Patterns of species occurrence as a response to environmental and biotic constraints: the uncertainty of correlative approaches

Padrões de ocorrência de espécies como resposta a constrangimentos ambientais e bióticos: a incerteza de abordagens correlativas

Pedro Segurado

Dissertação de tese apresentada à Universidade de Évora, Área Departamental de Ciências da Natureza e Ambiente, para obtenção do grau de Doutor em Biologia.

Orientador

Prof. Doutor Diogo Caeiro Figueiredo
Departamento de Biologia, Universidade de Évora

Co-orientadores

Doutor Miguel Araújo
Científico titular, Museo Nacional de Ciencias Naturales

Doutor William Edward Kunin
Faculty of Biological Sciences, University of Leeds

171843

Esta tese não inclui as criticas e sugestões feitas pelo juri

Évora 2007

À Susana e ao André

Table of Contents

Agradecimentos i
Abstract iii
Resumo iv
Resumo alargado v
Declaration viii

General introduction 1
1.1 Correlative models in ecology 2
1.2 Species patterns of occurrence as a consequence of environmental requirements: species distribution models 4
1.3 Species' patterns of occurrence as the consequence of biotic interactions: analysis of species co-occurrence 7
1.4 Issues of model uncertainty 9
1.5 The case study 12

The combined effect of modeling technique and species macroecological traits on the accuracy of species distribution models 15
2.1 Summary 16
2.2 Introduction 17
2.3 Materials and methods 20
2.4 Results 29
2.5 Discusion 37

The effect of autocorrelation_on species distribution models 42
3.1 Summary 43
3.2 Introduction 44
3.3 Materials and methods 48
3.4 Results 55
3.5 Discussion 63

The effect of spatial scale on patterns of species co-occurrence 68
4.2 Summary 69
4.2 Introduction 69
4.3 Materials and methods 73
4.4 Results 80
4.5 Discussion 88

Dealing with species rarity when analyzing patterns of species co-occurrence 95
5.1 Working example 1: coexistence of two freshwater turtles along a stream 96
    5.1.1 Summary 96
    5.1.2 Introduction 97
    5.1.3 Materials and methods 101
    5.1.4 Results 106
    5.1.5 Discussion 114
5.2 Working example 2: Coexistence of two freshwater turtles at the habitat and biogeographical scales 121
    5.2.1 Summary 121
    5.2.2 Introduction 122
    5.2.3 Materials and methods 123
    5.2.4 Results 128
    5.2.5 Discussion 136

Final discussion 140
6.1 Is the "best" distribution modelling approach contingent to species macroecological traits? 141
6.2 Is the effect of spatial autocorrelation contingent to the goals and the kind of modelling approaches? 143
6.3 How to deal with spatial autocorrelation? 144
6.4 Is the potential contribution of biotic interactions on species distributions contingent to spatial scale of analysis? 145
6.5 How to deal with small sample sizes? 146
6.6 Concluding remarks 148

Literature cited 149

Agradecimentos

Muitas pessoas contribuíram das mais variadas formas para a concretização deste trabalho. Começo por agradecer ao Miguel Araújo por se ter disponibilizado prontamente como orientador efectivo desta tese e pelo incentivo, discussões frutuosas e sugestões sempre oportunas, apesar da distância física que nos separou durante grande parte da duração da tese. Ao Diogo Figueiredo agradeço o seu papel como orientador oficial da tese e pelas facilidades concedidas e disponibilidade que sempre demonstrou durante estes anos. Também apesar da distância, gostaria ainda de agradecer ao William (Bill) Kunin o seu importante contributo como co-orientador, a sua infinita amabilidade e por me ter proporcionado agradáveis e muito proveitosas estadias na Universidade de Leeds.

Esta tese foi financiada por uma bolsa concedida pela Fundação para a Ciência e Tecnologia (SFRH/BD/8493/2002).

A todos aqueles que de uma maneira ou de outra contribuíram com os seus neurónios. Gostaria de agradecer aos meus colegas de Leeds, em especial ao grupo do Chris Thomas – e principalmente às ibéricas Aldina Franco, Rosa Menéndez e Adela González-Megías - pelos conselhos e questões pertinentes e em especial por me terem integrado em algumas discussões do grupo. Ao Stephen Roxburgh, Ana Bio e Wilfried Thuiller agradeço os comentários e sugestões ao terceiro capítulo desta tese. Agradeço também a todos os referees anónimos que de uma maneira ou de outra contribuíram para o resultado final desta tese. Ao Rui Rebelo agradeço o seu papel de conselheiro científico independente e por me ter revisto o texto da introdução e discussão.

Gostaria de agradecer a todos os investigadores que contribuíram com trabalho de campo para os atlas utilizados nesta dissertação ou outra informação relevante, em especial à Paula Rito, Lélia, Rita Alcazar, Paulo Cardoso e ao pessoal do CIBIO do Porto.

Não poderia deixar de agradecer aos colegas de laboratório, gabinete ou alojamento por me terem proporcionado condições ambientais propícias ao desenrolar desta tese. Começo por agradecer alfabeticamente aos colegas de trabalho de Évora, António, Manuela, Miguel, Né, Paulo, Pedro, Rui, pela ajuda e amizade. Ao pessoal de Leeds que ainda não mencionei - Ahmed, Mario e Pedro - por terem tornado mais agradável a minha estadia em Inglaterra. Faço uma viagem para o presente e agradeço ao pessoal do laboratório 2.13 do DEF - Luís Lopes, Luís Reino, Paulo Branco e Rosário Fernandes - pelo bom ambiente de trabalho na última fase da tese. Agradeço ainda à Teresa Ferreira por me ter dispensado algumas horas do projecto para isto...

O trabalho de campo não teria sido possível sem o apoio logístico da Liga para a Protecção da Natureza e da minha irmã. Agradeço em particular ao pessoal da Herdade do Vale Gonçalinho pela companhia e facilidades que me proporcionaram durante o trabalho de campo em Castro Verde. À minha irmã agradeço as estadias na casa da Longueira. Agradeço ao Paulo Cabrita do Parque Natural da Costa Sudoeste pelas informações e disponibilidade. Ao Sr. António Idálio Santos Silva e Sr. José Francisco Santos Silva um especial agradecimento por terem permitido realizar amostragens nas suas propriedades. Agradeço ainda à Paula Rito e à Lélia pela cedência e empréstimo de armadilhas para a captura de tartarugas.

Um agradecimento a todos os que me ajudaram no campo: aos amigos Rui Rebelo, Paulo Catry (e respectivas alunas de mestrado), Inês e Paulo. Espero não me ter esquecido de ninguém.

Aos meus pais, claro, pelo apoio (alguns clichés são mesmo verdades).

E por fim agradeço à minha nova família, sem a qual nada disto faria grande sentido, e à qual dedico esta tese (apesar de merecerem melhor): à Susana pelo apoio, ajuda a todos os níveis e principalmente por me ter aturado durante estes últimos anos, e ao André pelos poucos minutos diários que teve a amabilidade de me emprestar e que tão preciosos foram para que esta tese fosse levada a bom termo.

Abstract

Correlative modelling approaches have been used as first approximations to infer the role of environmental and biotic factors underlying species occurrences. However correlative models cannot demonstrate unequivocally causal relationships. Furthermore a number of artefacts may cause some inferred relationships to be spurious. In this thesis I focus my work on these sources of uncertainty. First I demonstrate the joint effect of the kind of modelling technique and species macroecological traits in the accuracy of models of species distribution. Then, by means of simulated spatial patterns I assess the overestimation of species distribution models' accuracy due to spatial autocorrelation. Third, I address the issue of the effect of spatial scale on patterns of species co-occurrence showing that evidences of spatial segregation at finer scales can be inversed at wider scales. In the last part I propose a simple resampling technique to overcome problems derived from sampling artefacts due to species rarity.

Resumo

Como primeira aproximação para inferir o papel de factores ambientais e bióticos na ocorrência de espécies, tem-se recorrido frequentemente a abordagens correlativas. Porém, os modelos correlativos não demonstram inequivocamente relações causais. Diversos artefactos estatísticos podem ainda tornar espúrias algumas relações inferidas. Nesta tese foco o meu trabalho nestas fontes de incerteza. Começo por demonstrar um efeito combinado da técnica estatística e de características macroecológicas das espécies na precisão de modelos de distribuição. Através de padrões espaciais simulados avalio depois a sobrestimação da precisão de modelos de distribuição de espécies devido à autocorrelação espacial. Abordo posteriormente a questão do efeito da escala espacial nos padrões de co-ocorrência de espécies em que demonstro que evidências de segregação a escalas finas podem ser invertidas a escalas grosseiras. Por último, proponho uma técnica simples de reamostragem como forma de contornar artefactos de amostragem de espécies raras.

Resumo alargado

Em parte porque os edifícios teóricos em ecologia estão longe de estar consolidados e devido à necessidade de tomar medidas eficazes a curto prazo para a preservação da diversidade biológica, muitos investigadores têm-se focado no desenvolvimento de metodologias que permitem identificar os factores que, mesmo que indirectamente, influenciem a ocorrência e frequência de espécies e, em última análise, a biodiversidade. Três mecanismos principais são geralmente apontados como sendo responsáveis pelos padrões de ocorrência não aleatórios das espécies: (1) a distribuição de condições ambientais propícias, (2) a interacção biótica entre espécies e (3) factores históricos. Esta dissertação debruça-se sobre a utilização de abordagens de carácter correlativo para estudar os dois primeiros factores referidos, nomeadamente modelos de distribuição de espécies e de co-ocorrência de espécies.

Os modelos de distribuição de espécies constituem actualmente técnicas familiares para um grande número de investigadores, com inúmeras aplicações em ecologia fundamental e aplicada. Estes modelos assentam essencialmente na procura de relações entre a distribuição ou abundância de espécies e variáveis ambientais. Estão de certa forma relacionados com a teoria do nicho ecológico uma vez que assumem que os padrões espaciais da ocorrência e abundância das espécies resultam do efeito multivariado de factores ambientais e bióticos. Para estudar os padrões de co-ocorrência de espécies têm prevalecido abordagens baseadas em modelos nulos em que se testa a independência da relação entre padrões com base na geração de padrões ecológicos aleatórios que são de esperar ocorrer na ausência de determinado mecanismo, ou em que se mantém fixo algum artefacto estatístico cujo efeito se pretende eliminar.

Ambos os modelos de distribuição e de co-ocorrência de espécies permitem fornecer primeiras pistas sobre eventuais mecanismos mas não permitirem demonstrar inequivocamente relações de causalidade, sendo por isso frequentemente entendidos como métodos geradores de hipóteses. Por outro lado existem inúmeras fontes de incerteza a eles associados que devem ser também devidamente acautelados pelos ecólogos, para que os seus modelos não só possuam robustez do ponto de vista teórico, como apresentem alguma utilidade do ponto de vista da sua aplicação. O presente trabalho aborda o tema da incerteza nos modelos correlativos, em particular dos modelos de distribuição e dos modelos de co-ocorrência de espécies.

Em primeiro lugar é abordado o efeito conjunto da selecção da técnica estatística e das características macroecológicas das espécies sobre a precisão dos modelos de distribuição de espécies. Com base em 7 técnicas estatísticas e 44 espécies de anfíbios e répteis demonstra-se que diferentes técnicas são mais ou menos adequadas consoante determinadas características macroecológicas das espécies.

No capítulo seguinte é pela primeira vez avaliado o efeito da autocorrelação espacial em diferentes abordagens de modelação de distribuição de espécies. Para tal recorreu-se a modelos nulos para gerar padrões geográficos em que a autocorrelação era mantida fixa por forma a permitir retirar conclusões sobre o efeito deste artefacto estatístico. Como caso de estudo utilizou-se a distribuição ibérica de duas tartarugas-de-água-doce ibéricas, o cágado-mediterrânico *Mauremys leprosa* e o cágado-de-carapaça-estriada *Emys orbicularis*. Em todas as situações experimentais verificou-se que a autocorrelação provoca uma forte sobre-estimação da precisão dos modelos. Este efeito verifica-se independentemente do tipo de técnica estatística (paramétrica, semi-paramétrica e não paramétrica) e do objectivo da modelação (explicativo ou preditivo). No final do capítulo discutem-se formas de atenuar os efeitos da autocorrelação espacial.

No terceiro capítulo experimental é avaliado o efeito da escala espacial sobre os padrões de co-ocorrência de espécies. Para tal é analisada a co-ocorrência das mesmas duas espécies de tartarugas de água doce ibéricas considerando diferentes resoluções e extensões de território: desde a escala de um troço de ribeira até à escala da Península ibérica. A co-ocorrência é analisada, quer considerando a semelhança entre espécies na resposta a variáveis ambientais, quer a associação puramente espacial. Verifica-se que as evidências de segregação espacial entre as espécies a escalas finas não se mantêm a escalas mais grosseiras, podendo mesmo inverter-se. Este estudo sugere que eventuais interacções bióticas podem não ser necessariamente detectadas a escalas mais grosseiras e que para explorar de forma completa os padrões de co-ocorrência de espécies são necessárias abordagens a escalas múltiplas.

A raridade das espécies é a última fonte de incerteza analisada neste trabalho através de dois casos de estudo relacionados. Nestes estudos utilizaram-se também dados relativos às duas espécies de tartarugas de água doce ibéricas que diferem grandemente na frequência de ocorrência: *E. orbicularis* é uma espécie muito pouco abundante com uma ocorrência muito dispersa enquanto que *M. leprosa* é uma espécie muito abundante com uma ocorrência muito contínua ao longo do território. Em ambos os estudos propõe-se uma técnica de reamostragem como método para evitar artefactos de amostragem em estudos de co-ocorrência entre duas espécies relacionadas mas com frequências de ocorrência muito dispares. Este método foi utilizado com sucesso para testar diferenças no uso do espaço e na resposta a variáveis ambientais entre as duas espécies a diferentes escalas espaciais, bem como para comparar o grau de marginalidade das distribuição de *E. orbicularis* e *M. leprosa* ao longo de gradientes ambientais univariados.

As fontes de incerteza abordadas nesta tese estão de uma forma ou de outra relacionadas entre si. Por exemplo, o efeito dos atributos macroecológicos das espécies na precisão dos modelos de distribuição está muito provavelmente relacionada com o efeito da autocorrelação espacial. Um importante desafio será desenvolver estudos mais aprofundados sobre a inter-relação entre diferentes factores de incerteza de modelos de ocorrência de espécies com vista à produção de modelos mais realistas e com maior utilidade prática.

Declaration

I declare that most of the work presented in this thesis is my own. I wrote most of the text, except for chapter 2 where Miguel Araújo, as the co-author, also contributed significantly to the writing. All the statistical analyses were carried out by myself. Chapter 2 was published in *Journal of Biogeography* under the title "an evaluation of methods for modelling species distributions". (Segurado, P. & Araújo, M.B., 2004. *Journal of Biogeography*, 31: 1555–1568). Chapter 3 was published in *Journal of Applied Ecology* under the title "Consequences of spatial autocorrelation for niche-based models" (Segurado, P, Araújo, M.B. & Kunin, W.E., 2006. *Journal of Applied Ecology*, 43: 433–444). The first case study of chapter 5 was partially presented in the IV International Symposium on *Emys orbicularis* held in Valencia, Spain (8-10th June 2006) and has been published in *Acta Oecologica* under the title "Coexistence of two freshwater turtle species along a Mediterranean stream: The role of spatial and temporal heterogeneity" (Segurado, P. & Figueiredo, D., 2007. *Acta Oecologica*, 32: 134-144). The second case study included in Chapter 5 was presented during the III International Symposium on *Emys orbicularis* held in Slovak Republic (18–20th April 2002) and was published in *Biologia* under the title "Coexistence of *Emys orbicularis* and *Mauremys leprosa* in Portugal at two spatial scales: is there evidence of spatial segregation?" (Segurado, P. & Araújo, A.P.R., 2004. *Biologia*, 59 (Suppl. 14): 61-72).

During my PhD, apart from the papers presented in the current thesis I have also submitted a paper to *Revista Española de Herpetologia* under the title "Skewed age structures of *Emys orbicularis* populations at habitats shared by *Mauremys leprosa*: evidence of ecological interaction between species?" (Segurado, P. & Araújo, AP.R., *Revista Española de Herpetologia*, In press.). This paper was based on a work presented in the IV International Symposium on *Emys orbicularis* held in Valencia, Spain (8-10th June 2006). I have also participated in the Bioclimatic Modelling Workshop held in CEFE, CNRS, Montpelier, France (14-16th November) that resulted in a paper published in *Journal of Biogeography* under the title "Model-based uncertainty in species range prediction" (Pearson R.G., Thuiller W., Araújo M.B., Martinez-Meyer E., Midgley G.F., Brotons L., McClean C., Miles L.J., Segurado P., Dawson T.E. & Lees D.C., 2006. *Journal of Biogeography* 33: 1704-1711).

Chapter 1

General introduction

"The science of ecology is in its infancy."

B.A. Maurer (1999)

1.1 Correlative models in ecology

The mechanisms that govern patterns of species occurrence and abundance have been a central issue in ecology. In the past this issue was typically addressed at local or regional scales in view of processes acting either at the level of single species (e.g. demography and resource selection) or species assemblages (e.g. biotic interactions and resource partition). In the last decade, the broader and more integrative perspective of macroecology has emerged, benefiting from advances on analytical tools that were made possible by new technologies to compile and handle spatial information, such as remote sensing and geographical information systems (Maurer 1994, 1999, Brown 1995). This approach is broader in the sense that it focuses the analyses at the scale of the entire geographical range of species, and it is more integrative in the sense that aims to relate patterns and processes acting at different spatial scales (Maurer 1999).

The issue of species occurrence has been largely addressed in light of the classical theory of niche differentiation between species that, in essence, assumes that each species has its own fitness that depends on environmental and biotic constraints (Chase & Leibold 2003). Recently, a whole new theoretical framework has emerged - the unified neutral theory - which opposes the classical niche theory by assuming an ecological equivalence between species and suggesting that most community properties such as the abundance structure are the result of stochastic events of extinction and speciation (Hubbel 2001). According to its proponents the niche perspective paradigm needs to be completely rethought. Despite the large controversy among ecologists (Chase & Leibold 2003, Holyoak *et al.* 2006), the two theories are not necessarily mutually exclusive (Holt 2006, Leibold & Mcpeek 2006). In fact it is possible that the niche theory and the neutral theory are simply describing two extreme situations along a broader continuum of different community types (Gravel *et al.* 2006).

In part because theoretical frameworks in ecology are yet to be consolidated and due to the need for undertaking efficient short term actions to reduce the extinction pace (Lawton & May 1995), many ecologists have focused their research on the development of methodologies that aim to identify the factors that, even if acting indirectly, have an influence on species occurrence and frequency and, ultimately, on biodiversity itself. Such studies are essentially based on correlative approaches since their aim is to seek for relationships between phenomena and do not necessarily assume a causal relationship. Correlative models can in fact provide first clues about the underlying mechanisms but cannot demonstrate unequivocally causal relationships as mechanistic models do (Olden & Jackson 2002, Diniz-Filho *et al.* 2003, Gaston 2003, Pearson & Dawson 2003, Peek *et al.* 2003, Kearney & Porter 2004, Robertson *et al.* 2004, Kearney 2006) and, as such, are frequently referred to as hypothesis generator methods (Manel *et al.* 2000). The use of such models in ecology has proven to be of great convenience when manipulative experiments are very difficult or impossible to be undertaken, particularly when phenomena at the geographic scale (macroscale *sensu* Maurer 1999) are addressed. Although these models do not strictly rely on any ecological theory, there has been a recent concern to reinforce their link to niche theory (Soberón & Peterson 2005, Araújo & Guisan 2006, Kearney 2006).

Irrespective of which is the correct ecological theory, the fact is that species show non-random patterns of occurrence and abundance. In broad terms there are four main mechanisms that may contribute to the non-randomness of species occurrence at different spatial scales: (1) the distribution of appropriate environmental conditions, (2) the biotic interactions among species (3) historical or phylogenetic factors (Peres-Neto 2004, Sfenthourakis *et al.* 2005) and (4) stochastic processes (as predicted by the unified neutral theory). This dissertation will focus on the use of correlative approaches to address the first two families of mechanisms.

1.2 Species patterns of occurrence as a consequence of environmental requirements: species distribution models

With the advent of Geographical Information Systems (Haslett 1990) in the eighties and after the pioneer works of Walker (1990) and Pereira e Itami (1991), the correlative models that assumes species occurrence as the result of environmental requirements have each year been increasingly used, tested and enhanced (Fig. 1.1). They are now familiar tools for many fundamental and applied researchers (for a review see e.g. Guisan & Zimmermann 2000, Rushton *et al.* 2004, Guisan & Thuiller 2005, Araújo & Guisan 2006, Guisan *et al.* 2006, Austin 2007).

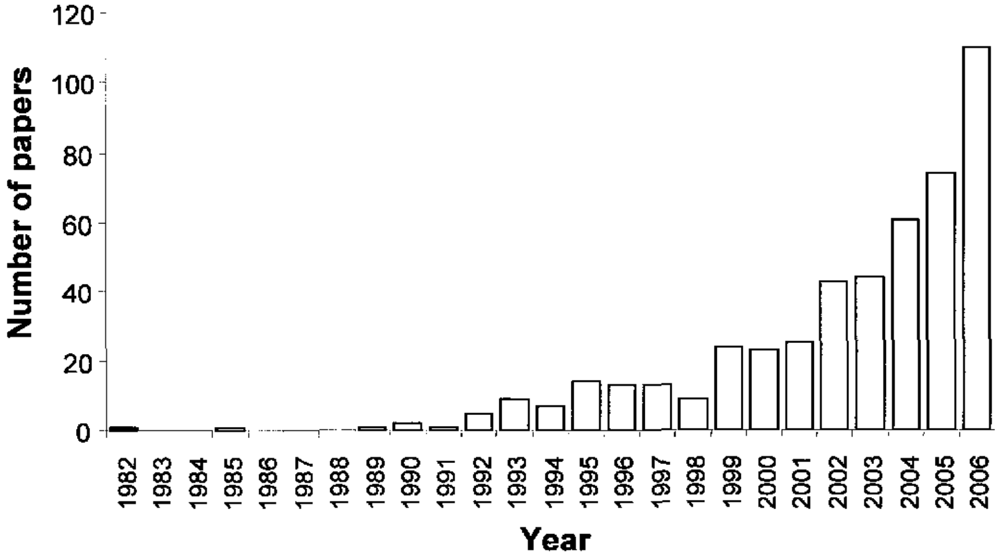


Fig. 1.1 – Number of papers included in the Science Citation Index Expanded list (1945 – present) containing either the terms "species distribution models", "predictive distribution models", "predictive models of distribution", "predictive models of species distribution", "niche-based models", "niche modelling", "bioclimatic envelope", "bioclimatic modelling", "cimatic envelope", "environmental envelope", "habitat models" or "habitat modelling" (source: ISI web of knowledge).

Different designations for this family of models can be found in the literature, which often cause some confusion among ecologists. Some of the most common designations include *species distribution models* (SDM; Guisan &

Thuiller 2005), *habitat models* (Guisan & Zimmermann 2000), *ecological niche models* (Peterson *et al.* 1999), *niche-based models* (Feria & Peterson 2002, Thuiller *et al.* 2006), *bioclimatic envelope* (Walker 1990, Pearson & Dawson 2003), and more recently, *resource selection functions* (RSF; Meyer & Thuiller 2006). Although some designations are more general than others, all refer to a common family of models that relies on a given statistical approach and assumption, to seek for relationships between the species patterns of distribution or abundance and the environment. *Ecological niche model* or *niche-based model* designations are somehow more comprehensive and have the advantage of explicitly relating this kind of models to the theoretical framework of niche theory. Indeed, these models are closely related to the niche theory since they assume that the spatial patterns of species occurrence and abundance are the result of the multivariate effect of environmental and biotic factors (Soberón & Peterson 2005, Araújo & Guisan 2006).

Although there have been conflicting views of what the output of niche-based models truly represent - the realised niche (e.g. Pearson & Dawson 2003) or the fundamental niche (e.g. Soberón & Peterson 2005) – it has been at a great measure a false question since both views are valid (see Kearney & Porter 2004 and Araújo & Guisan 2006 for a discussion). On one hand niche-based models are calibrated with presence and absence samples that are already influenced by biotic factors and therefore model output is closer to the realised niche. On the other hand these models can predict species presence well beyond species' distribution ranges and therefore can be viewed as representing the fundamental niche. In fact, it can be assumed that, at most, they provide a spatial representation of the 'potential' realised niche, i.e., areas where species are more likely to occur, given the biotic constraints at sites where they were sampled (that may be different from non-sampled sites).

Species distribution models have been essentially used on the production of predictive models of individual species, including plants (Franklin 1998, Thuiller *et al.* 2003), fishes (Mastrorillo *et al.* 1997, Leathwick *et al.* 2005), invertebrates (Berg *et al.* 2004), amphibians (Teixeira *et al.* 2001), reptiles (Brito *et al.* 1996, 1999, Guisan & Hofer 2003), birds (Osborne *et al.* 2001, Muñoz *et al.* 2005) and mammals (Fischer *et al.* 2001, Illoldi-Rangel *et al.* 2004). These models have also contributed to important advances on several issues such as: area selection for conservation (Williams & Araújo 2000, Williams & Araújo 2002, Cabeza *et al.* 2004, Moilanen 2005, Rondinini *et al.* 2006), invasion ecology (Silva *et al.* 2002, Peterson *et al.* 2003, Arriaga *et al.* 2004, Mau-Crimmins *et al.* 2006), identification of environmental drivers of species richness (Cumming 2000, Jetz & Rahbek 2002, Lehmann *et al.* 2002, Luoto *et al.* 2004, Diniz-Filho & Bini 2005, Nogués-Bravo & Araújo 2006), inferences of biotic interactions (Anderson *et al.* 2002), temporal dynamics of spatial patterns (Augustin *et al.* 2001) and enhancement of null models in community ecology (Peres-Neto *et al.* 2001). In particular, in recent years, these models have been widely used to predict future shifts of species distributions under different scenarios of climatic changes (Buckland *et al.* 1996, Guisan & Theurillat 2000, Peterson *et al.* 2002, Thomas *et al.* 2004, Araújo *et al.* 2005b, Bomhard *et al.* 2005, Araújo *et al.* 2006, for a revision see Walther *et al.* 2002, Araújo & Rahbek 2006).

However, a great proportion of the literature focused on species distribution models have been essentially methodological, in which the relative performance of alternative statistical techniques and model building procedures were assessed (Walker & Cocks 1991, Mastrorillo *et al.* 1997, Franklin 1998, Brito *et al.* 1999, Manel *et al.* 1999, Ozesmi & Ozesmi 1999, Elith 2000, Hirzel *et al.* 2001, Olden & Jackson 2002, Stockwell & Peterson 2002, Brotons *et al.* 2004, Elith *et al.* 2006, Meyer & Thuiller 2006, Pearson *et al.* 2006) or methods of model validation were enhanced (Buckland & Elston 1993, Fielding and Bell 1997, Cumming 2000, Pearce & Ferrier 2000, Vaughan & Ormerod 2005, Hirzel *et al.* 2006). As a whole these works contributed very significantly for the improvement of modelling techniques and experimental designs in the last decade.

In general, the main objective of these works has been primarly to optimise the models' usefulness for concrete questions. Models have been mainly assessed according to their performance, and therefore the "function" has prevailed relatively to the "form". It seems that ecologists have been taking too literally the famous citation "all models are wrong, but some models are useful" (Box 1976).

Recently this concern has been replaced by an increasing demand of building models that express natural phenomena more realistically (Guisan & Thuiller 2005, Araújo & Guisan 2006, Guisan *et al.* 2006, Austin 2007). In particular, it has been advocated that more process-based understanding of how species distributions form is crucial for the urgent need of predicting species responses to climate change (Buckley & Roughgarden 2006). For example, modelling procedures assuming a more realistic hierarchical organization of environmental effects have been proposed (Pearson *et al.* 2004, Heikkinen *et al.* 2005, Anadón *et al.* 2006), although they suffer from the same inferential limitations. There has been also a growing concern of niche modellers to incorporate biotic interaction into their models (e.g. Leathwick and Austin 2001, Gutiérrez *et al.* 2005). Recent advances on the field of metabolic ecology and the development of approaches based on individual responses are promising for the development of mechanistic models of species occurrence and abundance (Kearney & Porter 2004, Kearney 2006, Buckley & Roughgarden 2006). In spite of these concerns a good agreement of predictions between correlative and mechanistic predictive modelling techniques has been documented (Robertson *et al.* 2003).

1.3 Species' patterns of occurrence as the consequence of biotic interactions: analysis of species co-occurrence

In most situations species do not occur independently from each other. As mentioned earlier, according to the deterministic view of ecology (as opposed to the neutral view), this non-random pattern is essentially the consequence of historical factors, species' environmental requirements and/or biotic interactions. Therefore, patterns of species co-occurrence have been the topic of much ecological research as a first approach to infer the mechanisms underlying species distributions and community structure.

For the last thirty years many attempts have been made to develop methodologies to unequivocally test if two or more species are spatially associated, that is, if their co-occurrence is more or less frequent than random expectations. Simple statistical tests, such as simple correlation and analysis by contingency tables, have proven to be inadequate because they do not account for many sources of statistical artefacts that are intrinsic to species patterns of occurrence, such as non-random spatial distributions (Roxburgh & Chesson 1998) or the effects of geometric constraints (Colwell & Lees 2000). On the other hand there are many stochastic factors that can lead to the non-randomness of species co-occurrence (Gotelli & Graves 1996). Furthermore these simple approaches say little about the processes underlying the observed patterns.

In the late seventies, Connor & Simberloff (1979) proposed a method that compared presence-absence matrices with those generated by Monte Carlo randomisations of real data that corresponded to expected matrices in the absence of biotic interactions. Such approaches are known as null models (Gotelli & Graves 1996). Since then, although being the issue of much debate among community ecologists (Weiher & Keddy 1995, Sanderson *et al.* 1998, Gotelli 2000), this approach has been widely used. Null models are pattern-generating models in which certain elements of the data are held constant while others are allowed to vary stochastically (Gotelli & Graves 1996). The production of a large number of new patterns that would be expected in the absence of the ecological mechanism of interest allows estimating the probability of non-randomness of the observed pattern in respect to that mechanism. The validity of this approach to infer processes at the community level was greatly contested, partially because some of the results questioned the role of competition in community structure, which contradicted the ecological paradigm of niche and competition theory that dominated ecology during the seventies (Chase & Leibold 2003).

Null models have been typically used with presence/absence matrices of entire species assemblages at discrete and relatively spatially independent locations (e.g. islands). Separate procedures have been developed for occurrence records based on transects or grids, in which spatially explicit null models are required to correct for spatial dependency among observations (Palmer & Van der Maarel 1995, Roxburgh & Matsuki 1999, Roxburgh & Chesson 1998).

Similarly to models of species distribution, many ecologists are sceptical about the capabilities of null models when analysing species co-occurrence, to unequivocally infer the mechanisms underlying the observed patterns (Roxburgh & Chesson 1998, but see Rejmanek & Leps 1996, Sfenthourakis *et al.* 2006). Furthermore, there is still no consensus about the assumptions that lead to the right "null space" (Manly 1995, Sanderson *et al.* 1998, Gotelli 2000, Manly & Sanderson 2002). This is a crucial issue because the significance of the observed patterns greatly depends on the assumptions underlying null models. For example, Roxburgh and Matsuki (1999) compared different techniques to produce fixed spatially structured null models and showed that the adoption of different approaches would greatly affect the p-values and even the sign of spatial association. Nevertheless, several frameworks, either based on species distribution modelling or null model approaches, have been proposed to infer mechanisms from the analysis of spatial patterns alone (Anderson *et al.* 2002, Apps *et al.* 2006, Sfenthourakis *et al.* 2006, Cadena & Loiselle 2007).

1.4 Issues of model uncertainty

Besides the difficulty to unequivocally demonstrate causal relationships, models that relate patterns of occurrence to environmental or biotic factors have many sources of uncertainty due to statistical limitations and artefacts that ecologists should be aware of in order to confer robustness and an applied value to their models. This dissertation addresses the issue of model uncertainty, focusing in particular those models that deal with the response of patterns of species occurrence to environmental and biotic factors. Four sources of uncertainty are addressed in the following chapters: (1) choice of the modelling technique, (2) spatial autocorrelation of ecological patterns, (3) spatial scale of analysis and (4) small sample sizes or rarity. The main objective of this work is to provide some guidelines on how to handle each of these problems.

Chapter 2 starts by asking whether there is an interaction between general macroecological traits of species and the accuracy of different modelling techniques. The choice of the technique is a decisive step in the model building process because different statistical approaches, with distinct assumptions on the kind of species' response to gradients, may lead to very different conclusions about relationships between variables. On the other hand, the species' ecological traits may influence the performance and adequacy of different modelling techniques. For example, the way species respond to an environmental gradient is expected to affect the detection and the strength of species-environmental relationships by a given statistical method. This chapter also intends to serve as a rough guide to choose the best modelling procedure for species with given geographical and environmental kinds of distribution.

The spatial dependency or autocorrelation of observations and/or of environmental variables is another important source of model uncertainty that has been well documented (Legendre 1993, Koenig & Knops 1998, Lennon 2000, Dale & Fortin 2002, Fortin & Payette 2002, Legendre *et al.* 2002). One main problem of inferring species-environment relationships is that the relative effect of space and environment on species distribution is in most cases entangled due to spatial autocorrelation of both species occurrences and environmental variables. Another main problem with this property is that it leads to an inflation of the significance of measured spatial relationships, due to an over-estimation of the degrees of freedom (Lennon 2000). In the context of species distribution models, the extent to which spatial autocorrelation has an influence on model performance, whether its effect is conditional to the goal of the modelling procedure (explanatory versus predictive) and whether it affects in the same way the output of different modelling techniques, has never been assessed. This provides the topic for the third chapter.

A third important source of uncertainty in correlative models is the spatial scale of analysis. At each spatial scale different mechanisms may be responsible for the observed patterns and therefore different spatial relationships may be observed at different scales (Levin 1992, May 1994). This issue has been frequently addressed in the context of species distribution models (Mackey & Lindenmayer 2001, Thuiller *et al.* 2003, Pearson *et al.* 2004, Guisan & Thuiller 2005), although much research is still needed to fully understand the relationships between patterns and processes at different spatial scales. For example, robust quantitative testing on the generally accepted idea that the effect of local ecological processes, such as biotic interactions, is weaker on species patterns of distribution at broader geographical extents and coarser resolutions is still lacking (Brändle & Brandl 2001, Pearson & Dawson 2003, Guisan & Thuiller 2005). The fourth chapter of this dissertation addresses this issue using occurrence data of two related species compiled at different spatial scales.

Species rarity is the fourth and last issue, which is addressed using two related working examples. Because rarity is a common feature of most endangered and protected species (Rodrigues & Gaston 2002), the challenge of sampling and estimating ecological traits of rare species has been the topic of recent research (Engler *et al.* 2004, Cunningham & Lindenmayer 2005, Mackenzie *et al.* 2005, Guisan *et al.* 2006, Pearson *et al.* 2007). The main difficulty of studying rare species is to get a sufficient amount of data in order to get an unbiased sampling set and enough degrees of freedom for conclusive test statistics to be obtained. Indeed, it is inevitable that rare species will be recorded at much fewer sites than those where they actually occur, and therefore measures of ecological traits such as niche breadth and niche position will most certainly be biased (Gaston 1994). In Chapter 5 an approach to confront the pattern of occurrence of a rare and a related common species and its response to environmental factors is proposed using two related working examples. Although the main focus of these studies is not directly related to the issue of rarity, I opted to include them in a separate chapter since they use common approaches to deal with small sample sizes. In fact, the main subjects of these working examples are closely related with the work presented in chapter 4 since common datasets are used and the issue of species coexistence is also addressed.

