps improved to 0.96 and 0.67, respectively (Table 4.1). The most accurate distribution map for the short-toed eagle was achieved by combining the land cover, elevation, and slope evidential maps (AUC = 0.69). The discrimination capacity of distribution maps for the Bonelli's eagle improved from 0.71 for the support combination of land cover and elevation, to 0.78 for when slope, aspect, and temperature were also combined.

We examined the calibration of the final distribution maps for both species to evaluate the reliability of the output. The support and plausibility models of the short-toed eagle were more calibrated between the probability ranges 02 to 0.8 compared to the Bonelli's eagle (Figure 4.6a). The support and plausibility maps of the Bonelli's eagle generally underestimated the probability of occurrence (Figure 4.6b). Overall, in the case of short-toed eagle the discrimination capacity of the distribution maps were low, though the models were well-calibrated, and therefore the transferability of the evidential functions was high. However, in the case of the Bonelli's eagle predictive power of the evidential functions was higher and the favourable habitats were well-discriminated, but the models were not well-calibrated. The uncertainty propagation assessment revealed that the overall uncertainty of the combined maps were diminished when the Dempster's combination rule were applied recursively for both species (Figure 4.7). This reflects the convergence of the support and plausibility maps over the combination phase.

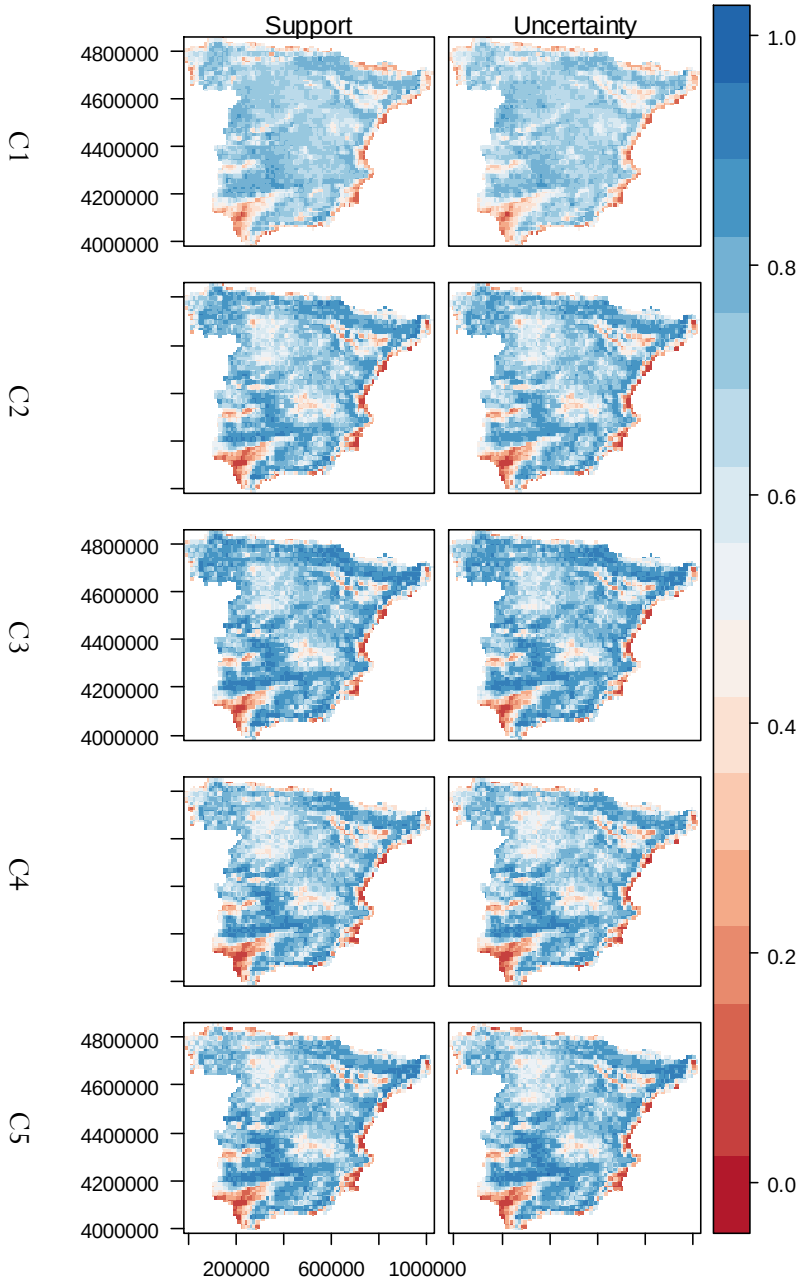


Figure 4.5a Combination of evidential layers for Short-toed eagle. Columns represent support (*Bel*) and plausibility (*Pls*). Species distribution maps are presented on a graduated blue-red scale; blue indicating favourable and red indicating unfavourable locations. Map codes follow Table 1.

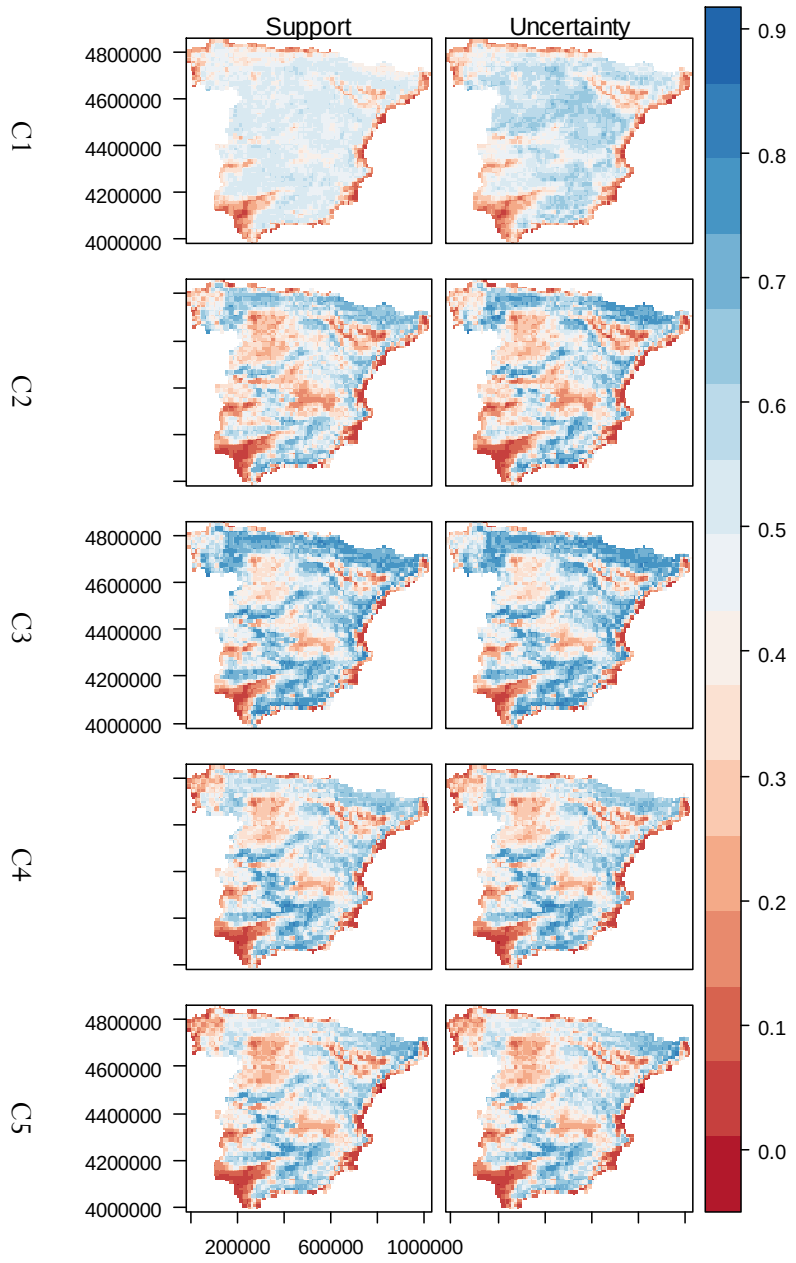


Figure 4.5b Combination of evidential layers for the Bonelli's eagle. Columns represent support (*Bel*) and plausibility (*Pls*). Species distribution maps are presented on a graduated blue-red scale; blue indicating favourable and red indicating unfavourable locations. Map codes follow Table 1.

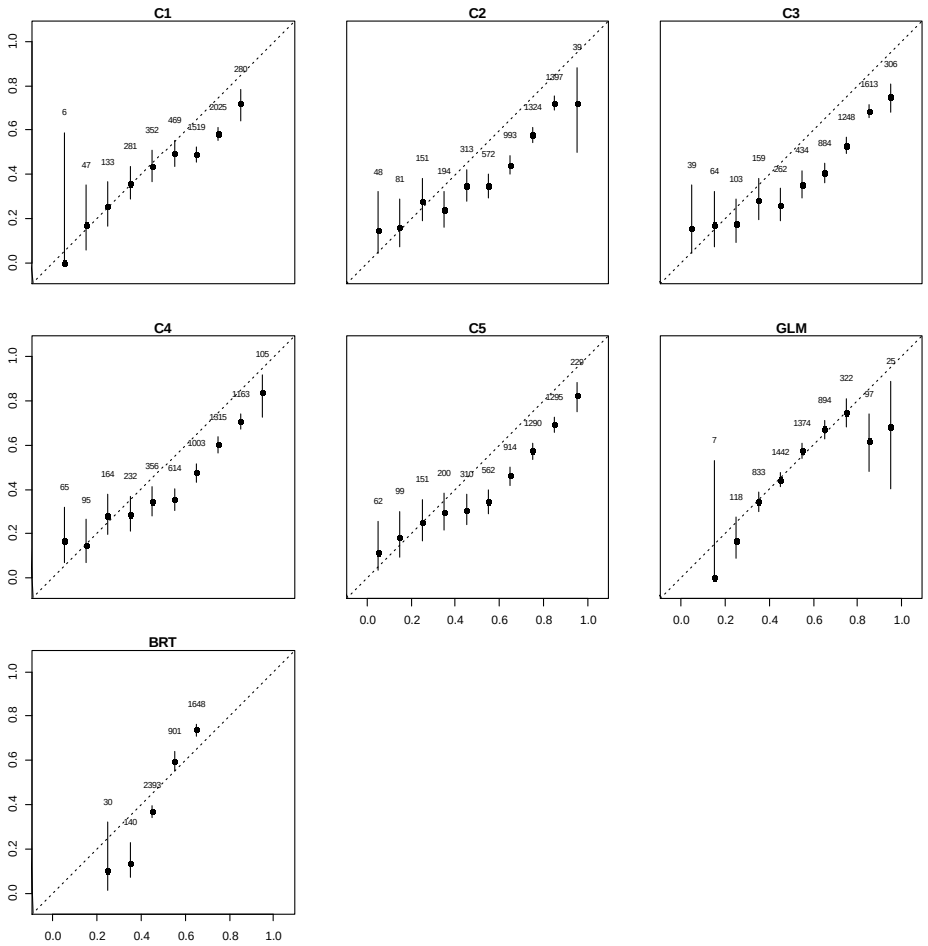


Figure 4.6a Calibration plots of the combined evidential maps for the short-toed eagle showing the mean model estimated probability (x-axis) against the mean observed proportion of positive cases (y-axis) for ten equally sized probability intervals. Figure labels follow Table 1. The value above each point represents total number of observation records.

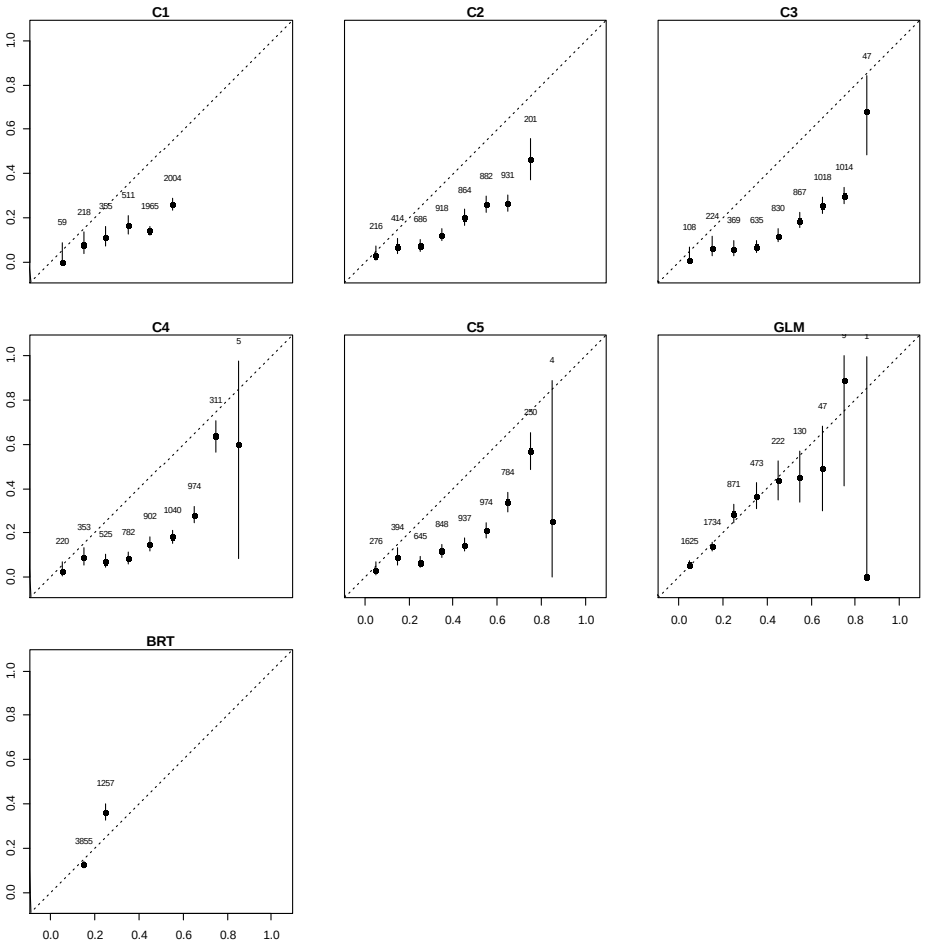


Figure 4.6b Calibration plots of the combined evidential maps for Bonelli's eagle (b) showing the mean model estimated probability (x-axis) against the mean observed proportion of positive cases (y-axis) for ten equally sized probability intervals. Figure labels follow Table 1. The value above each point represents total number of observation records

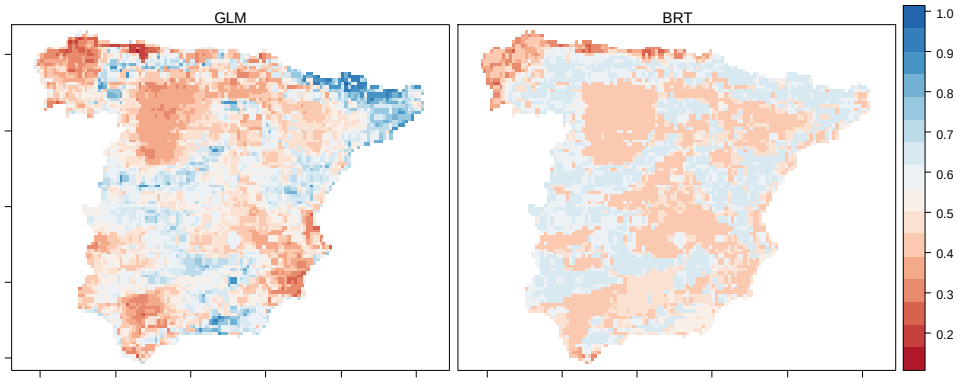
4.6.3 Performance of the inductive models

The average AUC values as a result of the Monte Carlo simulation for Bonelli's eagle were 0.74 in GLM and 0.78 in BRT with the standard deviation of 0.08, and 0.07 respectively. These values for the short-toed eagle were 0.66 in GLM and 0.72 in BRT with relatively higher standard deviation (i.e. 0.11 and 0.09 respectively) compared to those of the Bonelli's eagle. In case of the short-toed eagle, GLM performed a very well-calibrated specifically over the probability range between 0.3 and 0.8, while this range was limited to 0.4 and 0.7 in BRT. Discrimination power of the GLM models were slightly lower than the DST outputs; 0.74 and 0.66 compared to 0.78 and 0.69 for Bonelli's Eagle and the short-toed eagle respectively, while performed higher calibrated models compared to DST outputs and BRT models. Discrimination power of the BRT model for Bonelli's Eagle was similar to DST output, while it was higher in case of the short-toed eagle. However, calibration of DST outputs was higher than BRT models for both species (Figures 4.7 & 4.7).

4.6.4 Variability in knowledge domain

To have a better insight into the variability among experts, we compared the belief functions assigned by the expert panel. It was revealed that two of the experts were consistently assigned relatively higher probability, while one other was consistently assigned relatively lower probability for both species. We also noticed in our experiments that direct observations and field experiments made experts more confident, resulting in them assigning a relatively lower uncertainty level to their estimates. In contrast, experts who obtained their knowledge indirectly and deductively, self-evaluated a lower level of confidence in their estimates.

a. short-toed eagle



b. Bonelli's eagle

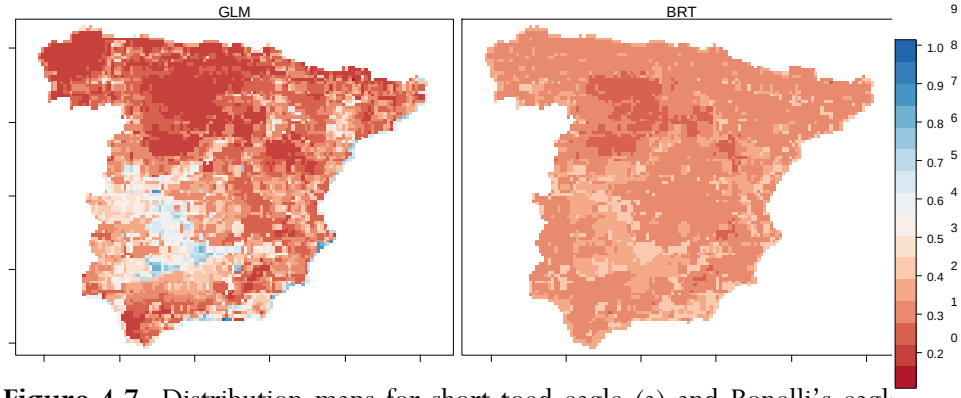


Figure 4.7 Distribution maps for short-toed eagle (a) and Bonelli's eagle (b) generated by GLM (left) and BRT (right). Species distribution maps are presented on a graduated blue-red scale; blue indicating favourable and red indicating unfavourable locations.

In general the expert's self-acknowledged confidence level – and therefore the difference between support and plausibility functions for the Bonelli's eagle, was relatively lower compared with the short-toed eagle. This was in accordance with the fact that the Bonelli's eagle is a better known species in Spain than the short-toed eagle. Experts were more consistent in expressing their knowledge when using land cover classes for both species

than when using other environmental predictors. There were higher levels of inconsistency among experts when considering topographic variables, though their levels of confidence were relatively high. Experts were less confident when assigning belief functions to bioclimatic variables. We also noticed that experts were more confident in expressing their knowledge based on visible and geographically detectable phenomena (e.g. land cover class, aspect) than those with implicit characteristics (e.g. elevation, temperature).

4.7 Discussion

Inductive models are useful for predicting species distribution where there are sufficient occurrence data and little knowledge of species-environment interactions. Utilizing expert knowledge, as an alternative modelling approach, may allow models to be more reliable and thus applicable in situations where occurrence data are scarce or unreliable. However, expert knowledge is subject to uncertainty and this should be accounted in the modelling procedure. The results demonstrate the successful implementation of expert knowledge elicitation and combination process through DST evidential belief functions in considering knowledge uncertainty across multiple experts.

Our approach was entirely based on the knowledge of experts with a variety of backgrounds, i.e. academics and technical fieldworkers. The evidential functions effectively represent an expert's epistemic uncertainty. Experts' replies may be biased, producing optimistic or pessimistic responses (Cooke 1991), or they may say "don't know" and assign 100% uncertainty to a class within in a single variable or to an entire set of variables. Furthermore, the rule of combination offers a basis for both aggregation as well as the propagation of uncertainties (Baraldi & Zio 2010). Our results show a decrease by their mutual combination in each iteration for uncertainties, as well as for the difference between the support and the plausibility functions (Figures 4.4 and 4.6). In his third

general principle about probability Laplace (1825) defined that if the events are independent of one another, then the probability of their combined occurrence is the product of their probabilities. This suggests that the number of pair-wise combinations may compensate for the overall gap between the degree of support and the degree of plausibility, and diminish the level of uncertainty per iteration. The number of experts and the variety of explanatory variables may be modified based on the required level of confidence.

We evaluated the final distribution maps with two conventional inductive approaches (i.e. GLM and BRT) using independent atlas data. In the case of Bonelli's Eagle our proposed approach demonstrated almost similar discrimination power compared to inductive approaches, while inductive approaches outperform DST in the case of the short-toed eagle. Bonelli's eagle is a cliff-nesting species, evidence of nesting is generally more obvious than it is for the short-toed eagle, which is a forest species. Indeed, cliffs are easier to monitor than forested areas; the reproduction of Bonelli's eagles is usually confirmed by direct observation of the chicks in the nest, whereas in the case of the short-toed eagle, even if there is reasonable confidence that they are breeding in an area, the nest is rarely observed, leaving doubt. Although evidential functions better discriminated favourable habitats for the Bonelli's eagle than the short-toed eagle, the distribution map of the short-toed eagle was well-calibrated, while this was not the case for Bonelli's eagle. This might be an artefact of direct observations and inductive sources of knowledge - being able to identify highly favourable areas and to a lesser extent highly unfavourable areas, while not being able to identify areas where probability is intermediate. This may be due to the scale of our study, which is relatively coarse, making it difficult for experts to link their local observations to the national scale of 10 ×10 km. For the short-toed eagle, which is a relatively lesser known species, experts provided more general beliefs and less specific classes than for the Bonelli's eagle. This resulted in a well-calibrated model (Figure 6a), though with a lower discrimination capacity

(AUC = 0.69), mainly failing to identify the most favourable areas. This also suggests that areas with intermediate probability are more abundant than areas with high or low probability, with it being difficult to discriminate presence from absence in areas with intermediate probability.

Although the DST is a powerful tool for probabilistic reasoning and has been applied widely in engineering and computer science, it has yet to reach the ecological modelling mainstream. Our proposed approach using the DST would be practical where knowledge of a species' geographic distribution is needed for conservation purposes, especially in the case of poorly-sampled species. However, the use of knowledgeable experts and, well-structured elicitation processes, are pre-requisites for maximizing the reliability of expert-based models (Murray *et al.* 2009; Smith *et al.* 2012). The particular strength of the developed modelling approach in this study is that it is explicit in accommodating uncertainty in knowledge for model prediction by using evidential functions. It also yields similar results to conventional inductive methods and provides a framework to propagate and aggregate uncertainty, and it capitalizes on the range of data sources usually considered by an expert.

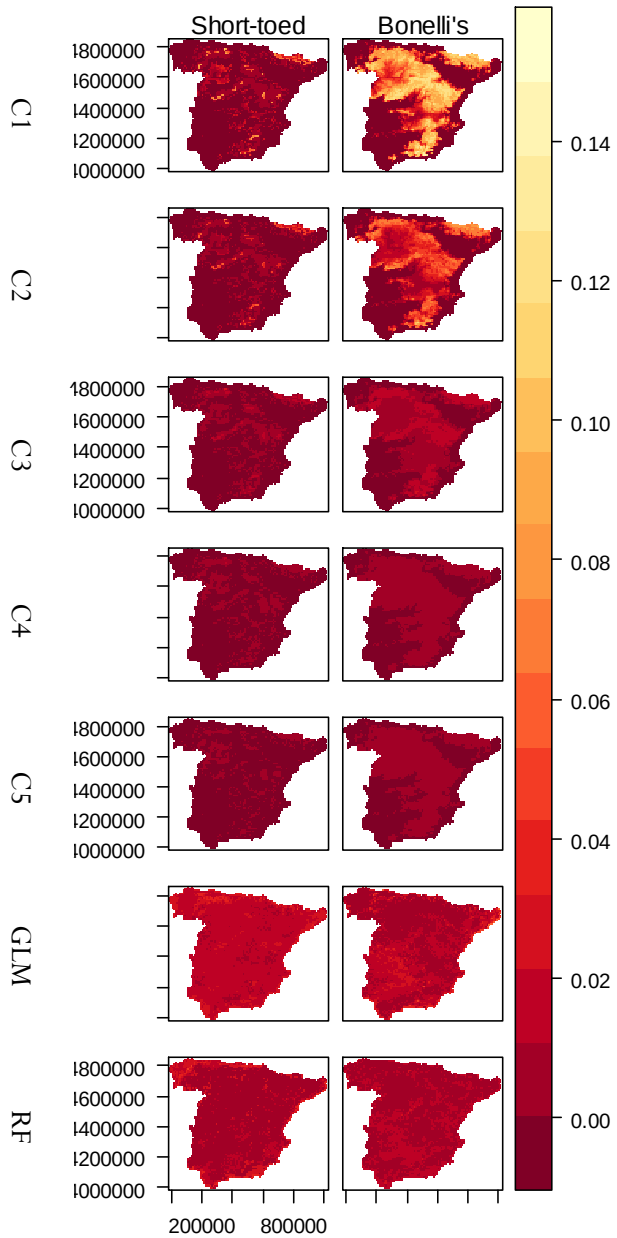


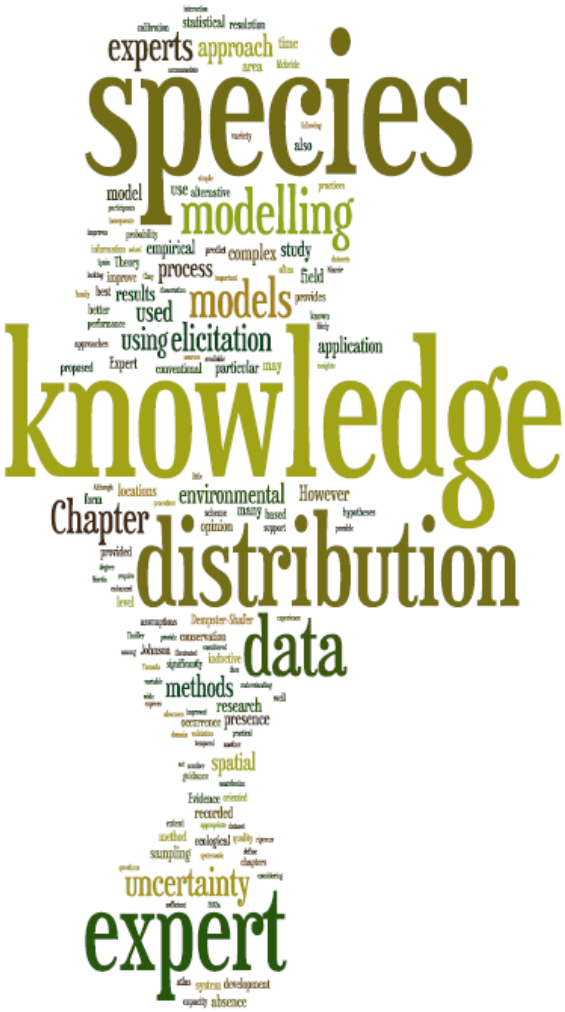
Figure 4.7 Uncertainty propagation in combined evidential maps (C1 – C5) and distribution maps generated by GLM and BRT for short-toed eagle (left) and Bonelli's eagle (right). Map codes follow Table 1. Uncertainty maps are presented on a graduated red-yellow scale.

4.8 Acknowledgements

The authors highly appreciate the contributions by Dr. José Antonio Sánchez-Zapata, Dr. Santi Mañosa, Dr. Pascual López-López, Dr. Juan Carlos del Moral, and three other experts who prefer anonymity. They generously shared their knowledge. The foundation of this study was laid during the short-term scientific mission of AN to the University of Malaga in September 2011 founded by the MOVE Action of the European Cooperation in Science and Technology, and further developed in April 2013 founded by the HarmBio Action of the European Cooperation in Science and Technology. AN was supported by the European Union Erasmus Mundus program External Cooperation Windows 2008/2441/001 MUN ECW 15-7-2008. RR was supported this study through the project CGL2009-11316/BOS from the Spanish Government and FEDER. The language editing was done by Eva Skidmore. We thank to David Nogues-Bravo, François Guilhaumon, and the five anonymous reviewers for their critical constructive comments on a previous versions of this paper.

Chapter 5:

KNOWLEDGE ORIENTED SPECIES DISTRIBUTION MODELLING; SYNTHESIS


statistical
resolution
experts
approach
species
also
model
use
modelling
empirical
complex
study
process
field
results
models
used
using
elicitation
application
Expert
particular
may
knowledge
locations
environmental
However
Chapter
many
based
opinion
support
distribution
Dempster-Shafer
data
conservation
Johnson
inductive
methods
research
presence
occurrence
spatial
recorded
method
ecological
sampling
uncertainty
expert
system
development
absence

5.1 Introduction

There is an increasing demand for species distribution models to address numerous questions in conservation biology and evolution (Guisan & Thuiller 2005), to test biogeographical hypotheses (Graham *et al.* 2004b), to predict species' invasion and reproduction (Peterson 2003), and to predict new occurrences of rare species (Williams *et al.* 2009). Distribution models have also been used to assess the impact of climate, land use, and other environmental changes on species distribution (Thuiller *et al.* 2005), as well as to support conservation planning and reserve selection (Seoane *et al.* 2005). The demand for such information has led to a variety of statistical and machine-learning methods for modelling species' distributions (Ferrier *et al.* 2002; Guisan & Thuiller 2005; Elith & Leathwick 2009; Franklin 2009). Several researchers have criticized the species distribution modelling community for diverging from theoretical groundings (Austin 2002; Huston 2002; Jiménez-Valverde & Lobo 2007), and warned about the lack of robust linkage between models' assumptions and underlying ecological theories (Elith & Leathwick 2009). The majority of modelling practices have focused on detecting functional relationships between the species' locations and the surrounding environment, and considering species' biological knowledge tends to be of secondary importance (Austin 2007).