1.5 The case study

A national dataset comprising the whole set of amphibian and reptile species was used in the second chapter. In the following chapters, analyses were based on the patterns of distribution and abundance at several spatial scales, from both compiled and personal databases, of two native freshwater turtles co-occurring in the Iberian Peninsula: the Mediterranean pond turtle, *Mauremys leprosa* (Schweiger 1812), and the European pond turtle, *Emys orbicularis* (L. 1758).

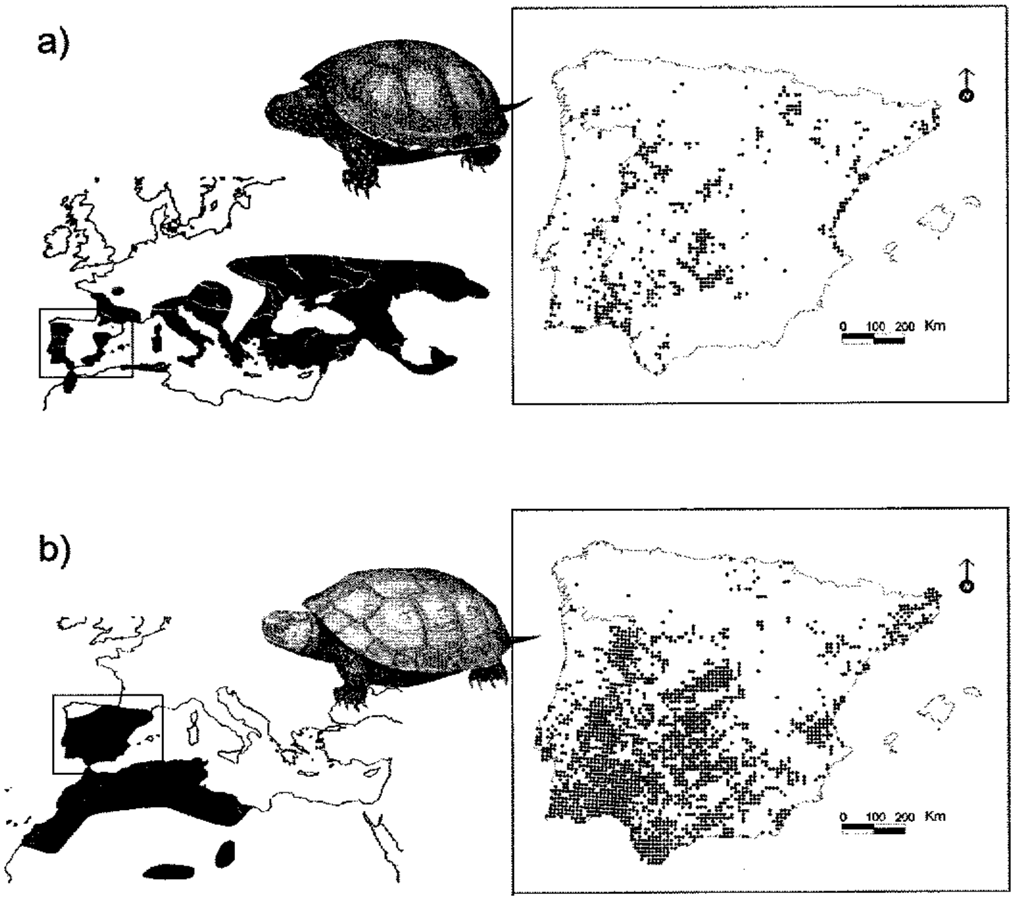


Fig. 1.2 – a) Geographic range (left) and distribution in the Iberian Peninsula (right) of *E. orbicularis* (a) and *Mauremys leprosa* (b). Adapted from Bonin *et al.* 1998 (left maps) and Pleguezuelos *et al.* 2002 (right maps).

E. orbicularis is one of the world's most widespread freshwater turtle. Its geographic range includes Northwest Africa, Iberian Peninsula, most of Europe south of Scandinavia, Asia Minor and eastwards to the Caspian and Aral Sea (Ernst & Barbour 1989; Fritz 1998; Fig. 1.2a). The distribution area of *M. leprosa* is more restricted, ranging from Southwest Europe, including Portugal, Spain and Southwest France, and most of the Maghreb region, including Morocco, Algeria, Tunisia, Western Libya, Niger, Southern Mauritania and Mali (Ernst & Barbour, 1989; Fig. 1.2b). The geographic range of the two species overlaps in the Iberian Peninsula, where the species widely coexist. Both species are restricted to aquatic habitats, including streams, ponds, swamps, marshes, lakes and drainage canals, although they often migrate and disperse throught land (Ernst & Barbour, 1989).

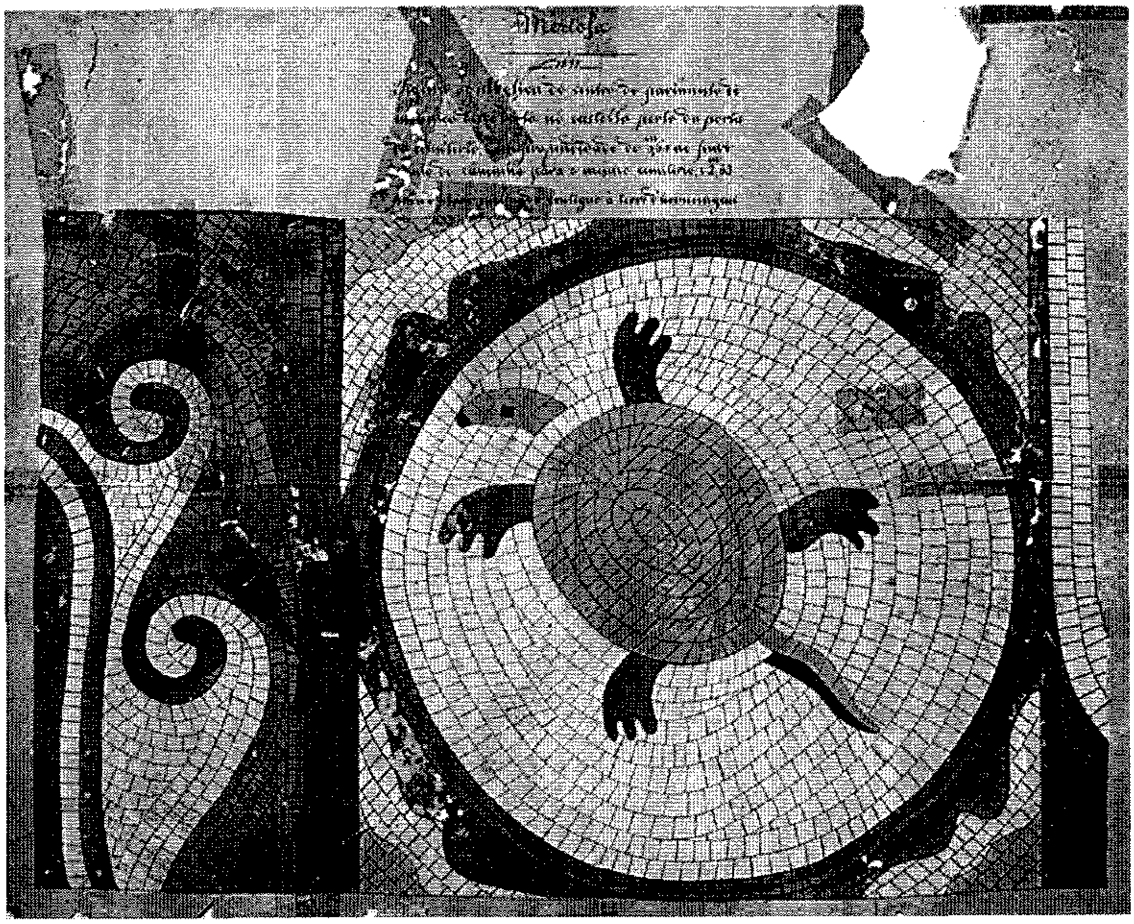
Although being ecologically related, these species show very distinct patterns of occurrence in the Iberian Peninsula. While *E. orbicularis* is a rare species, with a sparse distribution (Fig. 1.2b) and most often with very low local abundances, *M. leprosa* is a very common species with a more continuous distribution (Fig 1.2d) and typically with very high local abundances (Araújo *et al.* 1997).

This thesis also aimed to provide insights for implementing adequate conservation measures for these protected freshwater turtle species. Whether *M. leprosa* is in part responsible for the rarity of *E. orbicularis*, as suggested by some empirical observations (Segurado & Araújo, in review), and whether adequate management efforts are compatible between species are important conservation issues. Indeed, it is possible that *E. orbicularis*, being less abundant and more specialist, will experience a stronger effect than *M. leprosa* from potential interactions between the two species, as predicted by the abundance-asymmetry hypothesis (Vázquez *et al.* 2007). This issue is particularly relevant for the conservation of *E. orbicularis*, which has been suffering severe declines all over its wide geographical range. In fact, although both freshwater turtles are protected under Annexes II and IV of the European Habitats Directive (92/43/CEE), the status of *E. orbicularis* is by far more critical.

This species is currently classified as "endangered" in many European countries (Fritz *et al.* 1998, Buskirk & Servan 2000, Fritz & Havas 2004), including Portugal (Cabral *et al.* 2005).

Chapter 2

The combined effect of modeling technique and species macroecological traits on the accuracy of species distribution models



Late-roman mosaic V century A.C., Basílica do Forum, Mértola, Southern Portugal. After Estácio da Veiga (the image was gently provided by J.L. Cardoso).

Note – This chapter has been published in Journal of Biogeography under the title "an evaluation of methods for modelling species distributions". (Segurado, P. & Araújo, M.B., 2004. *Journal of Biogeography*, 31: 1555–1568). The text has been slightly modified for layout harmonization (e.g. citation format) or for some minor English corrections.

2.1 Summary

Aim Various statistical techniques have been used to model species probabilities of occurrence in response to environmental conditions. This paper provides a comprehensive assessment of methods and investigates whether errors in model predictions are associated to specific kinds of geographical and environmental distributions of species.

Location Portugal, Western Europe.

Methods Probabilities of occurrence for 44 species of amphibians and reptiles in Portugal were modelled using seven modelling techniques: Gower metric, Ecological Niche Factor Analysis, classification trees, neural networks, generalized linear models, generalized additive models and spatial interpolators. Generalized linear and additive models were constructed with and without a term accounting for spatial autocorrelation. Model performance was measured using two methods: sensitivity and Kappa index. Species were grouped according to their spatial (area of occupancy and extent of occurrence) and environmental (marginality and tolerance) distributions. Two-way comparison tests were performed to detect significant interactions between models and species groups.

Results Interaction between model and species groups was significant for both sensitivity and Kappa index. This indicates that model performance varied for species with different geographical and environmental distributions. Artificial neural networks performed generally better, immediately followed by generalized additive models including a covariate term for spatial autocorrelation. Nonparametric methods were preferred to parametric approaches, especially when modelling distributions of species with a greater area of occupancy, a larger extent of occurrence, lower marginality and higher tolerance.

Main conclusions This is a first attempt to relate performance of modelling techniques with species spatial and environmental distributions. Results indicate a strong relationship between model performance and the kinds of species distributions being modelled. Some methods performed generally better, but no method was superior in all circumstances. A suggestion is made that choice of the appropriate method should be contingent on the goals and kinds of distributions being modelled.

Keywords

Classification and regression trees, conservation planning, Ecological Niche Factor Analysis, generalized additive models, generalized linear models, Gower metric, neural networks, Portugal, spatial interpolators.

2.2 Introduction

Models exploring the relationships between species' occurrences and sets of predictor variables produce two kinds of useful outputs. The first are estimates of the probability that species might occur at given unrecorded locations. The second are estimates of an area's suitability for species. The use of empirical models of occurrence in conservation planning has been increasingly advocated (e.g. Margules & Nicholls 1987, Cocks & Baird 1989, Araújo & Williams 2000, Williams & Araújo 2000, 2002, Polasky & Solow 2001, Araújo *et al.* 2004). For example, Polasky & Solow (2001) suggested that probabilities of occurrence could be used as attributes for species-based reserve selection. These probabilities would be regarded as estimates of the likelihood that a species might occur at a given unrecorded location. From another perspective, Araújo & Williams (2000) suggested that probabilities of occurrence could be interpreted as estimates of the probability that species might find suitable habitat in a given area. They also argued that suitability estimates from probabilistic models would be inversely correlated to probabilities of extinction in a near future. Observations of local extinctions in British breeding birds in relation to estimates of probability of occurrence have provided empirical support for this idea (e.g. Gates & Donald 2000, Donald & Greenwood 2001, Araújo *et al.* 2002). Nevertheless the question remains as to what method to choose and whether the choice of modelling technique should depended on the goals, region or species being modelled.

There are many modelling techniques available to explore the correlation between occurrence of a species and sets of predictor variables and there is a possibility that techniques may differ in their ability to summarize useful relationships between response and predictor variables. But model performance may also be contingent on the goals, assumptions and data used for particular analysis. If the goal of a model is to predict occurrences of species outside their known range, interest could be focused on methods that optimize the overall fit, i.e. the balance between false positives (absent, but predicted to be present) and false negatives (present, but predicted to be absent) (Fielding & Bell 1997). However, if the goal is to minimize unexplained variation by models (i.e. minimize false negatives), then a measure of performance would be required that accounted for the proportion of false negatives alone (e.g. Araújo & Williams 2000).

Although some studies have compared the performance of different modelling techniques (Walker 1990, Bio *et al.* 1998, Manel *et al.* 1999, Araújo & Williams 2000, Pearce & Ferrier 2000, Vayssières *et al.* 2000, Vetaas 2000, Thuiller *et al.* 2003), none has, to our knowledge, investigated systematically how variation in species geographical and environmental distributions affect model' performance. Furthermore, the number of techniques compared has often been small (four or less techniques compared at once) and the number of species studied in such comparisons limited (Table 2.1). Here, the performance of seven techniques in modelling the occurrence of 44 reptile and amphibian species in Portugal was compared using nine model runs for each species. The aim was to investigate whether observed errors are consistently related to descriptions of spatial and environmental distributions of species. The initial idea was that the frequency and magnitude of errors in the models might not be independent of the kinds of distributions species have within geographical and environmental space. If the behaviour of error could be predicted from analysis of species distribution alone, then modellers could determine the circumstances where empirical models of occurrence were likely to provide adequate outputs. The results should also indicate which models would be more robust for specific sets of data, and when results should be generally less likely to be useful for specific conservation applications. This could provide users of models with a better understanding of model uncertainty and further insights into how to reduce it.

Table 2.1 – Previous studies comparing model performance, number of species modelled, number and description of techniques used

Reference to paper	No of species modelled	No of models compared	Models used *
Walker 1990	3	2	GLM, CART
Pereira & Itami 1991	1	3	GLM, SI, BM
Lek *et al.* *1996*	1	2	GLM, NNETW
Bio *et al.* *1998*	156	2	GLM, GAM
Franklin 1998	20	3	GLM, GAM, CART
Manel *et al.* *1999*	6	3	GLM, DA, NNETW
Mastrorillo *et al.* *1997*	3	2	NNETW, DA
Araújo & Williams 2000	187	2	GLM, SI
Pearce & Ferrier 2000	24	2	GLM, GAM
Vayssières *et al.* 2000	3	2	GLM, CART
Hirzel *et al.* 2001	1 (virtual)	2	EE, GLM
Dettmers *et al.* 2002	6	4	GLM, DA, MD, CART
Elith & Burgman 2002	8	4	GLM, GAM, BP, GARP
Fertig & Reiners 2002	1	2	GLM, CART
Thuiller *et al.* 2003	4	3	GLM, GAM, CART

* DA, discriminant analysis; BM, Bayesian Model; BP, bioclimatic profiles; CART, classification and regression trees; GAM, generalized additive models; GLM, generalized linear models; GARP, genetic algorithm for rule-set production; MD, mahalanobis distance method; NNETW, neural networks; SI, spatial interpolation.

2.3 Materials and methods

Data

Species data

Data included 9939 occurrence records for 44 species of reptiles and amphibians in Portugal (Table 2.2). Data were located in 993 Universal Transverse Mercator (UTM) 10 by 10 km grid cells, and were compiled from a recently updated Atlas of the Portuguese mainland herpetofauna (Godinho *et al.* 1999). The minimum number of records for a species was four, the median number was 184, and the maximum number was 787.

Environmental and habitat data

Sixteen variables were compiled from various digital sources and treated as predictor variables in the models (Table 2.3). Environmental variables were compiled from available digital layers of the 'Atlas do Ambiente' (http://www.dga.pt). Habitat data included eight variables selected from the Portuguese EU CORINE land-cover data base (http://www.snig.cnig.pt). Human-population data were also compiled from published sources (CNA 1983). Data were converted from available vector maps and resampled to raster UTM 10 by 10 km grid cell maps, using zonal functions in ArcView version 3.2 (ESRI 1999) to calculate the mode of each variable in each cell.

Table 2.2 – List of species modelled, prevalence, area of occupancy, extent of occurrence, marginality and tolerance (Units: occupancy – number of 10 km grid squares; Extent of occurrence – km).

Species	Prevalence	Occupancy	Extent of occurrence (Km)	Marginality	Tolerance
Acanthodactylus erythrurus	0.062	62	595.06	0.589	0.624
Alytes cisternasii	0.271	270	590.93	0.75	0.713
Alytes obstetricans	0.198	198	446.54	0.896	0.699
Anguis fragilis	0.097	97	568.59	0.945	0.557
Blanus cinereus	0.126	126	608.36	0.617	0.754
Bufo bufo	0.541	540	630.71	0.317	0.922
Bufo calamita	0.275	274	617.41	0.450	0.872
Chaemaleo chaemaleo	0.022	22	104.40	1.016	0.172
Chalcides bedriagai	0.056	56	594.81	0.595	0.505
Chalcides striatus	0.166	166	613.27	0.579	0.868
Chioglossa lusitanica	0.207	207	304.63	1.025	0.538
Coluber hippocrepis	0.167	167	572.71	0.552	0.814
Coronella austriaca	0.031	31	272.03	0.939	0.314
Coronella girondica	0.136	136	604.15	0.607	0.810
Discoglossus galganoi	0.211	211	608.36	0.443	0.909
Elaphe scalaris	0.295	294	621.69	0.403	0.906
Emys orbicularis	0.065	65	610.98	0.706	0.584
Hemidactylus turcicus	0.031	31	252.39	0.779	0.378
Hyla arborea	0.156	156	568.59	0.479	0.780
Hyla meridionalis	0.117	117	416.77	0.739	0.632
Lacerta lepida	0.423	422	635.06	0.485	0.909
Lacerta monticula	0.004	4	14.14	0.756	0.102
Lacerta schreiberi	0.291	290	590.93	0.852	0.768
Macroprotodon cucullatus	0.061	61	542.31	0.751	0.484
Malpolon Monspessulanus	0.492	491	630.71	0.347	0.940
Mauremys leprosa	0.436	435	613.27	0.637	0.838
Natrix maura	0.492	491	622.41	0.401	0.934
Natrix natrix	0.192	192	622.41	0.607	0.848
Pelobates cultripes	0.245	245	599.33	0.656	0.826
Pelodytes punctatus	0.104	104	523.93	0.808	0.624
Pleurodeles waltl	0.220	220	554.71	0.686	0.776
Podarcis bocagei	0.078	78	583.10	1.003	0.516
Podarcis hispanica	0.368	367	626.10	0.572	0.909
Psammodromus algirus	0.591	590	630.71	0.325	0.943
Psammodromus hispanicus	0.141	141	604.15	0.582	0.804
Rana iberica	0.264	263	399.62	0.923	0.681
Rana perezi	0.788	786	626.50	0.270	0.939
Salamandra salamandra	0.484	483	621.69	0.445	0.928
Terentola mauritanica	0.176	176	581.81	0.678	0.816
Triturus boscai	0.369	368	617.41	0.546	0.881
Triturus helveticus	0.019	19	230.87	0.776	0.313
Triturus marmoratus	0.382	381	621.69	0.531	0.907
Vipera latastei	0.088	88	617.41	0.744	0.688
Vipera seoanei	0.014	14	136.01	0.812	0.148

Table 2.3 – Predictor variables in the models.

Variables	Units	Parameters	Type
Environmental			
Drainage	mm	mode	ordinal
Acidity	pH	mode	ordinal
Evapo-transpiration	mm	mode	ordinal
Humidity	%	mode	ordinal
Precipitation	mm	mode	ordinal
Solar radiation	kcal	mode	ordinal
Temperature	°C	mode	ordinal
Socio-economic			
Human population density	n/km^2	mode	ordinal
Land use			
Urban areas	-	% area	ordinal
Annual crops	-	% area	ordinal
Permanent crops	-	% area	ordinal
Pastures	-	% area	ordinal
Heterogeneous crops	-	% area	ordinal
Florests	-	% area	ordinal
Schrub lands	-	% area	ordinal
Bear soil	-	% area	ordinal

Modelling techniques

There is a broad range of modelling techniques available to explore the correlation between response and predictor variables (Guisan & Zimmermann 2000). These techniques include the Gower-similarity model (e.g. Carpenter *et al.* 1993), Ecological Niche Factor Analysis (ENFA) (e.g. Hirzel *et al.* 2001), classification trees (e.g. Breiman *et al.* 1984, Clark & Pregibon 1992), neural networks (e.g. Mastrorillo *et al.* 1997, Manel *et al.* 1999, Özesmi & Özesmi 1999), generalized linear models (McCullagh & Nelder 1989), generalized additive models (Hastie & Tibshirani 1990), and spatial interpolation techniques (e.g. Bailey & Gatrell 1995). Here the performance of seven modelling techniques was compared and applied to the same data. Nine models were obtained for each of the 44 reptiles and amphibian species. Overall, 396 models were compared. To make outputs from models comparable, the whole set of variables in all models was used (e.g. Fielding & Haworth 1995).

Gower-similarity models

Modelling approaches based on similarities between data points only use presence data (ignoring absences) to create species' environmental envelopes. The Gower-similarity approach, as implemented in DOMAIN V1.3 for Windows (http://www.cifor.cgiar.org/domain/index.htm), was used. The DOMAIN algorithm assigns each cell in the output layer an average multivariate distance, termed the Gower metric, between that cell and the closest presence cell in the training set (Carpenter *et al.* 1993). These distance values were rescaled from zero to one (value - minimum/maximum - minimum) to provide values comparable with that obtained with probability- based techniques.

Ecological Niche Factor Analysis

A related approach, using presence data alone, is based on ordination of data in a multivariate space of environmental variable (ENFA; Perrin 1984). This technique is based on the computation of the factors explaining the major part of species environmental distribution. Extracted factors are uncorrelated and have biological significance: the first factor is the marginality factor, which describes how far the species optimum is from the mean environmental profile in the study area; the second is the tolerance factor, which is sorted by decreasing amount of explained variance and describe how specialized the species is by reference to the available range of environments in the study area (Hirzel 2001). This approach was implemented using BIOMAPPER (Hirzel 2001) software to produce habitat suitability maps. A Habitat Suitability Index (HSI) of each cell is a value inversely proportional to the weighted mean distance of the cell to the median of each ENFA factor. This value is normalized in such a way that the suitability index ranges from zero to one (Hirzel 2001). The number of retained factors for computing habitat suitability maps was chosen using a 'broken stick advice' implemented in BIOMAPPER (Hirzel *et al.* 2001). Here, the distribution of the eigenvalue of each factor is compared with the distribution of MacArthur's (1957) brokenstick, which is the expected distribution when breaking a stick randomly. The eigenvalues that are larger than expected according to the broken stick distribution may be considered 'significant'.

Classification trees

Classification trees (TREE; Breiman *et al.* 1984, Clark & Pregibon 1992) consist of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of the response. The tree is built by repeatedly splitting the data into two exclusive groups, defined by a simple rule based on a single explanatory variable at each step. Classification trees were fitted using the RPART library of tree routines developed by Therneau & Atkinson (1997, 2000) for S-Plus (Statistical Sciences 1999). The RPART function generates pruned trees based on the results of crossvalidations (Therneau & Atkinson 1997, 2000). For each tree, a series of 50 10-fold cross-validations were run and the most frequently occurring tree size was chosen using the 1-SE rule (De'Ath & Fabricius 2000).

Neural networks

Artificial neural networks were computed using the NNET library of S-Plus (NNETW; Venables & Ripley 2002). NNETW is a feed-forward neural network, parameterized using seven hidden units in a single layer (selected by cross-validation), with a weight decay equal to 0.03. Given the heuristic nature of NNETW, each simulation gives slightly different results. For this reason, NNETW solutions were calculated 10 times for each species and the mean was used to provide predictions. The procedure described is commonly used for modelling presence/absence data using neural networks (e.g. Mastrorillo *et al.* 1997, Manel *et al.* 1999, Özesmi & Özesmi 1999, Thuiller 2003).

Generalized linear models

Generalized linear models (GLM) (McCullagh & Nelder 1989) assuming a binomial error distribution – logistic regression – were adjusted using S-Plus functions and routine facilities (Statistical Sciences 1999). In order to account for autocorrelation in the observations, models were also fitted in which contagion (see below: spatial interpolators) was included as an autocovariate term in the initial variable set (AGLM). These models are termed autologistic (Smith 1994, Augustin *et al.* 1996, Araújo & Williams 2000).

Generalized additive models

To allow consideration of more complex response shapes than those possible through the linear responses of GLM, generalized additive models (GAM; Hastie & Tibshirani 1990) were fitted using logit as the link function and binomial error distribution. As for GLM, a set of models were fitted that included an autocovariate term in the initial set of variables. Hence, two GAM models were obtained for each species (GAM and AGAM).

Spatial interpolators

Measures of aggregation for point and lattice data, such as Kernel estimation and nearest neighbour measures (e.g. Bailey & Gatrell 1995), can be used to model species' probabilities of occurrence. This uses the idea of positive spatial autocorrelation (Legendre 1993), in which the occurrence of a species in one area is expected to be more likely if the species occurs in many surrounding areas (Araújo & Williams 2000, 2001, Araújo *et al.* 2002). A measure of contagion (CONT) for each cell, based on a two-order neighbourhood, was used to estimate a distance-based probability of occurrence. Contagion is measured as a weighted average of the number of occupied grid-cells among a set of ka neighbours of a central grid-cell y_a, so that:

UNIVERSIDADE DE EVORA

$$Contagion = \left(\frac{\sum_{b=1}^{k_a} w_{ab} y_b}{\sum_{b=1}^{k_a} w_{ab}} \right) \qquad (1)$$

where the weight given to the grid-cell y_b is $w_{ab} = 1/d_{ab}$, and d_{ab} is the distance between grid-cells y_a and y_b. Two orders of neighbours, assigning a weight of $d = 1$ to the first-order and a weight of $d = 2$ to the second-order neighbours were used. Neighbours in the first order were the eight adjacent cells touching the central cell along the edges and at the corners within a rectangular grid. The second-order neighbours were the next group of cells concentric to the first order with 16 grid cells.

Measures of model performance

Models can be assessed qualitatively, or quantitatively. Qualitative assessments measure how well models fit the data, while quantitative assessments measure how well models predict real events (Myers 1997). In most circumstances, model performance can only be estimated through qualitative assessments (but for example of quantitative assessment see Feria & Peterson 2002). Hence, because models are optimised for the training set, the 'goodness-of-fit' to these data is likely to be an over-optimistic estimate of predictive skill outside the training set (Beutel *et al.* 1999). Here model' performance was assessed with two measures: sensitivity and the Kappa statistic.

Sensitivity is based on the concept of true-presences misclassification (false negatives or type II error), and is calculated as 1 – percentage of false negatives (Fielding & Bell 1997, Cumming 2000). The number of false negatives is particularly useful because it measures the number of residuals, or amount of unexplained variation in the data; the greater the number of false negatives, the more models are likely to be unrealistic. The main problem using standard indexes of model performance, such as sensitivity, is that a cut-off level of probability has to be defined a priori. Some authors have used the questionable rule of thumb of a 0.5 cut-off probability (e.g. Austin *et al.* 1996, Franklin 1998, Manel *et al.* 1999, Manel *et al.* 2001), but this has been shown to be inadequate because variation in prevalence (i.e. the number of occurrences in relation to the number of samples) makes optimal cut offs for species vary (e.g. Manel *et al.* 2001). In this paper sensitivity was based on the cut-off level for which the Cohen's Kappa was maximum. The Kappa statistic assesses the extent to which models predict occurrence at a rate higher than expected by chance (Monserud & Leemans 1992). Hence, it was used both as a performance measure and as a criterion for cut-off level selection.

Since the goal of this study was to investigate variation of model performance in relation to attributes of species' distribution, the adoption of standard procedures across species to assess model performance was required. A bootstrap procedure (Efron & Tibshirani 1993) was used to estimate the standard error of sensitivity as a measure of model stability (Guisan & Zimmermann 2000). Ideally this should be performed in a training set checked against a validation set. However, the number of records for some species was so low that it was unfeasible to split the records into training and validation sets for a great proportion of them. An alternative would have been to exclude some species with restricted ranges from the analysis (e.g. Thuiller 2003, Araújo *et al.* 2004). The consequence would be that some of the most important species for conservation would be excluded. In order to avoid this problem, model performance was estimated using the training set alone. This represents a severe limitation for models trying to predict distributions outside known ranges. However for models trying to explain current occurrences and assign suitability scores to areas this limitation is less severe.

Spatial and environmental distributions

When comparing the performance of models for large numbers of species it becomes difficult to make interpretations on a species-by-species basis. In such circumstances, it might be useful to group species, for example, according to their spatial or environmental distributions. These groupings are expected to mirror ecological and historical factors. Hence, they are likely to provide useful insights for post-hoc interpretation of results. Here spatial distribution of species was described with two measures: (1) area of occupancy, given by the number of 10 km grid cells where the species occur; and (2) extent of occurrence, given by the straight-line distance between the two most distant occupied grid cells (Gaston 1996). Species-environmental distributions were described with two measures: (1) marginality (or niche position), which reflects how far the species optimum is from the mean environmental conditions in the study area; and (2) tolerance (or niche breath), which describes how variable the species association to environmental factors is with reference to the available range in the study area. These two environmental measures were calculated using the ENFA approach implemented in BIOMAPPER (Hirzel *et al.* 2001).

Species were then grouped according to their distributional profile, using hierarchical agglomerative clustering (group average). The statistical significance of the resulting groups was tested with ANOSIM (analysis of similarity), which performs permutation tests for the null hypothesis that there are no assemblage differences between groups (Clarke & Gorley 2001). Both cluster analysis and ANOSIM were performed using primer for Windows, version 5.2.0 (Clarke & Gorley 2001).

Testing differences in model performance

Since the same measurement of model performance was calculated several times (different modelling procedures) on each subject (species), a repeated measures factorial design was used for testing differences in model performance. Differences between overall performances of each modelling procedure, assuming no interaction between model procedure and species groups, were tested using Friedman test, which is a nonparametric version of one-way repeated anova measures (Sprent & Smeeton 2001). Dunn tests (Klockars & Sax 1986) were then performed for post-hoc multiple comparisons.

In order to test differences between groups of species, twoway repeated measures anova-tests were also performed using the modelling procedure as the within-subject factor and the species group as the between-subject factor. The main goal of this test was to check for interactions between modelling procedures and species groups, that is, to test if distinct modelling procedures performed differently for distinct groups of species. Data were normalized using an arcsine transformation (Sokal & Rohlf 1995). Tests were carried out independently for each partition of species and for each performance measure, i.e. for the Kappa index and sensitivity. The same analyses were carried out to test for differences in model stability, measured as the bootstrap standard error estimates, between species groups and modeling techniques.

2.4 Results

Comparison of the overall model performance

The results of Friedman tests (Table 2.4) showed that there is significant effect of the modelling technique in model performance for both measures considered (Kappa index and sensitivity). These differences were lower for sensitivity. The rank order of model performance did not vary much with the measure used (Fig. 2.1a,b). On the other hand, not all rank order differences were significant according to multiple comparisons Dunn tests (Tables 2.5–2.7). Neural networks (NNETW) provided the best results as measured both with Kappa and sensitivity indices ($P < 0.001$) (Table 2.5). However, NNETW performance as measured with Kappa index was not significantly different to the results achieved with generalized additive models with a covariate term accounting for spatial autocorrelation (AGAM). AGAM also showed better performances in terms of Kappa index ($P < 0.001$, except for GAM where $P < 0.05$) than the remaining models. However, AGAM distribution of sensitivity was no different than GAM, without a covariate term accounting for spatial auto-correlation, generalized linear model (AGLM), classification tree analysis (TREE) and contagion (CONT). GAM, AGLM and TREE had similar performances showing no significant differences amongst them. The poorest performing methods in terms of Kappa index and sensitivity measurements were DOMAIN and BIOMAP. There were no significant differences of model performance among these methods. GLM also performed poorly, showing no significant differences from DOMAIN and BIOMAP in terms of the Kappa index, although being significantly better ($P < 0.01$) than the worse method in terms of sensitivity, i.e. BIOMAP. Spatial interpolation with CONT had a performance similar to GLM but, unlike GLM, did not show significant differences in the Kappa index from AGLM and TREE.

Table 2.4 – Results of the Friedman tests for global differences between the performance measures of the modelling procedures.

Performance measure	F	p-level
Kappa	292.54	< 0.0000
Sensitivity	145.84	< 0.0000
SE of sensitivity	107.67	< 0.0000

The results also show that model stability, measured by bootstrap estimates of sensitivity standard errors, was also affected by modelling technique. NNETW showed the lowest bootstrap standard errors (Fig. 2.1c) and according to Dunn tests (Table 2.7) were significantly different from all the other methods ($P < 0.001$, except for DOMAIN where $P < 0.05$). DOMAIN also showed significantly lower bootstrap standard errors than BIOMAP ($P < 0.01$), GLM ($P < 0.001$) and CONT ($P < 0.01$). GLM had the lowest overall model stability, although differences were only significant in relation to AGAM ($P < 0.01$) and NNETW ($P < 0.001$).

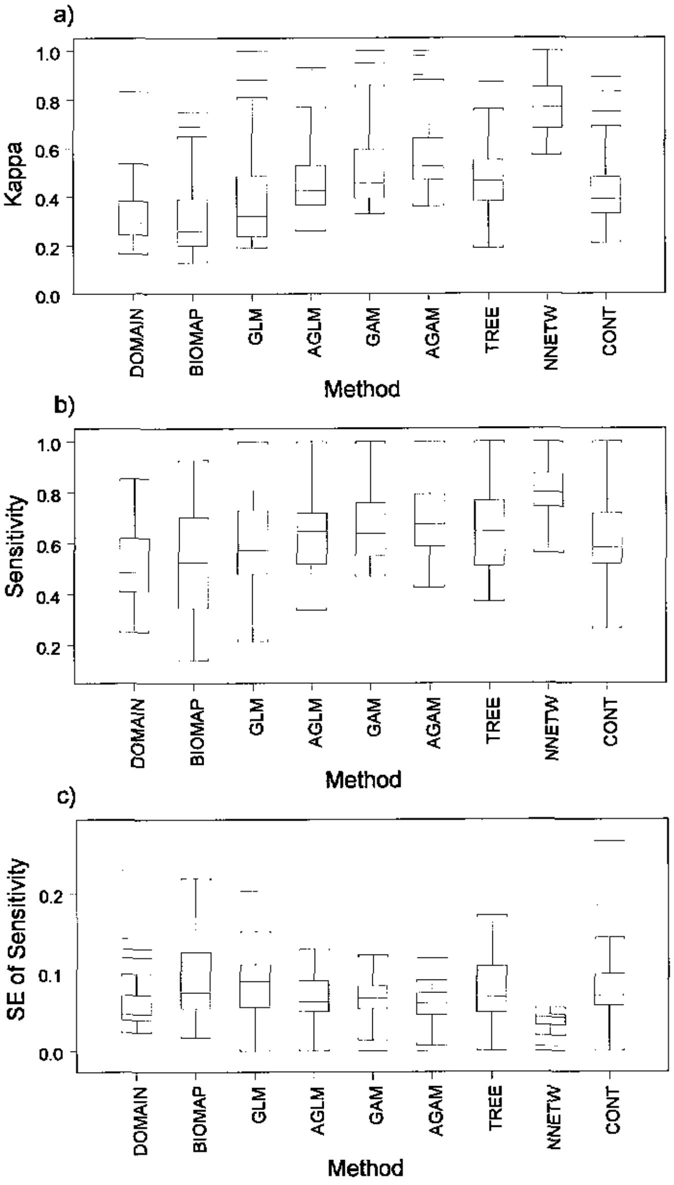


Fig. 2.1 - Boxplots of overall performance measures for each 12 modelling procedures (a, Kappa; b, sensitivity). Abbreviations: DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, autologistic general linear model, GAM, general additive model; AGAM , autologistic general *additive model*; TREE, classification tree; NNETW, neural network; CONT, contagion.