Expert knowledge has been increasingly incorporated in management recommendations and practices in a wide variety of fields (Yamada *et al.* 2003; Martin *et al.* 2005; Seoane *et al.* 2005). Species distribution modelling is another field of research that can benefit from the possibilities offered by expert knowledge to bridge the gap between ecological theories and model's assumptions. Given the complex spatial and temporal aspects of species distribution, many researchers struggle to set up appropriate inductive solutions. This is, in particular, apparent when they only rely on traditional approaches and observational data to understand species' interactions with their abiotic, as well as biological environments. By

exploring alternatives to overcome the limitations of conventional species distribution modelling, we propose and use expert knowledge to complement poor data, to replace missing empirical data, and to cope with the complexity and uncertain nature of species distribution. We argue that incorporation of expert knowledge bridges the gap between ecological theory and model's assumptions in species distribution modelling, and is beneficial for developing more robust techniques and their consistent use. For many, the first application of expert knowledge is a response to the challenge of having little or no empirical data to support conservation decisions. We illustrate that considering expert knowledge as a component of a data-driven species distribution model improves the performance and the outcome, particularly where several assumptions are made to satisfy inductive statistical methods. An example could be considering unknown locations as pseudo-absences. A wide variety of statistical methods specially designed to model species distribution are currently available. Therefore, sound knowledge of the performance of these models helps researchers to set the most appropriate experimental settings for their purposes. The value of this knowledge may be of particular importance, since usually the only reliable information on the distribution of organisms is their recorded presence, while reliable absence data are expensive and rare (Gu & Swihart 2004; Tsoar *et al.* 2007). In this dissertation we argue that the interpretation of species distribution models may vary if theoretical considerations are taken into account. We demonstrate alternative and reasonable knowledge oriented settings that improve the performance of widely used species distribution models. We highlight that advances in computational facilities along with the development of rigorous statistical approaches in species distribution modelling should be led by corresponding development of ecological realms in methods. In this chapter we summarize the main findings of the previous chapters, and provide a framework to discuss the contribution of expert knowledge to species distribution models.

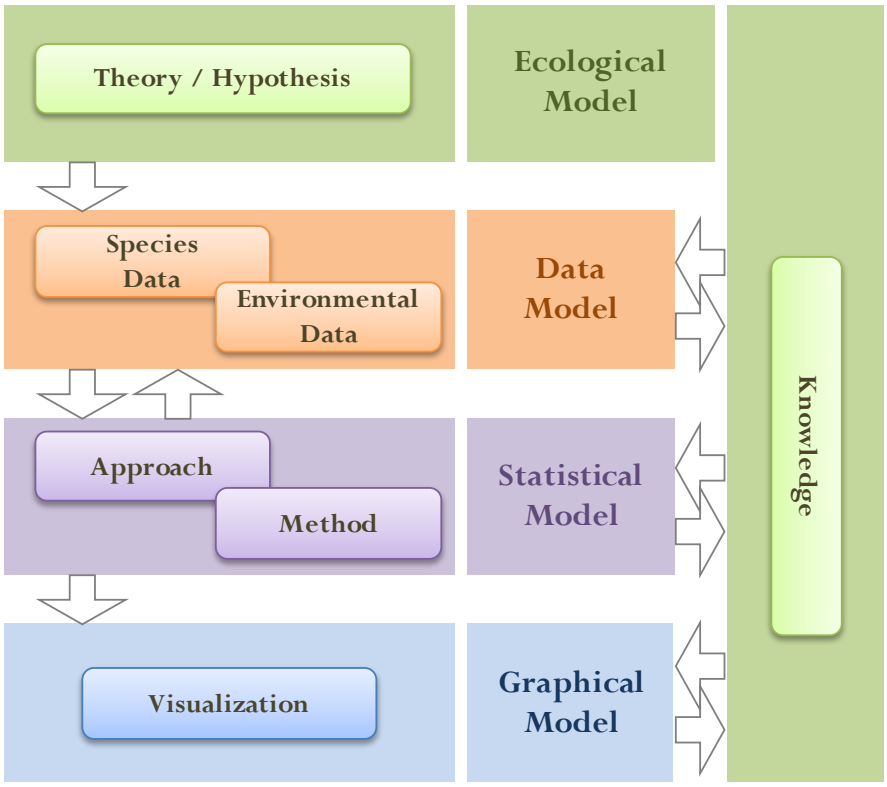


Figure 5.1 Revised framework for species distribution modelling

We have revised the framework for species distribution modelling (Figure 5.1) and propose a parallel contribution of the ecological model represented by expert knowledge to other components of the framework. Ecological knowledge needs to be present in every step of species distribution modelling. In the data model the decision regarding how data are collected and estimated would benefit from interactions with existing knowledge on species biology (Chapters 2 and 3). The contribution of knowledge to the statistical model (Chapter 4) is not only to set deductive rules, but to also accommodate its corresponding uncertainty. The dynamic interaction between the components of a species distribution model and an ecological model is vital for the reliability. We would like to

emphasise that in using rigorous and highly advanced techniques the increasing disassociation from solid conceptual thinking and profound biological knowledge may not be sustainable for the advancement of species distribution research and conservation science in the long run.

5.2 Revisiting the research objectives

5.2.1 Finessing spatial quality of species occurrence data

Efficiency of expert knowledge use for improving the spatial quality of species occurrence data is highlighted in Chapter 2 (Niamir *et al.* 2011). The spatial resolution of the majority of species occurrence datasets and consequently resulting model predictions are often too coarse for local application. Collecting distribution data at a finer resolution for large numbers of species requires a comprehensive sampling effort, making it impractical and expensive. In our study, species occurrence data were obtained from the Spanish Atlas of Breeding Birds (Marti & Del Moral 2003) with a spatial resolution of 10 x 10 km. We employed a Bayesian expert system utilizing existing knowledge to yield the probability of a species being recorded at a finer 1 x 1 km resolution. The recorded probability was then used as a weight vector to generate a sampling scheme from the species atlas to enhance the accuracy of the modelling procedure. We used MaxEnt (Phillips *et al.* 2006) as the species distribution model. A comparison was made between the results of the MaxEnt using the enhanced and the random sampling scheme. In the field of species distribution modelling this study provided the following insights:

1. Expert knowledge is an efficient alternative to enhance the spatial quality of species data in regions that are poorly known or where fine resolution species information is too costly and time consuming to gather (Chapter 2).

2. The Bayesian expert system is a promising approach for incorporating existing knowledge into the disaggregation of species atlas data. Such a system provides the possibility to take the importance of environmental explanatory variables into account based on expert knowledge by assigning weights to form the relationship between the proxy variable and the recorded probabilities (Chapter 2).
3. A sampling scheme enhanced by expert knowledge better informs a species distribution model than a random sampling scheme does; consequently, the distribution maps using the knowledge oriented approach are more specific regarding nest locations, and show more contrast than those of the conventional model (Chapter 2).
4. Our proposed approach is a feasible substitute for comprehensive field work; however, the usefulness of this approach may be limited to well-known species (Chapter 2)..

5.2.2 Delimiting environmental extent

Distribution models rely either on systematic sampling data for species occurrences (including both presence and absences records), or on data exclusively denoting species presence and environmental conditions in the area, known as background or pseudo-absence (Barbet-Massin *et al.* 2012). Species distribution models actually do require data on both presence and absence of species. Collecting systematic field data can be expensive and it is not always possible to acquire appropriately designed samples within the time and budget constraints of a project (Field *et al.* 2004). Absence data in particular require higher levels of sampling effort compared to presence data, especially for mobile species. Furthermore, with advances in animal tracking equipment, data on species presence are more obtainable. Thus distribution models increasingly use artificial absence data, which can significantly mislead the outcome.

In the Chapter 3, the use of expert knowledge about the phylogeny and taxonomy of species is investigated to define the area that is more likely to represent absence of the species. We demonstrate that delineating the study area and excluding irrelevant absences based on species taxonomy significantly improves the calibration of species distribution models and thus their capacity to predict environmental potential. Using atlas data, two taxonomically enhanced datasets were created for 356 terrestrial species from mainland Spain. As well as a species' presence, its absence was recorded. However, an absence was only recorded for localities where another species of the same family or order was known to be present. We then assessed the effects of environmental extent on the predictive performance of five conventional species distribution models, comparing the results to those obtained with the original atlas dataset. Using family and order datasets significantly improved the calibration of the species distribution models. While the discrimination capacity decreased for several species when using the family dataset, this drop was significant when the models were trained with the order dataset. This decline in discrimination capacity of the models explains the similarity of environmental conditions in absence locations of the target species to the presence locations of its ancestors. In the field of species distribution modelling this study provided the following insights:

5. Expert knowledge is an efficient alternative to delineate the geographical extent of modelling practices in regions that are not systematically sampled and where absences are compilations of non-observed locations (Chapter 3).

6. Expert knowledge on excluding or including species taxonomical extent significantly improves the calibration of species distribution models and thus their capacity to predict environmental potential (Chapter 3).

5.2.3 Accounting for self-acknowledged ignorance

Habitat preference of a species may be delineated from the description of environmental features and are often based on field experience and beliefs regarding what is important in determining the distribution of the species (Yamada *et al.* 2003). These beliefs are important resources that may improve the performance of distribution models, specifically where species observation data are inadequate or unreliable (Janssen *et al.* 2010; Kuhnert *et al.* 2010; Niamir *et al.* 2011; Martin *et al.* 2012; Mcbride *et al.* 2012a).

In Chapter 4, we demonstrate the use of the Dempster-Shafer Theory (DST) of Evidence to accommodate knowledge uncertainty through its belief functions. This approach offers a practical alternative where knowledge of a species' geographic distribution is needed, and the distribution data are not sufficient. We used the Dempster-Shafer Theory of Evidence to model the distribution of a well-known and a poorly-studied species in mainland Spain. We invited experts to form a knowledge domain, and asked them through online questionnaires to express their beliefs about the habitat of the target species by assigning a probability value for a given environmental variable. Experts were also asked to acknowledge their confidence level regarding the assigned probability values. We then calculated evidential belief functions and combined them using Dempster's rules of combination. We evaluated and compared the calibration and discrimination capacity of the DST models with two other conventional inductive models. In the field of species distribution modelling this study provided the following insights:

7. The Dempster-Shafer Theory of Evidence yields similar results to conventional data-driven methods for modelling the distribution of well-known species. (Chapter 4).
8. Our proposed knowledge oriented approach offers a practical alternative where knowledge of a species' geographic distribution is needed, and the distribution data is not sufficient.

9. The particular strength of the proposed approach is that it explicitly accommodates knowledge uncertainty. It also provides a framework to propagate and aggregate uncertainty, and it capitalizes on the range of data sources usually considered by an expert (Chapter 4).

5.3 Lesson learned

5.3.1 Broad application of expert knowledge

Expert knowledge can no longer be considered as a secondary information resource (Johnson *et al.* 2012) in species distribution modelling (Niamir *et al.* 2011). Although application of expert knowledge in biodiversity conservation and particularly in species distribution modelling has a long history, we are now observing a greater level of interest in this approach due to its increasing degree of strength (Sutherland 2006), and the complexity of modelling hypotheses. Several studies have been dedicated to the assessment of bias and uncertainty in ecological knowledge (Boussabaine & Duff 1996; Reddy 1996; Janssen *et al.* 2010; Mcbride *et al.* 2012a), to the ability to transfer expert knowledge to predicting the distribution of a species within and beyond the region of expertise (Murray *et al.* 2009), to the comparison of expert knowledge and biological empirical data (Edward & Rykiel 1989; Skidmore 1989; Store & Kangas 2001), and to best practices for eliciting expert knowledge (Yamada *et al.* 2003; Choy *et al.* 2009; James *et al.* 2010; Kuhnert *et al.* 2010; Martin *et al.* 2012; Mcbride *et al.* 2012b) in conservation biology. Our contribution provides insight and guidance in the application of expert knowledge for specific challenges in species distribution modelling. We illustrate the application of expert knowledge to improve the spatial quality of species occurrence data (Chapter 2) and to help modellers to better define their study area where non-sampled locations are considered pseudo-absences (Chapter 3). In addition, these chapters provide the reader with an

overview of a wide range of merits and drawbacks of knowledge oriented approaches in species distribution modelling. In Chapter 4 we focus less on the application, and instead report the development of a novel method to elicit and accommodate the ignorance inherent in expert knowledge. We have used the Dempster-Shafer Evidence Theory to utilize expert knowledge in species distribution modelling and to characterize the full state of knowledge of a species. This approach may be used to prioritize conservation when managers believe that inaction while waiting for better empirical data is not an option.

5.3.2 Method for expert knowledge elicitation

In Chapter 2 and 4 we discuss that expert knowledge elicitation should promote and support transparent methods and provide for an assessment of uncertainty results. Knowledge elicitation is beyond a simple open invitation to discuss a particular subject with little development of the approach, documentation, or use of rigorous methods (Johnson & Gillingham 2004). Such an approach would result in knowledge that has little internal consistency or external validity. Working properly with experts is not necessarily a simple process (Drescher *et al.* 2012; Johnson *et al.* 2012). We allocated nearly a year to prepare for the elicitation of the expert knowledge in Chapter 4. We would like to emphasise that allocating enough time to refine research questions, to test and revise the pilot elicitation process, and to motivate and maintain participation throughout the process are critical for a successful elicitation practice (Choy *et al.* 2009; Knol *et al.* 2010; Johnson *et al.* 2012). We provided experts with a sufficient space to freely explain themselves in each section of the questionnaire, while directing them to express their opinion using the provided format. We found the design and planning phase of elicitation process quite critical, as it decreases the chance of misunderstanding and variety types of errors. We used a generic survey tool; an online questionnaire customized to our need to communicate with experts, as it was not possible to meet in person. Throughout the elicitation process, the

consistency of the method, elicitation score and preliminary results, should be continuously verified. This can be done by real-time feedback from the experts and/or by cross check among experts. Uncertainty is inseparable from the nature of expert knowledge and generally represented in form of the degree of variation among experts (Burgman *et al.* 2011; Mcbride *et al.* 2012a). However, another source of uncertainty is in expert himself. Elicitation procedure should facilitate self-acknowledge level of uncertainty as well and this possibility should be utilized and documented rather than suppressed (Aspinall 2010).

5.3.3 Expert knowledge has more to offer

Although expert knowledge in biodiversity conservation has often been used as temporary or substitute for empirical species data, it may offer several other advantages in complex hypothesis generation or in locations that are still not known (Chapter 3). Species distribution modelling addresses questions that often have a broad spatial and temporal domain with many interacting processes and variables. Such complex relationships may be difficult to set up using empirical data and rigorous inductive statistical methods. We would like to emphasise that a rigorous and complicated method is not necessarily the best way to approach natural and complex systems. A species' interaction with its environment is complex, thought complex is not automatically complicated. A complex system consists of several elements and their interaction according to simple, individual rules, resulting in emergent properties (Perony 2013). Experts can indeed facilitate the understanding of a complex system by making it simpler by, for example, debating issues like environmental explanatory variable selection, or formulating parsimonious hypotheses that conform to their broader combined experience.

Expert knowledge is not without error, bias and inaccuracy. As we observed in Chapter 2: where *knowledge* is lacking, experts tend to contribute their very own *opinion*. The difference between expert

knowledge and expert opinion is not always obvious. The experts themselves are also not aware of the limit of their knowledge until they are asked to quantify their degree of uncertainty. The Dempster-Shafer Evidence Theory provides the opportunity for experts to express themselves in the form of belief functions (e.g. support and plausibility, see Chapter 4 for further information), thus discriminating knowledge from opinion. Where direct knowledge is lacking, experts may come up with a useful opinion on a particular question (Cooke 1991; Ayyub 2010). However, uncertainty is likely to be much higher where it is based on opinion rather than knowledge (Walley 1996).

5.4 Recommendations

Application of expert knowledge in species distribution modelling is rapidly expanding. The contribution of this dissertation demonstrates that the use of expert knowledge in species biology and ecological processes has moved to a formalized and systematic set of methods. However, there is indeed room for refinement, improvement, and innovation. Below we suggest areas for further research that will be necessary to improve the practice of application when developing species distribution models based on expert knowledge.