Table 2.5 - Multiple comparisons of model's Kappa index

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	ns							
GLM	ns	ns						
AGLM	***	***	**					
GAM	***	***	ns	ns				
AGAM	***	***	***	***	*			
TREE	***	***	***	ns	ns	***		
NNETW	***	***	***	***	***	ns	***	
CONT	*	*	ns	ns	**	***	ns	***

Dunn tests; *** - $p<0.001$; ** - $p<0.01$; * - $p<0.05$; ns – $p>0.05$

Table 2.6 - Multiple comparisons of model's sensitivity

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	ns							
GLM	*	ns						
AGLM	***	*	ns					
GAM	***	*	ns	ns				
AGAM	***	***	*	ns	ns			
TREE	***	**	ns	ns	ns	ns		
NNETW	***	***	***	***	***	***	***	
CONT	**	ns	ns	ns	ns	ns	ns	***

Dunn tests; *** - $p<0.001$; ** - $p<0.01$; * - $p<0.05$; ns – $p>0.05$

Table 2.7 - Multiple comparisons of model's standard errors of sensitivity (Dunn tests; *** - $p<0.001$; ** - $p<0.01$; * - $p<0.05$; ns – $p>0.05$).

	DOMAIN	BIOMAP	GLM	AGLM	GAM	AGAM	TREE	NNETW
BIOMAP	**							
GLM	***	ns						
AGLM	ns	ns	ns					
GAM	ns	ns	ns	ns				
AGAM	ns	ns	**	ns	ns			
TREE	ns	ns	ns	ns	ns	ns		
NNETW	*	***	***	***	***	***	***	
CONT	**	ns	ns	ns	ns	ns	ns	***

Dunn tests; *** - $p<0.001$; ** - $p<0.01$; * - $p<0.05$; ns – $p>0.05$

Model performance between methods and species

groupings – spatial distributions Four groups were created on the basis of species-spatial distribution (Fig. 2.2a) and they were all significantly different from each other (ANOSIM, $P < 0.05$): (1) narrowly distributed species, i.e. species with low area of occupancy and narrow extent of occurrence; (2) species with aggregated distributions, i.e. species with both intermediate areas of occupancy and extent of occurrence; (3) widespread but locally rare species, i.e. species with low to intermediate areas of occupancy, but with large extent of occurrence; and (4) truly widespread species, i.e. species with high area of occupancy and large extent of occurrence.

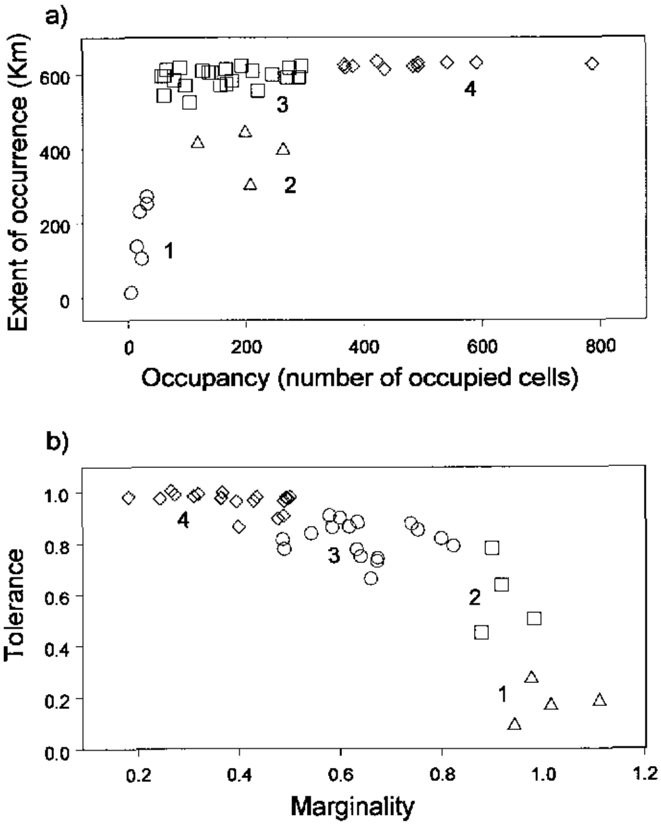


Fig. 2.2 - Representation of species and groups according to heir geographical attributes (a) and their distribution in the environmental space (b).

Model performance varied consistently between species with different spatial distributions (anova $P < 0.001$; Fig. 2.3a,b). Differences in standard error of sensitivity estimates and species groupings were also significant at P < 0.05. Variations in Kappa index revealed a clear trend towards increasing model performance with the restricted-range species of group 1, i.e. species with low area of occupancy and narrow extent of occurrence (Fig. 2.3a). At the opposite extreme we found truly widespread species of group 4 to have the lowest performance from the models. Variations in sensitivity values were similar to Kappa index, except for widespread species of group 4 that had generally good performances with sensitivity (Fig. 2.3b). This discrepancy is due to the numbers of false positives –included in Kappa index measurement, but not in sensitivity – that are particularly high among widespread species models causing Kappa index values to drop.

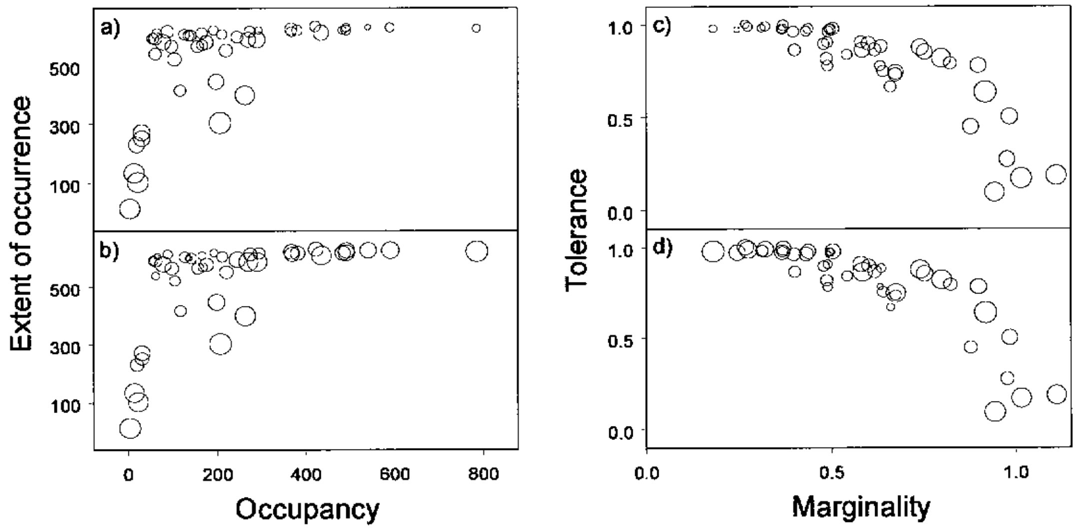


Fig. 2.3 - Representation of the mean values of Kappa index (a and c) and Sensitivity (b and d) of each species according to the species distribution in the geographical (a and b) and the environmental spaces.

Interactions between the effects of method and species groupings on measures of model performance and stability were also highly significant (anova, P < 0.00001). The relative performances of each method varied between species with different spatial distributions (Figs 2.4a, 2.5a & 2.6a): (1) lower relative performance and stability measures of DOMAIN and BIOMAP were more evident for species of group 1; (2) highest performances of NNETW are more evident for species of groups 3 and 4 (more marked for the Kappa index); (3) AGAM showed performances comparable with those of NNETW for species of groups 1 and 2.

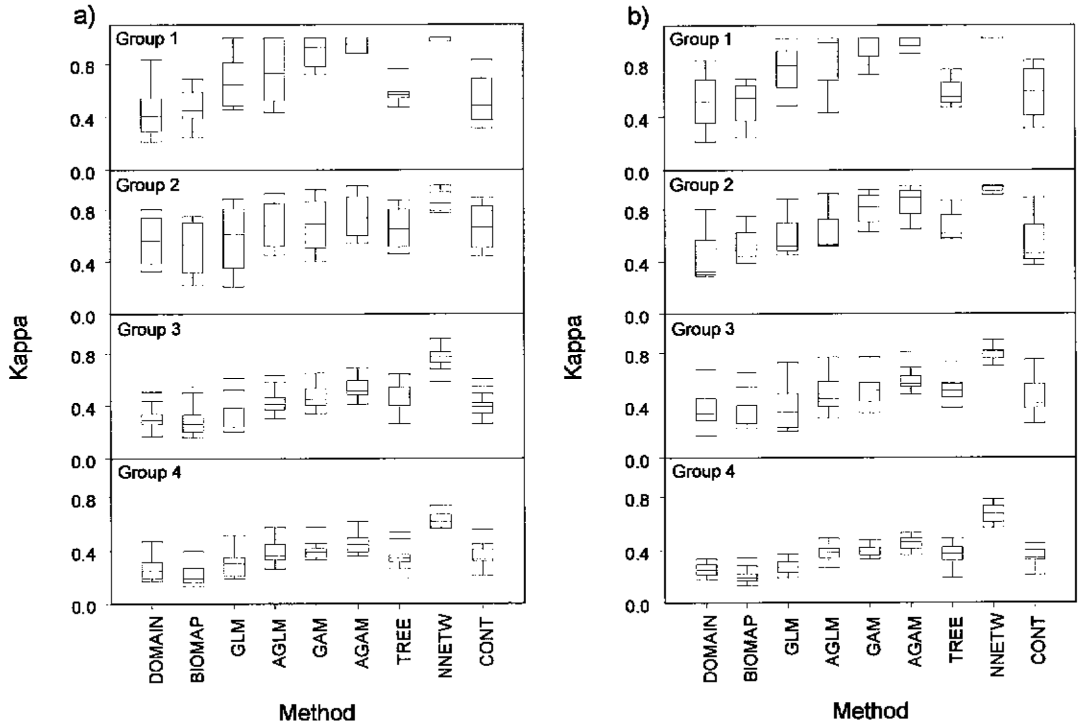


Fig. 2.4 - Boxplots of Kappa index for each model procedure and each species according to the species distribution in (a) geographical space and (b) environmental space (DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, with a covariate term accounting for spatial autocorrelation general linear model, GAM, general additive model; AGAM, with a covariate term accounting for spatial autocorrelation general additive model; TREE, classification tree; NNETW, neural network; CONT, contagion).

Model performance between methods and species groupings – environmental distributions

As for spatial groupings, there were also four groups created on the basis of species' environmental distribution (Fig. 2.2b) and they were also significantly different from each other (ANOSIM, $P < 0.05$): (1) species with high marginality and low tolerance; (2) species with high marginality and intermediate levels of tolerance; (3) species with intermediate marginality and high tolerance; (4) species with low marginality and high tolerance.

Model performance also varied consistently between species with different environmental distributions (anova, $P < 0.01$; Fig. 2.3c,d). Differences in standard error of sensitivity estimates and species groupings were also significant at $P < 0.05$. Model performance was higher for specialist species of group 1, i.e. species with high environmental marginality and low tolerance, and lower for generalist species of group 4, i.e. species with low marginality and high tolerance (Fig. 2.3c). Variations in sensitivity values were similar to Kappa index, except for species of group 4 that had generally good performances with sensitivity (Fig. 2.3d).

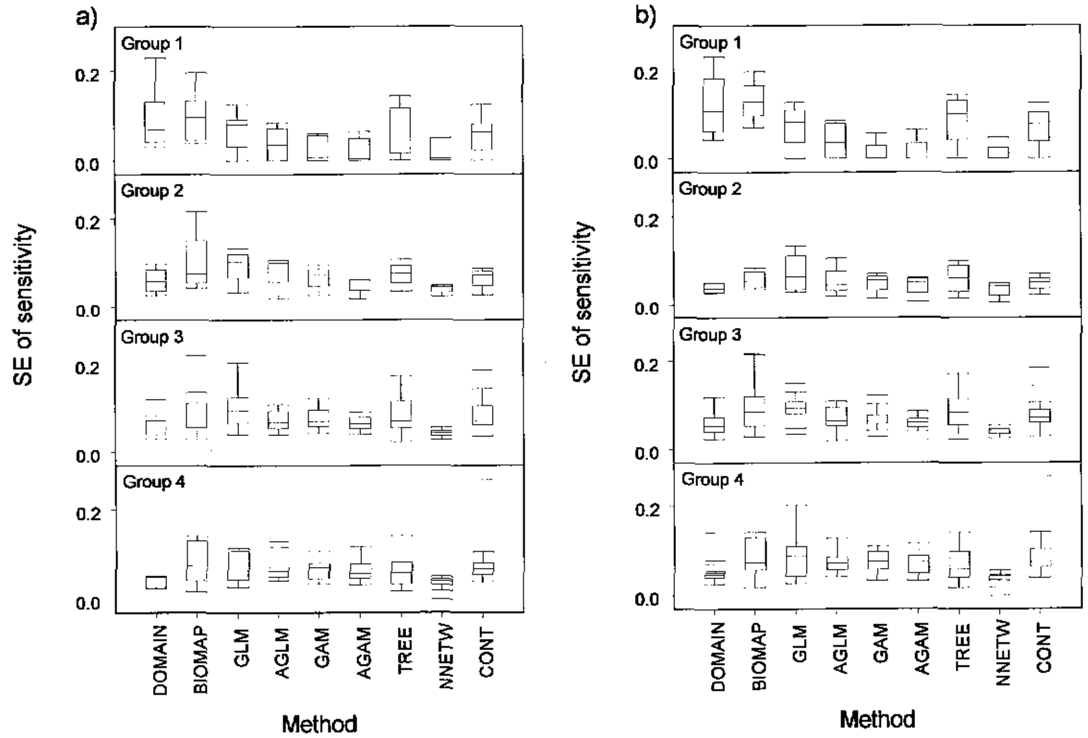


Fig. 2.5 - Boxplots of sensitivity for each model procedure and each species according to the species distribution in (a) geographical space and (b) environmental space. Abbreviations: DOMAIN, Gower metric; BIOMAP, Ecological Niche Factor Analysis; GLM, general linear model; AGLM, general linear model, with a covariate term accounting for spatial autocorrelation GAM, general additive model; AGAM, with a covariate term accounting for spatial autocorrelation general additive model; TREE, classification tree; NNETW, neural network; CONT, contagion.

Interactions between the effects of method and species groupings on measures of model performance and stability were also highly significant (anova, $P < 0.00001$). The relative performances of each method varied between species with different environmental distributions (Figs 2.4b, 2.5b & 2.6b). As for spatial distributions, the same trend was observed for environmental distributions: (1) lower measures of performance and stability of DOMAIN and BIOMAP were more evident for species of group 1; (2) highest performances of NNETW were more evident for species of groups 3 and 4; (3) AGAM showed performances comparable to those of NNETW for species of groups 1 and 2.

2.5 Discusion

Most studies assessing probabilistic models of occurrence for species have compared the performance of techniques for one or few species (e.g. Walker 1990, Pereira & Itami 1991, Manel *et al.* 1999, Dettmers *et al.* 2002, Thuiller *et al.* 2003). Amongst these, only a few have investigated how variation in predictor (e.g. environmental) variables affected the ability of the models to predict correctly variation in the response (e.g. species) variable (e.g. Manel *et al.* 1999). There have also been attempts to investigate how variation in the response variable would affect the outcome of the models (Araújo & Williams 2000, Pearce & Ferrier 2000, Manel *et al.* 2001, Pearce *et al.* 2001, Hepinstall *et al.* 2002, Karl *et al.* 2002). This latter approach was investigated further and the error in models shown to be associated consistently with simple quantitative descriptors of the response variable, i.e. species distributions. In other words, the original idea that model performance was not independent of the kinds of geographical or environmental distributions of species was shown to be correct.

Generally, species with large areas of occupancy and great extents of occurrence (i.e. truly widespread) had greater overall errors (i.e. Kappa statistic), although the amount of unexplained variation (i.e. sensitivity) was not larger for species of this group than for species with other distribution profiles. This coincides with observations made by Araújo & Williams (2000), who explored the relationship between model performance (sensitivity and specificity) and the total number of records per species (i.e. area of occupancy). They found that sensitivity (proportion of false negatives) was higher for widespread and lower for restricted-range species, while specificity (proportion of false positives) was lower for widespread species and higher for restricted-range species. Similarly, Manel *et al.* (2001) investigated how model performance varied with the number of records per species relative to the number of samples (i.e. prevalence). They found sensitivity to be related positively to prevalence and negatively to specificity. Since the number of samples was identical for all species, in both studies, area of occupancy and prevalence were perfectly covariant. Karl *et al.* (2002) related avian species commonness and model accuracy using field and simulated data and concluded that habitat relationships for many rare species were likely to be as accurate as for common species, despite increasing error estimates with decreasing sample sizes. Similarly, Elith & Burgman (2002) did not find clear associations between modelling success and species characteristics such as rarity.

In this study, species with low marginality and high tolerance had lower overall performances. They were also generally better modelled with non-parametric techniques and benefited most from the inclusion of a term accounting for spatial autocorrelation. This is because autocorrelation reduces the amount of false positives (specificity), which in turn contributes to a reduction in the Kappa statistic. Hepinstall *et al.* (2002) also observed that performance of avian habitat models was negatively correlated with the proportion of habitats used by a species, although they used total correct predictions as the accuracy measure, and did not take marginality into account. They also noted an opposite trend with measures of sensitivity but in the current study no clear trend was observed for this measure.

Overall, the NNETW showed the highest model performance whereas similarity and ordination-based models (DOMAIN and BIOMAP, respectively) showed the lowest performances. While some authors (e.g. Mastrorillo *et al.* 1997, Pearson *et al.* 2002) also consider NNETW to be advantageous to model species occurrences, these observations are not supported by other studies where NNETW showed overall performances comparable to GLM (Manel *et al.* 1999). In the few studies where species distribution models using GAM were compared with their GLM counterparts, the former approach also resulted in better model performances (e.g. Bio *et al.* 1998, Franklin 1998, Pearce & Ferrier 2000, Thuiller *et al.* 2003). Studies comparing TREE and GLM reached variable conclusions. For example, Dettmers *et al.* (2002) found that TREE and GLM produced comparable results. Franklin (1998) showed TREE to perform better than GLM, while Thuiller *et al.* (2003) showed GLM to perform better than TREE. Since these studies did not always use the same parameterization for TREE and GLM, they are, however, not fully comparable. Other studies also showed that similarity and ordination-based methods perform less well than other techniques, namely GAM and GLM (Elith & Burgman 2002). Hirzel *et al.* (2001) compared the performance of GLM and BIOMAP using virtual species and concluded, as here, that GLM produced better results for overabundant species.

These apparently divergent patterns of model' performance are likely to be related to variations in the methods' abilities to recover useful relationships between species with different distributions and environmental factors with different strengths and lengths of gradients. Different modelling techniques have distinct approaches regarding adjustment to data (Guisan & Zimmermann 2000). Indeed, modelling techniques used in this paper can be grouped into four categories according to their kinds of adjustment to data: (1) techniques that give priority to the empirical behaviour of species' presence/absence response to environmental variables (e.g. non-parametric models such as GAM, classification trees and neural networks); (2) techniques that focus on general trends of species' presence/absence response (e.g. parametric models such as GLM); (3) techniques that use only species presence data to seek relationships with the environmental predictors (e.g. Gower-similarity and ENFA); (4) techniques that use only species presence data and their geographic positions to develop predictions of species occurrence (e.g. spatial interpolators). The first and fourth approaches are expected to provide better models for species with complex distribution patterns, i.e. where occurrences do not respond to environmental variables according to a predefined 'shape'. This is frequently the case among widespread species that do not appear to respond to clear environmental gradients, although studies at varying spatial scales might provide further insights on species' governing factors (Thuiller *et al.* 2003). Conversely, the second approach is expected to provide reasonable models for species responding to environmental gradients as predicted by simple response curves. The third and fourth approaches were both expected to provide models with high sensitivity (low misclassification of true presences). However, they were also expected to yield models with low overall performances, since their formulation disregards the response of absence data to environmental variables. The third approach, in particular, should be considered when no reliable data on species absences are available.

Factors other than species ecological attributes can also be a source of variation in model performance. For example, environmental variables may be adequate for some species but not for others. Data quality can also strongly affect model performance. This may include imprecise location of species occurrences, but also error propagation during generation of response variables in GIS environments (Corsi *et al.* 2000). Moreover, if another spatial scale or resolution is adopted, then a different pattern of model performances might be observed (Mackey & Lindenmayer 2001, Thuiller *et al.* 2003).

It is unlikely that a single best habitat modelling procedure will ever be identified. Different methods have different strengths and weaknesses and the choice of the appropriate method depends on the data, assumptions and goals of the exercise. This study provides some insights into the structure of the error in the models and how they vary for species with different kinds of distribution. Although NNETW and AGAM provided generally better results than other methods tested, an unequivocal pattern of model performance that worked for all species was not found. This leaves the modeller with two choices. Either (1) to use expert systems (e.g. GARP, Stockwell & Peters 1999; BIOMOD, Thuiller 2003) that compare methods in an automatic fashion and choose the best method for each species run, or else choose a single method that is seen to be generally robust (e.g. NNETW and AGAM) or; (2) choose a method thought to be particularly robust to the kinds of data and goals pursued (e.g. DOMAIN and ENFA when limited records of presence, alone, are available). The first strategy seeks to optimize the fit of the models given available data. It is completely data driven and no a priori hypotheses are made about data and the nature of species' responses to gradients. The second option ensures reasonable results while preserving accountability, given that model assumptions are made clear and applied for all species modelled. This last option would exclude NNETW because ecological principles underlying the fit of these models are poorly understood, although progress is being made towards eliminating this limitation (e.g. Olden & Jackson 2002). From the relatively extensive literature on the subject, it becomes clear that model adequacy depends greatly on each particular situation. Studies that attempt to evaluate and compare model performances for various techniques should always consider the effect of the many factors involved, such as geographical scale, adequacy of predictor variables and, as shown in this paper, the spatial and environmental distribution of species.

Acknowledgments

We thank the many volunteer fieldworkers who contributed to atlas records; Alexander Hirzel for support with ENFA; Wilfried Thuiller, Richard Pearson, Susana Rosa and two anonymous referees for comments on the manuscript. This work benefited from research grants by the Portuguese FCT/ICN research programme (PNAT/1999/BIA/15051/P) and by the European ATEAM project (EVK2-CT-2000-00075); MBA and PS were also funded by FCT (SFRH/BPD/5547/2001 and SFRH/BD/8493/2002, respectively).

Chapter 3

The effect of autocorrelation on species distribution models

"...professional ecologists have become stuck upon the issue of how much real biology ought to be put in a null model. If we construct a very realistic null model, should we be surprised when we find that nature conforms to it? And if nature does conform to a realistic null model, are we therefore justified in concluding that there are no pattern in nature?"

P. Keddy (1998) commenting the book by Gottelli and Graves *Null models in ecology* (Gottelli & Graves 1996).

Note – This chapter has been published in Journal of Applied Ecology under the title "Consequences of spatial autocorrelation for niche-based models" (Segurado, P, Araújo, M.B. & Kunin, W.E., 2006. *Journal of Applied Ecology*, 43: 433–444). The text has been slightly modified for layout harmonization (e.g. citation format) or for some minor English corrections.

3.1 Summary

1. Spatial autocorrelation is an important source of bias in most spatial analyses. We explored the bias introduced by spatial autocorrelation on the explanatory and predic-tive power of species' distribution models, and make recommendations for dealing with the problem.

2. Analyses were based on the distribution of two species of freshwater turtle and two virtual species with simulated spatial structures within two equally sized areas located on the Iberian Peninsula. Sequential permutations of environmental variables were used to generate predictor variables that retained the spatial structure of the original variables. Univariate models of species' distributions using generalized linear models (GLM), generalized additive models (GAM) and classification tree analysis (CTA) were fitted for each variable permutation. Variation of accuracy measures with spatial autocorrelation of the original predictor variables, as measured by Moran's I, was analysed and compared between models. The effects of systematic subsampling of the data set and the inclusion of a contagion term to deal with spatial autocorrelation in models were assessed with projections made with GLM, as it was with this method that estimates of significance based on randomizations were obtained.

3. Spatial autocorrelation was shown to represent a serious problem for niche-based species' distribution models. Significance values were found to be inflated up to 90-fold.

4. In general, GAM and CTA performed better than GLM, although all three methods were vulnerable to the effects of spatial autocorrelation.

5. The procedures utilized to reduce the effects of spatial autocorrelation had varying degrees of success. Subsampling was partially effective in avoiding the

inflation effect, whereas the inclusion of a contagion term fully eliminated or even overcompensated for this effect. Direct estimation of probability using variable simulations was effective, yet seemed to show some residual spatial autocorrelation effects.

6. ***Synthesis and applications***. Given the expected inflation in the estimates of significance when analysing spatially autocorrelated variables, these need to be adjusted. The reliability and value of niche-based distribution models for management and other applied ecology purposes can be improved if certain techniques and procedures, such as the null model approach recommended in this study, are implemented during the model-building process.

Key-words

AUC, classification tree analysis, freshwater turtles, generalized additive models, generalized linear models, Iberian Peninsula, likelihood ratio statistics, model validation, toroidal shifts

3.2 Introduction

Niche-based models are familiar tools used to explain and predict species' spatial distributions (Walker 1990, Pereira & Itami 1991, Austin *et al.* 1996, Manel *et al.* 1999). Recently, niche models have received increased attention, in part because of the need to predict species' range shifts under future climate-change scenarios (Guisan & Theurillat 2000, Midgley *et al.* 2002, Peterson *et al.* 2002, Thomas *et al.* 2004, Thuiller *et al.* 2005, 2006, Araújo *et al.* 2006). However, there are a number of unresolved methodological issues requiring further enquiry. One example is the problem of non-independence between data used for calibration of models and that used for validation (Araújo *et al.* 2005a). Non-independence is often the result of using spatially autocorrelated data to calibrate and validate the models, and one of the consequences is that the perceived ability of models to make realistic predictions in space (Randin *et al.* 2006) and time (Araújo *et al.* 2005a) may be inflated. This problem may be greater than previously anticipated, as illustrated by studies showing high levels of intermodel variability in projections of species' range shifts under climate change scenarios (Thuiller 2004, Thuiller *et al.* 2004, Araújo *et al.* 2005c, Araújo *et al.* 2006). We have addressed the biases in model predictions that arise from using different procedures of model adjustment and validation under varying levels of spatial dependencies among predictor and response variables.

Sample size is a crucial parameter in the outcome of classical hypothesis testing as it determines the necessary degrees of freedom for pattern detection. In spatial analyses the simple count of sample units is not always an adequate estimator of effective sample size. For example, if the values of a variable depend on the distance between sample points, a set of closely spaced observations effectively provides less information than the same number of observations more widely separated in space. Such spatial dependency between values is termed spatial autocorrelation (SA; Cliff & Ord 1973) and its causes and consequences have been the focus of much research (Legendre 1993, Koening & Knops 1998, Lennon 2000, Dale & Fortin 2002, Fortin & Payette 2002, Legendre *et al.* 2002). SA leads to an overestimation of the effective sample size (leading to pseudoreplication), inflating the statistical significance of measured spatial relationships and consequently increasing the likelihood of type I errors (false positives). There is a serious possibility that previous analyses that used correlative approaches might be flawed because of 'red herrings' generated by SA (Lennon 2000), with both the estimated predictive power and the choice of variables being seriously biased. Because of the potential importance of such biases, many methods have been developed to help account for SA within models, including a priori procedures at the level of sampling design (Harrison 1997, Legendre *et al.* 2002, Legendre *et al.* 2004), modifications at the level of model adjustment (Keitt *et al.* 2002, Lichstein *et al.* 2002) and a posteriori procedures, such as the use of correction factors, to improve statistical accuracy of models (Dutilleul 1993, Legendre *et al.* 2002).

Niche-based models use several alternative techniques to summarize relationships between species occurrences and environmental variation (Guisan & Zimmermann 2000, Segurado & Araújo 2004b – Chapter 2), usually in the context of spatial and temporal predictions. Other designations for this family of models can be found in the literature, such as habitat models (Guisan & Zimmermann 2000), species distribution models (Olden *et al.* 2002), bioclimatic envelope models (Pearson & Dawson 2003) and presence/absence models (Fielding & Bell 1997). Although authors acknowledge the importance of SA, they often disregard or minimize the extent to which the presence of spatially autocorrelated data might affect the explanatory power and predictive accuracy of models (i.e. the ability in which models calibrated in one set represent observations in an independent set; for discussion see Araújo *et al.* 2005a, Randin *et al.* 2006). This could be a serious shortcoming in models as species' occurrences tend to be aggregated at most spatial scales, and the more aggregated species' occurrences are, the more likely it is that environmental variables will show some explanatory power simply because of the fact that environmental conditions tend to be more similar at neighbouring sites. Indeed, the strength of the correlation between variables has been shown to be increasingly more pronounced as SA grows stronger, whereas unbiased correlations are produced when at least one variable exhibits no SA (Lennon 2000).

Model generalization (e.g. variable selection in stepwise logistic regression, pruning of classification trees and stopping rules in artificial neural networks) are common procedures to avoid overadjustments to calibration data and are designed to increase the predictive power of models (Franklin 1998, Pearce & Ferrier 2000, Thuiller 2003). In the case of regression-based techniques this procedure implies that an assessment of the explanatory power of variables is made. The problem is that the inflation of explanatory power for spatially autocorrelated variables makes them, a priori, disproportionately likely to be selected in the final models. This is made at the expense of selecting potentially more important variables with lower SA. Therefore, variable selection procedures can be an additional source of bias in model fitting.

Patterns of species' distributions may be spatially autocorrelated because of contagious population dynamics and historical factors, but they may also be the result of spatial structure among environmental predictors (Storch *et al.* 2003).

In fact, species and the environment may share spatial structure because of the effect of spatially structured environmental predictors and non-environmental contingencies that may or may not be related amongst them (Borcard *et al.* 1992). If part of the spatial structure in the species' data is shared by the environmental data, knowing the relative weight of each item that contributes to the observed spatial structure is an important challenge when testing causal hypotheses to data (Borcard *et al.* 1992, Storch *et al.* 2003). A common procedure to cancel the effect of spatial structure of species' occurrences is to incorporate a term for SA into the analysis (Smith 1994, Augustin *et al.* 1996, Araújo & Williams 2000, Keitt *et al.* 2002, Segurado & Araújo 2004b – Chapter 2), usually a measure of contagion that encompasses the effect of spatial neighbourhood in the statistical test.

An alternative procedure to avoid pseudoreplication is to subsample the original species' distribution data, usually by adopting a systematic scheme that constrains observations to be spaced far enough from each other (Gates *et al.* 1994, Brito *et al.* 1999, Guisan & Theurillat 2000). This method has the disadvantage of not using all the available information and thereby artificially limiting sample size, a procedure that may have serious consequences for the predictive performance of models (Araújo *et al.* 2005c).

The effect of SA on correlation and linear regression significance values has been tested elsewhere using artificially generated variables with known spatial structures (Lennon 2000). In the context of niche-based modelling the extent to which SA in the response and predictor variables influences model performance is poorly known. In particular, the effect of SA in the validation data set used to estimate predictive power of models has never been assessed. In this study the effects of SA in species' distribution models were quantified using a null model approach with test variables generated from randomizations of spatially structured environmental data. First, the overall effect of SA on measures of model performance was assessed. Secondly, we explored whether different modelling techniques, differing in the response functions used, were equally sensitive to the effect of SA. We also quantified the effect of spatial autocorrelation in the predictive ability of covariates entering models of species' distributions using both resubstitution and data set partitioning (Olden *et al.* 2002, Araújo *et al.* 2005a). Finally, the effectiveness of different approaches to reduce undesirable effects of SA was evaluated.

3.3 Materials and methods

Data

We used the distributions of two species of freshwater turtles, the Mediterranean pond turtle *Mauremys leprosa* (Schweiger, 1812) and the European pond turtle *Emys orbicularis* (Linnaeus, 1758), in two equally sized rectangular areas on the Iberian Peninsula (Fig. 3.1). In each rectangular area, data on species' distributions were located in 66 × 21 universal transverse mercator (UTM) 10 × 10-km grid cells. The main criteria for delimitation of the two rectangular areas were to maximize (1) the geographical extent of the rectangles and (2) the geographical distance between them, in order to ensure the greatest feasible amount of information while retaining high levels of spatial independence between rectangles. Distribution data were compiled from four different sources: the updated atlas of the Portuguese herpetofauna (Godinho *et al.* 1999), the atlas and Red Data Book of the amphibian and reptiles of Spain (Keller & Andreu 2002, da Silva 2002), the UNIBA database (Alentejo's Biodiversity Database Unit; www.cea.uevora.pt/umc, 2006) and P. Segurado (unpublished data). The spatial structure of the two species' distributions in the study region is distinct: *E. orbicularis* is found in fewer grid cells and it is widely scattered compared with *M. leprosa*, which has more clumped distributions. There are also regional differences in occupancy patterns: in the western rectangle occur-rences of *M. leprosa* are more densely distributed and the occurrences of *E. orbicularis* are slightly more scattered than in the eastern rectangle. Simulated distributions of two species with distinctive spatial structures were also generated from the *M. leprosa* database creating a random distribution and a clumped distribution. The random distribution was obtained by randomly assigning the position of species' presences, while the clumped distribution was obtained by joining all species' presences into a single contiguous block. The species' distributions, as well as the simulated distributions, were used as response variables in the analyses.

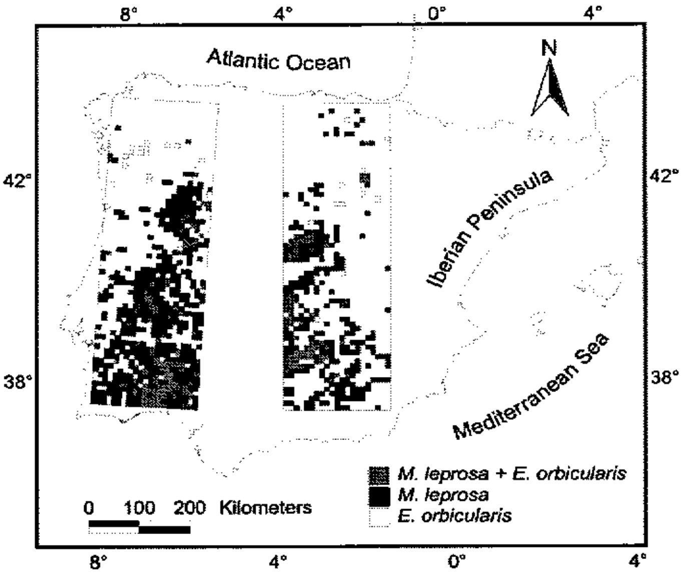


Fig. 3.1 - Location of the rectangular areas used in the analysis (10-km linear resolution).