5.4.1 Tools to support knowledge elicitation

We demonstrated an innovative method for knowledge elicitation using the Dempster-Shafer Theory of Evidence that accommodates self-acknowledged uncertainty. However, refinement of elicitation methods is a key area in need of further research. Integration of real-time statistical analysis and GIS data interactions within the elicitation process will be more intuitive for biologists, making the knowledge expression less abstract. In Chapter 4 we asked 8 experts to form the knowledge domain for the target species. Combining their beliefs using Dempster's rules

decreased the level of uncertainty in each step, indicating the sufficiency of the number of participants. The social science literature provides some guidance based on group dynamics and biases to set the optimal number of experts (Aspinall 2010). In species distribution modelling it seems likely that the heterogeneity across a study area, and the variation in knowledge among expert participants will define the appropriate number of participants. More research on understanding the optimal size of the expert domain is clearly warranted. Development of tools that better match spatial and temporal experience and knowledge of an expert to the proposed question would be an important direction for future research. These tools should allow experts to document their knowledge using an easy to understand, transparent, and repeatable process over a specific time period, or self-defined geographical boundaries.

5.4.2 Critically evaluate expert knowledge

Every expert-based project should incorporate a critical analysis regarding the reliability of the results. Expert knowledge can be cross-validated against other expert sources (Johnson *et al.* 2012). However, there is no formal guidance on how best to proceed with validation. Although validation may not be possible in all cases, every project should entail evaluation and verification of the process. Further research is required to define the expected outcome of such validation. Uncertainty is not an unknown concept to species distribution modelers who incorporate expert knowledge. However, the approach to documenting uncertainty has remained quite ad hoc. The science of eliciting and using expert knowledge would be improved by methods directly categorizing, measuring and incorporating uncertainty inherent in the knowledge. Achieving this goal will require consistent and standardized measures of uncertainty regarding knowledge and a better understanding of the sources of this uncertainty, especially in the context of expert selections. It will also require guidance on how best to manage and accommodate uncertainty. Throughout the chapters of this dissertation we have illustrated exciting and practical

examples of improved species distribution models using expert knowledge. Where experts are available and a proper method is employed, neither the expert's discipline, the spatio-temporal characteristics of the hypotheses, nor the subject of the study should prevent advancement of our understanding or the development of a solution to the complex problems faced by species distribution modeller.

5.4.3 Forwarding species distribution models in time

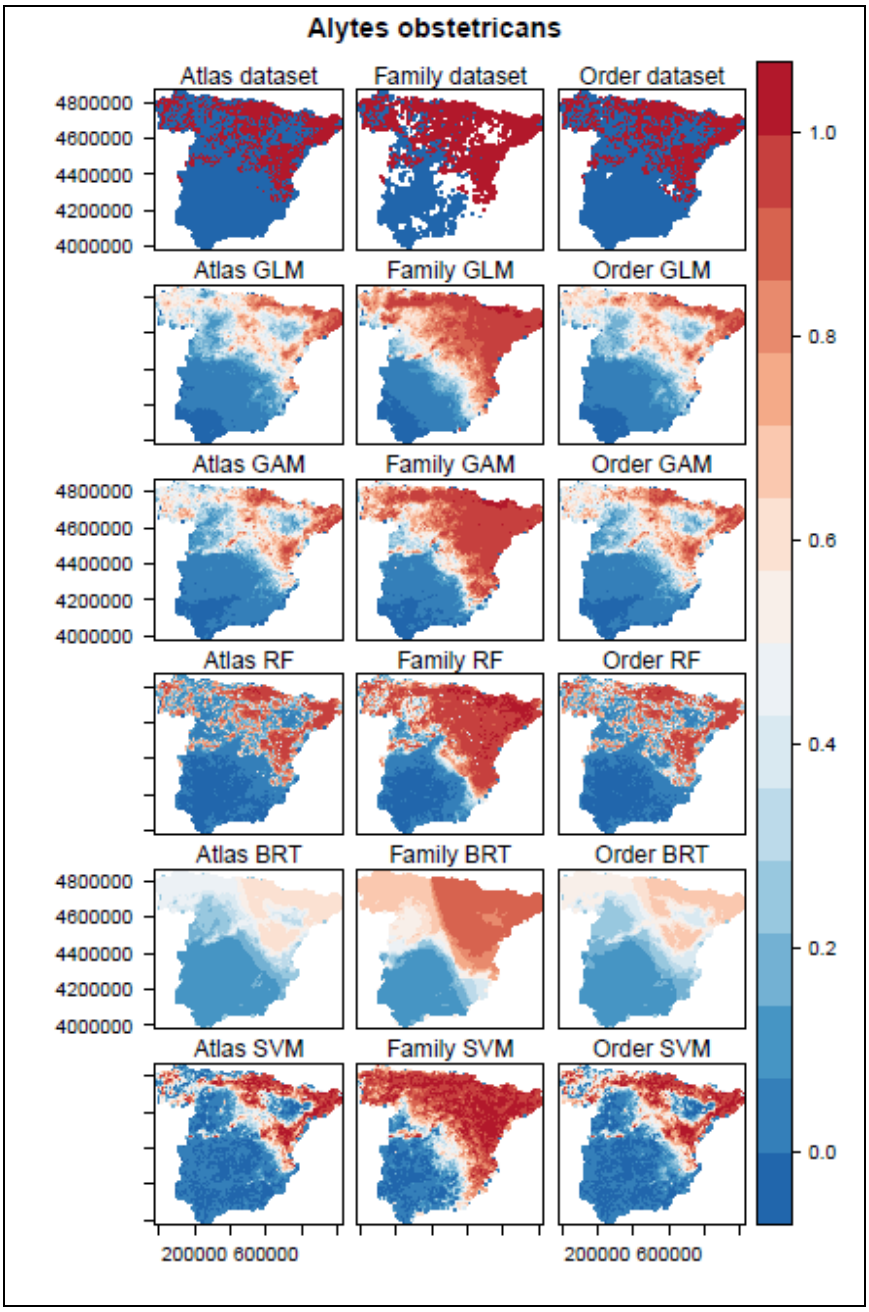
Predicting the future status of species distribution is a challenge in conservation science, since any validation of models projecting the future would be unlikely. Another challenge is dealing with the complexity of nature and the factors determining occurrence, survival, reproduction or dispersal of living organisms. However, predicting the future distribution of species is a fundamental goal of ecological research in the era of global change and the biodiversity crisis. This combination of challenges calls for rigorous scientific approaches and careful interpretation of results. All studies aiming to predict future distributions of species should carefully accommodate the uncertainty in nature, and assess the uncertainty of the data used (Chapter 4). This provides estimates of the consistency and the reliability of the projections. Projection of a species model onto future scenarios may be improved by the integration of different methodological approaches, as the strength of one method may compensate for the weakness of another. Multi-species distribution models may further improve by taking into account processes such as speciation, dispersal and extinction (Rangel *et al.* 2010), and overcome the general neglect of biotic interactions in species distribution models (Anderson & Raza 2010). Insights into past dynamics in species distribution may be used for an improved calibration of models for future projection (Chapter 3). Expert knowledge is an efficient source of information on species dispersal over time and experts' inputs may improve species distribution models by acknowledging unrealistic assumptions. We are confident that application

of expert knowledge in species distribution modelling will continue to expand and will continue to contribute to improved results.

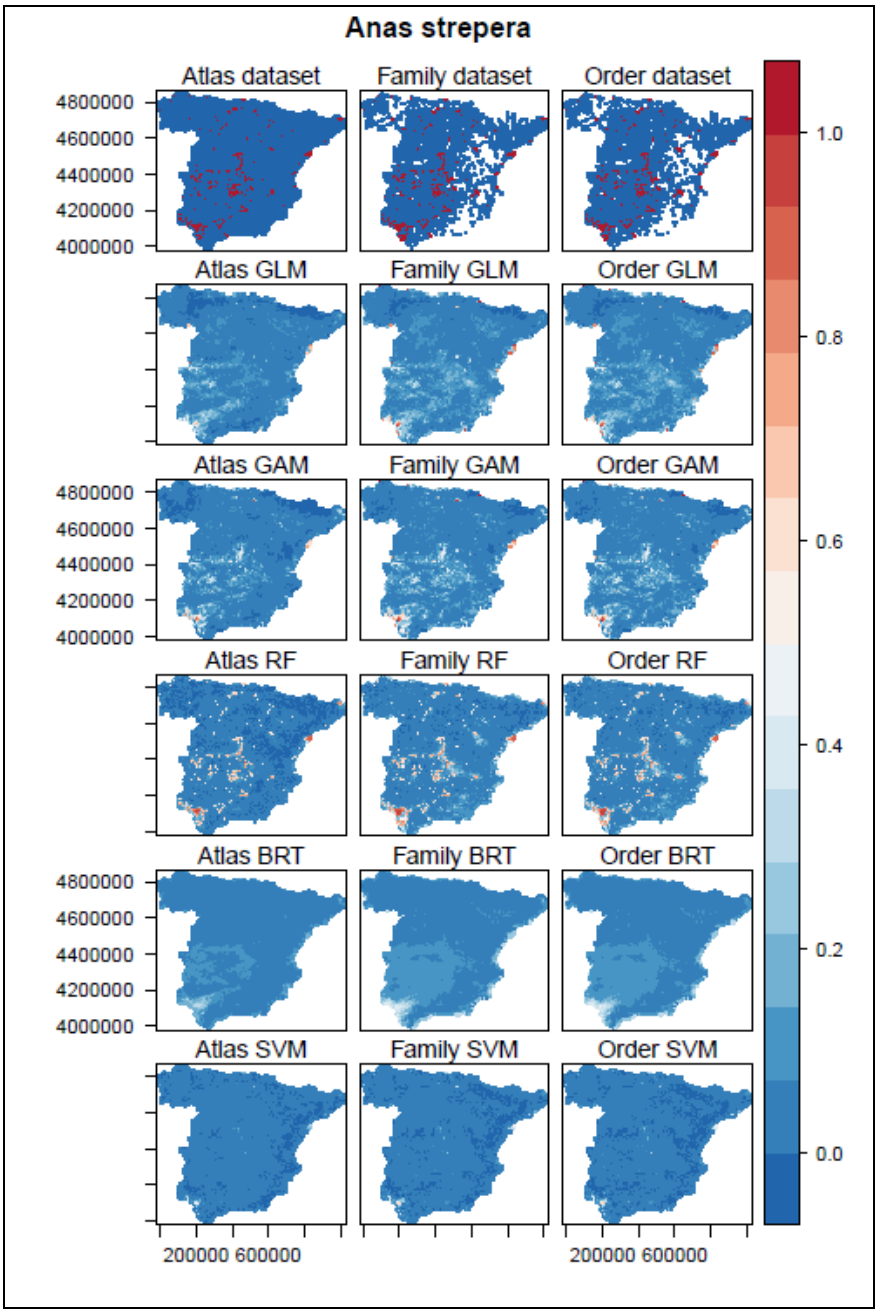
Appendix I:

EXAMPLES OF SPECIES DISTRIBUTION MAPS OBTAINED USING ATLAS DATASET AND TAXONOMICAL EXTENTS

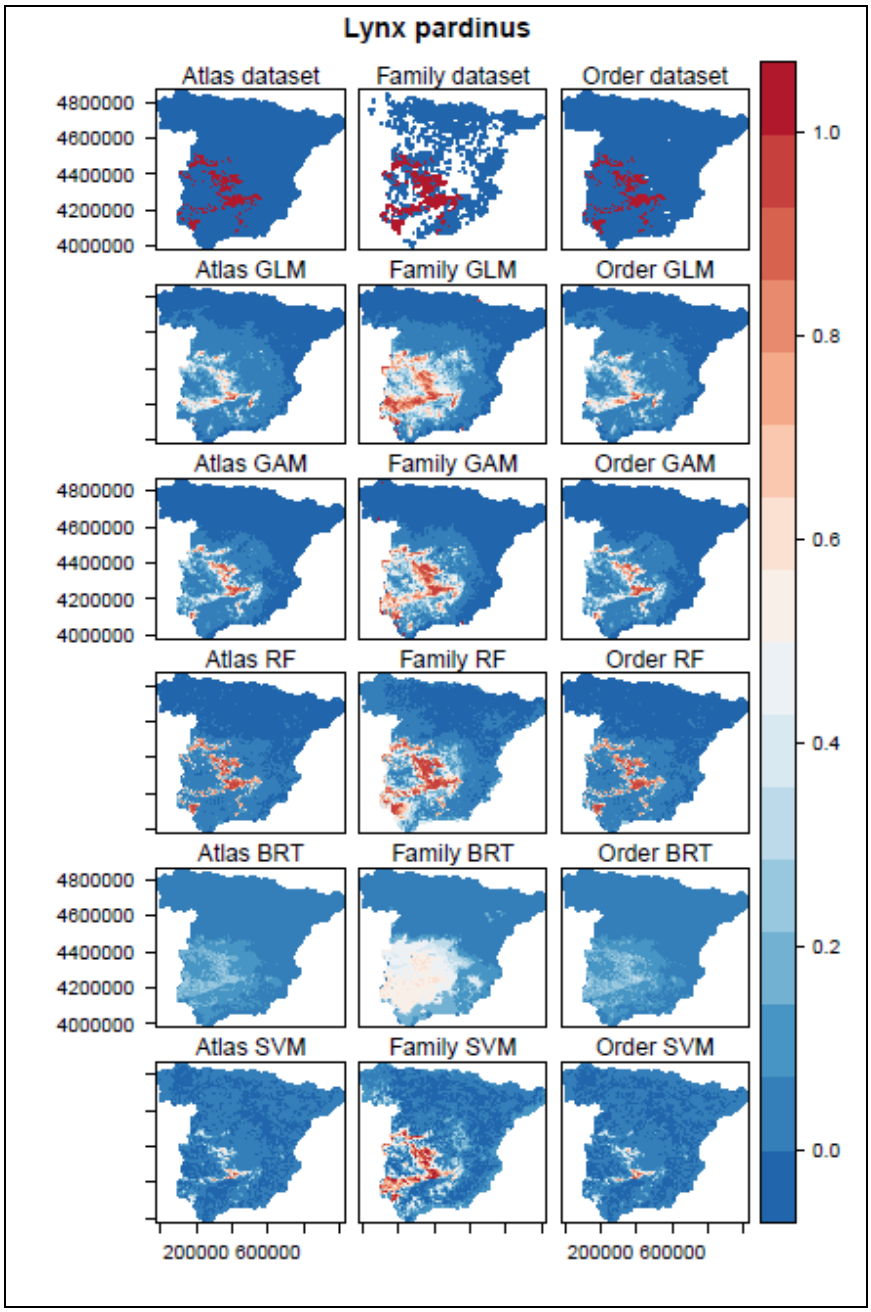
Species distribution maps obtained using atlas dataset as well as taxonomical extents (family and order) to refine five species models for *Alytes obstetricans*. In the top row the red dots indicate species presences and the blue dots indicate species absences per dataset.



Species distribution maps obtained using atlas dataset as well as taxonomical extents (family and order) to refine five species models for *Anas strepera*. In the top row the red dots indicate species presences and the blue dots indicate species absences per dataset.



Species distribution maps obtained using atlas dataset as well as taxonomical extents (family and order) to refine five species models for *Lynx paradinus*. In the top row the red dots indicate species presences and the blue dots indicate species absences per dataset.



Appendix II:

KNOWLEDGE ELICITATION ONLINE QUESTIONNAIRE

This version is presented in rich-text format as an appendix of the fourth chapter of this dissertation. Only part of the questionnaire was used in the current manuscript, other questions meant for other studies.