Environmental data included climatic and topographic information (Table 3.1) resampled at the same grid resolution as the species' occurrence data. There is a danger that species might respond indirectly to topography, which would limit the models' predictive powers. However, topographic variables have the potential to summarize important surrogate predictor variables, such as habitat availability for freshwater turtles, that are not captured by the available variables. For example, both species show a preference for still and slow-moving water habitats and therefore their distributions might respond directly to slope. Climate data included 11 variables compiled from point data with 10-minute resolution (New *et al.* 2000). A randomly generated predictor variable showing negligible SA was also included in the analyses. The two rectangular areas did not differ considerably in their environmental range, although the western rectangle included a slightly wider gradient range for some variables.

Pattern generation

The effect of SA was evaluated by generating simulated patterns with known and fixed spatial structures. There are two main categories of such pattern generation: (1) fully synthetic patterns, generated purely from mathematical principles such as the method based on the inverse discrete Fourier transform (Lennon 2000); and (2) patterns generated from real data using restricted or sequential permutations of real patterns (Fortin & Jacquez 2000) or using more elaborate approaches such as the random patterns implemented by Roxburgh & Chesson (1998) and the patch model proposed by Watkins & Wilson (1992).

In this study we employed sequential permutations of environmental variables based on toroidal shifts (Palmer & Van der Maarel 1995, Fortin & Jacquez 2000, Dale & Fortin 2002, Fortin & Payette 2002, Storch *et al.* 2003) to generate patterns from the original environmental variables. With this randomization technique, coordinates of the original variable are moved by a common random factor in every geographical direction; cells that are shifted beyond one side of the range of real coordinates are moved to the opposite side of the range. This randomization assures that the main spatial structure is maintained. This technique is more straightforward, easier to implement and computationally more efficient than alternative methods. Its main drawback is that it can create unrealistic environmental patterns with abrupt orthogonal lines originated by the shifted edges. We assumed that this feature would not have an effect in the analysis because the existence of linear edges in species' distributions is unlikely. As toroidal shifts can be too liberal (Fortin & Jacquez 2000), an image reflection and a 180-degree rotation were initially performed for each variable. This procedure ensured that even toroids involving only small shifts would differ substantially from the original pattern, thus making the pattern simulation more conservative.

Moran's I statistics was used to estimate general patterns of spatial dependency of variables. In order to evaluate the degree to which the original spatial structure was maintained after randomizations, 1000 toroidal shifts were run and for each permutation the Moran's I was calculated. The distribution of the resulting values was compared with the Moran's I-value of the original variables.

Assessing the effect of spatial autocorrelation

Three modelling techniques, differing in their ability to model complex response shapes, were used to relate species' distributions to each of the 1000 toroidal shifts of the environmental variables: (1) generalized linear models, (2) generalized additive models and (3) classification tree analysis. Generalized linear models (GLM; McCullagh & Nelder 1983) are generalizations of the classical linear regression allowing error distributions other than the normal distribution; here a binomial error distribution was assumed (logistic regression). Generalized additive models (GAM; Hastie & Tibshirani 1990) are semi-parametric forms of GLM that use smooth functions instead of the usual regression coefficients. GAM were fitted using cubic splines as the smooth function and assuming a binomial error distribution. Classification tree analysis (CTA; Clark & Pregibon 1992) is a non-parametric technique that is based on recursive partitions of the dimensional space defined by the predictor variables into groups that are as homogeneous as possible for the response variable. We used a recursive algorithm that successively splits the data into binary branches by choosing the splits that cause the maximum reduction of the residual deviance.

For each permutation, the model classification accuracy was measured by calculating the receiver operational characteristic (ROC) curve and summing the area under that curve (AUC; Fielding & Bell 1997). The AUC assesses whether model predictions differ from that expected by chance, varying from 0·5 (random classification) to 1 (perfect classification). GLM performance was also measured using the likelihood ratio test statistics (LRS; Hosmer & Lemeshow 1989), which correspond to the reduction of model residual deviance in relation to null model deviance.

Models were calibrated on both rectangles and model accuracy was measured using the whole calibration set (i.e. resubstitution). Accuracy (AUC) of GLM was also measured by application to the second rectangular area, which was interpreted as providing an independent validation. Explanatory power was expressed by meas-ures of model accuracy using the calibration set, while predictive power was expressed by measures of model accuracy using the validation set.

The overall effect of SA was assessed exploring the relationship between the 95th percentile and the standard deviation of measures of accuracy for each null pattern set with Moran's I of the original predictor variables. The variation of the 95th percentile was analysed because it represents a common threshold in most statistical hypothesis testing. This parameter is expected to be inflated by SA.

In GLM, the LRS is assumed to be chi-square distributed and therefore it is also possible, for each run of 1000 permutations, to calculate the number of LRS tests on the calibration set that are found to have values above the expected number according to a chosen type I significance level. The ratio between this number and the expected number of significant tests according to the significance level (e.g. 50 out of 1000 for P = 0.05) expresses the type I inflation ratio (Lennon 2000). Here, a significance level of 0.01 was considered in order to compare with other results found in the literature.

All data analyses were performed with S-PLUS 2000 (Statistical Sciences 1999) using the default functions for model adjustments. Random toroidal shifts were performed using a modified function from the Splancs library of S-PLUS (Rowlingson & Diggle 1993).

Table 3.1 – Variables included in the analysis, Moran's I values for the original variables and mean Moran's I values for the 1000 toroidal permutations (standard deviation values inside brackets).

Variable	Eastern area		Western area	
	Original Moran's I	Mean Moran's I	Original Moran's I	Mean Moran's I
Response variables:				
Mauremys leprosa	0.416	-	0.543	-
Emys orbicularis	0.386	-	0.264	-
Species with clumped distribution	0.959	-	0.975	-
Species with a random distribution	-0.016	-	0.009	-
Predictor variables:				
Altitude – mean (Alt)	0.817	0.764 (0.012)	0.882	0.819 (0.019)
Altitude – minimum (Altmin)	0.843	0.796 (0.013)	0.847	0.775 (0.024)
Altitude – maximum (Altmax)	0.784	0.739 (0.011)	0.889	0.837 (0.015)
Slope – mean (Slope)	0.733	0.729 (0.011)	0.832	0.787 (0.012)
Hillshade – mean (Hill)	0.292	0.284 (0.007)	0.181	0.171 (0.007)
Hillshade – minimum (Hillmin)	0.657	0.658 (0.008)	0.752	0.716 (0.013)
Hillshade – maximum (Hillmin)	0.626	0.623 (0.006)	0.716	0.681 (0.012)
Mean annual temperature (Tann)	0.973	0.943 (0.016)	0.981	0.937 (0.013)
Mean temperature of the coldest month (Mtc)	0.945	0.929 (0.014)	0.971	0.909 (0.015)
Mean temperature of the warmer month (Mtw)	0.987	0.952 (0.016)	0.981	0.943 (0.014)
Mean annual growing degree days (Gdd)	0.975	0.944 (0.017)	0.983	0.938 (0.012)
Mean annual global net radiation (Rann)	0.957	0.906 (0.022)	0.958	0.861 (0.016)
Mean annual evapotranspiration / potential evapotranspiration (A2P)	0.950	0.915 (0.017)	0.975	0.914 (0.011)
Mean annual precipitation sum (Pann)	0.901	0.877 (0.023)	0.968	0.908 (0.023)
Mean winter precipitation sum (Pwin)	0.895	0.890 (0.021)	0.966	0.898 (0.025)
Mean summer precipitation sum (Psum)	0.924	0.879 (0.020)	0.971	0.932 (0.016)
Mean spring precipitation sum (Pspr)	0.916	0.872 (0.018)	0.966	0.914 (0.020)
Mean autumn precipitation sum (Paut)	0.888	0.879 (0.021)	0.968	0.900 (0.024)
Null environment (Null)	0.001	0.000 (0.003)	-0.018	-0.014 0.003)

Dealing with spatial autocorrelation

Two methods that address the effects of SA were compared using toroidal shifts and the procedures described above. The first method consisted in subsampling the original data set by eliminating cells in a systematic manner. All cells with even coordinates were eliminated from the original data sets. The second method included the incorporation of an autocovariate term accounting for the SA of observations. Autologistic models use contagion as an autocovariate term in the logistic regression equation. The measure of contagion was based on a two-order neighbourhood as the weighted average of the number of occupied grid cells among a set of k_a neighbours of a central grid cell y_a, so that:

$$Contagion = \left(\frac{\sum_{b=1}^{k_a} w_{ab} y_b}{\sum_{b=1}^{k_a} w_{ab}} \right) \qquad \text{eqn 1}$$

where the weight given to the grid-cell y_b is $w_{ab} = 1/d_{ab}$, and d_{ab} is the distance between grid-cells y_a and y_b. Two orders of neighbours, assigning a weight of d=1 to the first-order and a weight of d=2 to the second-order neighbours, were used. Neighbours in the first order were the 8 adjacent cells touching the central cell along the edges and at the corners within a rectangular grid. The second order neighbours were the next group of cells concentric to the first order with 16 grid cells.

These two methods were assessed for their ability to produce unbiased estimates of the explanatory power as measured by LRS. The contribution of contagion terms was removed from the LRS estimates, in order to assess exclusively the contribution of each variable to the explanatory power. This assessment was based on the distribution of *M. leprosa* only. The predictive power of models was not assessed using the autologistic approach because the spatial autocovariate term is a function of the calibration data and cannot be predicted for independent validation data.

Finally, corrected univariate variable significances were estimated for distribution models of *M. leprosa* and *E. orbicularis*. A Monte Carlo simulation approach was used to compare the AUC computed using the original predictor variables with the test statistics generated from 1000 toroidal shifts. Significance (probability of rejecting a true null hypothesis) was defined as the fraction of the AUC statistics of the 1000 simulated variables that fell above the AUC statistics of the original variable.

3.4 Results

Spatial autocorrelation of variables

Spatial dependency is stronger for *M. leprosa* than for *E. orbicularis*, as shown by the overall SA of variables measured by the Moran's I index (Table 3.1). In the western rectangle the SA of *M. leprosa* occurrence is more pronounced while that of *E. orbicularis* is less pronounced in the western than in the eastern rectangle. Among environmental variables, the topographical descriptors are generally less spatially autocorrelated than climatic variables. Hill shade variables have the lowest Moran's I-values. For most environmental variables, with the exception for mean hill shade and mean temperature of the warmest month, the western rectangle shows slightly higher SA values than the eastern rectangle.

The mean Moran's I-values for each set of 1000 toroidal shifts of the environmental variables tend to be slightly lower than the original Moran's I, probably because of the effect of the shifted edges, which can slightly disturb the spatial structure. However, toroidal displacement did not shift significantly the relative order of SA between variables. On the other hand, the variability of values among the 1000 simulations, as measured by the standard deviation, was reasonably low, representing a small fraction of the mean values (SD/mean ranging from 0.010 to 0.042), which means that spatial structure was maintained among permutations (Table 3.1).

Effect on the explanatory power of variables

Amongst sets of 1000 toroidal permutations the 95th percentile of the model accuracy measures increased as the SA of the predictor variables increased (Fig. 3.2). This effect was more pronounced for species' distributions that were more spatially autocorrelated and was eliminated in a simulated random species distribution. The 95th percentile of the AUC distribution of the 1000 toroidal shifts per variable increased in an approximately linear fashion with SA for each modeling approach (Fig. 3.2). In the western section of the studied rectangle the variation of AUC values with SA of predictor variables was more pronounced than in the eastern rectangle for *M. leprosa* and less pronounced for *E. orbicularis* (Fig. 3.2).

There were slight differences on the effect of SA for GAM and CTA compared with GLM, as shown by the variation of the 95th percentile of AUC values with Moran's I of the environmental variables (Fig. 3.2). The variation of AUC 95th percentiles was less marked for GAM and CTA than for GLM models; this was particularly evident for species with more autocorrelated distributions. This weaker variation was because of the better performances of GAM and CTA compared with GLM models for species' distributions and predictor variables with reduced SA while showing comparable performances for those with higher SA (Fig. 3.2). CTA tended to perform better than GAM, especially for the less autocorrelated distributions. These results were consistent between rectangles.

The 95th percentile of the LRS distribution of GLM models using variable permutations was more sensitive than AUC to the increase of Moran's I. This parameter tended to increase in an exponential fashion with SA (Fig. 3.3).

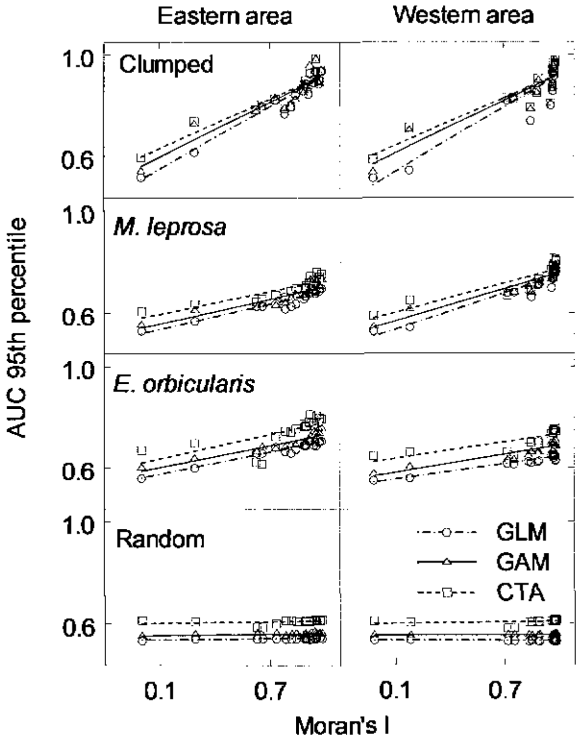


Fig. 3.2 - Variation of the 95th percentile of AUC with Moran's *I* of environmental variables using three modelling techniques (GLM, GAM and CTA), four distributions (*M. leprosa*, *E. orbicularis*, a simulated distribution with totally clumped occurrences and a random distribution) and 1000 simulated surfaces (lines represent linear fits).

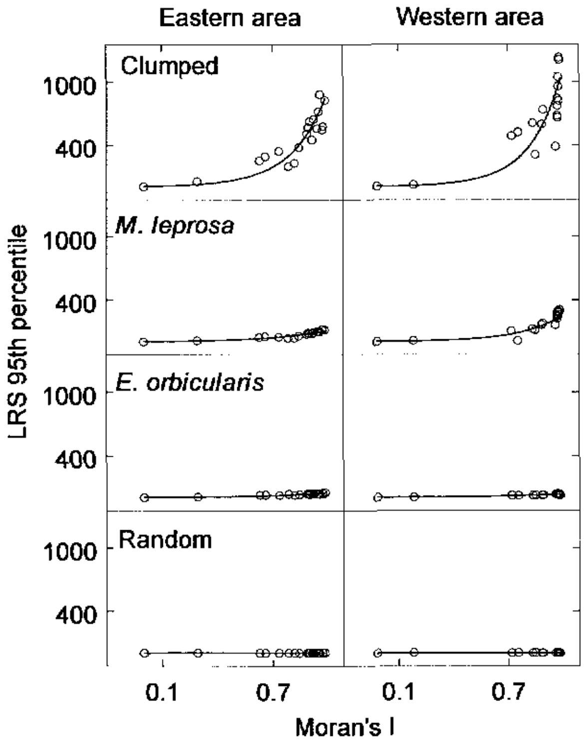


Fig. 3.3 - Relationship between spatial autocorrelation and the 95th percentiles of LRS statistics for GLM, regressing each simulated surface with presence/absence data of *M. leprosa*, *E. orbicularis*, a simulated distribution with totally clumped occurrences and a random distribution.

Effect on the predictive power of variables

The effect of the increase in Moran's I using an independent validation data set to evaluate the predictive power of GLM models was as strong as using the calibration set. The 95th percentile of deviance also tended to increase with SA in a linear form (Fig. 3.4). This trend was more marked for models of *M. leprosa* calibrated in the eastern rectangle and for models of *E. orbicularis* calibrated in the western rectangle.

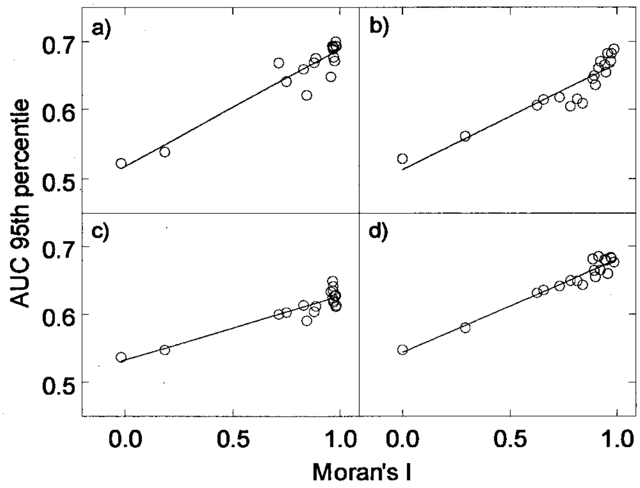


Fig. 3.4 - Model validation: relationship between Moran's I for the validation data set and the 95th percentiles of *M. leprosa* model's AUC, for each set of 1000 permutations of the original variables. Lines represent linear fits. (a) *M. leprosa*, eastern rectangle model validated with the western rectangle data set; (b) *M. leprosa*, western rectangle model validated with the eastern rectangle data set; (c) *E. orbicularis*, eastern rectangle model validated with the western rectangle data set; (d) *E. orbicularis*, western rectangle model validated with the eastern rectangle data set.

Inflation ratios

Inflation ratio increased with Moran's I-values (Fig. 3.5) of the environmental variables, especially at lower Moran's I-values. Thus there was a substantial inflation of the apparent predictive power of analyses using variables with even a modest amount of SA. Predictor variable significance could be inflated by a factor up to 90 for the clumped distribution. At higher Moran's Ivalues, inflation ratios tended to stabilize (Fig. 3.5). Inflation ratios also increased where species' distributions displayed increased SA; indeed, in the absence of SA in the distribution (Fig. 3.5, Random distributions) even highly autocorrelated environments did not cause inflated predictive power estimates. Overall, the higher the level of SA in both species' distributions and environmental variables used, the higher the inflation ratios. Consequently, it was unsurprising to find that inflation ratios for *M. leprosa* were more pronounced in the western rectangle (mean inflation ratio of 75.3 vs. 60.2 for the eastern rectangle) while those for *E. orbicularis* were more pronounced in the eastern rectangle (mean inflation ratio of 40.4 vs. 38.3 for the eastern rectangle). The observation that random distributions did not suffer variations in the inflation ratio supported the interpretation that it was SA rather than alternative unmeasured factors that caused inflation of the significance estimates in the models.

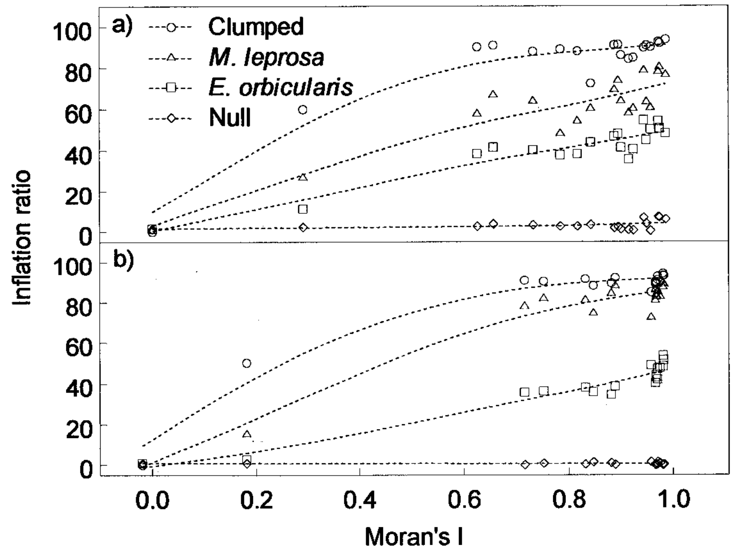


Fig. 3.5 - Variation of the inflation ratio (number of times a significant test was found in relation to the expected number of significant tests, according to the significance level adopted, in this case $P < 0{\cdot}01$, with the Moran's I-values of the original variables). Dashed lines are spline fits. (a) Eastern rectangle; (b) western rectangle.

Dealing with spatial autocorrelation

Comparison of two methods for dealing with SA within models showed that including an autocovariate term in the regression was more effective than systematically subsampling the area (Fig. 3.6). Indeed, the autologistic procedure may slightly overcompensate. For example, in the eastern rectangle there was a slight decrease in the location measure of the LRS test significances for an increase of Moran's I (Fig. 3.6a). The subsampling procedure was only partially effective in avoiding the inflation effect of model performance because of SA. The increase of LRS test significance with an increase of Moran's I was less pronounced using this procedure (Fig. 3.6; note that the axes have different ranges).

When surface permutations were used to produce Monte Carlo significance values of AUC accounting for SA, no obvious relationship with SA was observed, which suggested that unbiased estimates were produced. However, the most significant variables had consistently higher Moran's I-values, especially for *M. leprosa* (Table 3.2). Variables tended to have greater explanatory power for *M. leprosa* than for *E. orbicularis* and in the western rectangle there were more variables with significant ($P < 0.05$) effects (e.g. for GLM 16 significant variables for *M. leprosa* and 13 for *E. orbicularis*) than there were in the eastern rectangle (11 significant variables for *M. leprosa* and two for *E. orbicularis*). The majority of significant variables for *M. leprosa* distribution were common to the three modelling approaches. In the eastern rectangle nine significant variables were common and two uncommon, while in the western rectangle 15 significant variables were common and three uncommon. For *M. leprosa* seven significant variables were simultaneously common to both rectangles and the three modelling approaches. For *E. orbicularis* in the eastern rectangle, only one significant variable was common to the three modelling approaches and three variables were uncommon, while in the western rectangle five significant variables were common and eight uncommon. No significant variables were simultaneously common to both rectangles and the three modelling approaches for *E. orbicularis*.

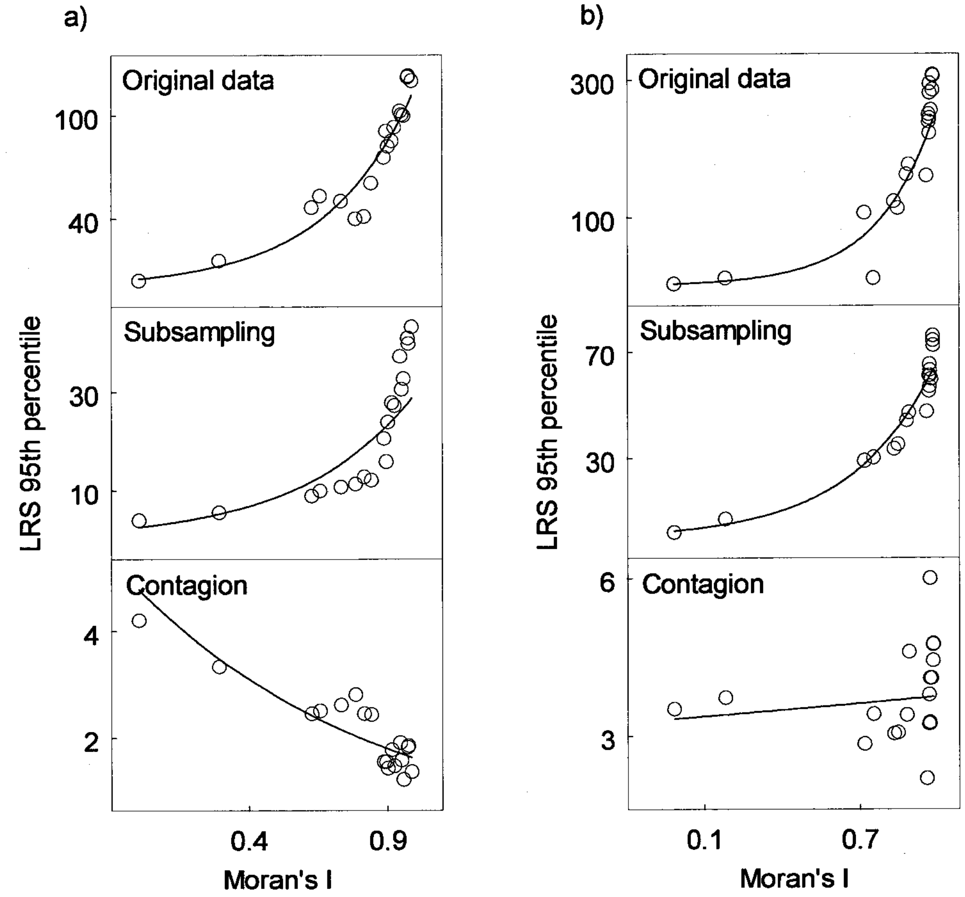


Fig. 3.6 – Relationship between spatial autocorrelation and the 95th percentiles of the LRS tests generated by regressing each simulated surface with presence/absence data of *M. leprosa*. Comparison between results using the original *M. leprosa* data on presence/absence and two different procedures usually adopted to avoid effects spatial autocorrelation: subsampling of the original dataset and forcing the inclusion of a contagion term in the model. Lines represent exponential fits. a) Eastern rectangle; b) Western rectangle.

Table 3.2 - Significance of each predictor variable using a Monte Carlo simulation approach (*P < 0·05; **P < 0·01). Variables are displayed in a increasing order of autocorrelation as measured by Moran's.

	Mauremys leprosa			*Emys orbicularis*		
Variables	GLM	GAM	TREE	GLM	GAM	TREE
a) Eastern rectangle						
Noise	0.753	0.596	0.936	0.582	0.986	0.977
Hill	0.013*	0.201	0.283	0.004**	0.011*	0.044*
Hillmax	0.717	0.928	0.937	0.283	0.418	0.028*
Hillmin	0.959	0.869	0.922	0.997	0.369	0.003**
Slope	0.988	0.852	0.795	0.613	0.412	0.552
Altmax	0.153	0.313	0.319	0.454	0.773	0.847
Altmed	0.017*	0.059	0.077	0.160	0.419	0.548
Altmin	0.018*	<0.001	0.012*	0.164	0.137	0.389
Paut	0.957	0.357	0.351	0.982	0.157	0.072
Pwin	0.940	0.412	0.557	0.988	0.122	0.273
Pann	0.185	0.418	0.366	0.724	0.698	0.791
Pspr	<0.001	<0.001	<0.001	0.134	0.320	0.440
Psum	<0.001	<0.001	<0.001	0.077	0.289	0.358
Mtc	<0.001	<0.001	<0.001	0.107	0.368	0.638
A2p	0.005**	0.018*	0.048*	0.235	0.678	0.436
Rann	<0.001	<0.001	0.008*	0.140	0.790	0.460
Tann	<0.001	0.003**	<0.001	0.091	0.173	0.340
Gdd	<0.001	<0.001	<0.001	0.090	0.149	0.289
Mtw	<0.001	0.001**	0.013*	0.042*	0.191	0.421
b) Western rectangle						
Noise	0.426	0.383	0.204	0.882	0.982	0.982
Hill	0.013*	0.022*	0.049*	0.410	0.091	0.551
Hillmax	0.036*	0.033*	0.023*	0.044*	0.035*	0.242
Hillmin	0.013*	0.008**	0.008**	0.004**	0.021*	0.019*
Slope	0.015*	0.020*	0.041*	0.034*	0.070	0.033*
Altmin	0.109	0.050	0.048*	0.551	0.840	0.857
Altmed	0.068	0.041*	0.062	0.285	0.333	0.289
Altmax	0.028*	0.017*	0.009**	0.146	0.159	0.132
Rann	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Pwin	0.001**	<0.001	<0.001	0.011*	0.006**	0.018*
Pspr	0.002**	0.002**	0.010*	0.008**	0.024*	0.010*
Paut	0.009**	<0.001	<0.001	0.040*	0.027*	0.037*
Pann	0.007**	0.007**	<0.001	0.012*	0.053	0.113
Psum	0.005**	0.006**	0.002**	0.020*	0.231	0.289
Mtc	0.045*	0.110	0.078	0.252	0.459	0.568
A2p	0.005**	0.001**	<0.001	0.006**	0.071	0.038*
Tann	0.006**	0.004**	<0.001	0.012*	0.033*	0.078
Mtw	<0.001	<0.001	<0.001	<0.001	0.060	0.001**
Gdd	0.002**	0.001**	0.012*	0.006**	0.029*	0.169

3.5 Discussion

In this study we used a simple and straightforward method of pattern simulation to explore the effect of SA on niche-based models. The results reinforce the idea that conclusions from niche-based models could be compromised because of the autocorrelated nature of both predictor variables and species' occurrences (Lennon 2000, Hampe 2004).

When plotted against a measure of SA (Moran's I), model performance showed clear trends. For example, we found that significant log-likelihood ratio tests in GLM need to be exponentially larger as SA of predictor variables are stronger. This means that even slight changes in the degree of SA have a strong effect on the probability of a predictor variable being chosen as significant. The same trend was observed with measures of models' predictive accuracy (here AUC), although this measure was less sensitive to SA on both the response and predictor variables.

Results also suggest that sensitivities of model accuracy (AUC) to SA using either validation or calibration data sets are nearly identical. Model performances are commonly measured using either resubstitution or by splitting data into calibration and validation sets (Olden *et al.* 2002, Araújo *et al.* 2005a). The last approach does not fully avoid an overestimation of model accuracy because two sources of bias may arise: (1) SA between the calibration and validation data sets and (2) SA within data sets. In fact, even when using more complex approaches such as cross-validation, bootstrapping and jack-knifing (Guisan & Zimmermann 2000), most procedures involve splitting the original data set into calibration and validation subsets using random assignments. These two subsets have inevitably some degree of non-independence because of SA (Araújo *et al.* 2005a). In this study we used two data sets that were spatially separated in order to avoid an overestimation of accuracy because of a lack of independence between data sets. This allowed an assessment of the role of SA within each data set when predicting distributions in other regions, i.e. its effect on model transferability (Randin *et al.* 2006). Despite differences in the environments of two rectangles, results suggest that the models' transferability depends greatly on the SA of the environmental variables.

As demonstrated before with correlation and linear regression analysis (Lennon 2000), SA of covariates inflates their statistical significance in predictive models. In the present study, the maximum inflation ratio was 94 for the clumped distribution, i.e. variables were found to be significant 94 times more often than expected by chance. Inflation ratios reached 89 for *M. leprosa* and 54 for *E. orbicularis*. These inflation values are similar to those estimated by Lennon (2000) using synthetically generated spatial patterns.

The two a priori procedures used to minimize the effect of SA partially cancelled (systematic subsampling) or fully cancelled (inclusion of the contagion term) the effect of SA. However, it should be noted that testing for relationships between environmental variables and species' occurrences after forcing the inclusion of a contagion term might represent a problem. This procedure cancels the SA of species' distributions without differentiating environmental from demographic/historical contributions. By cancelling the environmental contribution to the SA, models tend to underestimate the importance of environmental variables that co-vary with species' occurrences (Araújo & Williams 2000). Variables that enter in models with an autocovariate term are most likely to explain the non-autocorrelated aspects of patterns, which may represent only a minor fraction of total variance. Moreover, such a model would most probably lack predictive power outside the calibration set or under future climate-change scenarios. Indeed, models that incorporate a term for SA cannot be extrapolated to regions where no occurrence data are available. Hence, the inclusion of an autocovariate term is only possible when projections are made within the geographical range of the calibration data set.

When sample size is not a limiting factor, Subsampling of the original data matrix is a possibility. Even though it does not completely eliminate the inflation effect, it reduces it substantially. This can be done by simply selecting samples in a systematic manner as done here, or by using geostatistical tools such as variogram and correlogram plots (Maurer 1994, Diniz-Filho *et al.* 2003), to analyse the overall pattern of spatial dependency and to help establish a minimum distance between samples that will reduce SA at a given amount (Catry *et al.* 2003). None the less, as populations and environmental variables tend to be autocorrelated at all scales, it seems likely that the spacing out of samples will never fully eliminate SA effects.

Another type of procedure is to account a posteriori for SA using a Monte Carlo approach, such as that used here, to estimate variable significance based on null spatial patterns. However, even when P-values are estimated using this approach, variables that exhibit higher SA tend to be more significant. Nevertheless, these are likely to be the better candidates as actual causative factors of species' distributions; they are more likely to contribute to the autocorrelated nature of species' patterns of occurrence.

The use of semi-parametric modelling techniques, such as GAM, and non-parametric techniques, such as classification trees, slightly reduces the effect of SA in our analyses. These techniques place fewer constraints on the shape of species' responses to their environments, and stronger adjustments to predictions are consequently produced in comparison with parametric techniques, such as GLM. Our results suggest that this is particularly evident for response or environmental variables that display only moderate levels of SA. Thus these techniques are slightly less sensitive to variations in the variables' patterns of SA than traditional parametric approaches such as GLM.

The results described above, and thus the recommendations given for dealing with SA, apply only to univariate modelling. When there is more than one candidate variable to explain a species' distribution, the assessment of the effect of SA on models requires a more complex and thorough approach. Indeed, the impact of SA on multivariate model building is an urgent issue for future investigation. The first step of multivariate modelling is variable selection. Methods based on stepwise variable selection are still widely used by ecologists although it has been demonstrated that automated model building procedures can result in the selection of a subset of predictor variables with no direct effect on the response variable (Derksen & Keselman 1992). Nevertheless, even if SA inflates variable significance this does not mean that the final model configuration will exclusively include the most autocorrelated variables. If an autocorrelated variable is included in the model, it may explain a substantial fraction of the SA in the species' pattern of occurrence. If that is the case the remaining variables may then be related to the less autocorrelated aspects of the distribution, probably driven by factors that act at finer spatial scales (Diniz-Filho *et al.* 2003).

Some recommendations for the variable selection process can, however, be drawn from our results. Variable selection based on simulated patterns is not cost effective. We recommend, instead, before model adjustment running univariate tests based on null patterns such as the procedure described in this study. Strong autocorrelated variables that loose explanatory power when their significance is adjusted for SA should be handled with special care because they could inflate the models' significance. An alternative would be simply to exclude such variables from the analysis. Employing methods based on the recent information–theoretic paradigm (Burnham & Anderson 2002, but see Stephens *et al.* 2005) during model building, such as variable selection procedures using the Akaike information criterion (Akaike 1974), is always preferable because it does not fully rely on significance thresholds. For modelling species' distributions that show a high degree of SA it might be adequate to use semi-parametric or non-parametric techniques, as well as avoiding the use of statistics such as LRS, which are bound to be more sensitive to SA than other accuracy measures (e.g. AUC).

The value of niche-based distribution models for planning and management purposes greatly depends on their ability to overcome different sources of biases that are inherent with biological data. SA is just one among other sources of bias, yet probably the most challenging one. It is our belief that simple procedures such as the ones discussed in this study would help to enhance ecological reliability of models and therefore to increase their applied value during the decision-making process.

Acknowledgements

We thank Stephen Roxburgh, Ana Bio and Wilfried Thuiller for their comments on the manuscript. Pedro Segurado was funded by FCT (SFRH/BD/8493/2002). This research is partly funded by the EC Integrated FP6 ALARM (GOCE-CT-2003-506675) project.

Chapter 4

The effect of spatial scale on patterns of species co-occurrence

GRUPO DESPORTIVO DE SÃO MANÇOS

Illustration of a freshwater turtle in the insignia of the sport club of São Manços, Évora, Portugal

Note – The work presented in this chapter is a paper that is still in preparation under the title "Patterns of coexistence of two Iberian freshwater turtles at varying spatial scales" (Segurado, P., Kunin, W.E., Filipe, A.F. & Araújo, M.B. In prep.).

4.2 Summary

Inferring biotic interactions from the examination of patterns of species occurrences has been a central issue of much community ecology research, and recently it has gained interest in the context of single-species modelling as well. However, there are a number of methodological problems that still limit the reliability of many of such inferences. For example researchers seldom consider in their analysis the spatial extent and the grain size at which their studies are carried out. Here, we test the spatial and environmental association between two closely related species of freshwater turtles in the Iberian Peninsula using five datasets compiled over different geographical extents (from the whole Iberian Peninsula to a stream reach) and at different spatial resolutions (from 50 km grid cells to microhabitat features within a stream). The main goals were to assess whether biotic interactions can be inferred from broad scale spatial patterns of coexistence and whether coexistence recorded at broader spatial scales translate into coexistence at finer spatial scales. Results showed a shift from positive association between the two species at broader scales to negative association at finer scales. We demonstrate the need to factor in spatial scale when investigating biotic interactions from the examination of spatial patterns of occurrence.

4.2 Introduction

Several studies have attempted to infer biotic interactions from the analysis of spatial patterns using datasets compiled at broad spatial scales (Letcher *et al.* 1994, Bullock *et al.* 2000, Anderson *et al.* 2002, Mönkkönen *et al.* 2004, Sfenthourakis *et al.* 2005, Apps *et al.* 2006, Cadena & Loidelle 2007). Analyses usually seek to infer competitive exclusion or competitive release from occurrence data. If two species show an extensive overlap in the use of resources and simultaneously show complementary geographical distributions it is assumed that competitive exclusion might have an influential role governing their distributions (Sfenthourakis *et al.* 2005). Species are therefore said to occupy only a portion of their fundamental niche, i.e., the realized niche (Hutchinson 1957, but see Pulliam 2000) and a possible consequence is that weaker competitors will occupy a broader set of conditions in the absence of stronger competitors (competitive release). Geographical patterns coupled with niche modelling approaches can also provide clues about the existence of competitive release if there is evidence that species occupy a wider range of niches in the absence of their competitor than in its presence (Anderson *et al.* 2002). In this paper we analyse the patterns of coexistence between two Iberian freshwater turtles and measure the sensitivity of the inferences made to the spatial extent and resolution of the data.

The study of the mechanisms controlling species occurrence has been a central issue of ecological modelling. Three contrasting mechanisms have divided ecologists: those that perceive species occurrences resulting mainly from single-species responses to environmental variables (niche or habitat modellers; e.g. Guisan & Zimmerman 2000), those that are concerned with testing whether patterns of occurrence among multiple species are the result of biotic interactions (community ecologists; e.g. Gotelli & Graves 1996) and those that are focused on the importance of demographic and dispersal factors (meta-population ecologists; e.g. Hanski & Gilpin 1997). The first two groups share the common goal of inferring causal processes from the analysis of spatial patterns that represent a snap-shot in time. They both focus on particular aspects of local population success that affect distributions indirectly through their impacts on demography and dispersal, which ultimately govern species occurrence. Recently there has been concern among niche modellers that biotic interaction should be included within models of species responses to climate in order to increase their realism (Guisan & Thuiller 2005, Araújo & Guisan 2006). The simplest procedure is to include the distribution of interacting species as a predictor variable in niche models (e.g. Leathwick & Austin 2001, Gutiérrez *et al.* 2005). Moreover, probabilities of occurrence from niche models have been proposed as additional constraints in null models' generation used by community ecologists to infer biotic interactions (Perez-Neto *et al.* 2001).

The analysis of spatial pattern is not sufficient to infer causal relationships and therefore it must be regarded as providing conditional support for hypothesis testing rather than providing unequivocal support about the processes that generate the patterns (e.g. Connor & Bowers 1987, Roxburgh & Chesson 1998, Anderson *et al.* 2002). Furthermore, demonstrating that biotic interactions have spatial consequences becomes more difficult at broader and coarser spatial scales, since other factors such as spatial heterogeneity within the geographical unit may allow pairs of competing species to coexist (e.g. Connor & Bowers 1987; Araújo & Guisan 2006).

On the other hand there are a number of uncertainties associated to the analyses of spatial patterns that must be taken into account when performing such analyses. The issue of how to accurately test spatial association between species has been a matter of debate (Weiher & Keddy 1995, Sanderson *et al.* 1998, Gotelli 2000). Species co-occurrence is often analysed using null pattern-generating models where a mechanism of interest is excluded (Sanderson 2000; Gotelli 2001). However the significance of the observed patterns greatly depends on the assumptions underlying such null models. Models of coexistence among species are particularly prone to type I errors (significant test in the absence of a real pattern) in particular when observations show a strong spatial autocorrelation (Palmer & van der Maarel 1995, Roxburgh & Chesson 1998). Even the adoption of different techniques to produce spatial structured null models can greatly affect the resulting p-values and even the sign of spatial association (Roxburgh & Matsuki 1999).

Scale is also an important issue that is seldom considered in the analysis of co-occurrence patterns (Bullock *et al.* 2000; Araújo & Guisan 2006). Changes in the scale of processes can shift the relative importance of different mechanisms of spatial coexistence (Snyder & Chesson 2004). Three main categories of mechanisms are responsible for patterns of coexistence among pairs of species: competition, habitat requirements and historical factors (Sfenthourakis *et al.* 2005). However these factors can show different relative impacts at distinct spatial scales; for example competition may shape co-occurrence patterns at finer scales, although a positive covariation between species occurrences at broader scales may occur due to common historical or climatic constraints. On the other hand there is evidence that the interaction between pairs of species with wide-ranging overlapping distributions can switch from competition to facilitation along an environmental gradient (Mönkkönen *et al.* 2004); these results show that the geographical extent in the analysis can be decisive for the detection of biotic interaction patterns. Although there is a generally accepted idea that competition or facilitation have less of an effect on species occurrences at broader geographical extents and coarser resolutions (Bullock *et al.* 2000, Huston 2002; Pearson & Dawson 2003), this inference still lacks robust quantitative testing (Guisan & Thuiller 2005).

In the present paper we investigate the patterns of co-occurrence of two freshwater turtle species, the European pond turtle, *Emys orbicularis*, and the Mediterranean pond turtle, *Mauremys leprosa*, considering multiple spatial scales within the Iberian Peninsula. The geographical range of the two species overlaps widely in the Iberian Peninsula. In some regions of northern Iberia *E. orbicularis* extends its range beyond that of *M. leprosa* (Ayres & Cordero 2001, Keller & Andreu 2002) while being absent from vast regions of southern Spain where *M. leprosa* also occurs (da Silva 2002). Whilst *E. orbicularis* is distributed throughout Europe, *M. leprosa* is restricted to Iberia and North Africa. The wide zone of coexistence in the Iberian Peninsula can be partially explained by similarities between the historical constraints faced by the two populations, such as their retreat to similar glacial refuges and similar subsequent expansion patterns that occurred in the past (Fritz *et al.* 2006). On the other hand species show general similar responses to climatic and habitat factors, although *M. leprosa* occupies a wider range of available habitats (Segurado & Araújo 2004a - chapter 5, section 5.2).

Despite wide overlap in the geographic ranges of the two species, some evidence suggests that biotic interactions may occur between the two focal species at finer spatial scales. In habitats where the species co-occur, the abundance of *E. orbicularis* is very low and tends to show skewed age structures with a predominance of adults (Segurado & Araújo in review). Indeed, *E. orbicularis* is much rarer than *M. leprosa*, both in terms of its local abundance and site occupancy, being absent from many apparently favourable habitats in the Iberian Peninsula (Keller & Andreu 2002, Segurado & Araújo 2004a - chapter 5, section 5.2).

We begin by testing the hypothesis that complementary distributions between species become increasingly more marked when moving from broad to fine spatial scales. This is accomplished in two ways: testing the pure spatial association between species and testing whether the two species have similar responses to environmental variables, using different sets of data taken from distinct spatial extents and resolutions. We then analyse species coexistence throughout the geographical range of the two species in the Iberian Peninsula. We explore whether the spatial patterns of coexistence and exclusive occupancy of each species are related to any particular environmental gradient. Finally, we discuss some hypotheses for the mechanisms generating patterns of coexistence of the two freshwater turtles at the different spatial scales.

4.3 Materials and methods

Data

Distribution data for *Emys orbicularis* and *Mauremys leprosa* was compiled from five different datasets, varying in their spatial extent, spatial resolution and information content (Fig. 4.1). Detailed information on each dataset is listed in Table 4.1. The two broader scale datasets include information for the whole Iberian Peninsula and are based on the UTM grid system considering two resolutions: 50 km and 10 km. Data for the 50 km resolution resulted from the aggregation of information available at the 10km resolution. Since categorical information on the sampling effort per 10x10-km grid cell was available for Spain (Pleguezuelos *et al.* 2002), only those absence cells where a maximum effort was undertaken were considered as absence data. Absence data for Portugal was also restricted to cells that were prospected by the first author.

Only those 50x50-km cells containing more than ten adequately surveyed 10x10-km cells were considered for the analysis. The third dataset included information on the presence/absence of species gathered throughout the Portuguese geographical range of *Mauremys leprosa* (see Segurado & Araújo 2004a, or chapter 5, section 5.2, for further details). The fourth dataset was based on preliminary results of an ongoing study in the Southwest Coast of Portugal (Segurado, unpublished data). The study area corresponds to a 30x30 km square where an almost complete spatial segregation between species occurs. The fifth dataset consists of relative abundance data and microhabitat variables measured along a 1.3-km mediterranean stream reach. The stream is located within a geographical region of Southern Portugal where the two turtle species widely coexist in many habitats (See Segurado & Figueiredo 2007, or Chapter 5, section 5.1, for further details). Species sampling consisted mainly in the detection of basking individuals. For some locations - particularly those containing habitats with few available basking sites - and for the stream reach dataset, sampling consisted of trapping sessions using baited hoop nets. Forty-seven variables extracted at several spatial scales were considered for the analyses, each dataset including a different combination of variables according to their relevance for each spatial scale (Appendix 1). Environmental variables at the broader spatial scales were resampled at the same grid resolution as the species' occurrence data using average value for the continuous variables and class percentage for the land cover variables. Most of the information was processed and integrated in a GIS environment using ArcGis 9.1 (ESRI Inc., Redlands, CA, USA).

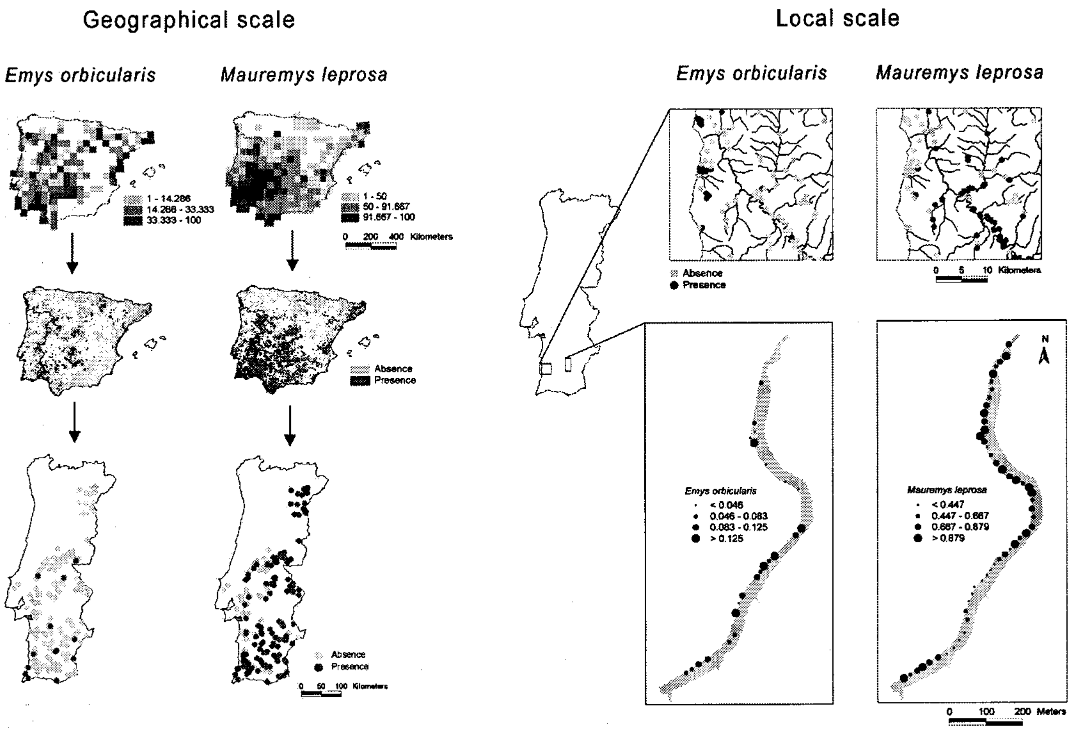


Fig. 4.1 – The five datasets used in the analysis.

Table 4.1 – The five datasets used for the analysis and the correspondent kind of information. Sources of information: [1] Godinho *et al.* (1999), Silva (2002), Keller & Andreu (2002) and the UNIBA database (www.cea.uevora.pt/umc); [2] Segurado & Araújo (2004a - chapter 5); [3] Segurado (unpublished data); [4] Segurado & Figueiredo (2007 - chapter 5); [5] New *et al.* (2000; see Segurado *et al.* 2006 - chapter 3 - for more detailed information); [6] Corine Landcover 2000 data set (Bossard *et al.* 2000); [7] grid theme with 2.5 minutes resolution (available at: http://www.diva-gis.org/data/DataServer.htm).

Extent	**Resolution**	**Type of data**	**Number of records**	**Environmental variables**	**Number of variables**
Geographical range	50 km grid cells (UTM)	Occupancy[1]	6111	Climate[5]	14
Geographical range	10 km grid cells (UTM)	Presence/Absence[1]	182	Climate[5], hydrology, land cover[6], population density[7]	30
Geographical range	200 m line transects	Presence/Absence[2]	217	Habitat[2]	29
Regional (30 x 30 km)	5 to 250 m radius	Presence/Absence[3]	99	Land cover[7], habitat[3], microhabitat[3]	26
Local (1.2 km)	5 m radius	Abundance[4]	52	Microabitat[4]	16

Data analysis

Multiscale analysis of species associations

Species coexistence were investigated considering both similarities on the use of space (i.e. spatial coordinates) and similarities on the response to environmental variables. The pure spatial association between species was tested using either Gaussian Regression (GR) for the species' abundance data or Logistic Regression (LR) for presence-absence data. Since *E. orbicularis* occupies a broader environmental range within the Iberian Peninsula we opted to regress *M. leprosa* against *E. orbicularis*, although the resulting relationship between species can be interpreted in both directions. We used a variable selection procedure based on the Akaike's Information Criterion (AIC). The nine terms of a third-degree polynomial of latitude and longitude of the cell centroids (trend surface analysis; Pereira & Itami 1991, Mönkkönen *et al.* 2004) were also included as explanatory variables in the final equation in order to control for the effect of geographical gradients at large extent (Legendre & Legendre 1998, Guisan *et al.* 2006). Except for the Iberian extent with 50 km resolution, finer scale spatial patterns were also controlled using a contagion term based on a two-order neighbourhood (see Segurado *et al.* 2006 for further details). For the Gaussian approaches both species abundance and its squared term were included in the analysis to account for a possible unimodal relationship between species (Mönkkönen *et al.* 2004).

We opted to use regression techniques since the methods for inferring spatial association based on null models did not apply for all datasets and moreover the incorporation of terms accounting for spatial autocorrelation and geographical trends is straightforward in regressions. Nevertheless, when possible, we also used two null models in order to examine the consistency of our results: a randomization approach implemented in ECOSIM (Gotelli 2000, Gotelli & Enstminger 2006) and a randomization algorithm based on toroidal random shifts (Palmer & Van der Maarel 1995, Fortin & Jacquez 2000, Segurado *et al.* 2006). For the first analysis we used the C-score of co-occurrence and performed 5000 randomisations maintaining a fixed number of occurrences of each species and assuming the sites as equiprobable (see Gotelli 2000, for further details). This method was not applicable for the stream reach dataset because one of the species (*M. leprosa*) occurred at all sampling sites. Inflation of type I errors are expected to occur with this method since it does not account for the spatial autocorrelation of data (Roxburgh & Matsuki 1999). Toroidal shifts were run only for the Iberian Peninsula wide data (10 km resolution) and for the stream reach datasets, since this method requires data recorded at regular grid systems. For running toroidal shifts at the scale of the Iberian Peninsula, a rectangular area - as required by the algorithm - with 44 by 60 cells was selected (see Segurado *et al.* 2006 for further details on the method). One thousand null spatial patterns were obtained for one species and each one was regressed against the spatial pattern of the other species. The observed *t*-value of the regression coefficient was then located in the expected distribution of *t*-values obtained under the null hypothesis in order to obtain the non-randomness probability of the relationship.

Univariate analyses were performed to estimate the strength and sign of species responses to each environmental variables and to assess the environmental association between the species. This was also accomplished using GR or LR according to the kind of response considered (abundance or presence-absence, respectively) and also controlling for the effect of spatial structure using the terms of a third-degree polynomial of coordinates and a contagion term. Regressions were adjusted both excluding and including these spatial terms to assess the influence of space on the environmental variables. For these analyses no variable selection procedure was used in order to full control for the spatial component of variation. This was accomplished by forcing the spatial terms to be included in the regression equation. The resulting *t*-values of the regression coefficients for each environmental variable were compared between species in order to assess the similarity of the response to the environment.

Except for the randomisation approach implemented in ECOSIM, all analyses were performed using S-Plus 2000 (Statistical sciences 1999).

Exploring patterns of coexistence

We used linear Discrimination Analysis (DA) to measure the degree to which available predictor variables accounted for the observed pattern of species coexistence, considering the Iberian extent and 10 km resolution. Three groups of turtle records were considered for analysis (Fig. 4.2): (1) exclusive distributions of *E. orbicularis*, (2) coexistence of the two species and (3) exclusive distributions of *M. leprosa*. Since the variables did not meet the assumptions of normal distribution and equal within-group variances required by most tests involved in the DA diagnosis, we used a canonical model to reduce dimensionality in the data and to minimize the importance of the overall test statistics of the analysis. No variable selection procedure was used in order to avoid using further test statistics based on such assumptions. Nevertheless, to help satisfy the DA assumptions and improve the discrimination between groups, the environmental variables were previously transformed using either square-root (Climatic and hydrologic variables) or arcsine (Land cover variables) transformations (Quinn & Keough 2002). Furthermore, to avoid collinearity between variables we used Pearson correlations as a preliminary step to eliminate redundant variables; correlations higher than 0.8 were not allowed.

Multivariate distances between each group were calculated using Mahanalobis distances based on the whole set of environmental variables. Finally, the same distances were also calculated considering only those variables that were more directly related to habitat features (i.e., land cover and hydrology) in order to explore changes in the relative positions of groups in the multivariate space when the variation attributed to climate is eliminated.

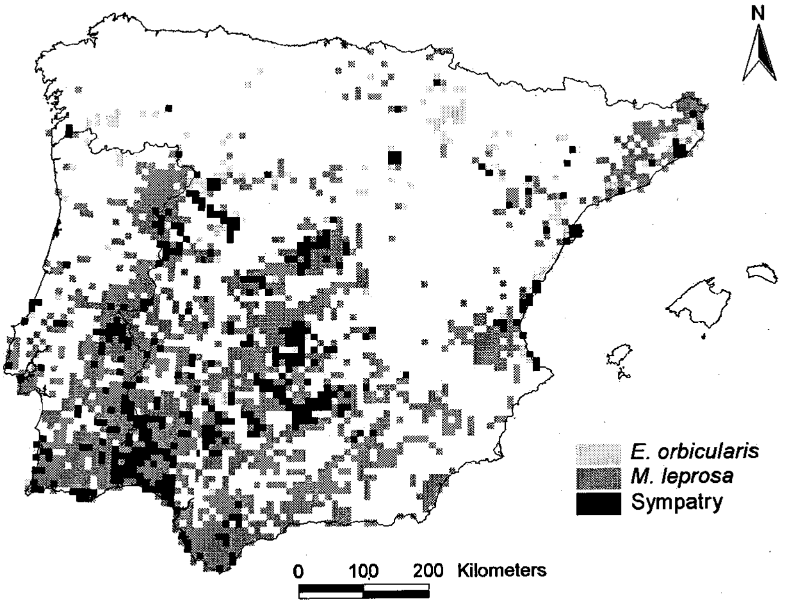


Fig. 4.2 – Distribution of *E. orbicularis* and *M. leprosa* in the Iberian Peninsula according to a 10 x 10 km resolution grid.

4.4 Results

Multiscale analysis of species' coexistence

As evidenced by the results of the multiscale analysis of species spatial association (Table 4.2, Fig. 4.3) there is a shift from positive association at broader spatial scales to a negative or non-significant association between the species at finer scales. The overall sign and magnitude of the spatial association was generally consistent between the different alternative methods (ECOSIM, toroidal shifts and regression-based approaches; Table 4.2).

The spatial association between species varied among spatial scales. For the Iberian dataset compiled at 50 km resolution there is a significant unimodal relationship between species, whether or not the spatial pattern is controlled for (Fig. 4.3a, 4.3b). Nevertheless a positive relationship is the general trend as shown by the positive sign of the non-quadratic term of species occupancy. Both at the Iberian extent with 10km resolution and at the Portuguese extent with habitat resolution there is a positive association that is maintained even after controlling for the spatial pattern and contagion (Fig. 4.3a, 4.3b). For the Southwestern coast dataset there is a negative association that turns into a positive association when controlling for the spatial pattern and contagion (Fig. 4.3a, 4.3b). This indicates that a considerable amount of the explained environmental variation is shared by spatial variation. For the stream reach dataset, although no significant relationship was found (Table 4.2), there is a slight tendency for a negative association between species that is maintained after controlling for spatial pattern and contagion (Fig. 4.3a, 4.3b).

Table 4.2 - P-values for different alternative association tests: LR/GR* - Logistic Regression/Gaussian Regression (t-test significances of the regression coefficients); LR/GR** - same analysis but controlling for the spatial pattern and contagion; ECOSIM (probability of the observed C-score to be equal or higher than the expected under the null pattern); TOROID (probability of the observed t-value of the regression coefficient to be equal or higher than the expected under the null spatially structured pattern). Blank cells means that the method was not applicable for the dataset.

Dataset	LR/GR*	LR/GR**	ECOSIM	TOROID
Iberian Peninsula 50km	< 0.001 (+); 0.010 (-)[1]	0.002 (+); 0.042 (-)[1]	< 0.001 (+)	-
Iberian Peninsula 10 km	< 0.001 (+)	< 0.001 (+)	< 0.001 (+)	< 0.001 (+)
Portugal habitat	0.638	0.634	< 0.005 (+)	-
South-western coast habitat	0.003 (-)	0.052 (+)	< 0.001 (-)	-
Stream reach microhabitat	0.783; 0.493[1]	0.387; 0.299[1]	-	0.764

[1] – P-value of the quadratic term of species abundance (see text for details).

The shift from positive to negative association between species with decreasing spatial extents and resolutions is more evident when it is based on the comparison of species response to environmental association (Fig. 4.4). For the three broader spatial extents and coarse resolutions there is a significant positive relationship between species in their response to the environment given by the absolute value and sign of the univariate relationship (*t*-value) between species occurrence and each environmental variable. This positive association is maintained even after controlling for spatial pattern and contagion (Fig. 4.4a, 4.4b). For the Southwestern coast dataset there is a negative association that becomes non-significant after controlling for the spatial pattern and contagion (Fig. 4.4a, 4.4b). For the stream reach dataset there is a consistent negative association between the occurrence patterns of the two species.

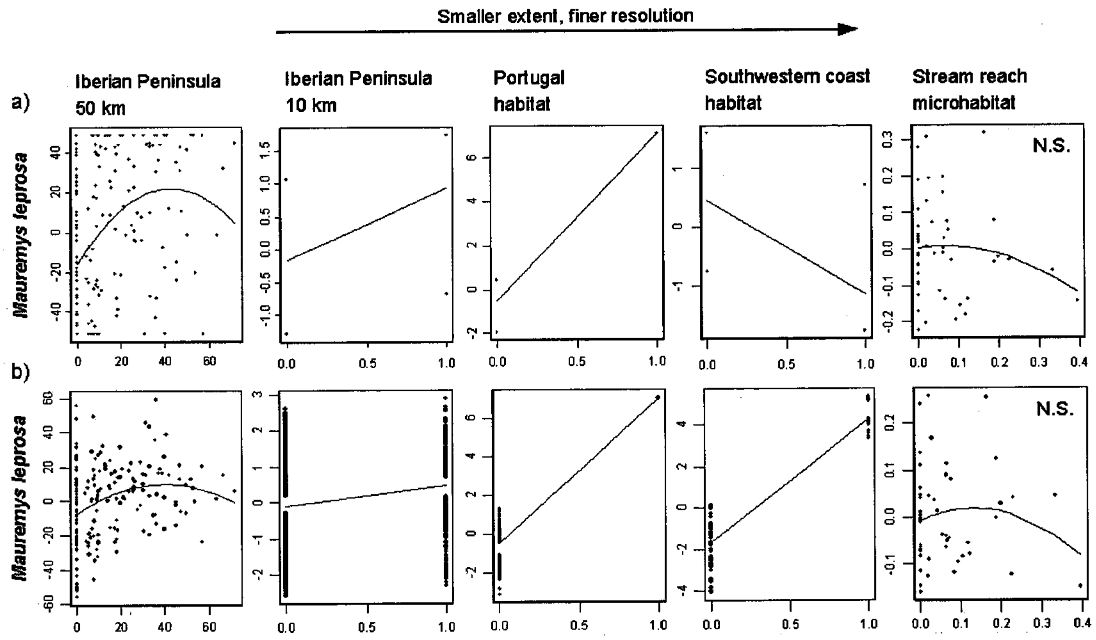


Fig. 4.3 – Spatial association between *E. orbicularis* and *M. leprosa* for each five spatial extents and resolutions considered; a) response curves not controlling for space, b) partial response curves controlling for spatial pattern and contagion (see Material and Methods section for further details).

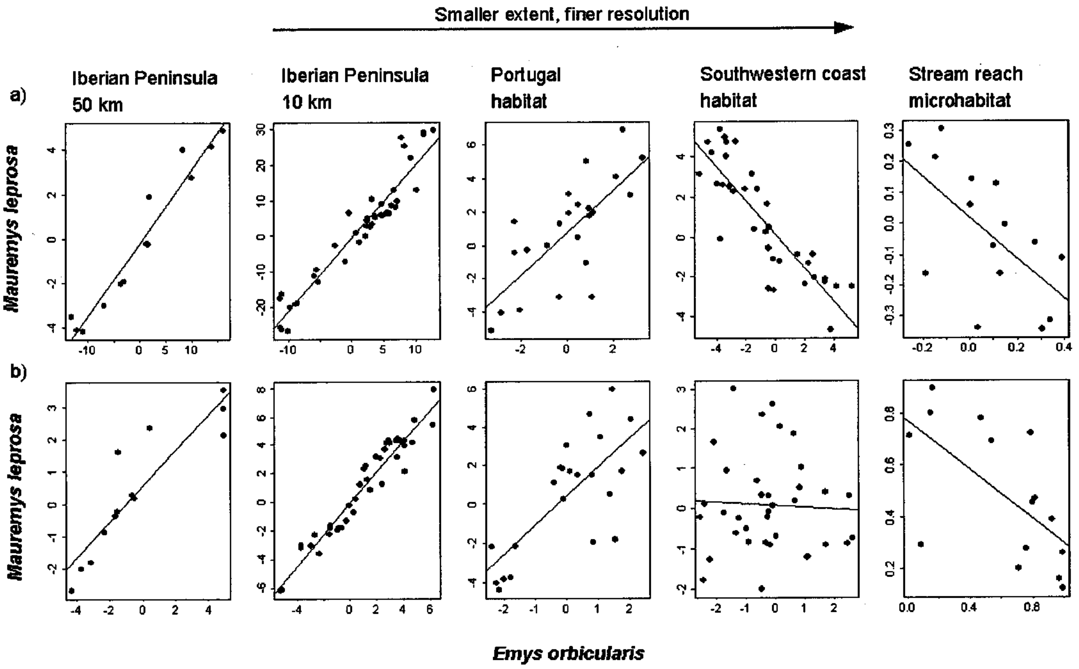


Fig. 4.4 – Relationship between species' response to environmental variables based on the t-values of univariate regressions of species occurrence versus each environmental variable; a) relationship not controlling for space, b) relationship controlling for spatial pattern and contagion (see Material and Methods section for further details).

Patterns of coexistence in the Iberian Peninsula

The three possible combinations of occurrence patterns, i.e. coexistence versus exclusive distributions of each species, were poorly differentiated by the linear discriminant functions as shown by the low percentage of the explained variance (38.9%; Table 4.3a) and high classification errors (overall error = 0.33; Table 4.3b). As shown by the Mahanalobis distances between the three groups of species occurrences (two exclusive occupancy zones and one coexistence zone), considering all independent variables together, the exclusive zone of *E. orbicularis* was the most distant group from the two others (Table 4.4a).

Furthermore, sites occupied solely by *E. orbicularis* were slightly more similar to sites occupied by *M. leprosa* alone than to the sites where species coexist. When the multivariate Mahanalobis distances between the three groups of species occurrences are calculated exclusively with habitat-related variables (i.e. land cover and hydrology) the exclusive occupancy sites of *E. orbicularis* becomes even more similar to those of *M. leprosa* (Table 4.4b).

Table 4.3 – Diagnosis of the discriminant analysis considering three groups for species co-occurrence: non-coexistence sites of *E. orbicularis*, coexistence sites, non-coexistence sites of *M. leprosa*. a) Overall diagnosis; b) classification matrix.

a)

Function	Eigenvalue	% of explained variance	Canonical Correlation
1	0.257	66.070	0.452
2	0.132	33.930	0.342

b)

		classified			
		E. orbicularis	Coexistence	*M. leprosa*	Omission error
Observed	*E. orbicularis*	57	16	83	0.635
	Coexistence	26	90	289	0.778
	M. leprosa	56	67	944	0.115
	Commission error	0.590	0.480	0.283	0.330*

Table 4.4 – Mahanalobis distances between the three groups of species occurrences (two exclusive occupancy zones and one coexistence zone) considering: a) the whole set of variables and b) only the habitat-related variables.

a)

	E. orbicularis	Coexistence	*M. leprosa*
E. orbicularis	0	3.441	2.982
Coexistence		0	0.735
M. leprosa			0

b)

	E. orbicularis	Coexistence	*M. leprosa*
E. orbicularis	0	1.303	0.547
Coexistence		0	0.567
M. leprosa			0

The plot of the discriminant analysis coefficients (Fig. 4.5) shows that the exclusive sites of *E. orbicularis* are separated from the other groups mainly by the first discriminant function while the exclusive sites of *M. leprosa* are mainly separated by the second discriminant function. The first discriminant axis is predominantly associated with climate variables while the second discriminant function axis is predominantly associated with habitat-related variables (hydrology and landcover), as shown by the pooled within-groups correlations between environmental variables and standardized canonical discriminant functions (Table 4.5). Sites exclusive to *E. orbicularis* tend to be positively associated with summer precipitation and negatively associated with precipitation seasonality and mean temperature. Exclusive sites of *M. leprosa* sites tend to be positively associated with steeper stream slopes and with agricultural lands and negatively associated with 'Montado' agro-forestry systems (Fig. 4.6).

Table 4.5 – Structure matrix of the discriminant analysis - pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.

	1st DF	2nd DF
Psum	-0.734	0.084
Pseas	0.692	-0.051
T_T_WARM	0.604	-0.194
T_T_COLD	0.503	0.027
t_Prec_range	0.426	0.045
T_EVAP	-0.392	0.176
T_RADS	0.356	-0.184
t_cor_agrrega	-0.289	0.022
t_Prec_ann	-0.187	0.145
t_Flowacc	-0.172	-0.093
t_Flowlen_f1	-0.116	-0.079
t_Stream_alt	-0.116	0.111
T_LENGTH	-0.110	0.010
T_FDIR	0.014	0.004
t_Stream_slope	0.033	0.427
t_cor_agro-for	0.282	-0.419
t_cor_agr	-0.287	0.326
t_cor_agrseq	-0.073	0.182
t_Streamo	-0.052	-0.164
t_cor_matos	-0.040	0.152
t_T_range	-0.055	-0.150
t_cor_past	0.069	-0.145
t_cor_for	0.080	-0.111
t_cor_artif	-0.023	-0.053
T_DENPOP	-0.030	0.035

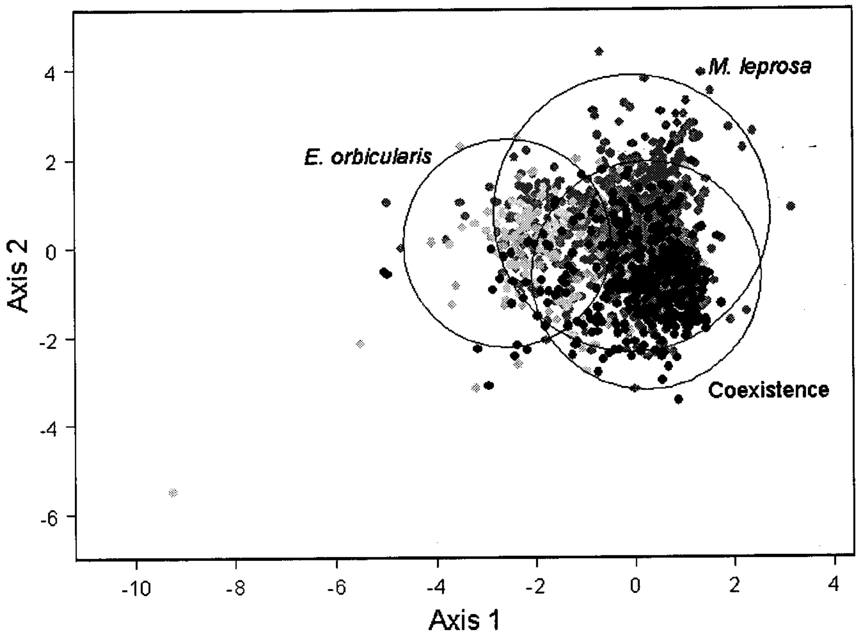


Fig. 4.5 – Discriminant scores for each group of observations (sites with *E. orbicularis*, *M. leprosa* and both species) in the 2-D space defined by the two discriminant functions.

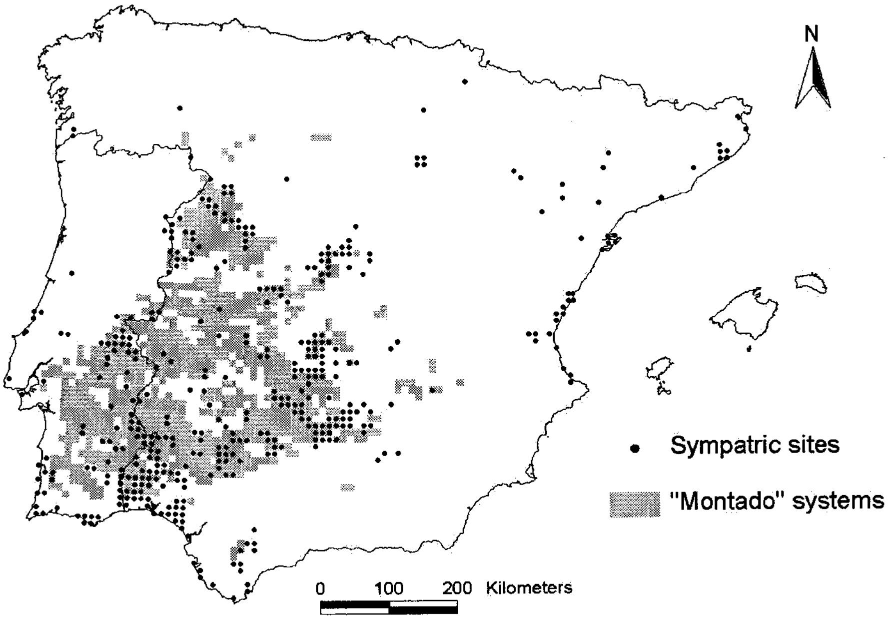


Fig. 4.6 – Geographical coincidence between the "Montado" agro-foretry systems and the coexistence sites of *E. orbicularis* and *M. leprosa*.