1. **How familiar are you with the habitat preferences of Bonelli's Eagle (Aquila fasciata)?**
   A little bit [1-2-3-4-5] Very well
2. **What were the main sources of your knowledge?**
   - Direct observations and fieldwork
   - Indirect sources; text books, articles, atlases, discussion forum, news …
   - Common sense
   - Other [TEXT BOX]
3. **There are some questions on the distribution of Bonelli's Eagle specifically related to Andalusia. Would you like to answer them? You will be asked about the protected areas and the provinces.**
   - Yes, go to question 6
   - No, go to question 9
4. **We assumed that "the presence of an active nest, indicates a favorable habitat". Do you agree with our assumption?**
   [TEXT BOX]
5. **How confident are you to express your knowledge on the habitat preferences of Bonelli's Eagle using formal boundaries?**
   No sure at all [1-2-3-4-5] Quite sure
6. **Given the province name, how likely it is that Bonelli's Eagles nest in that province?**
   [I do not know-less favourable-favourable-more favourable]
   - Almería
   - Cádiz
   - Córdoba
   - Granada
   - Huelva
   - Jaén
   - Málaga

- Seville

7. **Given the protected area name, how likely it is that Bonelli's Eagles nest in that area?**

   [I do not know-less favourable-favourable-more favourable]

   - Cabo de Gata-Níjar
   - Sierra María-Los Vélez
   - Bahía de Cádiz
   - De la Breña y Marismas del Barbate
   - Del Estrecho
   - Los Alcornocales
   - Sierra de Grazalema
   - Sierra de Cardeña y Montoro
   - Sierra de Hornachuelos
   - Sierra Subbética
   - Sierra de Baza
   - Sierra de Castril
   - Sierra de Huétor
   - Sierra Nevada
   - Sierra de Aracena y Picos de Aroche
   - Entorno de Doñana
   - Despeñaperros
   - Sierra Mágina
   - Sierra de Andújar
   - Sierras de Cazorla, Segura y Las Villas
   - Montes de Málaga
   - Sierra de las Nieves
   - Sierras de Tejeda, Alhama y Almijara
   - Sierra Norte de Sevilla

8. **Any remarks on this section?**

   [TEXT BOX]

9. **How confident are you to express your knowledge on the habitat preferences of Bonelli's Eagle using land cover classes?**

   No sure at all [1-2-3-4-5] Quite sure

10. **Given the land cover class, how likely it is that Bonelli's Eagles prefer to nest in?**

    [I do not know-not at all-less likely-more likely-definitely]

- Urban fabric
- Industrial, comercial and transport units
- Mine, dump and construction sites
- Artificial, non-agricultural vegetated areas

11. Given the agriculture area, how likely it is that Bonelli's Eagles nest in that area?

[I do not know-not at all-less likely-more likely-definitely]

- Non-irrigated arable land
- Permanently irrigated land
- Rice fields
- Vineyards
- Fruit trees and berry plantations
- Olive groves
- Pastures
- Annual crops associated with permanent crops
- Complex cultivation patterns
- Principally agriculture with significant natural vegetation
- Agro-forestry areas

12. Given the land cover class, how likely it is that Bonelli's Eagles prefer to nest in?

[I do not know-not at all-less likely-more likely-definitely]

- Broad-leaved forest
- Coniferous forest
- Mixed forest
- Natural grasslands
- Moors and heathland
- Sclerophyllous vegetation
- Transitional woodland-shrub
- Beaches, dunes, sands
- Bare rocks
- Sparsely vegetated areas
- Burnt areas

13. Any remarks on this section?

[TEXT BOX]

14. How confident are you to express your knowledge on the habitat preferences of Bonelli's Eagle using topographical attributes?

No sure at all [1-2-3-4-5] Quite sure

15. **Given the topographic characteristic, how likely it is that Bonelli's Eagles prefer to nest in?**

    [I do not know-not at all-less likely-more likely-definitely]

    - Lowlands (below 1000 meter a.s.l)
    - Midlands (1000 - 2000 meter a.s.l)
    - Highlands (above 2000 meter a.s.l)
    - South-looking slopes
    - North-looking slopes
    - Flat area (below 30%)
    - Steep slopes (30 to 100%)
    - Very steep slopes (above 100%)

16. **Do you prefer to modify the above mentioned classes?**

    [TEXT BOX]

17. **Do you prefer to express your judgment with fuzzy statements?**

    [TEXT BOX]

18. **Any remarks on this section?**

    [TEXT BOX]

19. **How confident are you to express your knowledge on the habitat preferences of Bonelli's Eagle using biological interactions?**

    No sure at all [1-2-3-4-5] Quite sure

20. **Given the dominant vegetation species, how likely it is that Bonelli's Eagles prefer to nest in?**

    [I do not know-not at all-less likely-more likely-definitely]

    - Oak
    - Pines
    - Spanish Fir
    - Eucalyptus
    - Populus
    - Ulmus
    - Olive
    - Almond

21. **Do Bonelli's Eagles select their nest location, based on presence or absence of a specific physical state?**

    [TEXT BOX]

22. **Do Bonelli's Eagles select their nest location, based on presence or absence of a specific biological state?**
    [TEXT BOX]
23. **What are Bonelli's Eagles' main prey species? What do they feed on?**
    [TEXT BOX]
24. **Is there any species that symbiotically lives with Bonelli's Eagles? Symbiotically means; when you see that specific species you will conclude that Bonelli's Eagle is there as well.**
    [TEXT BOX]
25. **Is there any species that competes with Bonelli's Eagles? Competition means; when you see that specific species you will conclude that there will be no Bonelli's Eagle.**
    [TEXT BOX]
26. **Is Bonelli's Eagle a territorial species?**
    - Yes
    - No
    - I do not know
27. **Did you think that the presented distribution map of Bonelli's Eagle at the Spanish Atlas of Breeding Birds, is accurate?**
    - Yes, it is quite accurate
    - The observed locations are quite accurate, but i have some doubt on other areas
    - It is an out-dated map
    - I do not know

    Other [TEXT BOX]
28. **Is there any climatic condition that helps us to better define the favourable habitat of Bonelli's Eagle?**
    [TEXT BOX]
29. **Your judgment matters the most. If you have other environmental explanatory variables in mind to express your knowledge on the habitat preferences of Bonelli's Eagle, please explain them here.**
    [TEXT BOX]

30. **May we mention your name and affiliation in acknowledgement of your contribution?**
    - Yes, no problem.
    - No, keep me anonymous.
31. **Name**
32. **Affiliation**
33. **Email Address**
34. **General comments. This is the last question before you submit the form. Do you like to add anything else?**

Bibliography

Abrams, M., Bailey, B., Tsu, H. & Hato, M. (2010) The aster global dem. *Photogrammetric Engineering and Remote Sensing*, **76**, 344-348.

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

An, P., Moon, W.M. & Bonham-Carter, G.F. (1994) Uncertainty management in integration of exploration data using the belief function. *Nonrenewable Resources*, **3**, 60-71.

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on gis models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus nephelomys) in venezuela. *Journal of Biogeography*, **37**, 1378-1393.

Anselin, A., Meire, P.M. & Anselin, L. (1989) Multicriteria techniques in ecological evaluation: An example using the analytical hierarchy process. *Biological Conservation*, **49**, 215-229.

Aranda, S.C. & Lobo, J.M. (2011) How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the tenerife flora. *Ecography*, **34**, 31-38.

Araujo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *In: Journal of biogeography, 33(2006), pp. 1677-1688*,

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

Armstrong, J.S. & Collopy, F. (1992) Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, **8**, 69-80.

Aspinall, W. (2010) A route to more tractable expert advice. *Nature*, **463**, 294-295.

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

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

Ayyub, B.M. (2010) *Elicitation of expert opinions for uncertainty and risks*. CRC press.

Balbontín, J. (2005) Identifying suitable habitat for dispersal in bonelli's eagle: An important issue in halting its decline in europe. *Biological Conservation*, **126**, 74-83.

Baraldi, P. & Zio, E. (2010) A comparison between probabilistic and dempster-shafer theory approaches to model uncertainty analysis in the performance assessment of radioactive waste repositories. *Risk Analysis*, **30**, 1139-1156.

Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, **3**, 327-338.

Barbosa, A.M., Real, R. & Mario Vargas, J. (2009) Transferability of environmental favourability models in geographic space: The case of the iberian desman (galemys pyrenaicus) in portugal and spain. *Ecological Modelling*, **220**, 747-754.

Barbosa, A.M., Real, R. & Mario Vargas, J. (2010) Use of coarse-resolution models of species' distributions to guide local conservation inferences. *Conservation Biology*, **24**, 1378-87.

Barbosa, A.M., Real, R., Olivero, J. & Mario Vargas, J. (2003) Otter (lutra lutra) distribution modeling at two resolution scales suited to conservation planning in the iberian peninsula. *Biological Conservation*, **114**, 377-387.

Barbosa, A.M., Estrada, A., Márquez, A.L., Purvis, A. & Orme, C.D.L. (2012) Atlas versus range maps: Robustness of chorological relationships to distribution data types in european mammals. *Journal of Biogeography*, **39**, 1391-1400.

Bardossy, G. & Fodor, J. (2004) *Evaluation of uncertainties and risks in geology : New mathematical approaches for their handling*. Springer, Berlin.

Barve, N., Barve, V., Jimenez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberon, J. & Villalobos, F. (2011) The

crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810-1819.

Beaugrand, G., Lenoir, S., Ibanez, F. & Mante, C. (2011) A new model to assess the probability of occurrence of a species, based on presence-only data. *Marine Ecology-Progress Series*, **424**, 175-190.

Beynon, M., Cosker, D. & Marshall, D. (2001) An expert system for multi-criteria decision making using dempster shafer theory. *Expert Systems with Applications*, **20**, 357-367.

Bierman, S.M., Butler, A., Marion, G. & Kühn, I. (2010) Bayesian image restoration models for combining expert knowledge on recording activity with species distribution data. *Ecography*, **33**, 451-460.

Binaghi, E., Gallo, I. & Madella, P. (2000) A neural model for fuzzy dempster-shafer classifiers. *International Journal of Approximate Reasoning*, **25**, 89-121.

Birdlife-International (2000). European bird populations:estimates and trends. BirdLife International (BirdLife Conservation Series no. 10), Cambridge, UK.

Birdlife-International (2004) *Birds in europe: Population estimates,trends and conservation status.* BirdLife International, Cambridge, UK.

Birdlife-International (2008) *Conserving the world's birds*. Available at: http://www.birdlife.org/ (accessed November 2009).

Birdlife-International (2011) *Iucn red list for birds*. Available at: (accessed December 2011).

Bloch, I. (1996) Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, **17**, 905-919.

Boitani, L., Corsi, F., Biase, A.D., Carranza, I.D.I., Ravagli, M., Reggiani, G., Sinibaldi, I. & Trapanese, P. (1999) *A databank for the conservation and management of the african mammals*. IEA – Institute of Applied Ecology, Roma.

Bombi, P., Salvi, D. & Bologna, M.A. (2012) Cross-scale predictions allow the identification of local conservation priorities from atlas data. *Animal Conservation*, **15**, 378-387.

Borsuk, M.E. (2004) Predictive assessment of fish health and fish kills in the neuse river estuary using elicited expert judgment. *Human and Ecological Risk Assessment*, **10**, 415-434.

Bosch, R., Real, J., Tinto, 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.

Bossard, M., Feranec, J. & Otahel, J. (2000) *Corine land cover technical guide – addendum 2000*. European Enviromnet Agency, Copenhagen.

Boudraa, A.O., Bentabet, A. & Salzenstein, F. (2004) Dempster-shafer's basic probability assignment based on fuzzy membership functions. *Electronic Letters on Computer Vision and Image Analysis*, **4**, 1-9.

Boussabaine, A.H. & Duff, A.R. (1996) An expert-simulation system for construction productivity forecasting: Conformation method used to illustrate the uncertainty associated with expert's knowledge. *Building Research and Information*, **24**, 279-286.

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

Breiman, L. (2001a) Statistical modeling: The two cultures. *Statistical Science*, **16**, 199-231.

Breiman, L. (2001b) Random forests. *Machine Learning*, **45**, 5-32.

Brown, J.D. (2004) Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief. *Transactions of the Institute of British Geographers*, **29**, 367-381.

Brown, J.H. (1995) *Macroecology*. University Of Chicago, Chicago and London.

Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: Size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597-623.

Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145-1157.

Burgman, M.A., Mcbride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L. & Twardy, C. (2011) Expert status and performance. *PLoS ONE*, **6**, e22998.

Capel Molina, J.J. (1981) *Los climas de españa*. Oikos-Tau, Vilassar de Mar, Barcelona.

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

Carpenter, S.R. (2002) Ecological futures: Building an ecology of the long now. *Ecology*, **83**, 2069-2083.

Carranza, E.J.M. & Hale, M. (2003) Evidential belief functions for data-driven geologically constrained mapping of gold potential, baguio district, philippines. *Ore Geology Reviews*, **22**, 117-132.

Carranza, E.J.M., Woldai, T. & Chikambwe, E.M. (2005) Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, lundazi district, zambia. *Natural Resources Research*, **14**, 47-63.

Carrascal, L.M. & Seoane, J. (2009) Factors affecting large-scale distribution of the bonelli's eagle aquila fasciata in spain. *Ecological Research*, **24**, 565-573.

Carrete, M., Sanchez-Zapata, J.A., Martinez, J.E., Sanchez, M.A. & Calvo, J.F. (2002) Factors influencing the decline of a bonelli's eagle hieraaetus fasciatus population in southeastern spain: Demography, habitat or competition? *Biodiversity and Conservation*, **11**, 975-985.

Choy, S.L., O'leary, R. & Mengersen, K. (2009) Elicitation by design in ecology: Using expert opinion to inform priors for bayesian statistical models. *Ecology*, **90**, 265-277.

Chung, C.J. & Fabbri, A. (1993) The representation of geoscience information for data integration. *Natural Resources Research*, **2**, 122-139.

Clark, J.S. (2005) Why environmental scientists are becoming bayesians. *Ecology Letters*, **8**, 2-14.

Clements, A.C., Pfeiffer, D.U. & Martin, V. (2006) Application of knowledge-driven spatial modelling approaches and uncertainty management to a study of rift valley fever in africa. *Int J Health Geogr*, **5**, 57.

Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: Issues of spatial scale and extent. *Journal of Applied Ecology*, **37**, 13-27.

Comber, A., Fisher, P. & Wadsworth, R. (2005) Comparing the consistency of expert land cover knowledge. *International Journal of Applied Earth Observation and Geoinformation*, **7**, 189-201.

Comber, A.J., Law, A.N.R. & Lishman, J.R. (2004) A comparison of bayes', dempster-shafer and endorsement theories for managing knowledge uncertainty in the context of land cover monitoring. *Computers, Environment and Urban Systems*, **28**, 311-327.

Cooke, R.M. (1991) *Experts in uncertainty, opinion and subjective probability in science*. Oxford University Press, Oxford.

Costa, G.C., Nogueira, C., Machado, R.B. & Colli, G.R. (2010) Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot. *Biodiversity and Conservation*, **19**, 883-899.

Costa, M., Morla, C. & Sainz, H. (1997) *Los bosques ibéricos: Una interpretación geobotánica*.

Cutler, D.R., Edwards, T.C., Jr., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. & Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, **88**, 2783-92.

De Araujo, C.B., Marcondes-Machado, L.O. & Costa, G.C. (2014) The importance of biotic interactions in species distribution models: A test of the eltonian noise hypothesis using parrots. *Journal of Biogeography*, **41**, 513-523.

Del Hoyo, J., Elliot, A. & Sargatal, J. (1994) *Handbook of the birds of the world, vol. Ii. New world vultures to guineafowl*, Lynx edn, Barcelona.

Del Moral, J.C. (2006) *El águila perdicera en españa. Población en 2005 y método de censo*. SEO/BirdLife, Madrid.

Dempster, A.P. (1967) Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, **38**, 325-339.

Dempster, A.P. & Weisberg, H. (1968) A generalization of bayesian inference. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **30**, 205-247.

Dennis, B. (1996) Discussion: Should ecologists become bayesians? *Ecological Applications*, **6**, 1095-1103.