4.5 Discussion

Our results demonstrate that examining patterns of coexistence at different spatial extents and resolutions might lead to opposing conclusions about the nature of the biotic interactions between the studied species. *E. orbicularis and M. leprosa* coexist at broad spatial scale but segregate themselves when their occurrence is examined at smaller spatial extents and/or finer grain sizes. Consequently, the simple examination of spatial patterns at a given scale would suggest very different trends from broad to finer scales. Although there is a wide overlap of niches at biogeographical scales that would suggest no significant influence of negative biotic interactions on species geographical ranges, this does not hold at local scales where a marked spatial and environmental segregation between species is observed. The few studies that analysed patterns of co-occurrence of other species at multiple spatial scales have generally shown more consistency between scales (e.g. Bullock *et al.* 2000).

Broad scale evidences for biotic interactions

Conversely, as evidenced by the discriminant analysis, the resulting positive association between species at the biogeographical scale does not necessarily mean that negative biotic interactions are not detectable at that scale. The effect of negative biotic interactions may be masked by the effect of parallel historical factors or similar responses to broad environmental patterns. In fact, since the sites occupied exclusively by *M. leprosa* are environmentally closer to the coexistence sites than are the exclusive sites of *E. orbicularis*, arguably biotic interaction could be playing a role on species pattern of occurrence even at this coarse scale, suggesting that *M. leprosa* is the superior competitor in this two-species system. Where the two species coexist it seems to occur a shift in the niche of *E. orbicularis,* reflected in the substantial environmental shift between the exclusive occupancy sites of *E. orbicularis* and the coexistence sites. The resulting overall multivariate distances when only habitat-related variables (i.e. landcover and hydrology) are considered (i.e. not considering the effect of climatic patterns) somehow reinforce this hypothesis: the relative distance of coexistence sites becomes greater and exclusive sites of *M. leprosa* becomes closer to those occupied exclusively by *E. orbicularis*. This could be interpreted as an evidence of niche shift or competitive release: in the absence of the superior competitor (inferred to be *M. leprosa*), the inferior competitor (i.e. *E. orbicularis*) occupies a niche that is more similar to that of the superior competitor (Connor & Bowers 1987, Anderson *et al.* 2002). If that is the case, the absence of the superior competitor from these areas could be attributed to the effect of climatic factors, namely a greater intolerance to low temperatures by *M. leprosa*.

Gradients of co-occurrence patterns

Different species-specific mechanisms may be involved in the exclusion of each species since it appears that each species is excluded along two distinct environmental gradients. At the Iberian scale, exclusive occupancy sites of *E. orbicularis* are possibly beyond the ecological climatic tolerance of *M. leprosa* (physiological constraints), while many sites occupied exclusively by *M. leprosa* contain habitats that are probably less adequate for the presence of *E. orbicularis* (habitat selection). Areas of coexistence appear to lie on environmentally sustainable land covers, namely those dominated by the typical Iberian “Montado” agro-forestry-pastoral systems (Pereira & Fonseca 2003), corresponding probably to the climatic and the habitat optima of both species. These “Montado” landscapes probably maintain high levels of microhabitat heterogeneity within mediterranean streams, which appears to play an important role in species coexistence at the local scale (Segurado & Figueiredo 2007 - chapter 5).

Spatial and environmental segregation at finer scales

At smaller spatial extents and finer resolutions, the scales at which differences between species in the use of space are perceived seem to depend on specific local conditions. The two local scale datasets where a negative association between species was observed correspond to somehow distinct landscapes.

The Southwestern Portuguese coastal zone is a heterogeneous landscape with very diverse habitats available. Here species are segregated at the habitat level: *E. orbicularis* is almost entirely restricted to the temporary ponds in a coastal strip while *M. leprosa* is restricted to streams and permanent ponds in the inland zone. Since habitat gradients are coincident with a spatial gradient (inland versus costal zone), the effect of the environment was removed when the effect of space was included in the association test, such that it is very difficult to infer whether species have a true negative environmental association. In contrast to this first study area, the sampled stream reach lies within a homogeneous landscape, even though there is a high level of spatial heterogeneity along the stream at a finer scale. In this case species tend to be segregated at the microhabitat level and our results suggest that *E. orbicularis* also occurs preferentially on temporary and shallow reaches, while *M. leprosa* is less selective regarding microhabitat, although tending to occupy permanent and deeper reaches of the stream segment. This might indicate that the negative environmental association in the Southwestern Portuguese coast is not explained solely by spatial segregation.

Whether these differences are the result of spatial or environmental shifts by *E. orbicularis* induced by the presence of *M. leprosa* remains unanswered after the present study. Nevertheless, these results contrast with other quantitative studies on habitat selection by *E. orbicularis* where a preference for large permanent water bodies is documented, such as in the Po River delta of northern Italy (Ficetola *et al.* 2004). The preference for lower courses of large permanent rivers is also documented for the southern European Russia (Bozhansky & Orlova 1998). This may be due to different local adaptations of the species or may reflect different biotic constraints at distinct parts of its geographical range. Yet, the clear negative association between the two species at local scale further supports the above mentioned evidences for the role of biotic interactions on the observed pattern of species coexistence at the Iberian scale.

Concluding remarks

At broader spatial scales there is an increasing difficulty to infer the role of biotic interactions due to spatial resolution and/or spatial extent. In fact, larger resolutions imply that sample units will contain a greater heterogeneity of the environment which provides *per se* opportunities for sympatry even for competing species (Connor & Bowers 1987; Araújo & Guisan 2006). On the other hand, larger spatial extents augment the probability of biotic interactions being masked by the effect of similar responses to broad climatic patterns.

This difficulty becomes a challenge when incorporating biotic interactions onto niche modelling (Guisan & Thuiller 2005; Araújo & Guisan 2006). Niche models are usually employed for predicting or explaining species distributions at large spatial extents and, due to practical, logistic or data limitations, they are usually based on coarse scale data. Modelled niches using such coarse scales only partially describe the actual niche of a species and therefore it is probable that pairs of competing species show modelled niches that overlap more than would be expected if the full hierarchy of factors affecting species co-occurrence at several spatial scales would be properly assessed. Hence at such scales the distinction between fundamental and realised niche is not as straightforward as at local spatial scales, as the role of historical or dispersal factors gains importance relative to biotic interactions. That is the reason why the Volterra-Gause principle (Hutchinson, 1957) does not hold at the geographical scale, and helps explain how species that show intersecting modelled realised niches can geographically co-occur (Araújo & Guisan 2006).

Although there have been conflicting views of what the output of empirical correlative models truly represent (Araújo & Guisan 2006), we assume that, at most, such models provide a spatial representation of the 'potential' realised niche, i.e., areas where species may occur given the biotic and historical constraints at the sites where they were sampled (that may be different from non-sampled sites). Since our analyses for testing the environmental association between species are based on empirical correlative methods, the results correspond to the relationships between species' realised niches. If these realised niches are determined primarily by historical factors then negative associations are expected among pairs of species that have evolved by geographical speciation (Connor & Bowers 1987). On the other hand, positive associations might be expected among pairs of species that have evolved independently but which had similar historical constraints due to similar physiological, behavioural or dispersal limitations, which could be the case of the two studied Iberian freshwater turtle species.

This study suggests that potential biotic interactions may not necessarily be inferred at broad spatial scales and that multiscale approaches are cruicial to fully explore patterns of co-occurrence. Therefore, proposed frameworks to infer mechanisms underlying spatial patterns (e.g. Anderson *et al.* 2002; Sfenthourakis *et al.* 2006) should take into account that the relative importance of mechanisms of co-occurrence might change at different spatial extents and resolutions.

Acknowledgments

We are grateful to Susana Rosa, Rui Rebelo, Inês Rosário, Paulo Cardoso for their help during fieldwork. P. Segurado was supported by a grant from Fundação para a Ciência e Tecnologia (SFRH/BD/8493/2002).

APPENDIX 1

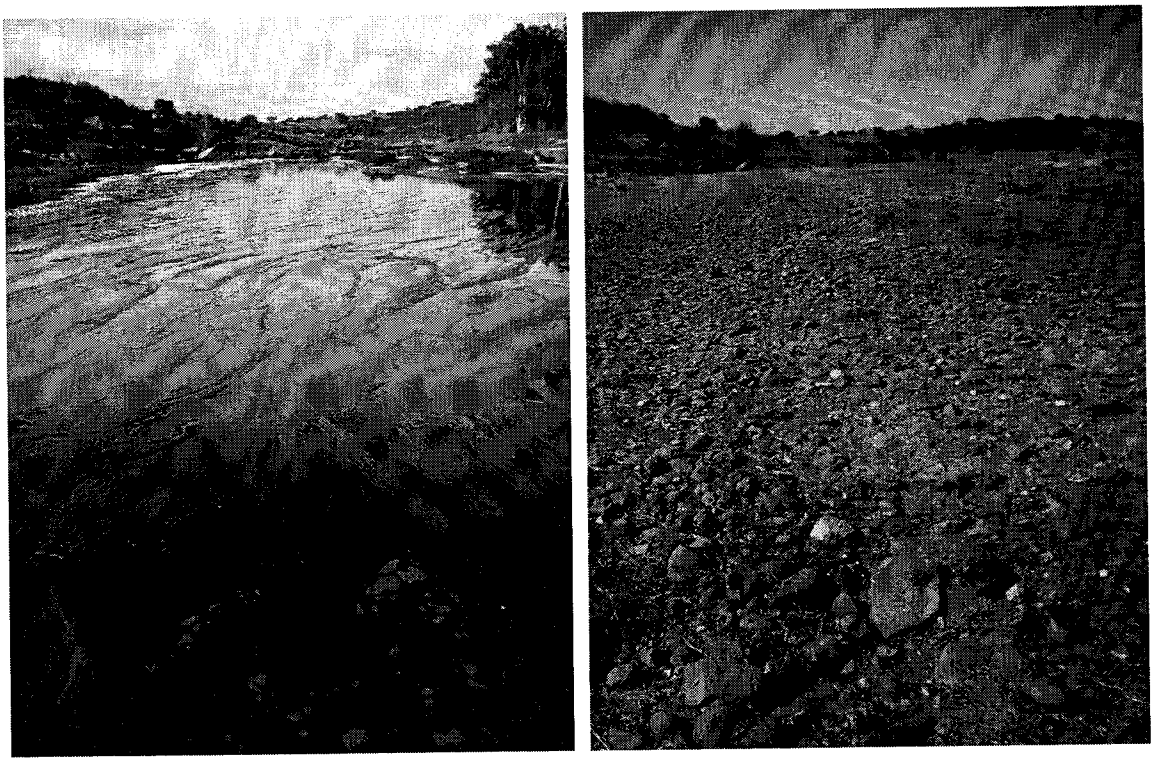
Variables	Databases
Climatic variables	
Mean annual temperature (Tann)	1, 2
Mean temperature of the coldest month (Mtc)	1, 2
Mean temperature of the warmer month (Mtw)	1, 2
Temperature annual range	1, 2
Mean annual global net radiation (Rann)	1, 2
Mean annual evapotranspiration / potential	1, 2
Evapotranspiration (A2P)	1, 2
Mean annual precipitation sum (Pann)	1, 2
Mean winter precipitation sum (Pwin)	1, 2
Mean summer precipitation sum (Psum)	1, 2
Mean spring precipitation sum (Pspr)	1, 2
Mean autumn precipitation sum (Paut)	1, 2
Precipitation annual range	1, 2
Precipitation seasonality	1, 2
Topographical variables	
Altitude	3
Slope	3
Hillshade	3
Hydrological variables	
Stream altitude	2
Stream slope	2, 4
Stream total length	2, 4
Flow accumulation	2
Flow direction	2
Stream order	2, 4
Socio-economic variables	
Human population (density)	2

APPENDIX 1 – cont.

Landcover variables	
Landcover richness	2, 3, 4
Forest	2, 3, 4
Shrub lands	2, 3, 4
Pasture	2, 3, 4
Inland water	2, 3, 4
Agriculture	2, 3, 4
Extensive crops	2, 3, 4
Intensive crops	2, 3, 4
Agro-forestry systems ("Montados")	2, 3, 4
Artificial landcover	2, 3, 4
Habitat / microhabitat variables	
Substratum (bedrock, gravel, sand, mud)	3, 4, 5
Stream width	3, 4, 5
Maximum depth	3, 4, 5
Current speed	3, 4, 5
Water turbidity	3, 4, 5
Sun exposure	3, 4, 5
Emergent vegetation	3, 4, 5

Chapter 5

Dealing with species rarity when analyzing patterns of species co-occurrence



Cobres River, Castro Verde.

5.1 Working example 1: coexistence of two freshwater turtles along a stream

Note – This study has been published in Acta Oecologica under the title "Coexistence of two freshwater turtle species along a Mediterranean stream: The role of spatial and temporal heterogeneity" (Segurado, P. & Figueiredo, D., 2007. *Acta Oecologica*, 37: 134-14). The text has been slightly modified for layout harmonization (e.g. citation format) or for some minor English corrections.

5.1.1 Summary

In the Iberian Peninsula the European pond turtle (*Emys orbicularis*) and the Mediterranean pond turtle (*Mauremys leprosa*) share many freshwater habitats, in particular Mediterranean streams. Whether and how these two species divide space within those habitats is poorly known in part due to the very low abundance of *E. orbicularis* at most syntopic sites. The spatial coexistence of these two species was studied along a 1.3 km reach of a typical Mediterranean stream based on data from trapping sessions and basking counts. The effect of the hydrological regime on differences in space use between species was also assessed. Spatial associations between species and between each species and microhabitat descriptors were estimated using a permutation procedure to account for spatial autocorrelation. Differences in the use of space were also estimated using a resample technique to account for the small sample sizes of *E. orbicularis*. Results indicate that *E. orbicularis* shows a preference for temporary, shallow, well vegetated and sandy reaches, while *M. leprosa* is less selective regarding microhabitat. Differences between *E. orbicularis* and juveniles of *M. leprosa* were less obvious. The high spatial heterogeneity of Mediterranean streams may be responsible for the persistence of viable populations of *E. orbicularis* as well as favouring the coexistence of the two turtle species. Therefore, stream habitat management and conservation plans for *E. orbicularis* should give priority to the maintenance of high levels of heterogeneity along Mediterranean streams.

Keywords: *Emys orbicularis*, *Mauremys leprosa*, Rarity, Resampling, Spatial autocorrelation, Spatial segregation

5.1.2 Introduction

Space is one of the niche axes where ecological differences may arise between species, making possible their coexistence at a given area (e.g. Connor & Bowers 1987, Chesson 2000, Amarasekare 2003). If spatial segregation can be one of the consequences of inter-specific competition, empirical evidences for such segregation between closely related species is not per se indicative of competition (Connor & Bowers 1987, Roxburgh & Chesson 1998). Further analyses on the spatial patterns of coexistence, such as evidences for competitive release (i.e., shifts on habitat selection between sympatric and allopatric populations, e.g. Anderson *et al.* 2002), are needed. Spatial heterogeneity is often regarded as one of the main driving forces for species coexistence (Petren & Case 1998, Chesson 2000, Amarasekare 2003), by providing opportunities for species that differ even slightly in their response to the environment to co-occur in the same area. Here we analyse the coexistence of two freshwater turtles, the European pond turtle *Emys orbicularis* (Linnaeus, 1758) and the Mediterranean pond turtle *Mauremys leprosa* (Schweiger, 1812), along a reach of a stream with a typical Mediterranean hydrological regime located in SW Iberian Peninsula.

Mediterranean streams follow a predictable sequence of regular and often extreme flooding and drying periods (Gasith & Resh 1999). When this temporal instability is combined with highly irregular geomorphic features of the river bed, an important consequence is the high spatial heterogeneity of microhabitats that typically occurs during the dry season (Magalhães *et al.* 2002). This heterogeneity most probably plays an important role in shaping spatial–temporal patterns of species coexistence. During the dry season streams usually turn into a chain of individual pools with varying areas and depths, each one lodging communities that probably differ in their composition and structure. On the other hand, the drying that follows the rainy season involves habitat contraction and consequently a gradual increase in biotic (e.g. competition and predation) and abiotic (e.g. dissolved oxygen and temporary habitat loss) controls in structuring biological communities (Gasith & Resh 1999).

In the Iberian Peninsula the two native freshwater turtles, *E. orbicularis* and *M. leprosa*, share many Mediterranean stream ecosystems and other wetlands. The geographical range of *E. orbicularis* in the Iberian Peninsula is very fragmented and greatly overlaps with that of *M. leprosa*, namely in southern Portugal, Extremadura and Western Andaluzia. However, in some regions of northern Spain *E. orbicularis* extends its range beyond that of *M. leprosa* (Ayres & Cordero 2001, Keller & Andreu 2002). On the other hand, *E. orbicularis* is absent from vast regions of southern Spain where *M. leprosa* occurs (da Silva 2002). While *E. orbicularis* has a predominant European distribution, *M. leprosa* has an Iberian and North African distribution, which probably explains these 'exclusion' areas. At the local scale, *E. orbicularis* is far less abundant than *M. leprosa* at sites where syntopic populations occur (Araújo *et al.* 1997, Segurado & Araújo 2004a – Chapter 5, section 5.2) and tend to show more skewed age structures towards a predomination of adults (Keller *et al.* 1998, P. Segurado & A.P. Araújo, unpublished data).

A strong decline of *E. orbicularis* in the past century has been documented throughout its geographical range, including North Africa (Bons & Geniez 1996), Southern Europe (e.g. Devaux & Bley 1998, Fattizzo 2004), Central Europe (e.g. Schneeweiss 1998, Quesada 2000), Eastern Europe (Szczerbak 1998, Kotenko 2004, Maciantowicz & Najbar 2004, Mazanaeva & Orlova 2004, Puky *et al.* 2004) and Asia Minor (Taskavak & Reimann 1998). Therefore, there is a growing concern on international cooperation for the conservation of this species (Fritz *et al.* 1998, Buskirk & Servan 2000, Fritz & Havas 2004). Factors related to habitat loss are the causes most often mentioned in the literature to explain species decline (Bozhansky & Orlova 1998, Mascort 1998, Lacomba & Sancho 2000, 2004, Cordero & Ayres 2004, Fattizzo 2004, Puky *et al.* 2004). Although both freshwater turtles are protected under Annexes II and IV of the European Habitats Directive (92/43/CEE), the conservation status of *E. orbicularis* populations in the Iberian Peninsula is far more critical than that of

M. leprosa. In fact, *E. orbicularis* is currently included in the "endangered" category according to the recent revision of the Portuguese Vertebrate Red Data Book (Cabral *et al.* 2005). Information on patterns of species coexistence is important to assess whether *M. leprosa* is an additional factor to be considered for *E. orbicularis* conservation in the Iberian Peninsula. Whether adequate management efforts for conservation are compatible between species is also an important issue that deserves attention. Due to the greater detectability and abundance of *M. leprosa*, reliable ecological data are very easily collected for this species. For that reason it is pertinent to assess whether its ecological attributes could be used as surrogates for those of *E. orbicularis*.

The pattern of coexistence of *E. orbicularis* and *M. leprosa* at the biogeographical and habitat scales has been analysed elsewhere (Segurado & Araújo 2004a – Chapter 5, section 5.2). The results of this study suggest that in Portugal the environmental optima seem to be similar among the two species, although *E. orbicularis* tends to occupy a more extreme and narrower position in the environmental gradient than *M. leprosa*; indeed, according to predictive models of species occurrence the presence of *E. orbicularis* tends to be associated with higher probabilities of occurrence and abundance of *M. leprosa* both at the biogeographical and habitat scales. This may be caused by a 'sample effect' (Gaston 2003) due to the low abundance of *E. orbicularis* that would result in fewer confirmed occupied habitats, mainly those corresponding to the species' optima. However, whether the two species show differences in the patterns of microhabitat use within these habitats is poorly known. The few studies on spatial partitioning within freshwater turtle assemblages have usually been conducted at wider spatial scales and were generally confined to North America (Kofron & Schreiber 1985, Fuselier & Edds 1994, DonnerWright *et al.* 1999, Bodie & Semlitsch 2000, Bodie *et al.* 2000, Lindeman 2000, Joyal *et al.* 2001, Bury *et al.* 2003).

Hypothesis testing of species coexistence is generally based on null models that are pattern-generating models where a mechanism of interest is excluded and randomization tests of ecological and biogeographic data are performed to determine whether a community ensemble differs from chance expectations (e.g. Gotelli & Graves 1996, Roxburgh & Matsuki 1999, Lindeman 2000, Manly & Sanderson 2002, Hofer *et al.* 2004). For example, spatial autocorrelation is an important spatial property of species occurrences that is important to maintain fixed (i.e., to exclude its effect) during pattern generation, when randomization tests are performed (Lennon 2000, Hofer *et al.* 2004, Segurado *et al.* 2006 – Chapter 3). Methods for testing spatial association (or dissociation) between species are largely skewed towards studies on plant communities (e.g. Roxburgh & Chesson 1998, Perry & Dixon 2002). However, these tests are usually too conservative for highly mobile animals such as freshwater turtles, because individuals may widely coexist in a zone between their preferred habitats or microhabitats into which they are continually diffusing (Pontin 1982).

In this study we hypothesise that the spatial–temporal heterogeneity of Mediterranean streams has an important role in structuring the use of space by the two species. To test that hypothesis we compare the use of space by *E. orbicularis* and *M. leprosa* along a typical Mediterranean stream based on simple microhabitat descriptors. We begin by testing the spatial association between species and their response to microhabitat using a null model approach based on randomization tests. However, the detection of differences between species can be strongly biased given the extremely low abundance of *E. orbicularis* compared to that of *M. leprosa*. Indeed, rarity represents in many cases a strong limitation for an adequate analysis of species' ecological traits, particularly those related to niche breadth and niche position (Gaston 1994). To deal with this difficulty we propose an approach based on a resample technique that allow quantifying differences in use of space between rare and common species. We also investigate whether differences in the use of space change as the drying season progresses and if they depend on individuals' sex and age class. This study is part of a wider framework to assess whether *M. leprosa* can be considered a limiting factor for *E. orbicularis*, i.e., if there are evidences for inter-specific interactions that would explain the absence of *E. orbicularis* from many apparently favourable habitats in the Iberian Peninsula.

5.1.3 Materials and methods

Study area

The study was conducted along a segment of a stream (Cobres River) near Castro Verde in southern Portugal (Fig. 5.1). Land cover is primarily a Mediterranean pseudo-steppe of nonirrigated cereal fields and pastures. The stream margins are used as pasture and therefore riparian vegetation is poorly developed at most of its length. The site is a typical Mediterranean stream that follows a predictable seasonal pattern: a wet period during which water flows throughout its length, followed by a dry season when the stream dries at most of its length and turns into a chain of individual pools. Seven sub-reaches were considered, each of them associated to a permanent pool (Fig. 5.1). The stream possesses high habitat heterogeneity, comprising pools and riffles of different size, morphology and vegetation cover. Some reaches have rocky and steep margins lacking riparian vegetation (e.g. sub-reach 5), while others have shallow margins with dense emergent vegetation (e.g. sub-reach 3).

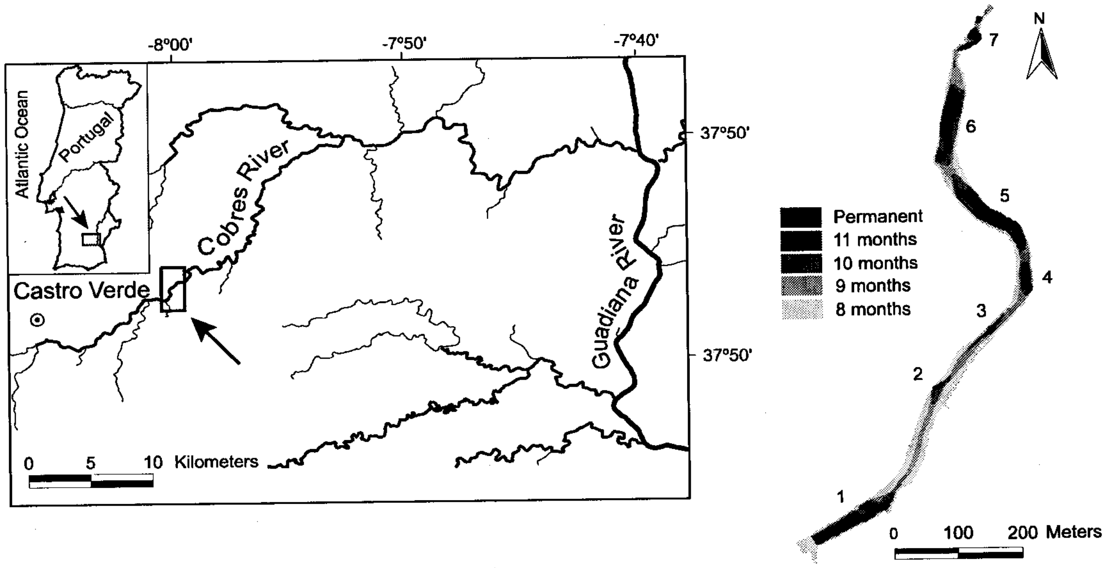


Fig. 5.1 – Map of the Cobres River and the study reach (numbers correspond to sub-reaches associated to each permanent pool).

UNIVERSIDADE

Sampling

Data were obtained from a capture–recapture program that took place from March 2003 to September 2005. Sampling was restricted to the period of high turtle activity, i.e., from March to September. Three hydrological seasons were considered. The first season, from March to April, corresponds to the period of increasing turtle activity, with a predomination of male individuals, when usually wet conditions prevail. The second season, from May to mid-July, corresponds to the reproductive period when the stream starts to dry and the water flow stops. The third season, from mid-July until the end of September, corresponds to the dry season. During the first year sampling took place monthly while in the second and the third years only one sampling session per hydrological season was carried out. In each sampling session traps were usually open during four or five consecutive days and were checked twice a day.

Table 5.1 – Variables recorded at each sampling point (cont – continuous variables, ord – ordinal variables).

Variables	Units	Type
Linear coordinate (x)	Meters	cont
Water permanence	Number of months	ord
Stream width	Meters	cont
Maximum depth	0: 0-0.25m; 1: 0.25 - 0.5m; 2: 0.5 - 1m; 3: > 1m	ord
Distance to permanent sectors	Meters	cont
Distance to basking sites	Meters	cont
Dominant shoreline inclination	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Bedrock	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Gravel	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Sand	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Mud	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Emergent vegetation	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Submerged vegetation	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Floating vegetation	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Grass cover (shoreline)	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Shrub cover (shoreline)	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Tree cover (shoreline)	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord

Turtles were captured with small baited hoop nets (approximately 30 x 60 cm) and permanently marked by marginal notching, according to a coding system. A non-toxic, waterproof paint was used as an aid for individual identification from a distance. Traps located at deeper water were kept afloat by attaching empty plastic bottles to them. Traps were checked early in the morning and late in the afternoon. The stream was sampled along an approximately 1.3 km stream reach at 52 fixed and regularly spaced sampling points. For each sampling point, 17 environmental variables were recorded (Table 5.1). Vegetation cover and substratum were visually assessed by the same observer for an area of approximately 5m radius surrounding the trap and the remainder variables were extracted for an approximately 10m side transversal strip of the stream. A strong collinearity (Pearson's r > 0.75) was detected between two pairs of environmental variables: submerged vegetation versus floating vegetation (r = 0.92) and water permanency versus maximum depth (r = 0.91). Since data analysis is based solely on univariate relationships there is no direct effect on the statistical estimates. Furthermore, although there is a spatial coincidence between submerged and floating vegetation cover, each variable showed a peak at different seasons. Nevertheless, interpretation of results should take this fact into careful consideration and floating vegetation cover, each variable showed a peak at different seasons. Nevertheless, interpretation of results should take this fact into careful consideration.

In the first sampling year (2003), from March to September, 16 transects along the margins of the stream reach were performed by the same observer, during which all turtles in basking activity were counted with the aid of a telescope (15–45x). The stream was divided into 51 sectors each one between two consecutive trapping points. These counts were used exclusively for between species comparisons, since their spatial distribution depends on many uncontrolled factors, such as the solar angle of the sector during the counts and the effects of habitat-related visibility (López *et al.* 2005).

Data analysis

To test for spatial associations between the distributions of species' abundance and to seek for relationships between those abundances and environmental variables along the stream reach, partial Spearman correlations were calculated removing the effect of water permanence. Spatial association between species was also calculated separately for each sex and age class of *M. leprosa.* Age classes were defined according to Keller's (1997) classification based on the carapace length. To avoid inflation of the coefficients' significance due to spatial autocorrelation (Lennon 2000, Segurado *et al.* 2006 - Chapter 3), a null model approach using one-dimension toroidal shifts (Fortin & Jacquez 2000, Dale & Fortin 2002) was performed to estimate the significance of correlations.

The procedure comprised the following steps: (1) the two-dimensional spatial coordinates were previously transformed into one-dimensional coordinates rescaled from 1 to 52, by stretching the stream reach into a line; (2) one thousand null-patterns were generated by moving spatial coordinates of the original variable using a common random integer factor in both geographical directions; cells that were shifted beyond one side of the range of actual coordinates were moved to the opposite side of the range (toroidal shifts); (3) for each permutation an inversion of coordinates was also allowed in order to attain a greater number of null-pattern configurations; (4) an adjusted P-value was obtained by calculating the proportion of the resulting 1000 correlations coefficients (Rho) that lied below the Rho of the correlation between the two original data sets.

This randomization procedure assures that the main spatial structure is maintained at each permutation. The adjusted P-value expresses the probability of the original Rho to be greater or lower than that expected from chance. The P-value was interpreted as a two-tailed test (values close to zero indicated negative associations and values close to one indicated positive associations).

We also assessed the microhabitat features that most differed between species at each hydrological season. Since there was a great asymmetry between species' number of captures (*E. orbicularis* represented only about 6% of the total turtles captured), a more robust approach was used. We used a procedure for testing whether the few *E. orbicularis* observations could be a random subsample of *M. leprosa* observations. For each independent variable, 10,000 random subsamples of *M. leprosa* observations were each tested for differences with the sample of *E. orbicularis* observations. The mean test statistic was computed and the associated P-value was used as the probability of the two samples to belong to the same population. Wilcoxon–Mann–Whitney (WMW) tests (Sprent & Smeeton 2001) were used for most variables to test for differences in the location of the two samples. Kolmogorov tests (Sprent & Smeeton 2001) were used to test for differences on linear spatial coordinates, since they are also sensitive to differences in the general shapes of the distributions in the two samples. Marked individuals caught within less than 30 days after the last capture were previously removed from the analysis in order to avoid temporal autocorrelation of data. This procedure was used separately for each season and also considering the whole sample. The same approach was also used for testing differences between species in the use of basking sites. Since only five basking individuals of *E. orbicularis* were observed during season 2 and none during season 3, the analysis was based on the merged data from the first two seasons. The asymmetry between species basking counts is even more marked than for captures (*E. orbicularis* represented only 2% of the total basking counts). This test was performed using only spatial coordinates, by allocating the number of observations in each sector to the coordinates of the first trapping point of the sector.

For both the approaches described above, no multiple testing adjustment of P-values were presented in the results, following recent objections for the use of such methods (Moran 2003). In fact, significances obtained from randomizations were interpreted simply as measures of spatial association/ dissociation between species and variables, and were mainly used for comparison purposes. Nevertheless, we reported how many tests remained significant after estimating adjusted P-values using both sequential Bonferroni adjustments and false discovery rate control (see e.g. Verhoven *et al.* 2005 for further details on the methods). We also used the Truncated Product Method

(TPM; Zaykin *et al.* 2002), as suggested by Neuhaüser (2004), as an indicator of the overall probability of type I errors amongst the significant P-values. The TPM is based on the product of those P-values that do not exceed the selected critical significance level (e.g. 0.05); if the P-value obtained with TPM is inferior to the significance level then we can declare that there is at least one truly false null hypothesis amongst the significant tests. All adjusted P-values were calculated separately for each species and season. All data analyses were performed using standard routines and programming capabilities of S-PLUS, 2000 (Statistical Sciences 1999), except for the TPM P-value that was computed using the program available at ftp://statgen.ncsu.edu/pub/zaykin/tpm.

5.1.4 Results

Patterns of abundance

At all three seasons, *M. leprosa* showed much higher abundances (980 'independent' captures of a total of 653 individuals, representing 0.727 mean captures per day - trap) than *E. orbicularis* (65 'independent' captures of a total of 42 individuals, representing 0.056 mean captures per day x rap) and for both species the highest abundances were observed on season 2 (Table 5.2). Only five captures of *E. orbicularis* corresponded to juvenile individuals. In basking counts the proportion of *E. orbicularis* observations was even lower (29 observations of *E. orbicularis* versus 1392 observations of *M. leprosa*; Table 5.2).

Table 5.2 - Total number of turtles captured per season, basking counts and mean abundance indexes (number of individuals per day x trap or number of basking counts per km) for each species (standard deviations are represented by numbers between brackets).

Data	*E. orbicularis*		*M. leprosa*	
	N	Abundance	N	Abundance
Captures - Season 1	27	0.059 (0.106)	288	0.491 (0.338)
Captures - Season 2	27	0.072 (0.151)	450	0.938 (0.655)
Captures - Season 3	11	0.031 (0.075)	242	0.761 (0.612)
Captures - Total	65	0.056 (0.118)	980	0.727 (0.577)
Basking counts	29	1.825 (1.748)	1392	87.000 (46.958)

As evidenced by the distribution of abundances (Fig. 5.2) the two species tend to show a complementary use of space along the stream. While *M. leprosa* is common throughout the stream segment except in sub-reach 2, *E. orbicularis* is more common on sub-reaches 2 and 3 (the most temporary sub-reaches) and fairly common on sub-reaches 1 and 6 (sub-reaches with emergent vegetation). The counts of basking individuals also suggest differences between species in the use of basking sites (Fig. 5.3). Basking individuals of *M. leprosa* were very abundant in sub-reach 5 (permanent and rocky sub-reach) while individuals of *E. orbicularis* tend to prefer basking sites near temporary sub-reaches.