Dormann, C.F., Mcpherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, **30**, 609-628.

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.

Drake, J.M., Randin, C. & Guisan, A. (2006) Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, **43**, 424-432.

Drescher, M., Buse, L., Perera, A. & Ouellette, M. (2012) Eliciting expert knowledge of forest succession using an innovative software tool. *Expert knowledge and its application in landscape ecology* (ed. by A.H. Perera, C.A. Drew and C.J. Johnson), pp. 69-85. Springer New York.

Edward, J. & Rykiel, J. (1989) Artificial intelligence and expert systems in ecology and natural resource management. *Ecological Modelling*, **46**, 3-8.

Elith, J. & Leathwick, J. (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*, **13**, 265-275.

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. & 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.

Elith, J., Burgman, M.A. & Regan, H.M. (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313-329.

Elith, J., Leathwick, J.R. & Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802-13.

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

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., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire,

R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

Ellison, A.M. (2004) Bayesian inference in ecology. *Ecology Letters*, **7**, 509-520.

Engler, R., Guisan, A. & Rechsteiner, L. (2004) Improved approach for predicting the distribution of rare and endangered species from occurence and pseudo - absence data. *Journal of Applied Ecology*, **41**

Estrada, A., Real, R. & Vargas, J.M. (2008) Using crisp and fuzzy modelling to identify favourability hotspots useful to perform gap analysis. *Biodiversity and Conservation*, **17**, 857-871.

European Environment Agency (2012) Corine land cover 2006 raster data. In. The European Topic Centre on Spatial Information and Analysis

Evangelista, P.H., Kumar, S., Stohlgren, T.J., Jarnevich, C.S., Crall, A.W., Norman Iii, J.B. & Barnett, D.T. (2008) Modelling invasion for a habitat generalist and a specialist plant species. *Diversity and Distributions*, **14**, 808-817.

Fazey, I., J. A. Fazey & Fazey, D.M.A. (2005) Learning more effectively from experience. *Ecology and Society*, **10**, 4.

Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast new south wales. I. Species-level modelling. *Biodiversity and Conservation*, **11**, 2275-2307.

Field, S.A., Tyre, A.J., Jonzen, N., Rhodes, J.R. & Possingham, H.P. (2004) Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, **7**, 669-675.

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.

Firestein, S. (2012) *Ignorance: How it drives science*. Oxford University Press, USA.

Font (1983) *Climatología de españa y portugal*. Universidad de Salamanca, Salamanca.

Font (2000) *Climatología de españa y portugal*. Universidad de Salamanca, Salamanca.

Forsyth, R. (1984) *Expert systems: Principles and case studies*. Chapman and Hall, London.

Franklin, J. (2009) *Mapping species distributions: Spatial inference and prediction*. Cambridge University, Cambridge.

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

Freeman, E.A. & Moisen, G. (2008) Presenceabsence: An r package for presence absence analysis. *Journal of Statistical Software*, **23**, 1-31.

Friedman, J.H. (2001) Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, **29**, 1189-1232.

Gaston, K.J. & Fuller, R.A. (2009) The sizes of species' geographic ranges. *Journal of Applied Ecology*, **46**, 1-9.

Gil-Sanchez, J.M., Moleon, M., Otero, M. & Bautista, J. (2004) A nine-year study of successful breeding in a bonelli's eagle population in southeast spain: A basis for conservation. *Biological Conservation*, **118**, 685-694.

Gil-Sánchez, J.M., Molino-Garrido, F. & Valenzuela, G. (1996) Selección de hábitat de nidificación por el águila perdicera (hieraaetus fasciatus) en granada (se de españa). *Ardeola*, **43**, 189–197.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004a) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, **19**, 497-503.

Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J., Moritz, C. & Cunningham, C. (2004b) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781-1793.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & Nceas Predect Species Working, G. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, **45**, 239-247.

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

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

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

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615-630.

Hartley, S., Kunin, W.E., Lennon, J.J. & Pocock, M.J.O. (2003) Coherence and discontinuity in the scaling of species' distribution patterns. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 81-88.

Hastie, T. & Tibshirani, R. (1990) *Generalized additive models*. Chapman and Hall, London.

Heikkinen, J. & Hogmander, H. (1994) Fully bayesian approach to image restoration with an application in biogeography. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **43**, 569-582.

Heikkinen, R.K., Marmion, M. & Luoto, M. (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, **35**, 276-288.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W. & Sykes, M.T. (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, **30**, 751-777.

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 modelling methods. *Ecography*, **29**, 773-785.

Heuvelink, G.B.M. (1998) *Error propagation in environmental modelling with gis*. Taylor & Francis, London etc.

Hijmans, R.J. & Van Etten, J. (2010) *Raster: Package for reading, writing, and manipulating raster (grid) type geographic (spatial) data*.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hobbs, N.T. (2003) Challenges and opportunities in integrating ecological knowledge across scales. *Forest Ecology and Management*, **181**, 223-238.

Hoffman, J.C. & Murphy, R.R. (1993) Comparision of bayesian and dempster-shafer theory for sensing: A practitioner's approach. *Neutral and stochastic methods in image and singnal processing*, pp. 266--279.

Hollander, M. & Wolfe, D.A. (1973) *Nonparametric statistical methods*. John Wiley & Sons, New York.

Hopkinson, P., Evans, J. & Gregory, R.D. (2000) National-scale conservation assessments at an appropriate resolution. *Diversity and Distributions*, **6**, 195-204.

Hortal, J., Lobo, J.M. & Jiménez-Valverde, A. (2012) Basic questions in biogeography and the (lack of) simplicity of species distributions: Putting species distribution models in the right place. *Natureza a Conservacao*, **10**, 108-118.

Hortal, J., Jimenez-Valverde, A., Gomez, J.F., Lobo, J.M. & Baselga, A. (2008) Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, **117**, 847-858.

Hosmer, D.W. & Lameshow, S. (2000) *Applied logistic regression*, 2nd edn. John Wiley & Sons, New York.

Huston, M.A. (2002) Critical issues for improving predictions. *Predicting species occurrences: Issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, M.G. Raphael and W.A. Wall), pp. 7–24. Island Press, Covelo, CA.

I.G.N. (1999) Mapa de carreteras. Península ibérica, baleares y canarias. In. Instituto Geográfico Nacional/Ministerio de Fomento, Madrid.

James, A., Gaston, K.J. & Balmford, A. (2001) Can we afford to conserve biodiversity? *Bioscience*, **51**, 43-52.

James, A., Choy, S.L. & Mengersen, K. (2010) Elicitator: An expert elicitation tool for regression in ecology. *Environmental Modelling & Software*, **25**, 129-145.

Janssen, J.a.E.B., Krol, M.S., Schielen, R.M.J., Hoekstra, A.Y. & De Kok, J.L. (2010) Assessment of uncertainties in expert knowledge, illustrated in fuzzy rule-based models. *Ecological Modelling*, **221**, 1245-1251.

Jeschke, J.M. & Strayer, D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, **1134**, 1-24.

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. (2014) Threshold-dependence as a desirable attribute for discrimination assessment: Implications for the evaluation of species distribution models. *Biodiversity and Conservation*, 1-17.

Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology*, **31**, 361-369.

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.

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.

Jiménez-Valverde, A., Acevedo, P., Barbosa, A.M., Lobo, J.M. & Real, R. (2013) Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*, **22**, 508-516.

Jiménez, J.J. & Muñoz, A.R. (2008) *Atlas de las rapaces diurnas de la provincia de málaga (reproducción, invernada y migración)*. Centro de Ediciones de la Diputación de Málaga, , Spain.

Johnson, C., Drew, C.A. & Perera, A. (2012) Elicitation and use of expert knowledge in landscape ecological applications: A synthesis. *Expert knowledge and its application in landscape ecology* (ed. by A.H. Perera, C.A. Drew and C.J. Johnson), pp. 279-299. Springer New York.

Johnson, C.J. & Gillingham, M.P. (2004) Mapping uncertainty: Sensitivity of wildlife habitat ratings to expert opinion. *Journal of Applied Ecology*, **41**, 1032-1041.

Kearney, M., Simpson, S.J., Raubenheimer, D. & Helmuth, B. (2010) Modelling the ecological niche from functional traits. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 3469-3483.

Kent, R. & Carmel, Y. (2011) Presence-only versus presence–absence data in species composition determinant analyses. *Diversity and Distributions*, **17**, 474-479.

Knol, A., Slottje, P., Van Der Sluijs, J. & Lebret, E. (2010) The use of expert elicitation in environmental health impact assessment: A seven step procedure. *Environmental Health*, **9**, 19.

Kuhnert, P.M. (2011) Four case studies in using expert opinion to inform priors. *Environmetrics*, **22**, 662–674.

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.

Laplace, P.S. (1825) *Philosphical essay on probabilities (translated from the fifth french edition of 1825)*. Springer -Verlag, New York.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014) Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54-64.

Le Hegarat-Mascle, S., Bloch, I. & Vidal-Madjar, D. (1997) Application of dempster-shafer evidence theory to unsupervised classification in multisource remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, **35**, 1018-1031.

Lee, T., Richards, J.A. & Swain, P.H. (1987) Probabilistic and evidential approaches for multisource data analysis. *Geoscience and Remote Sensing, IEEE Transactions on*, **GE-25**, 283-293.

Legendre, P. (1993a) Spatial autogorrelation: Trouble or new paradigm. *Ecology*, **74**, 1659-1673.

Legendre, P. (1993b) Spatial autocorrelation: Trouble or new paradigm? *Ecology*, **74**, 1659-1673.

Lehmann, A., Overton, J.M. & Leathwick, J.R. (2002) Grasp: Generalized regression analysis and spatial prediction. *Ecological Modelling*, **157**, 189-207.

Lemeshow, S. & Hosmer, D.W. (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, **115**, 92-106.

Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkönig, I.M.A. & Skidmore, A. (2007) Capturing the fugitive: Applying remote sensing to terrestrial animal distribution and diversity. *International Journal of Applied Earth Observation and Geoinformation*, **9**, 1-20.

Lloyd, P. & Palmer, A.R. (1998) Abiotic factors as predictors of distribution in southern african bulbuls. *Auk*, **115**, 404-411.

Lobo, J.M. (2008) *More complex distribution models or more representative data?*

Lobo, J.M. & Tognelli, M.F. (2011) Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation*, **19**, 1-7.

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

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

López-Iborra, G.M., Limiñana, R., Pavón, D. & Martínez-Pérez, J.E. (2010) Modelling the distribution of short-toed eagle (circaetus gallicus) in semi-arid mediterranean landscapes: Identifying important explanatory variables and their implications for its conservation. *European Journal of Wildlife Research*, 1-11.

Lopez-Lopez, P., Garcia-Ripolles, C. & Urios, V. (2007a) Population size, breeding performance and territory quality of bonelli's eagle hieraaetus fasciatus in eastern spain. *Bird Study*, **54**, 335-342.

Lopez-Lopez, P., Garcia-Ripolles, C., Aguilar, J., Garcia-Lopez, F. & Verdejo, J. (2006) Modelling breeding habitat preferences of bonelli's eagle (hieraaetus fasciatus) in relation to topography, disturbance, climate and land use at different spatial scales. *Journal of Ornithology*, **147**, 97-106.

Lopez-Lopez, P., Garcia-Ripolles, C., Soutullo, A., Cadahia, L. & Urios, V. (2007b) Are important bird areas and special protected areas enough for conservation?: The case of bonelli's eagle in a mediterranean area. *Biodiversity and Conservation*, **16**, 3755-3780.

Luck, G.W. (2002) The habitat requirements of the rufous treecreeper (climacteris rufa). 2. Validating predictive habitat models. *Biological Conservation*, **105**, 395-403.

Lukey, J.R., Crawford, S.S., Gillis, D.J. & Gillespie, M.G. (2011) Effect of ecological uncertainty on species at risk decision-making: Cosewic expert opinion as a case study. *Animal Conservation*, **14**, 151-157.

Luoto, M., Marmion, M. & Hjort, J. (2010) Assessing spatial uncertainty in predictive geomorphological mapping: A multi-modelling approach. *Computers & Geosciences*, **36**, 355-361.

Lutolf, M., Kienast, F. & Guisan, A. (2006) The ghost of past species occurrence: Improving species distribution models for presence-only data. *Journal of Applied Ecology*, **43**, 802-815.

Mackenzie, D.I. (2005) Was it there? Dealing with imperfect detection for species presence/absence data. *Australian and New Zealand Journal of Statistics*, **47**, 65-74.

Malpica, J.A., Alonso, M.C. & Sanz, M.A. (2007) Dempster-shafer theory in geographic information systems: A survey. *Expert Systems with Applications*, **32**, 47-55.

Marti, R. & Del Moral, J.C. (2003) *Atlas de las aves reproductoras de españa*. Sociedad Española de Ornitologıa, Madrid.

Martin, T.G., Kuhnert, P.M., Mengersen, K. & Possingham, H.P. (2005) The power of expert opinion in ecological models using bayesian methods: Impact of grazing on birds. *Ecological Applications*, **15**, 266-280.

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

Martinez, J.A., Calvo, J.F., Martinez, J.E., Zuberogoitia, I., Zabala, J. & Redpath, S.M. (2008a) Breeding performance, age effects and territory occupancy in a bonelli's eagle hieraaetus fasciatus population. *Ibis*, **150**, 223-233.

Martinez, J.E., Martinez, J.A., Zuberogoitia, I., Zabala, J., Redpath, S.M. & Calvo, J.F. (2008b) The effect of intra- and interspecific interactions on the large-scale distribution of cliff-nesting raptors. *Ornis Fennica*, **85**, 13-21.

Mcbride, M.F., Fidler, F. & Burgman, M.A. (2012a) Evaluating the accuracy and calibration of expert predictions under uncertainty: Predicting the outcomes of ecological research. *Diversity and Distributions*, **18**, 782-794.

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. (2012b) Structured elicitation of expert judgments for threatened species assessment: A case study

on a continental scale using email. *Methods in Ecology and Evolution*, no-no.

Mccarthy, M.A. (2007) *Bayesian methods for ecology*. Cambridge University Press, New York.

Mccullagh, P.a.N., J.A. (1989) *Generalized linear models*. Chapman and Hall, London.

Mcpherson, J.M., Jetz, W. & Rogers, D.J. (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions-possibilities and limitations. *Ecological Modelling*, **192**, 499-522.

Mertikas, P. & Zervakis, M.E. (2001) Exemplifying the theory of evidence in remote sensing image classification. *International journal of remote sensing*, **22**, 1081-1095.

Meyer, C.B. & Thuiller, W. (2006) Accuracy of resource selection functions across spatial scales. *Diversity and Distributions*, **12**, 288-297.

Miller, M.E., Hui, S.L. & Tierney, W.M. (1991) Validation techniques for logistic regression models. *Statistics in Medicine*, **10**, 1213-26.

Moleon, M., Bautista, J., Sanchez-Zapata, J.A. & Gil-Sanchez, J.M. (2009) Diet of non-breeding bonelli's eagles hieraaetus fasciatus at settlement areas of southern spain. *Bird Study*, **56**, 142-146.

Montgomery, D.C., & Peck, E. A. (1982) *Introduction to linear regression analysis*. John Wiley and Sons, New York.

Moon, W.M. (1989) Integration of remote sensing and geophysical - geological data using dempster - shafer approach. *In: IGARSS '89 : Geoscience and remote sensing symposium : 12th Canadian symposium on remote sensing, 10-14 Jul 1989, Vancouver, Canada. pp. 838-841*,

Moreno-Rueda, G., Pizarro, M., Ontiveros, D. & Pleguezuelos, J.M. (2009) The coexistence of the eagles aquila chrysaetos and hieraaetus fasciatus increases with low human population density, intermediate temperature, and high prey diversity. *Annales Zoologici Fennici*, **46**, 283-290.

Muñoz, A.-R. & Real, R. (2013) Distribution of bonelli's eagle aquila fasciata in southern spain: Scale may matter. *Acta Ornithologica*, **48**, 93-101.

Muñoz, A.-R., Márquez, A.L. & Real, R. (2013) Updating known distribution models for forecasting climate change impact on endangered species. *PLoS ONE*, **8**, e65462.

Muñoz, A.R., Real, R., Barbosa, A.M. & Vargas, J.M. (2005) Modelling the distribution of bonelli's eagle in spain: Implications for conservation planning. *Diversity and Distributions*, **11**, 477-486.

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.

Naimi, B., Skidmore, A.K., Groen, T.A. & Hamm, N.a.S. (2011) Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *Journal of Biogeography*, **38**, 1497–1509.

Naimi, B., Hamm, N.a.S., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2013) Where is positional uncertainty a problem for species distribution modelling? *Ecography*, no-no.

Naylor, C. (1984) How to build an inferencing engin. *Expert systems: Principals and case studies* (ed. by R. Forsyth). Chapman and Hall, London.

Newbold, T. (2010) Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, **34**, 3-22.

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Munoz, A.R. & Real, R. (2011) Finessing atlas data for species distribution models. *Diversity and Distributions*, **17**, 1173-1185.

Niels Raes, H.T.S. (2007) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727-736.

Oindo, B.O., Skidmore, A.K. & De Salvo, P. (2003) Mapping habitat and biological diversity in the maasai mara ecosystem. *International journal of remote sensing*, **24**, 1053-1069.

Olivero, J., Márquez, A.L. & Real, R. (2013) Integrating fuzzy logic and statistics to improve the reliable delimitation of biogeographic regions and transition zones. *Systematic Biology*, **62**, 1-21.

Ontiveros, D. & Pleguezuelos, J.M. (2000) Influence of prey densities in the distribution and breeding success of bonelli's eagle (hieraaetus fasciatus): Management implications. *Biological Conservation*, **93**, 19-25.

Ontiveros, D. & Pleguezuelos, J.M. (2003a) Physical, environmental and human factors influencing productivity in bonelli's eagle hieraaetus fasciatus in granada (se spain). *Biodiversity and Conservation*, **12**, 1193-1203.

Ontiveros, D. & Pleguezuelos, J.M. (2003b) Influence of climate on bonelli's eagle's (hieraaetus fasciatus v, 1822) breeding success through the western mediterranean. *Journal of Biogeography*, **30**, 755-760.

Ontiveros, D., Pleguezuelos, J.M. & Caro, J. (2005) Prey density, prey detectability and food habits: The case of bonelli's eagle and the conservation measures. *Biological Conservation*, **123**, 19-25.

Ortega-Huerta, M.A. & Peterson, A.T. (2008) Modeling ecological niches and predicting geographic distributions: A test of six presence-only methods. *Revista Mexicana De Biodiversidad*, **79**, 205-216.

Orton, T.G. & Lark, R.M. (2007) Accounting for the uncertainty in the local mean in spatial prediction by bayesian maximum entropy. *Stochastic Environmental Research and Risk Assessment*, **21**, 773-784.

Ozesmi, U. & Ozesmi, S.L. (2004) Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, **176**, 43-64.

Palma, L., Beja, P., Pais, M. & Da Fonseca, L.C. (2006) Why do raptors take domestic prey? The case of bonelli's eagles and pigeons. *Journal of Applied Ecology*, **43**, 1075-1086.

Palomino, D., Valls, J. & Bermejo , A. (2011) Large scale specific censuses in spain. In: *SEO/BirdLife monitoring programmes in 2009-2010* eds. V. Escandell, D. Palomino, B. Molina, A. Leal, C. Remacha, A. Bermejo, J. De La Puente and J.C. Del Moral), pp. 28-33, Madrid.

Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127-147.

Pearce, J.L. & Boyce, M.S. (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, **43**, 405-412.

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.

Pearman, P.B., Guisan, A., Broennimann, O. & Randin, C.F. (2008) Niche dynamics in space and time. *Trends in Ecology & Evolution*, **23**, 149-158.

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.

Pereira, J.M.C. & Itami, R.M. (1991) Gis-based habitat modeling using logistic multiple regression: A study of the mt graham red squirrel. *Photogrammetric Engineering and Remote Sensing*, **57**, 1475-1486.

Perera, A., Drew, C.A. & Johnson, C. (2012) Experts, expert knowledge, and their roles in landscape ecological applications. *Expert knowledge and its application in landscape ecology* (ed. by A.H. Perera, C.A. Drew and C.J. Johnson), pp. 1-10. Springer New York.

Perony, N. (2013) Puppies! Now that i've got your attention, complexity theory. In, TEDxZurich.

Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol*, **78**, 419-33.

Peterson, A.T., Papes, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: A comparison of garp and maxent. *Ecography*, **30**, 550-560.

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

Phillips, S.J., Dudik, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181-197.

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

R Development Core Team (2013) *R: A language and environment for statistical computing. R foundation for statistical computing*, Vienna.

Raes, N. & Ter Steege, H. (2007) A null-model for significance testing of presence-only species distribution models. *Ecography*, **30**, 727-736.

Rangel, T.F., Diniz, J.a.F. & Bini, L.M. (2010) Sam: A comprehensive application for spatial analysis in macroecology. *Ecography*, **33**, 46-50.

Real, J. (2003) *Águila-azor perdicera, hieraaetus fasciatus. Atlas de las aves reproductoras de españa. (ed. By r. Marti and j.C. Del moral)*. Ministerio de Medio Ambiente-Sociedad Española de Ornitología, Madrid.

Real, R., Barbosa, A.M. & Vargas, J.M. (2006) Obtaining environmental favourability functions from logistic regression *Environmental and Ecological Statistics*, **13**, 237-245.

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.

Reddy, S.G. (1996) Claims to expert knowledge and the subversion of democracy: The triumph of risk over uncertainty. *Economy and Society*, **25**, 222-254.

Refsgaard, J.C., Van Der Sluijs, J.P., Højberg, A.L. & Vanrolleghem, P.A. (2007) Uncertainty in the environmental modelling process – a framework and guidance. *Environmental Modelling & Software*, **22**, 1543-1556.

Reiss, H., Cunze, S., Konig, K., Neumann, H. & Kroncke, I. (2011) Species distribution modelling of marine benthos: A north sea case study. *Marine Ecology Progress Series*, **442**, 71-86.

Richards, C.L., Carstens, B.C. & Lacey Knowles, L. (2007) Distribution modelling and statistical phylogeography: An integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833-1845.

Rico, L., Sánchez-Zapata, J.A., Izquierdo, A., García, J.R., Morán, S. & Rico, D. (1999) Tendencias recientes en las poblaciones del águila real aquila chrysaetos y el águila-azor perdicera hieraaetus fasciatus en la provincia de valencia. *Ardeola*, **46**, 235–238.

Rocamora, G. (1994) *Bonelli's eagle hieraaetus fasciatus. Birds in europe, their conservation status (ed. By g.M. Tucker and m.F. Heath)*. BirdLife International (BirdLife Conservation Series no. 3), Cambridge, UK.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jimenez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011) Accounting for

uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography*, **35**, 211-226.

Rombaut, M. & Min Zhu, Y. (2002) Study of dempster-shafer theory for image segmentation applications. *Image and Vision Computing*, **20**, 15-23.

Rouget, M. (2003) Measuring conservation value at fine and broad scales: Implications for a diverse and fragmented region, the agulhas plain. *Biological Conservation*, **112**, 217-232.

Roy, K., Hunt, G., Jablonski, D., Krug, A.Z. & Valentine, J.W. (2009) A macroevolutionary perspective on species range limits. *Proceedings of the Royal Society B*, **276**, 1485-93.

Rüger, N., Schlüter, M. & Matthies, M. (2005) A fuzzy habitat suitability index for populus euphratica in the northern amudarya delta (uzbekistan). *Ecological Modelling*, **184**, 313-328.

Salski, A. (1992) Fuzzy knowledge-based models in ecological research. *Ecological Modelling*, **63**, 103-112.

Schmidt, K.S., Skidmore, A.K., Kloosterman, E.H., Van Oosten, H., Kumar, L. & Janssen, J.a.M. (2004) Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing*, **70**, 703-715.

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

Seoane, J., Bustamante, J. & Diaz-Delgado, R. (2005) Effect of expert opinion on the predictive ability of environmental models of bird distribution. *Conservation Biology*, **19**, 512-522.

Shafer, G. (1976) *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J.

Shafer, G. (1982) Belief functions and parametric models. *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**, 322-352.

Sillero, N., Brito, J.C., Skidmore, A.K. & Toxopeus, A.G. (2009) Biogeographical patterns derived from remote sensing variables: The amphibians and reptiles of the iberian peninsula. *Amphibia-Reptilia*, **30**, 185-206.

Skidmore, A.K. (1989) An expert system classifies eucalypt forest types using thematic mapper data and digital terrain model. *Photogrammetric Engineering and Remote Sensing*, **55**, 1449-1464.

Skidmore, A.K. & Turner, B.J. (1992) Map accuracy assessment using line intersect sampling. *Photogrammetric Engineering and Remote Sensing*, **58**, 1453-1457.

Skidmore, A.K., Gauld, A. & Walker, P. (1996) Classification of kangaroo habitat distribution using three gis models. *International Journal of Geographical Information Systems*, **10**, 441-454.

Smith, C., Van Klinken, R.D., Seabrook, L. & Mcalpine, C. (2012) Estimating the influence of land management change on weed invasion potential using expert knowledge. *Diversity and Distributions*, **18**, 818-831.

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19644-50.

Soberón, J. & Peterson, T. (2004) Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **359**, 689-698.

Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1-10.

Soutullo, A., Lopez-Lopez, P. & Urios, V. (2008) Incorporating spatial structure and stochasticity in endangered bonelli's eagle's population models: Implications for conservation and management. *Biological Conservation*, **141**, 1013-1020.

Stockwell, D. (1999) The garp modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

Store, R. & Kangas, J. (2001) Integrating spatial multi-criteria evaluation and expert knowledge for gis-based habitat suitability modelling. *Landscape and Urban Planning*, **55**, 79-93.

Sutherland, W.J. (2006) Predicting the ecological consequences of environmental change: A review of the methods*. *Journal of Applied Ecology*, **43**, 599-616.

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*, In press.

Tangestani, M.H. & Moore, F. (2002) The use of dempster-shafer model and gis in integration of geoscientific data for porphyry copper potential mapping, north of shahr-e-babak, iran. *International Journal of Applied Earth Observation and Geoinformation*, **4**, 65-74.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. (2005) Climate change threats to plant diversity in europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245-8250.

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.

U. S. Geological Survey (1996) Gtopo30. In. Land Processes Distributed Archive Center (http://edcdaac.usgs.gov/gtopo30/gtopo30.asp)

U. S. Geological Survey (2004) Shuttle radar topography mission. In. Global Land Cover Facility University of Maryland, College Park, Maryland.

Wade, P.R. (2000) Bayesian methods in conservation biology. *Conservation Biology*, **14**, 1308-1316.

Walker, W.E., Harremoës, P., Rotmans, J., Van Der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P. & Krayer Von Krauss, M.P. (2003) Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, **4**, 5-17.

Walley, P. (1996) Measures of uncertainty in expert systems. *Artificial Intelligence*, **83**, 1-58.

Wang, T., Ye, X., Skidmore, A.K. & Toxopeus, A.G. (2010) Characterizing the spatial distribution of giant pandas (ailuropoda melanoleuca) in fragmented forest landscapes. *Journal of Biogeography*, **37**, 865-878.

Wang, T.J., Skidmore, A.K. & Toxopeus, A.G. (2009) Improved understorey bamboo cover mapping using a novel hybrid neural network and expert system. *International journal of remote sensing*, **30**, 965-981.

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.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H. & Guisan, A. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763-773.

Wright, D.F. & Bonham-Carter, G.F. (1996) Vhms favourability mapping with gis-based integration models, chisel lake - anderson lake area. *Extech i: A multidisciplinary approach to massive sulphide research in the rusty lake - snow lake greenstone belts, manitoba* (ed. by G.F. Bonham-Carter, A.G. Galley and G.E.M. Hall). Geological Survey Canada Bulletin, Canada.

Xu, Z.W., Khoshgoftaar, T.M. & Allen, E.B. (2003) Application of fuzzy expert systems in assessing operational risk of software. *Information and Software Technology*, **45**, 373-388.

Yamada, K., Elith, J., Mccarthy, M. & Zerger, A. (2003) Eliciting and integrating expert knowledge for wildlife habitat modelling. *Ecological Modelling*, **165**, 251-264.

Yang, X., Skidmore, A.K., Melick, D.R., Zhou, Z. & Xu, J. (2006) Mapping non-wood forest product (matsutake mushrooms) using logistic regression and a gis expert system. *Ecological Modelling*, **198**, 208-218.

Zadeh, L.A. (1965) Fuzzy sets. *Information and Control*, **8**, 338-353.

Summary

Species distribution modelling is one of the most important tasks in current ecological research. There is an increasing demand for species distribution models to assess the impact of climate change on biodiversity, to address numerous questions in conservation biology and evolution, to test biogeographical hypotheses, to predict species' invasion and reproduction, and to anticipate new occurrences for rare species. The demand for such information along with advances in statistics and information technology has led to a variety of rigorous and complicated statistical methods to model the distribution of a species. As a result the consideration of underlying ecological aspects tends to become of secondary importance.

Experts provide valuable insight into species-environment interactions and can potentially optimise the use of ecological concepts in a species distribution modelling procedure. Experts may offer beliefs given specific environmental conditions in order to bypass complex systems and provide parsimonious solutions that focus on the key aspects of a situation.

This dissertation aims to contribute to understanding if and to what extent the incorporation of expert knowledge into species distribution models enhances the performance of those models. The key factor to successfully improve the performance of a species distribution model using expert knowledge is to get experts involved in each and every step of the modelling, from hypothesis generation to visualisation of the outputs.

Part of my dissertation is dedicated to providing evidence that incorporation of expert knowledge into a species sampling scheme significantly improves the discriminating capacity of models at a resolution 100 times finer than that of available species atlas data. This approach forms a feasible substitute for comprehensive fieldwork, however, may be limited to well-known species.

A common hypothesis is that species have relatively similar environmental preferences to their ancestors. Another part of the dissertation provides evidence for this hypothesis and puts it to use in species distribution modelling. I used species taxonomy to delineate environmental extent, and investigated how this affects the performance of species distribution models. Enhancement of species atlas data by excluding irrelevant absences based on species taxonomy significantly improves the calibration of species distribution models and thus their capacity to predict environmental potential.

Expert knowledge is particularly of value where information describing species distribution is scarce or unobtainable. However, it is always subject to uncertainty and accounting for this uncertainty in the modelling procedure poses a challenge. I propose using the Dempster-Shafer Theory of Evidence to incorporate expert knowledge and to accommodate its uncertainty regarding species distribution modelling. This approach explicitly accommodates knowledge uncertainty and is a practical alternative where the knowledge of a species' geographical distribution is needed.

In this dissertation I illustrate practical examples of improved species distribution models using expert knowledge. Where experts are available and a proper method is employed, there will be advancement in our understanding and in the development of a solution to the complex problems faced by the species distribution modeller. I am confident that the application of expert knowledge in species distribution modelling will continue to expand and improve.