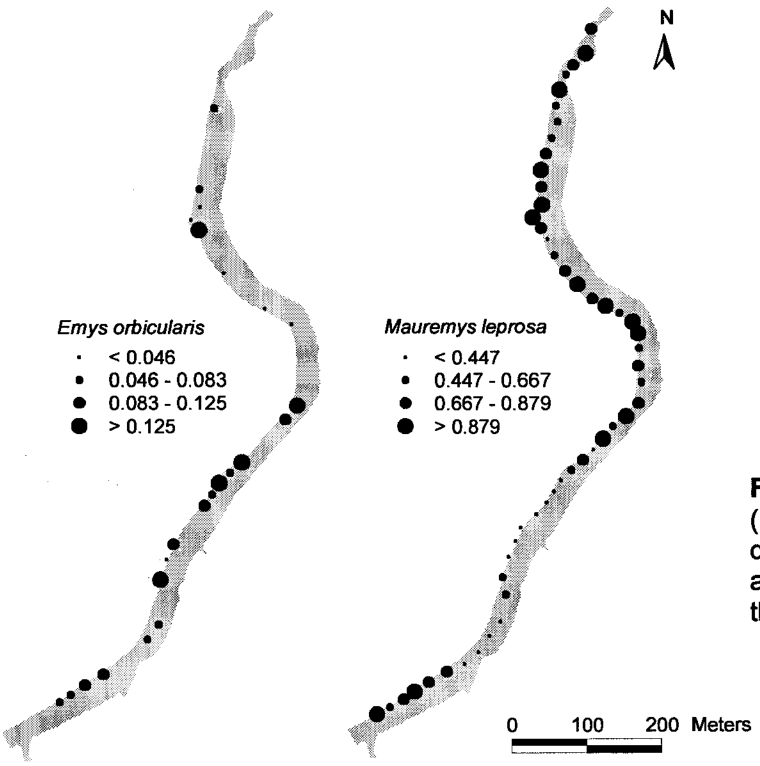


Fig. 5.2 – Mean abundance (number of captures per dayxtrap) of *E. orbicularis* (left) and *M. leprosa* (right) along the study reach.

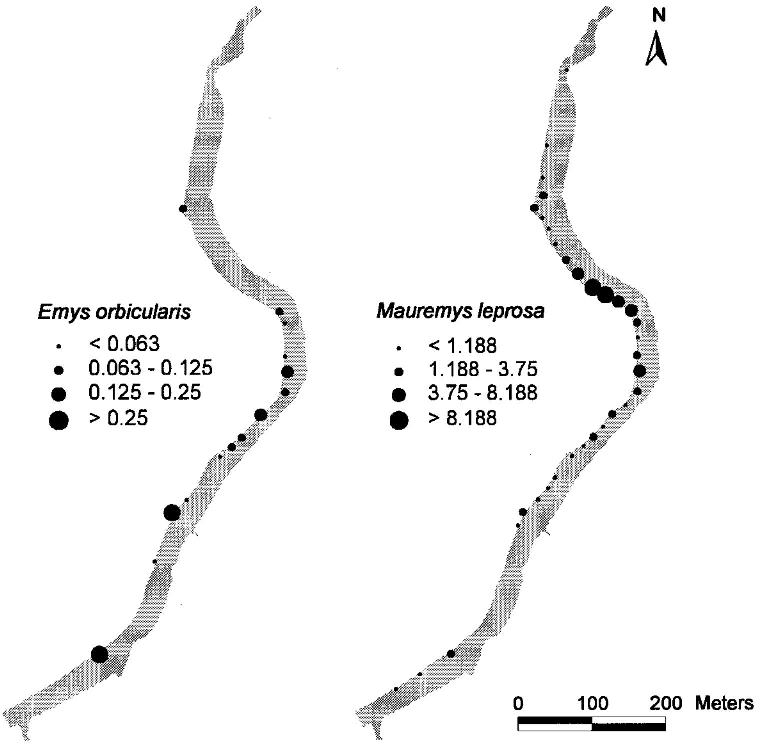


Fig. 5.3 – Mean basking counts of *E. orbicularis* (left) and *M. leprosa* (right) per transect along the study reach.

Spatial association

No evident positive or negative correlation was found between the two species' abundances along the stream segment (Table 5.3). However, when the correlations were computed separately for each sex and age class of *M. leprosa*, a fair negative correlation (adjusted significance close to zero) with males' abundance and a nearly significant positive correlation with juveniles' abundance were found (adjusted significance close to one; Table 5.3).

Table 5.3 - Adjusted significances of Spearman correlations between abundances of *E. orbicularis* and of each sex/age class of *M. leprosa*. Values bellow and above 0.5 represent, respectively, negative and positive relationships.

	Adjusted significances
E. orbicularis versus M. leprosa (Total)	0.236
E. orbicularis versus females of *M. leprosa*	0.156
E. orbicularis versus males of *M. leprosa*	0.080
E. orbicularis versus subadults of *M. leprosa*	0.543
E. orbicularis versus juveniles of *M. leprosa*	0.967

Association between species and microhabitat variables

Correlations between species and microhabitat clearly showed that *E. orbicularis* has stronger relationships with microhabitat variables than *M. leprosa* (Table 5.4). The abundance of *E. orbicularis* was significantly correlated with three variables (positive relationship with emergent vegetation and grass cover; negative relationship with bedrock), although showing a fair correlation with other three variables (positive relationship with distance to basking sites and sand; negative relationship with tree cover), while the abundance of *M. leprosa* only showed a fair correlation with one variable (positive relationship with dominant shoreline inclination). Considering sex and age class of *M.*

leprosa separately some variables gained more significance (Table 5.4), although only one significant correlation was found (negative relationship of subadults with distance to permanent sectors). The majority of the most significant variables were not common among different sex/age class of *M. leprosa*. The two most significant variables for juveniles of *M. leprosa* were common to two of the most significant for *E. orbicularis* (positive relationship with sand and negative relationship with bedrock). After computing sequential Bonferroni adjustments and false discovery rate control, no correlation remained significant. However, a significant TMP P-value was obtained for *E. orbicularis* ($P < 0.025$) indicating that at least one of the significant correlations is indeed significant. For subadults of *M. leprosa* the TMP P-value was not significant ($P > 0.025$) indicating that the single significant correlation found is most probably spurious.

When the adjusted significances of the relationships of *E. orbicularis* with microhabitat variables are plotted against those of each sex/age class of *M. leprosa*, some consistent patterns between species are revealed (Fig. 5.4). Correlations of the two species with microhabitat variables tend to have opposite signs, i.e., variables more positively correlated with one species tend to be more negatively correlated to the other species and vice-versa. Therefore a negative relationship between correlations' significances is evidenced (Fig. 5.4). This is particularly evident for adult females ($R^2 = 0.39$, $F = 8.45$, $P < 0.05$) and males ($R^2 = 0.32$, $F = 6.15$, $P < 0.05$) of *M. leprosa*, but less evident for subadults ($R^2 = 0.04$, $F = 0.56$, $P > 0.05$). In contrast, *E. orbicularis* and juveniles of *M. leprosa* showed a slightly positive relationship, although not significant ($R^2 = 0.08$, $F = 1.10$, $P > 0.05$).

Table 5.4 – Adjusted significances of Spearman correlations between abundances of each species (and also for each sex/age class for *M. leprosa*) and microhabitat variables (*significant correlations, P<0.025 or P>0.975)

Variables	*E. orbicularis*	*M. leprosa*				
		Total	Females	Males	Subadults	Juveniles
Stream width	0.817	0.473	0.362	0.631	0.361	0.237
Maximum depth	0.145	0.802	0.922	0.782	0.522	0.474
Distance to permanent sectors	0.711	0.201	0.278	0.194	0.015*	0.305
Distance to basking sites	0.927	0.388	0.393	0.454	0.491	0.801
Dominant shoreline inclination	0.160	0.902	0.686	0.655	0.973	0.646
Bedrock	0.012*	0.713	0.883	0.912	0.483	0.049
Gravel	0.474	0.781	0.310	0.923	0.802	0.620
Sand	0.971	0.162	0.129	0.135	0.396	0.791
Mud	0.758	0.277	0.205	0.184	0.873	0.655
Emergent vegetation	0.993*	0.262	0.249	0.111	0.770	0.756
Submerged vegetation	0.789	0.723	0.885	0.725	0.191	0.214
Floating vegetation	0.799	0.456	0.586	0.384	0.394	0.283
Grass cover (shoreline)	0.994*	0.125	0.028	0.113	0.685	0.964
Shrub cover (shoreline)	0.541	0.694	0.760	0.450	0.439	0.836
Tree cover (shoreline)	0.084	0.294	0.414	0.215	0.584	0.738

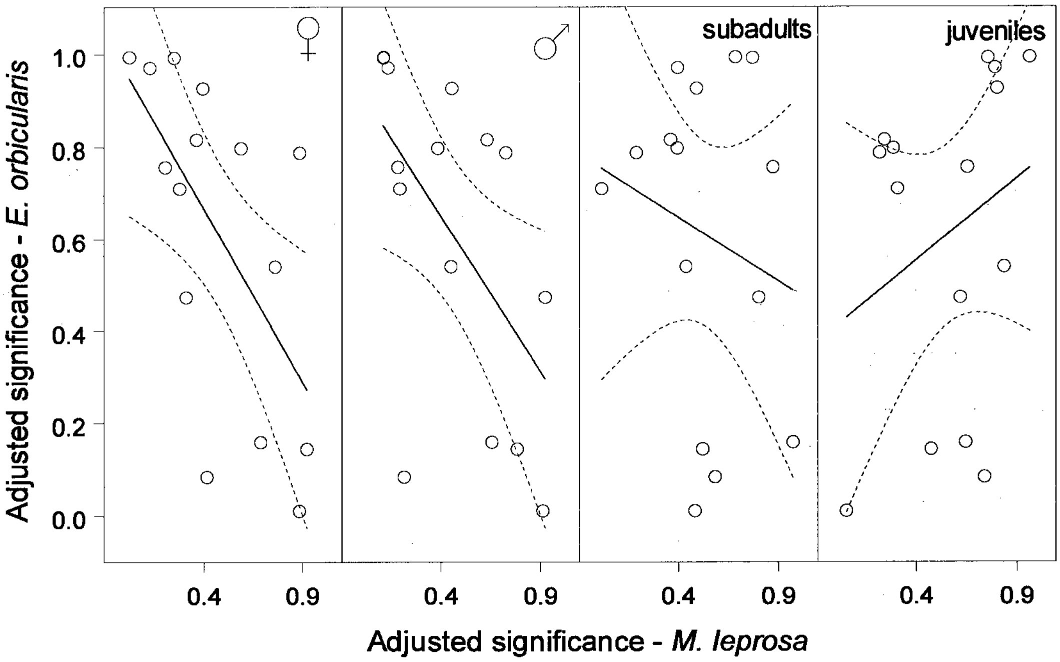


Fig. 5.4 – Relationship between *E. orbicularis* and each sex/age class of *M. leprosa* in the adjusted significances of each microhabitat variable (dashed lines represent 95% confidence intervals).

Differences between species in the use of space

Considering the whole sample, the variables for which species most differ were linear spatial coordinate, water permanency, maximum depth, sand, emergent vegetation, grass cover, dominant shoreline inclination and bedrock (Table 5.5). The most significant differences between species' use of space and environmental variables were observed during season 1 (Fig. 5.5). The most significant variables were common to those of the whole sample except for dominant shoreline inclination, bedrock (both less significant during the first season), distance to permanent sectors and tree cover (both more significant during the first season; Table 5.5). The estimated difference of spatial coordinates between species' basking sites resulted very significant (mean $P < 0.0003$). After the sequential Bonferroni adjustments only one test remained significant for the total sample (differences between spatial coordinates of locations during basking activity), while the false discovery rate control indicated two more significant variables (emergent vegetation cover and grass cover; Table 5.5). For season 1 the water permanency and maximum depth were the most significant variables (nearly significant after false discovery rate control). Highly significant TMP P-values were obtained for both the total and season 1 ($P < 0.001$) indicating that most significant tests are probably truly significant.

Table 5.5 – Mean significance of the tests for each one of the 10,000 resamples of *M. leprosa* and for each variable, expressing the probability of the *E. orbicularis* sample and the *M. leprosa* sample to belong to the same population (*mean P-values<0.05; **significant variables after false discovery rate control; ***significant variables after sequential Bonferroni adjustments)

Variables	Total	Season 1	Season 2	Season 3
Spatial coordinate (captures)	0.016*	0.020*	0.542	0.269
Spatial coordinate (basking)	0.001***	-	-	-
Water permanence	0.032*	0.006*	0.516	0.598
Stream width	0.524	0.520	0.162	0.419
Maximum depth	0.033*	0.006*	0.635	0.608
Distance to permanent sectors	0.143	0.026*	0.408	0.646
Distance to basking sites	0.115	0.056	0.613	0.257
Dominant shoreline inclination	0.082	0.112	0.611	0.508
Bedrock	0.095	0.360	0.402	0.240
Gravel	0.407	0.532	0.559	0.081
Sand	0.014*	0.066	0.412	0.091
Mud	0.393	0.650	0.283	0.114
Emergent vegetation	0.003**	0.037*	0.189	0.149
Submerged vegetation	0.357	0.176	0.437	0.601
Floating vegetation	0.339	0.131	0.418	0.513
Grass cover (shoreline)	0.006**	0.022*	0.278	0.221
Shrub cover (shoreline)	0.613	0.290	0.207	0.159
Tree cover (shoreline)	0.432	0.037*	0.550	0.166

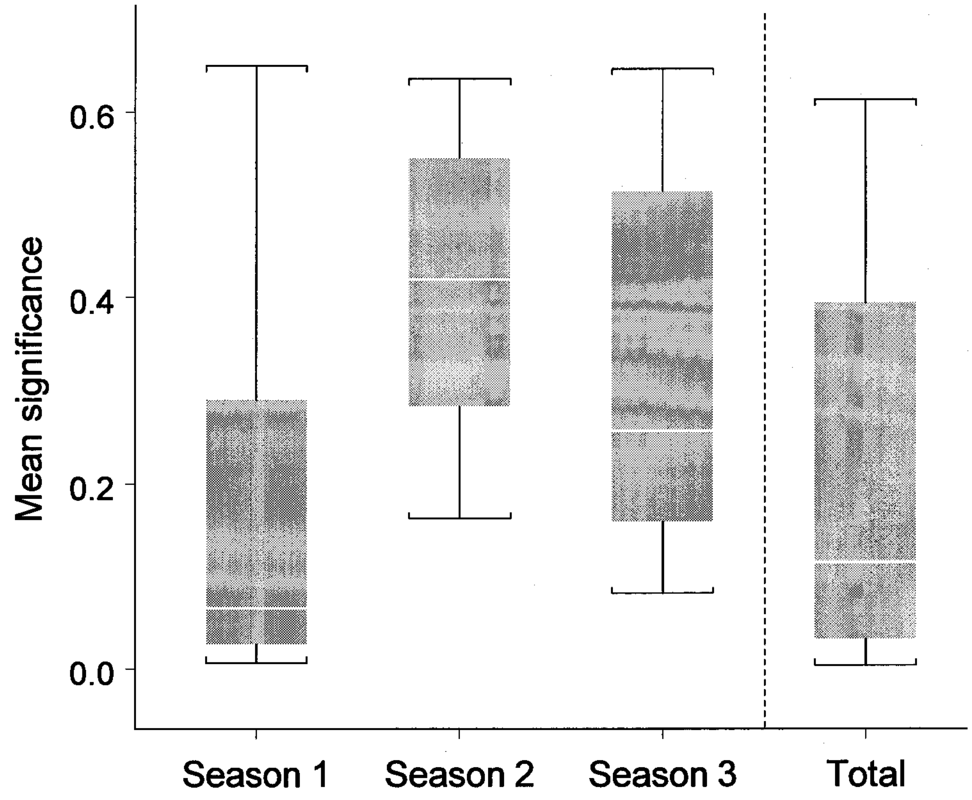


Fig. 5.5 – Boxplots showing the distribution of the mean significance of the tests for each one of the 10,000 resamples of *M. leprosa*, expressing the probability of the *E. orbicularis* sample and the *M. leprosa* sample to belong to the same population.

5.1.5 Discussion

A considerable amount of spatial segregation was detected between *E. orbicularis* and *M. leprosa*, which to our knowledge has never been demonstrated using an objective and quantitative approach. This spatial segregation is particularly relevant considering the small spatial extent of the study area in relation to turtles' mobility. Indeed, results suggest that spatial segregation at this scale is more evident than at wider spatial scales (Segurado & Araújo 2004a, see Section 5.2 of this chapter). Furthermore this segregation is not purely spatial: species also tend to respond to microhabitat variables in opposite directions. This divergence is particularly evident between *E.*

orbicularis and adults of *M. leprosa* during the first period of turtles' activity, when water generally flows throughout the full extent of the stream. On the other hand *E. orbicularis* shows a stronger response to microhabitat features than *M. leprosa*. Overall, the results suggest that *E. orbicularis* occurs preferentially on temporary, shallow, well vegetated and sandy reaches, while *M. leprosa* is less selective regarding microhabitat, although tending to occupy deeper and rocky reaches of the stream segment. The variables for which species' response most diverge were both related to vegetation cover: emergent vegetation cover and shoreline grass cover. As mentioned above, water permanency and maximum depth are greatly inter-correlated, yet it is most probable that species are responding to water depth rather to its permanency. Water depth might have an indirect effect on *E. orbicularis* since it has strong implications on vegetation growth; indeed, sites with dense vegetation cover have necessarily low water depths. Results also showed the importance of taking into account temporal variation throughout the hydrological cycle when analysing species coexistence. Analysing data without season discrimination may fail in detecting differences in the use of space.

Although lacking quantitative data analysis there are evidences of similar differences in the use of space at wider geographical scales in several regions of the Iberian Peninsula. For example, in SW Spain (e.g. Doñana Natural Park) and SW Portugal, *E. orbicularis* tends to occupy coastal temporary pond systems, while *M. leprosa* tends to prefer larger permanent ponds, rivers and streams (Keller 1997, P. Segurado, unpublished data). However, these results contrast with other quantitative studies on habitat selection by *E. orbicularis* where a preference for large permanent water bodies is documented, such as in the Po River delta of northern Italy (Ficetola *et al.* 2004). Furthermore, the preference for lower courses of large permanent rivers is documented for the southern European Russia (Bozhansky & Orlova 1998). This may be due to different local adaptations of the species or as a result of different biotic pressures (e.g. competition with other freshwater turtles) at distinct parts of the geographical range, which is very likely to occur in widespread species.

Possible mechanisms underlying spatial segregation

The complementary use of space in the study area can be due to different foraging behaviours and use of food resources by species. In fact, some authors suggest that *E. orbicularis* tends to have a more strictly carnivorous diet than *M. leprosa* (e.g. Keller and Busack 2001), although recent data show a high frequency of plant matter in the adults' diet of *E. orbicularis* (Ottonello *et al.* 2005). Nevertheless there are evidences that *E. orbicularis* selects smaller prey items such as macroinvertebrates, while *M. leprosa* favours larger animals like fish and crayfish, although feeding largely on plants (C. Keller & J.L. Garcia-Mudarra, unpublished data). On the other hand it was shown that tadpole consumption was much higher for *E. orbicularis* than for *M. leprosa* (Gomez-Mestre & Keller 2003). These differences in the use of food resources will most probably reflect differences in the use of space. For example, aquatic vegetation cover provides protection and oviposition sites for amphibians and macroinvertebrates, which in part could explain the preference for well vegetated reaches by *E. orbicularis*.

An alternative explanation is that spatial segregation between species is a consequence of inter-specific interactions, either direct (behavioural) or indirect (resource competition). Very few studies on behavioural interactions and resource competition between freshwater turtles have been carried out. One of such studies conducted at a similar spatial scale suggested that inter-specific competition was of little importance in structuring resource use by five sympatric turtle species in Kentucky, USA (Lindeman 2000). Behavioural interactions during basking activity were on the other hand evidenced in some studies (e.g. Lindeman 1999, Cadi & Joly 2003, Marco & Andreu 2005). Basking sites can be important limiting resources for freshwater turtles and therefore behavioural interactions between freshwater turtles' species during basking activity are likely to occur. In a controlled experiment, individuals of *E. orbicularis* were shown to shift their basking activity towards places considered of lower quality when the exotic turtle *Trachemys scripta* was present (Cadi & Joly 2003). Additionally, significant impacts of the exotic turtle T. scripta on the survival rates of *E. orbicularis* were shown to occur (Cadi & Joly 2004). In another experimental study, the interactions between *E. orbicularis*, *M. leprosa* and *T. scripta* were investigated and it was shown that spatial segregation during basking activity is very high for *E. orbicularis*, even for monospecific pairs (Marco & Andreu 2005). Factors affecting the outcome of species interactions during basking activity were investigated in a North America turtle assemblage comprising four species (Lindeman 1999). This study showed the occurrence of aggressive interactions between turtles during the basking activity and also that the outcome of this behaviour (displacements into the water) was exclusively related to turtles' body size: larger turtles displaced smaller turtles more often, independent of the species pair. Therefore, it is possible that *M. leprosa* will more often displace *E. orbicularis* during basking activity, since it attains larger body sizes. In fact, during our study most observations of basking individuals of *E. orbicularis* were made near the most temporary sectors of the study area where *M. leprosa* is less abundant. Furthermore, according to our results the distance of captured individuals to basking sites (exposed rocky shores with nearby aquatic refuges) was in average greater for *E. orbicularis*. This could be due to the very high abundance of *M. leprosa* at sites with better basking conditions.

Other mechanisms might be responsible for differences in aquatic microhabitat use by the species, such as differences in site selection for hibernation and aestivation, differences in the thermal efficiency of basking activity and different use of terrestrial habitats for feeding, migration or protection purposes. The clear preference of *E. orbicularis* for sites with dense emergent vegetation and shoreline grass cover might indicate a greater thermoregulation efficiency, since individuals are less exposed to the sunlight on these microhabitats. On the other hand it may be caused by differences in hunting behaviour and the use of terrestrial habitats. In fact, *E. orbicularis* often moves to land for different purposes, most probably including feeding activity (Ficetola & De Bernardi 2006), which in part could explain the preference for densely vegetated margins that would confer protection to turtles during movements between water and land. Further research is, however, needed to test the possible mechanisms underlying species' complementarity on the use of space.

Coexistence throughout the hydrological cycle

Independently of their intensity and influence on the spatial structuring of turtles' population, inter-specific interactions in the study area will most likely suffer great changes along the main activity period. Intra- and inter-specific interactions will probably increase from season 1 (March–April) to season 3 (mid-July–September), since according to our results spatial and environmental segregation between species is gradually cancelled and water availability becomes a strong limiting factor. Dietary shifts throughout the season were reported for *E. orbicularis* (Ottonello *et al.* 2005), suggesting that a shift from carnivorous to omnivorous diet occurs, probably linked to changes in habitat use. This could imply that the diet of the two species would be increasingly similar throughout the season. On the other hand it is possible that overall competition for food between *E. orbicularis* and *M. leprosa* is stronger for the juveniles of the latter species, since little difference in the use of space was detected between these two groups during all seasons. In fact, the affinities between *E. orbicularis* and juveniles of *M. leprosa* in the use of space may be explained by similar food requirements. Many freshwater turtles also switch diets between juvenile (often carnivorous) and adult (largely herbivorous) stages (McCauley & Bjorndal 1999, Burke *et al.* 2000, Siegel & Dodd 2000). This has been documented for *E. orbicularis* (Lebboroni & Chelazzi 1999, Ottonello *et al.* 2005) as well as for *Mauremys caspica* (Sidis & Gasith 1985), a species that is closely related to *M. leprosa*. Most likely the diet of *M. leprosa* will also follow this pattern.

The changes in patterns of coexistence between turtle species through the annual activity period can be explained in terms of changes in spatial heterogeneity. The maximum heterogeneity is probably attained during the first season since water is available throughout the stream reach and hence more diverse microhabitats are also available (temporary versus permanent sectors, vegetated versus non-vegetated sectors, etc.). During this period the two species have the opportunity to explore different space resources, according to their needs or biotic and physical constraints. Furthermore, juveniles of some species are known to change habitat use and feeding habits as body size increases and therefore spatially heterogeneous freshwater ecosystems may be critical to the maintenance of self-perpetuating populations of freshwater turtles (Burke *et al.* 2000). This is actually the case with *E. orbicularis* (e.g. Paul & Andreas 1998, Lebboroni & Chelazzi 1999, Meeske & Muhlenberg 2004, Cadi *et al.* 2004). Finally the most important facet of this spatial heterogeneity is that it probably allows the two turtle species to coexist in this particular stream ecosystem. Contrasting with this scenario, during the driest season turtles are forced to occupy the same kind of microhabitats, consisting usually of muddy water bodies lacking any kind of vegetation cover.

In opposition to the effect of hydrological cycle, competition for basking places may have a stronger impact just after wintering, i.e., during the first season of turtles' main activity period, when water temperature is low and the need for heat is high (Cadi & Joly 2004). Hence, this would be an additional explanation for the more obvious differences in the use of space observed during the first season.

Concluding remarks

As mentioned above spatial segregation between closely related species is not necessarily shaped by competition. To disentangle this issue, further research and testing should focus on other important differences between species of life history traits such as diet, basking activity and its relative efficiency, basking and nesting site selection and changes on the use of space throughout turtles' life. Furthermore studies comparing life history attributes between sympatric and allopatric populations would be important to detect eventual shifts on resource use, namely food, space and time, which could be indicative of inter-specific competition.

The results of this study suggest that adequate planning and habitat management actions for turtles' conservation may differ substantially between the two studied species at syntopic sites. Results also suggest that the ecological features of *M. leprosa* are not good surrogates for those of *E. orbicularis*. Since spatial requirements of the two species on Mediterranean streams seem to depart into opposite directions the most adequate management practices must be carefully implemented. If *E. orbicularis* occupies a much narrower portion of the environmental gradient at biogeographical and habitat scales than *M. leprosa* (Segurado & Araújo 2004a, section 5.2 of this chapter) and, as showed in this study, has a stronger response to microhabitat characteristics, priority of management practices should go towards the conservation of this species.

Maintaining high levels of spatial heterogeneity is probably the best approach for the conservation of both species at syntopic habitats not only because it enhances turtles coexistence but also because turtles' life history, and specially that of *E. orbicularis*, depends on the availability of different microhabitats for feeding, reproduction, growth and protection purposes.

Acknowledgments

We are grateful to Susana Rosa and Rui Rebelo for their review of the manuscript and help during fieldwork. We also thank two anonymous referees for their comments. P. Segurado was supported by a grant from Fundação para a Ciência e Tecnologia (SFRH/BD/8493/2002).

5.2 Working example 2: Coexistence of two freshwater turtles at the habitat and biogeographical scales

Note – This study has been published in Biologia, Bratislava under the title "Coexistence of *Emys orbicularis* and *Mauremys leprosa* in Portugal at two spatial scales: is there evidence of spatial segregation?" (Segurado, P. & Araújo, A.P.R., 2004. *Biologia*, In press). The text has been slightly modified for layout harmonization (e.g. citation format) or for some minor English corrections.

5.2.1 Summary

In the Iberian Peninsula the western limit of *Emys orbicularis* geographic range overlaps with the northern limit of that of *Mauremys leprosa*. Given their distinct distributions, which probably reflect distinct ecological adaptations, some geographic segregation would be expected between these species in the Iberian Peninsula. Coexistence patterns of both species in Portugal were analysed at biogeographic and habitat scales. Occurrences were related to environmental variables using predictive modelling procedures and a Geographic Information System (GIS), in order to compare the species in their response to environment. At the habitat scale, *E. orbicularis* occurrences were related to the abundance of *M. leprosa* estimated through visual surveys of 200 m transects along streams. Some of the most important variables affecting species presence at the two spatial scales studied were common to both species. At both scales, there was a positive correlation for the two species in the modelled probabilities of occurrence. At the biogeographic scale, the highest probabilities of occurrence were predicted for the southeast of the country for both species. At the habitat scale, *E. orbicularis* tended to occupy sites where *M. leprosa* attained higher abundances. According to the results no marked spatial segregation between *E. orbicularis* and *M. leprosa* seems to occur in Portugal, at least regarding the scales analysed in this study.

Key words: *Emys orbicularis*, *Mauremys leprosa*, species coexistence, biogeography, distribution, habitat modelling, Portugal.

5.2.2 Introduction

In the Iberian Peninsula the European pond turtle, *Emys orbicularis* (L., 1758), occurs along with another freshwater turtle, the stripe-necked turtle, *Mauremys leprosa* (Schweigger, 1812). For both species, the Iberian Peninsula corresponds to the edge of their geographic range, namely, the western limit for *E. orbicularis* and the northern limit for *M. leprosa*. Although in Portugal the species ranges widely overlap, *E. orbicularis* is much rarer and shows a far more fragmented pattern of occurrence than *M. leprosa* (Araújo *et al.* 1997, Godinho *et al.* 1999) (Fig. 5.6).

Their distinct distributions would be expected to reflect differences in some ecological traits, especially in their tolerance to climatic extremes. Therefore, some kind of geographic segregation should be expected between the patterns of occurrence of the two species in the Iberian Peninsula. One piece of evidence of such segregation is that *E. orbicularis* occurs alone in a few isolated populations in the northwestern Iberian Peninsula, where an Atlantic climate prevails (Araújo *et al.* 1997, Ayres & Cordero 2000, 2001, Segurado, unpubl. data).

On the other hand, despite having distinct biogeographic traits, these species tend to show similar ecological requirements, namely regarding habitat use. In fact, in the wild they are often observed basking in mixed groups. Interaction between species would also be expected to produce some kind of spatial segregation between them at finer scales. Evidence of segregation at the habitat level was found in the water bodies of Doñana National Park (Southwest Spain), where *E. orbicularis* tends to occupy small temporary ponds while *M. leprosa* is more frequent in larger and permanent ponds (Keller *et al.* 1995, Keller 1997). In other regions of the Iberian Peninsula where both species occur, *M. leprosa* is more of a generalist and tends to occupy rather permanent water bodies and rivers (e. g. Gómez & Lizana 2000).

This study is a first attempt to test quantitatively the pattern of coexistence of the two species. Since coexistence is undoubtedly a scale dependent phenomenon (e. g. Levin 1992, May 1994), species occurrence was compared at two distinct spatial scales, namely at the (1) biogeographic scale, in which the species coexistence analysis was based on their distribution pattern in Portugal at a coarse resolution and at the (2) habitat scale, in which the analysis was based on species counts at a finer resolution. For both scales we used a predictive modelling approach (Walker 1990, Buckland & Elston 1993, Guisan & Zimmermann 2000) to examine species response to environmental variables and to test for relationships between species abundances.

5.2.3 Materials and methods

Biogeographic scale data

Data on species distribution was derived from 993 UTM (Universal Transverse Mercator) 10 by 10 km grid cells, and was compiled from three sources: an updated atlas of the Portuguese mainland herpetofauna (Godinho *et al.* 1999), the UNIBA database (www.cea.uevora.pt/umc), and personal unpublished data. The data matrix for species presence included 65 cells for *Emys orbicularis* and 435 cells for *Mauremys leprosa* (Fig. 5.6).

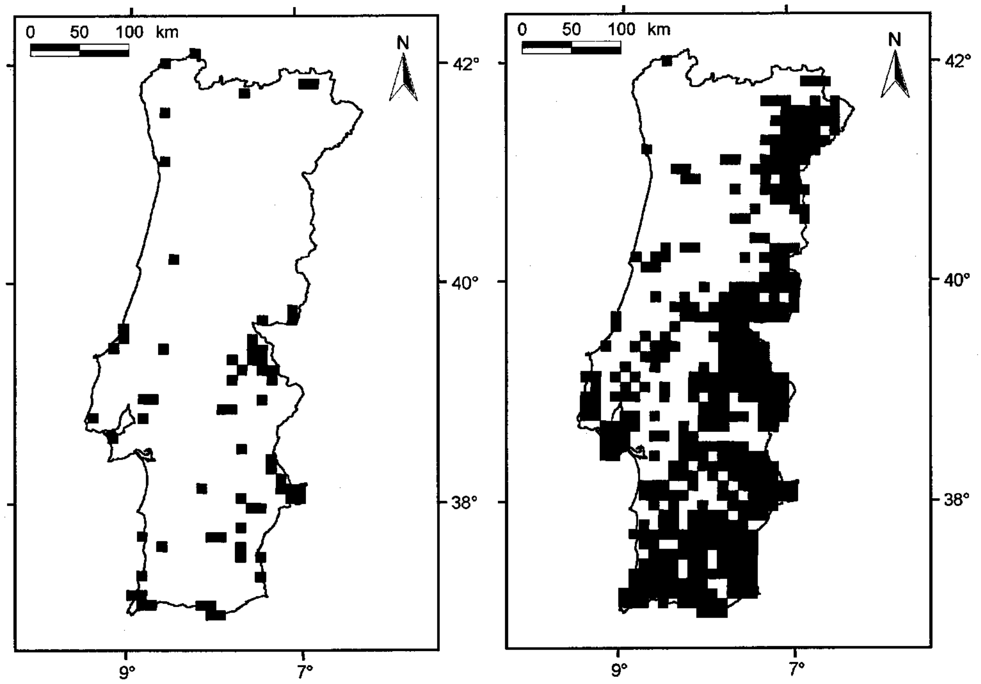


Fig. 5.6. - Distribution of *Emys orbicularis* (A) and *Mauremys leprosa* (B) in Portugal.

Forty-three macro-environmental variables, including climatic, geomorphologic, habitat, socio-economic, and landscape metric variables (Appendix 1), were compiled from various digital sources. Ten environmental, seven geomorphologic and human population density data were compiled from available digital layers of the Portuguese environmental atlas (http://www.iambiente.pt). Habitat data included nine variables selected from the Portuguese EU CORINE land cover database (http://www.snig.igeo.pt). Seven landscape metric variables were extracted using the Patch Analyst (Elkie *et al.* 1999) extension for ArcView version 3.2 (ESRI 1999). These include measurements related to landscape diversity, patch size, patch density, and patch shape (Forman 1995). Data were converted from available vector maps and resampled to raster UTM 10 by 10 km grid cell maps, using zonal functions (Tomlin 1990) in ArcView to calculate the mode of each variable at each cell.

Habitat scale data

A crude estimation of species abundance was carried out through visual surveys in approximately 200 m transects along the coastlines of streams, ponds or reservoirs. A total of 248 sites were studied from 1996 to 1998. The sites were distributed more or less evenly throughout the geographic range of *M. leprosa* (Fig. 5.7) in order to avoid mixing the effects of macroenvironmental variables in the habitat-abundance relationships. For each site, 15 environmental descriptors related both to water body and coastlines were measured (Appendix 2).

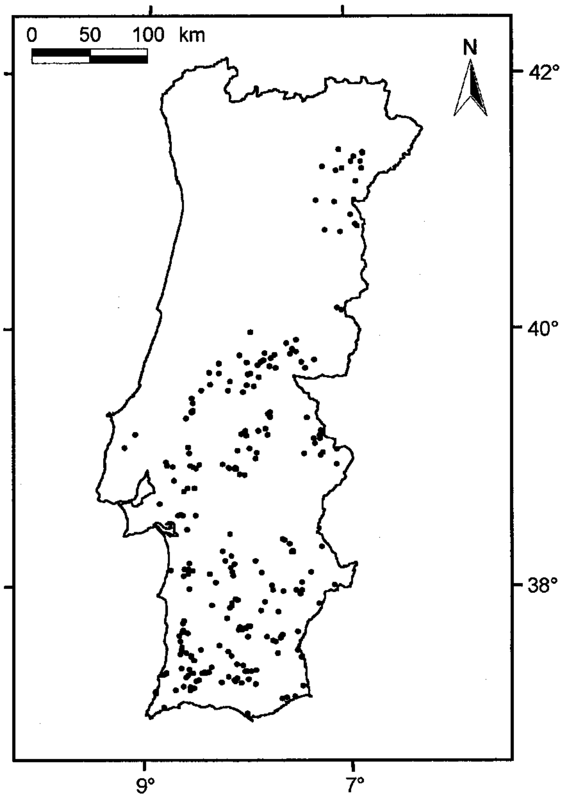


Fig. 5.7 - Distribution of *Emys orbicularis* (A) and *Mauremys leprosa* (B) in Portugal.