Samenvatting

Het modelleren van de verspreiding van plant- en diersoorten is een van de belangrijkste doelstellingen in huidig ecologisch onderzoek. Er is een toenemende behoefte aan verspreidingsmodellen om de consequenties van klimatologische veranderingen voor de biodiversiteit te kunnen meten, om de vele vragen met betrekking tot natuurbescherming en evolutie te kunnen beantwoorden, om bio-geographische vraagstellingen te kunnen toetsen, om invasies en uitbreiding van soorten te kunnen voorspellen en om na te gaan of uitbreiding van zeldzame soorten te verwachten is. De vraag naar bovenstaande informatie heeft, samen met nieuwe ontwikkelingen in de statistiek en informatie technologie, geleid tot tal van gecompliceerde methodes om de verspreiding van een soort te modelleren, waardoor inachtneming van ecologische factoren vaak van ondergeschikt belang is geworden.

Specialisten kunnen waardevolle informatie verstrekken met betrekking tot interacties tussen soorten en hun omgeving en kunnen eventueel het inpassen van ecologische concepten verbeteren wanneer de verspreiding van een soort wordt gemodelleerd. Specialisten kunnen zekere waarden toekennen aan specifieke ecologische condities of juist complexe systemen omzeilen en zo een oplossing bieden, gericht op de kern van een bepaalde situatie.

Dit proefschrift heeft als doel bij te dragen aan een beter begrip van in hoeverre de toevoeging van kennis van specialisten bij het modelleren van de verspreiding van soorten bijdraagt aan het verbeteren van de resultaten van het model. De belangrijkste factor in met goed resultaat de kennis van specialisten gebruiken is de specialisten bij elke stap in het modellerings proces te betrekken, van hypothese vorming tot het visualizeren van de resultaten.

Een deel van mijn proefschrift is eraan gewijd te bewijzen, dat het gebruik van de kennis van specialisten bij het samenstellen van een bemonsterings schema een significant verbeterd onderscheidingsvermogen oplevert voor modellen met een resolutie 100 keer nauwkeuriger dan in de soorten atlas voorkomt. Deze benadering vormt een geschikte vervanging voor uitgebreid veldwerk, maar mogelijk alleen voor bekende soorten.

Er wordt wel aangenomen, dat soorten een min of meer gelijke voorkeur voor een bepaald habitat hebben als hun voorouders. Een volgend onderdeel van mijn proefschrift levert hier het bewijs voor en past dit vervolgens toe in modellen voor de verspreiding van soorten. Ik heb de taxonomie van soorten gebruikt om hun natuurlijke verspreiding af te bakenen en uit te zoeken wat voor effect dit heeft op de resultaten van de verspreidings modellen. Het opwaarderen van atlas gegevens door het op basis van taxonomie uitsluiten van irrelevante gegevens over de afwezigheid van een soort, levert een aanzienlijke verbetering op in de kalibratie van het model en dus ook in de capaciteit om natuurlijke potentie te voorspellen.

Kennis van specialisten is vooral belangrijk, wanneer er weinig, of zelfs geheel geen, informatie bestaat over de verspreiding van een soort. Kennis is echter altijd onderhevig aan onzekerheid en het is een uitdaging om deze onzekerheid in het modellerings proces te betrekken. Ik stel voor de Dempster-Shafer Theory of Evidence te gebruiken om kennis van specialisten, inclusief de daaraan inherente onzekerheid, toe te kunnen voegen aan de modellen. Deze benadering erkent expliciet de onzekerheid in kennis van specialisten en vormt een praktisch alternatief wanneer deze kennis met betrekking tot de verspreiding van een soort van nut is.

In dit proefschrift heb ik met praktische voorbeelden geprobeerd duidelijk te maken dat, wanneer kennis van specialisten wordt gebruikt, de resultaten van de verspreidings modellen worden verbeterd. Waar specialisten beschikbaar zijn en een geschikte methode wordt toegepast zal, onafhankelijk van het vakgebied van de specialist of het karakter van de

vraagstelling, dit onze kennis ten goede komen en helpen een oplossing te vinden voor de complexe problemen van het ruimtelijke modelleren. Ik ben ervan overtuigd dat het inpassen van kennis van specialisten bij het modelleren van de verspreiding van plant- en diersoorten zal blijven toenemen en verbeteren.

چکیده

مدلسازی دانش‌محور پراگندگی گونه‌ها

برگردان به فارسی: سام خسروی فرد

مدل‌سازی پراکندگی گونه‌ها به ابزاری مهم در پژوهش‌های بوم‌شناختی تبدیل شده‌اند. نیاز روزافزون به این مدل‌ها را می‌توان در زمینه‌های مختلف پژوهشی مشاهده کرد. بررسی اثرات تغییرات اقلیمی بر تنوع زیستی، پاسخ به پرسش‌هایی در حوزه تکامل و حفاظت گونه‌ها، سنجش فرضیه‌های مربوط به زیست‌جغرافیا، پیش‌بینی رفتار جمعیتی، بازتولید و پراکندگی گونه‌های مهاجم و پیش‌بینی زیستگاه‌های جدید گونه‌های کمیاب از آن جمله‌اند. تحقیقاتی از این دست نیازمند تجزیه و تحلیل‌های آماری پیشرفته و نیز برخورداری از فناوری‌های نوین است. چنین نیازی گاه سبب می‌شود توجهی مضاعف به فناوری و روش‌های آماری پیچیده معطوف شود به طوری که جستارهای بوم‌شناختی اولویت خود را از دست می‌دهند و در جایگاه دوم اهمیت قرار می‌گیرند. با استفاده نظرات کارشناسان می‌توان پیچیدگی‌ها و دشواری‌های مدل‌سازی پراکندگی گونه‌ها را تسهیل کرد. در واقع دانش و بینش کارشناسی پیرامون روابط متقابل گونه – محیط داده‌هایی ارزشمند به شمار می‌روند که با مفاهیم بوم‌شناختی پیوندی تنگاتنگ دارند. به کارگیری نظر کارشناسان گاه می‌تواند میان‌بری باشد برای شناخت و حتی حل مسایل پیچیده طبیعی.

پایان‌نامه پیش‌رو با این هدف به رشته تحریر در آمده که نشان دهد با استفاده از دانش اهل فن چگونه و تا چه میزان دقت، صحت و کارایی مدل‌های پراکندگی گونه‌ها افزایش می‌یابد. نکته کلیدی برای رسیدن به هدف مذکور این است که کارشناسان از لحظه شروع تا پایان پژوهش حضوری چشمگیر داشته باشند و هر قدمی که برداشته می‌شود با نظر تعیین کننده آنان عجین باشد. بخشی از این پایان‌نامه توضیح می‌دهد که استفاده از نظر کارشناسان در طرح نمونه‌برداری گونه‌ها توانسته است ظرفیت افتراق مدل‌ها را به طرز چشمگیری ارتقا دهد. به طوری که نقشه‌های پراکندگی تولید شده با استفاده از این مدل‌ها نسبت به نقشه‌های موجود تا چندین برابر دقیق‌تر هستند. رویکرد به‌کارگیری نظرات اهل فن، جایگزینی مناسب برای عملیات صحرایی زمان‌بر و پرهزینه است، اما ممکن است برای شماری از گونه‌ها کارایی لازم را نداشته باشد.

بخش دیگر این پایان‌نامه به حضور و یا عدم حضور گونه های همتبار ژنتیکی برای مدلسازی دقیقتر پراکندگی گونه‌ها پرداخته است. با استفاده از نظر کارشناسان، از فهرست آرایه‌شناسی گونه‌ها استفاده شده که مشخص شود گستره محیطی تا چه میزان بر مدل‌های پراکندگی گونه‌ها اثر دارد. همچنین برای اصلاح

واسنجی مدلهای پراکندگی بر اساس اطلس جانداران، گونه‌های غایب و غیرمرتبط بر پایه نظر کارشناسان از فهرست رده‌بندی حذف شد. با این روش دقت نقشه‌های پراکندگی به طرز قابل توجهی افزایش پیدا کرد. نظر کارشناسان زمانی بیش‌تر اهمیت و ارزش پیدا می‌کند که داده‌های اندکی از پراکندگی گونه‌ها در دسترس است. با این حال استفاده از نظر اهل فن شامل خطا و نیز دربر دارنده عدم قطعیت است که می‌توانند چالشی اساسی در مدل‌سازی ایجاد کند. پیشنهاد این پژوهش برای فائق آمدن بر چالش پیش‌رو به کارگیری تئوری دمستر – شفر است. این رویکرد در مواقعی که دانش کارشناسی در ارتباط با پراکندگی جغرافیایی گونه‌ها مورد نیاز است کاربردی موثر دارد. به علاوه با استفاده از آن می‌توان عدم قطعیت موجود در نظر کارشناسان را به روشنی اصلاح کرد.

این پایان‌نامه با مثال‌هایی کاربردی نشان می‌دهد که چگونه دانش کارشناسی سبب اصلاح مدل‌های پراکندگی گونه‌ها می‌شود؛ وقتی که کارشناسان قابل اعتماد و اهل فن در دسترس هستند و روشی مناسب به کار گرفته می‌شود نه رشته تخصصی کارشناسان و نه ویژگی‌های فضایی – زمانی فرضیه‌ها می‌تواند جلوی پیشرفت و توسعه مدل‌سازی پراکندگی گونه‌ها را بگیرد. نویسنده این پایان‌نامه ایمان دارد که استفاده از دانش و بینش کارشناسی می‌تواند گسترش یابد و اصلاح و توسعه مدل‌سازی پراگندگی گونه ها را تداوم بخشد.

Biosketch

Aidin Niamir was born on 22 May 1980 in Tehran, Iran. He attended Azad University, Iran and obtained his Bachelor of Engineering degree, majoring Natural Resources in September 2002. After graduation, he started collaborating with several organizations as a GIS specialist. In 2004 he founded the Department of Environment and Geoinformatics of the RahShahr International Group to provide geo-information, earth observation, and environmental services to the national and international project of the group. He accomplished his first Master degree in June 2005 in Environmental Science from Tehran Science and Research Campus. He was awarded a full scholarship to study a joint MSc in Geo-information Science and Earth Observation for Environmental Modelling and Management at four leading European universities (University of Southampton, Lund University, University of Warsaw, and International Institute for Geo-information Science and Earth Observation). Aidin won the best MSc thesis prize of the International Institute for Geo-information Science and Earth Observation in 2009. Immediately after his graduation, he received a scholarship from European Union Erasmus Mundus Program to pursue PhD research at University of Twente. This opportunity was the starting point of his collaborations with International Union for Conservation of Nature (IUCN), and other professional societies.

Publications

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Real, R., Munoz, A.R. (2011) "Finessing species Atlas data for species distribution models", *Diversity and Distribution,* 17, 1173-1185.

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Real, R. "Use of species taxonomy to enhance atlas data for distribution modelling". Resubmitted, after revision to: *Global Ecology and Biogeography*.

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Munoz, A.R., Real, R. "Incorporating knowledge uncertainty of experts into species distribution modelling". Resubmitted, after revision to: *Journal of Biogeography*

Niamir, A., Skidmore, A.K., Toxopeus, A.G., Real, R. "How well does absence-only-based model predict species distribution?". In review: *Ecological Modelling*

Khosravifard, S., **Niamir, A.** "The lair of the lion in Iran". In press: *CAT news*, IUCN/SSC cat specialist group

Niamir, A. et al. (2014) "Using Dempster-Shafer's evidence theory for species distribution modelling". In the proceedings of The International Statistical Ecology Conference, Montpellier, France

Niamir, A. et al. (2014) "Accounting for species taxonomy improves distribution models". In the proceedings of The International Statistical Ecology Conference, Montpellier, France

Secades, C., O'Connor, B., Brown, C. and Walpole, M. and **Niamir, A.** (contributor) (2014). Earth Observation for Biodiversity Monitoring: A review of current approaches and future opportunities for tracking progress towards the Aichi Biodiversity Targets. Secretariat of the Convention on Biological Diversity, Montréal, Canada. Technical Series No. 72, 183 pages

Niamir, A. et al. (2011) "Downscaling species atlas data using expert-system", in the proceedings of the 5th International meeting of the International Biogeography Society, Greece

Niamir, A. et al. (2011) "Enhancing coarse-resolution species atlas data using an expert system", in the proceedings of the Netherlands Ecological Annual Meeting 2011, Lunteren, the Netherlands

Niamir, A. et al. (2010) "Generating finer resolution species distribution data using an expert system", in the proceedings of the Ecological Society of Australia annual Conference, Canberra, Australia

Niamir, A., Toxopeus, B., Groen, T. (2010) Distribution of Short-toed Eagle in Relation to Potential Food Availability, Book, Lambert Academic Publishing, Germany.

Niamir, A. et al. (2009) Distribution of Short-toed Eagle in southern Spain, in the proceedings of the 7th Conference of the European Ornithologists' Union, 2009, Zurich, Switzerland

Niamir, A. et al. (2008) State of the Environment of Iran. Department of Environment, Islamic Republic of Iran (in Farsi)

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- Distribution modelling of the short-toed eagle in relation food availability
- Study of artificial neural network and maximum entropy
- Review of the use of remotely-sensed data for monitoring biodiversity change and tracking progress towards the Aichi biodiversity targets

Writing of project proposal (4.5 ECTS)

- Knowledge oriented approaches in species distribution modelling

Post-graduate courses (6 ECTS)

- AniMove; Max Planck Institute for Ornithology, Germany (2013)
- Next generation data management in movement ecology; Leibniz Institute for Zoo and Wildlife Research, Germany (2012)
- Biodiversity modelling with GLOBIO; University of Twente, the Netherlands (2011)
- Geo-ecological data analysis; University of Amsterdam, the Netherlands (2010)

Laboratory training and working visits (4.5 ECTS)

- Visiting graduate fellows (2012, 2011, 2013); University of Malaga (Spain), Foundation Migres (Spain), Guelph University (Canada)

Invited review of journal manuscript (2 ECTS)

- Global Ecology and Biogeography: species distribution modelling (2014)
- IPCC Working Group II Contribution to AR5: climate change 2014: impacts, adaptation, and vulnerability (2013)

- Biodiversity and Conservation: biodiversity modelling (2014)
- International Journal of Geographical Information Science: remote sensing and biodiversity modelling (2010, 2012, 2013)
- Ecological Modelling: species distribution modelling (2012)
- Environmental Modelling & Software: species distribution modelling, biodiversity and knowledge (2011, 2012)

Deficiency, refresh, brush-up courses (3 ECTS)

- Geostatistics and open-source statistical (2010)
- Programming skills (2010)
- Animal movement analysis (2010)
- Systems analysis and modelling (2011)

Competence strengthening / skills courses (3.3 ECTS)

- Interview skills in English; University of Twente (2013)
- What's your story?; University of Twente (2013)
- Personal branding and pitching; University (2012)
- Technical writing and editing; University of Twente (2011)
- How to write a proposal?; University of Twente (2011)
- PhD Competence assessment; Wageningen University (2010)
- Techniques for writing and presenting scientific papers; Wageningen University (2010)

PE&RC weekend, Annual meetings, and seminars (1.5 ECTS)

- PE&RC Weekend (2009)
- Biodiversity Knowledge seminars; BELSPO, Brussels (2012)

Discussion groups and local seminars/scientific meetings (7.5 ECTS)

- ITC-NRS PhD tutorials (2009-2014)

International symposia, workshops and conferences (6.2 ECTS)

- International Statistical Ecology Conference; Oral presentation (2014)
- Dutch-Belgian Information Retrieval Workshop (2011, 2012)
- Conference of the International Biogeography Society NAEM (2011)
- Ecological Society of Australia conference (2010)

Lecturing / supervision of practical's / tutorials (0.6 ECTS)

- Spatial modelling of biological ecosystem properties (2010, 2011)

ITC Dissertation List

http://www.itc.nl/research/phd/phd_graduates.aspx