Data analysis

Multivariate regression techniques were used to detect the influence of environmental variables on species occurrence and compare the spatial patterns of probability of occurrence between species. For the habitat scale data, a Generalized Linear Model (GLM; Mccullagh & Nelder 1989) approach assuming a binomial error distribution (logistic regression) was performed. At the biogeographic scale the same comparison was based on the output of another related family of models – the Generalized Additive Models (GAM; Hastie & Tibshirani 1990) – also assuming a binomial error distribution. The advantage of GAMs is that allows the consideration of more complex response shapes than the linear responses of GLMs. The GAM output, like GLM, is a regression equation but it includes (not necessarily all) smoothed (non-linear) terms. A cubic spline with 4 degrees of freedom was adopted as the smoother function.

Both these approaches use presence/absence as dependent variables and allow the prediction of the species probability of occurrence at a given site (for a review see Guisan & Zimmermann 2000).

For the habitat scale, we opted not to employ models that use abundance as the response variable because turtle counts were very crude estimates of abundance. These were based on single runs, which made them heavily influenced by weather conditions and time of the day. For both approaches, a backward stepwise elimination procedure based on Akaike's Information Criterion statistics (AIC; Venables & Ripley 1997) was used to select variables to enter the models.

Model performances were quantified using the AUC index, which represents the area under the ROC (Receiver Operator Characteristic) curve (Fielding & Bell 1997, Cumming 2000). We used the Harrell *et al.* (1984) approach to estimate AUC. This method compares each predictive score associated with one outcome (say presence) to every score with the other (say absence). If the rank order of the predictive scores is in the same direction as the outcomes, the pair is concordant; if the scores are identical, the pair is tied; if the rank order of the scores is in the opposite direction as the outcomes the pair is discordant. The calculation of AUC is as follows:

AUC =(concordant +0.5 * tied) / total comparisons

This measure assesses how far from chance the model predicts occurrence, varying from 0.5 (random classification) to 1 (perfect classification). The biogeographic models were extrapolated to all Portuguese 10 by 10 km grid cells in order to produce surfaces of species probability of occurrence, using ArcView GIS version 3.2 (ESRI 1999).

Univariate approaches were also adopted in two cases: (1) to assess the relationship between the abundance (or modelled probabilities of occurrence) of the two species; (2) to compare the influence of the variables that were selected by both species' models. Since for both spatial scales the number of sites where *E. orbicularis* occurs is greatly inferior to those of *M. leprosa*, the following resample procedure was adopted: 10,000 random subsamples of *M. leprosa* sites of the same size as the *E. orbicularis* sample were taken and the median of the measure in question was calculated for each subsample; the distribution of the medians was then compared to the median for the *E. orbicularis* sample and a measure of the significance probability was obtained. This procedure allows testing if sites where *E. orbicularis* occurs are random samples of those where *M. leprosa* is present for a given measure. If not, it means that *E. orbicularis* tends to occur in sites corresponding to one of two extremes of *M. leprosa* occurrence: either where the measure in question (e.g. abundance) has very low values or where it has very high values. For the habitat scale data this procedure was used to compare *M. leprosa* absolute frequency at sites where *E. orbicularis* also occurs to random subsamples of all *M. leprosa* sites. For the biogeographic scale the same comparison was made using the probability of occurrence of *M. leprosa* instead (which ultimately reflects abundance) in 10 by 10 km grid cells. Finally, for both scales this procedure was also used to compare medians of the variables that were selected in both species' models. All statistical analyses were performed with Splus, version 2000 for Windows (Statistical Sciences 1999).

5.2.4 Results

Biogeographic scale

According to the AUC values (Table 5.6), the GAM predictive models performed well for both species, especially for *Mauremys leprosa*. According to Swets' (1988) rule of thumb (0.90–1 = excellent; 0.80–0.90 = good; 0.70–0.80 = fair; 0.60–0.70 = poor; 0.50–0.60 = null), AUC results are good for *Emys orbicularis* (AUC = 0.833) and excellent for *M. leprosa* (AUC = 0.924).

Model outputs in the form of probability maps show that the areas with higher probability values of *E. orbicularis* occurrence are largely included in regions where the probability of *M. leprosa* occurrence is also high (Fig. 5.8). For both species there is a general gradient of increasing probability of occurrence from Northwest to Southeast, that is, from the more Atlantic to the more Mediterranean region of Portugal, although the model predicts some dispersed occurrences in the northwest for *E. orbicularis*. Grid cells with higher values of probability for *E. orbicularis* occurrence tend to have higher values of probability for *M. leprosa* occurrence as evidenced in Figure 5.9.

There are four common variables to both species, selected in the regression equations of the GAM models (Table 5.6): (1) solar radiation; (2) edge density; (3) temperature; and (4) water drainage. Furthermore, the behaviour of the variables as shown by partial regressions of each variable in the model (Fig. 5.10) is very similar in both species. Both tend to show a belly shape response to the edge density of landscape patches, a monotonic negative response to water drainage, and a monotonic positive response to solar radiation and temperature.

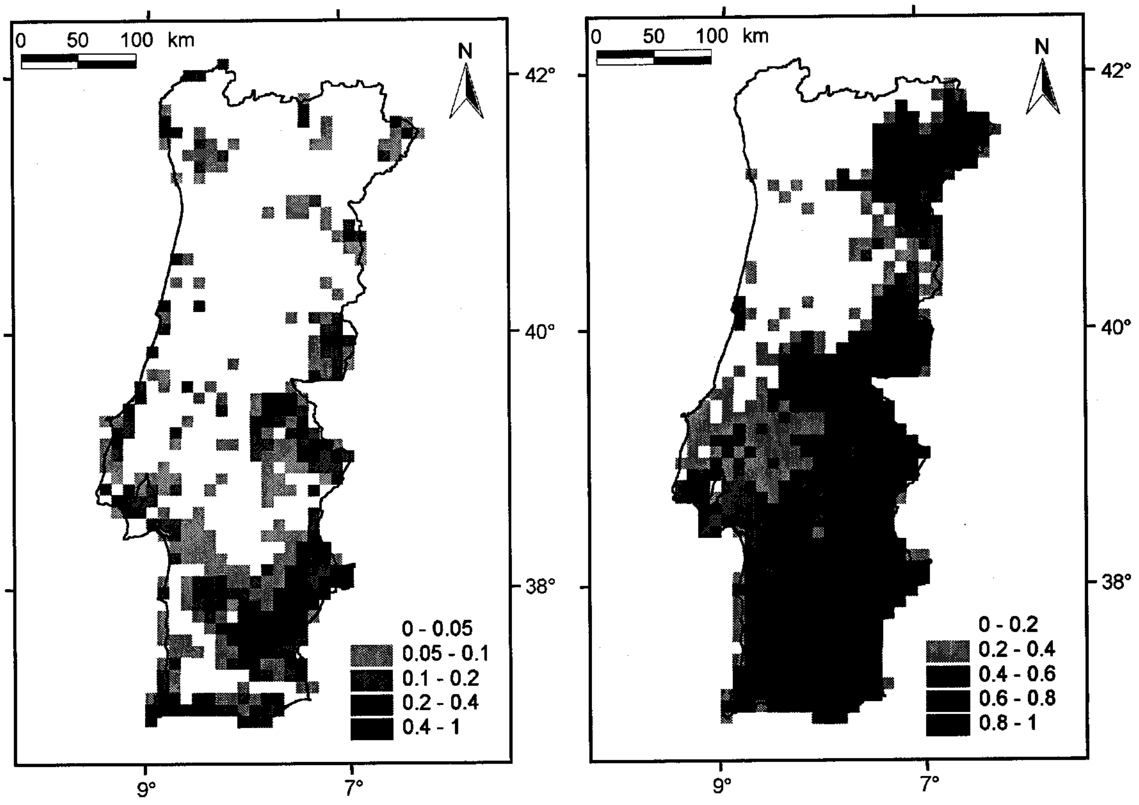


Fig. 5.8 - Modelled probabilities of occurrence of *Emys orbicularis* (A) and *Mauremys leprosa* (B).

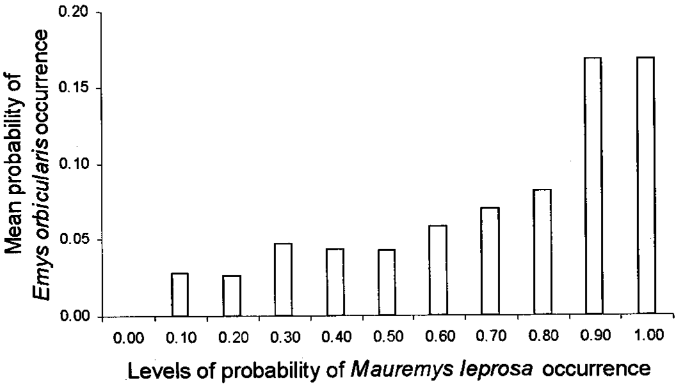


Fig. 5.9 - Mean probability of *Emys orbicularis* occurrence per 0.1 interval of probability of *Mauremys leprosa* occurrence.

Table 5.6 – Regression equations of the GAM models and the respective AUC index ("s" denotes the smooth functions for the terms that were smoothed).

Species	Equation	AUC
E. orbicularis	y = -3.225 - 0.075*drainage + 0.351*eco-regions + s(solar radiation) + s(edge density) + s(temperature) + s(human pop. density)	0.833
M. leprosa	y = -29.749 - 0.004*drainage + 0.251*solar radiation - 0.050*temperature + s(Evapotranspiration) + s(Altitude) + s(artif. landscape) + s(edge density)	0.924

The results of the univariate resampling technique (Table 5.7) show that there is no significant difference (less than 95% confidence) between the median of *M. leprosa* probabilities of occurrence for the grid cells where *E. orbicularis* occur and the medians of the random samples of the grid cells where *M. leprosa* is present. However, approximately 75% of the random subsamples of *M. leprosa* showed medians of probability lower than the median where *E. orbicularis* is present. The same analysis showed significant differences for two environmental variables: solar radiation and temperature. For both variables, in more than 95% of the random samples, sites where *E. orbicularis* occurs have higher median values than sites where *M. leprosa* is present. Figure 5.11 illustrates these results, showing clearly that *E. orbicularis* respond in the same direction as *M. leprosa* to these variables, but the median values of this species tend to be located in an extreme position among the median distribution of *M. leprosa* random samples.

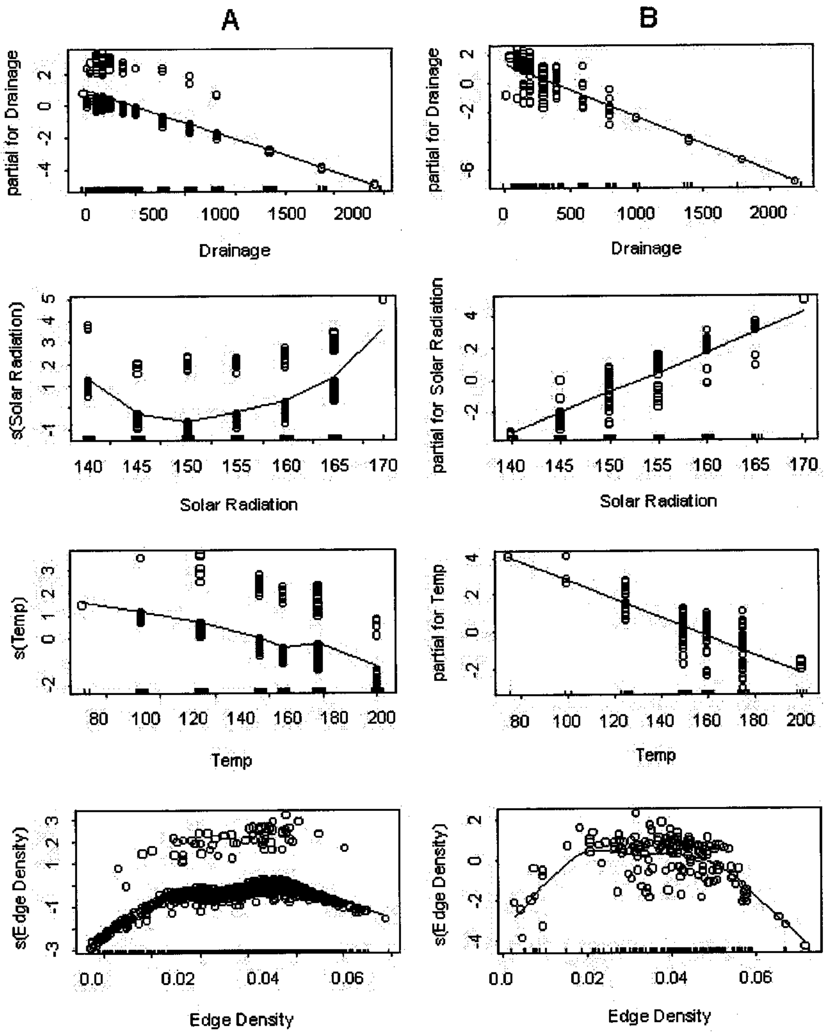


Fig. 5.10 - Partial regressions of the variables common to both species selected in GAM models (A: *Emys orbicularis*; B: *Mauremys leprosa*).

Table 5.7 – Results of the univariate resample analysis. The values are the percentage of medians of the *M. leprosa* random subsamples that are, respectively, inferior, equal and superior to the median of the *E. orbicularis* sample.

	% Inferior	% Equal	% Superior
Probability of *M. leprosa* occurrence	74.69	0.00	25.31
Solar Radiation	95.77*	4.23	0.00
Edge Density	85.02	1.01	13.97
Drainage	0.00	42.95	57.05
Temperature	99.72*	0.28	0.00

* - significant values

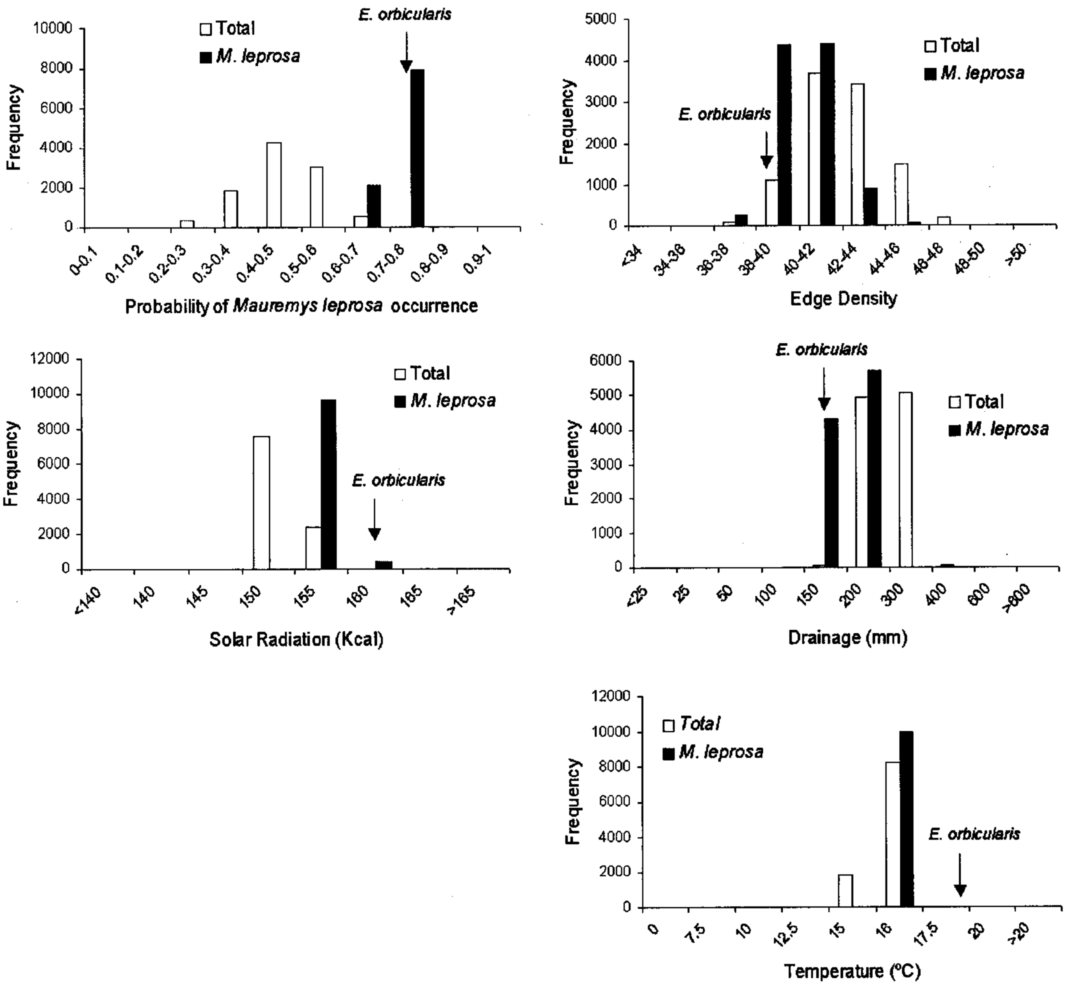


Fig. 5.11 - Distribution of the median values of 10,000 random subsamples of *Mauremys leprosa* observations for probability of *M. leprosa* occurrence, solar radiation, edge density, water drainage, and temperature. The median for the *Emys orbicularis* sample and the distribution of median of 10,000 random subsamples for the entire dataset are also shown.

Habitat scale

The AUC measures of the GLM for the habitat scale models (Table 5.8) can be classified as good for both species (AUC values included in the 0.80–0.90 interval). Although models were based only on species presence/absence data, the plots of Figure 5.12 show a fairly good agreement between the modelled probabilities of occurrence and the absolute frequency in the sample sites. Higher absolute frequencies were mainly observed at sites where the modelled probabilities of occurrence are also high.

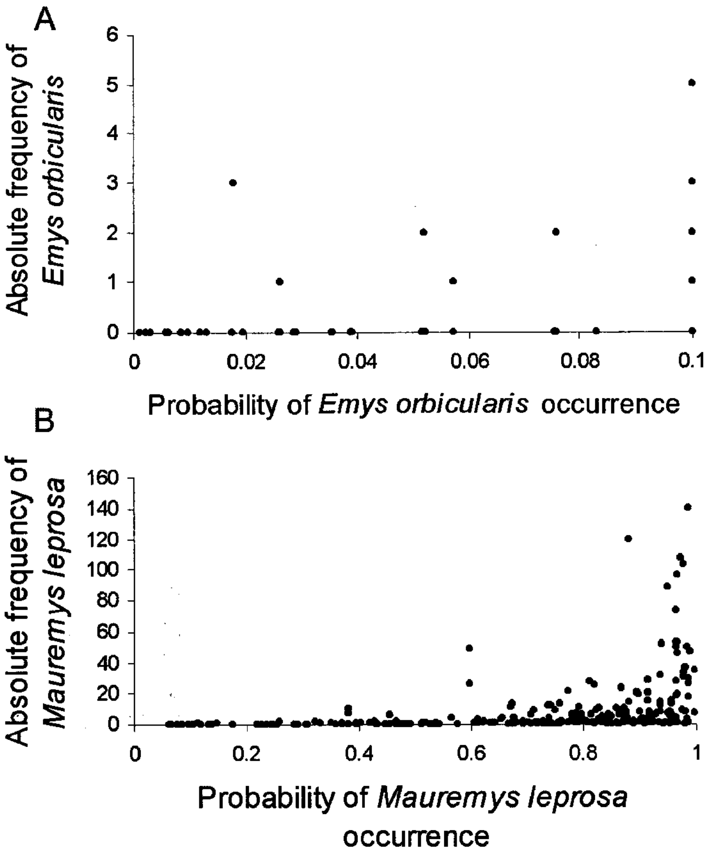


Fig. 5.12 - Absolute frequencies plotted against modeled probabilities of occurrence (A: *Emys orbicularis*; B: *Mauremys leprosa*).

Table 5.8 – Regression equations of the GAM models and the respective AUC index.

Species	Equation	AUC
E. orbicularis	y = - 3.621 + 0.713*emergent vegetation – 1.111*tree density – 1.012*gravel + 1.114*shrub	0.834
M. leprosa	y = - 0.136 + 0.634*depth – 3.82*current speed + 0.859*emergent vegetation + 0.685*submerged vegetation – 0.446*grass – 0.488*tree density	0.838

Sample sites with higher frequency of *E. orbicularis* occurrence tend to have higher frequency of *M. leprosa* occurrence. This pattern is obvious when the mean frequency of *E. orbicularis* is calculated per interval of *M. leprosa* frequency (Fig. 5.13).

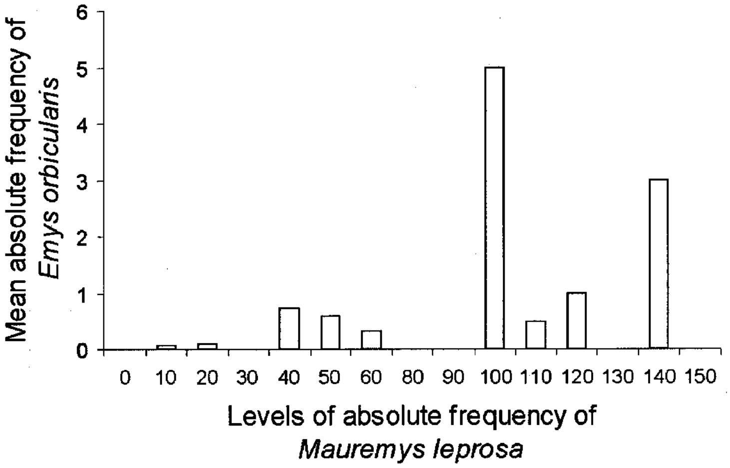


Fig. 5.13 - Mean absolute frequencies of *Emys orbicularis* per intervals of 10 of *Mauremys leprosa* absolute frequency.

There are two common variables to both species, selected in the regression equations of the GLM models (Table 5.8): emergent vegetation and tree density. Furthermore, the behaviour of the variables, as shown by partial regressions of each variable in the model (Fig. 5.14), is identical in both species. Both species tend to respond positively to the density of emergent vegetation cover and negatively to tree density on the water bodies coastlines.

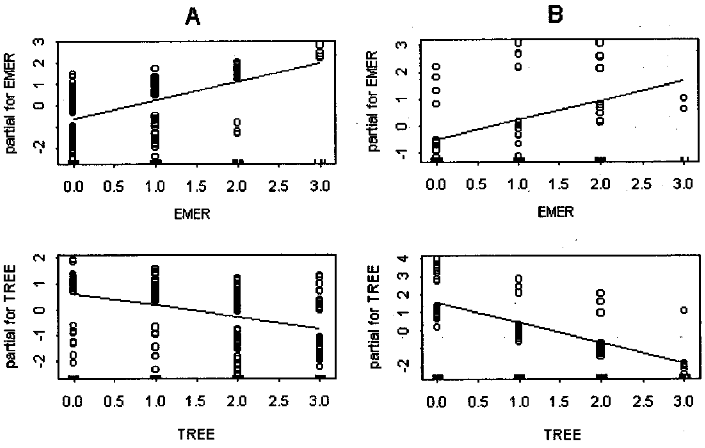


Fig. 5.14 - Partial regressions of the variables common to both species selected in GLM models (A: *Emys orbicularis*: B: *Mauremys leprosa*).

The results of the univariate resampling technique (Table 5.9) show that there are significant differences (95% confidence) between the *E. orbicularis* sample and the *M. leprosa* random samples regarding the medians of both *M. leprosa* absolute frequency and the modelled probabilities of occurrence. More than 95% of the random samples of *M. leprosa* sites have lower absolute frequencies and lower occurrence probabilities compared to the whole sample of sites where *E. orbicularis* is present. The same analysis did not reveal significant differences for the environmental variables. However, *E. orbicularis* tends to respond like *M. leprosa* to these variables but, as it was shown for the biogeographic scale, the median values of this species tend to be located in an extreme position among the median distribution of the random samples of *M. leprosa* (Fig. 5.15).

Table 5.9 – Results of the univariate resample analysis. The values are the percentage of medians of the *M. leprosa* random subsamples that are, respectively, inferior, equal and superior to the median of the *E. orbicularis* sample.

	% Inferior	% Equal	% Superior
Absolute frequency of *M. leprosa*	100.00*	0.00	0.00
Probability of *M. leprosa* occurrence	99.35*	0.00	0.65
Emergent vegetation	34.29	65.10	0.61
Tree density	4.13	3.60	92.27

* - significant values

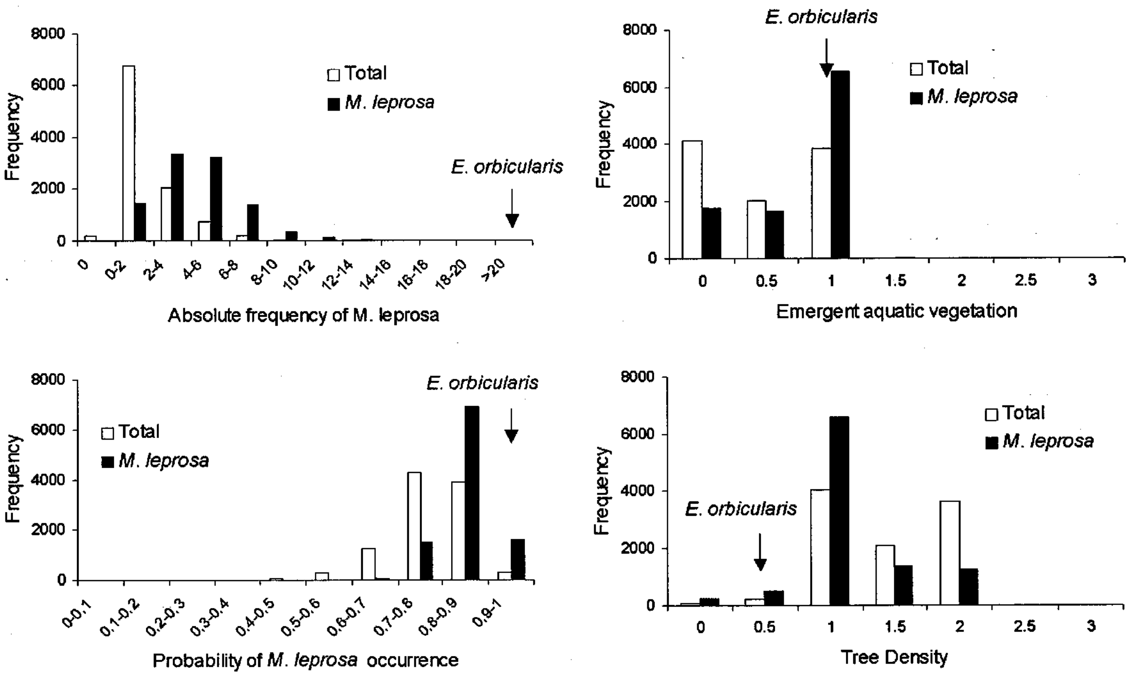


Fig. 5.15 Distribution of the median values of 10,000 random subsamples of *Mauremys leprosa* observations for absolute frequency of *M. leprosa*, probability of *M. leprosa* occurrence, emersed aquatic vegetation, and tree density. The median for the *Emys orbicularis* sample and the distribution of median of 10,000 random subsamples for the entire dataset are also shown.

5.2.5 Discussion

In Portugal *Emys orbicularis* is more commonly found in areas included in the geographic range of *Mauremys leprosa* and, among habitats within these areas, tends to occur more frequently at sites where *M. leprosa* attains higher abundances.

Both species are most common within areas with higher solar radiation, less water drainage and more homogeneous landscape. At areas where both species potentially coexist, both are most common in standing water habitats with highly emergent vegetation cover and low tree density on coastlines. Spatial coincidence of the major geographic variation (NW–SE gradient) among many of the macro-environmental variables makes it difficult to investigate possible causal relationships at the biogeographic scale, because the effect of one variable may be masked by the effect of another. On the other hand, macro-environmental variables can also act at the habitat scale. For example, northern Portugal is far more mountainous than the south, which in conjunction with higher values of precipitation can greatly limit habitat availability.

Our results also suggest that at both studied scales *E. orbicularis* tends to occupy a narrower and a more extreme position in the environmental gradient than *M. leprosa*. That means *E. orbicularis* is apparently absent from suboptimal macroenvironmental conditions and habitats. There are four, not mutually exclusive, possible reasons for such pattern: (1) *E. orbicularis* is more specialized regarding habitat selection; (2) interaction with *M. leprosa* excludes *E. orbicularis* from suboptimal habitats; (3) *E. orbicularis* is at the edge of its distribution and therefore occurs only in optimal habitat conditions; (4) *E. orbicularis* populations are in regression in Portugal and were removed from suboptimal habitats.

Where both species are present, the abundance of adults and the proportion of young age classes of *E. orbicularis* is generally much lower compared to *M. leprosa* (Araújo *et al.* 1997, Keller 1997, Segurado, unpubl. data). Habitats where *E. orbicularis* coexists with *M. leprosa* at higher abundances were never detected in Portugal. However, some allopatric populations of *E. orbicularis* showing comparable abundances to those of *M. leprosa* and in which all age classes are well represented were recently detected (Segurado, unpubl. data). This suggests that interaction with *M. leprosa* may play an important role in the occurrence pattern of *E. orbicularis*.

Further studies are required to answer these questions. They are crucial for a proper evaluation of the conservation status of *E. orbicularis* in Portugal. More detailed comparative information on life history parameters of species, such as diet, survival, fecundity, thermoregulation efficiency, and microhabitat use can help to elucidate some important issues regarding coexistence of both freshwater turtle species. According to our results no marked spatial segregation between the two species seems to occur, at least at the two spatial scales analyzed. Nevertheless, it is important to bear in mind that different sampling designs focusing at other spatial scales and resolutions could lead to distinct coexistence patterns.

Acknowledgements

We thank S. ROSA and two anonymous referees for their comments on the manuscript. This study was supported by a grant from Fundação para a Ciência e Tecnologia (SFRH/BD/8493/2002).

Appendices – Description of the environmental variables

Appendix 1 – Biogeographical scale environmental variables (cat. – categorical ; cont. – continuous ; ord. – ordinal).

Variables	Units	Type
Environmental		
Drainage	mm/year	ord
Acidity	pH	ord
Evapo-transpiration	mm/year	ord
Frost	days (mean)	ord
Humidity	% (annual mean)	ord
Insolation	hours (mean)	ord
Precipitation	mm/year	ord
Solar radiation	kcal/year	ord
Temperature	°C (annual mean)	ord
Ecological regions	ecological zones	cat
Geomorphological		
Altitude	m	cont
Slope	%	cont
Hillshade	Unknown (0 – 255)	cont
Soil	pedological units	cat
Socio-economic		
Human population density	n/km^2	ord
Land use		
Urban areas	% area	cont
Artificial landscape	% area	cont
Annual crops	% area	cont
Permanent crops	% area	cont
Pastures	% area	cont
Heterogeneous crops	% area	cont
Florests	% area	cont
Schrub lands	% area	cont
Bear soil	% area	cont
Landscape metrics		cont
Land use diversity	-	cont
Land use richness	-	cont
Number of patches	n	cont
Patch Size	ha	cont
Edge density	Km/ha	cont
Shape Index	-	cont
Patch Fractal Dimension	-	cont

Appendix 2 – Habitat scale environmental variables (bin – binary ; cont. – continuous ; ord. – ordinal).

Variables	Units	Type
Substratum (bedrock, gravel, sand, mud)	0 / 1	bin
Stream width	m	cont
Maximum depth	0: 0-0.25m; 1: 0.25 - 0.5m; 2: 0.5 - 1m; 3: > 1m	ord
Current speed	0: null; 1: low; 2: moderate; 3: high	ord
Water turbidity	0: null; 1: low; 2: moderate; 3: high	ord
Sun exposure	0: null; 1: low; 2: moderate; 3: high	ord
Emergent vegetation	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Submerged vegetation	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Dominant shoreline inclination	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Shoreline covered by grass	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Shoreline covered by shrub	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord
Shoreline covered by trees	0: 0-25%; 1: 25-50%; 2: 50-75%; 3: 75-100%	ord

Chapter 6

Final discussion

Otto Vaenius (1612), *Quinti Horatii Flacci Emblemata*.

http://emblems.let.uu.nl.

Note: this image is based on the legend surrounding the death of Aeschylus (525~65 BC), which is considered the father of Greek drama. Apparently, a vulture, mistaking the playwright's bald crown for a stone, dropped a turtle on his head. This is most probably the only documented case of human death directly attributed to a turtle.

In this thesis I have basically aimed to answer the following questions related to issues of uncertainty in the context of correlative models of species occurrence: (1) whether the performance of different modelling approaches is contingent to species macroecological traits, (2) how the bias induced by spatial autocorrelation is affected both by the main purpose of modelling - explanatory or predictive - and the kind of modelling approach - from purely parametric to purely non-parametric, (3) which is the best approach to deal with spatial autocorrelation, (4) whether the potential contribution of biotic interactions on species patterns of occurrence is contingent to the spatial scale and (5) how to estimate the biotic and abiotic constraints/requirements of rare species. As we will see, all these questions are in one way or another related to each other.

6.1 Is the "best" distribution modelling approach contingent to species macroecological traits?

According to the results presented in Chapter 2, a significant interaction between the effect of the kind of modelling approach and the effect of species macroecological trait was detected. However, there was too much variation on the relative model performances within each species group and, therefore, the trends that were detected were not strong enough to allow the establishment of simple rule of thumb for recommending a particular technique for modelling the distribution of a species showing a given macroecological trait. Such variation on model performances is possibly related to statistical artefacts due to varying prevalence's within each species group. In fact, prevalence has been shown to have an important effect on model accuracy (Manel *et al.* 2001, Brotons *et al.* 2004, McPherson *et al.* 2004, Hernandez *et al.* 2006). Nevertheless, among the modelling techniques assessed in chapter 2, those that allowed for more complex responses - i.e. generalized additive models and artificial neural networks - provided models with best overall performance. Indeed, according to most comparative studies on species distribution modelling, the same modelling techniques were shown to be the most effective (Heikkinen *et al.* 2006).

During the last year, and after the publication of our work, more thorough studies assessing the effect of species ecological traits on the accuracy of distribution models have further confirmed some of our results (Hernandez *et al.* 2006, McPherson & Jetz 2007). For example, it was shown that habitat tolerance and range size, which are directly related to the macroecological traits used in Chapter 2, were among those characteristics that most affected accuracy of distribution models even after controlling for methodological artefacts such as sampling prevalence (McPherson & Jetz 2007). According to the same authors other ecological traits, such as association to infrequent habitats and level of endemism, also influence accuracy via methodological artefacts.

It has been argued that species with countrywide distributions should be excluded from exercises such as that of chapter 2 (e.g. Arntzen 2006). In fact, since all empty cells were used as absence data in the binary response modelling approaches, that may well be the reason for lower accuracies found for species with larger occupancy and extent of occurrence, since there is the possibility that most of these empty cells represent lack of data instead of true absences. Nevertheless, a negative effect of global range size on model accuracy has been shown elsewhere (McPherson & Jetz 2007), which has been attributed to a possible spatial variability in occurrence-environment relationships of wide-ranging species due to varying local adaptations or constraints (Osborne & Suarez-Seoane 2002, McPherson & Jetz 2007). On the other hand, the absence of a countrywide species from a particular location might be due to the lack of adequate local scale features (e.g. habitat availability) or human induced factors that were not necessarily captured by the available set of environmental variables.

After the paper included in chapter 2 was published in 2004, the issue of modelling species occurrence has been the focus of much literature, in part due to the increasing demands of forecasting the effects of climate change (Araújo & New 2006, Botkin *et al.* 2007). More thorough studies have been devoted to model assessment (Hernandez *et al.* 2006, McPherson & Jetz 2007, Meynard & Quinn 2007) and model validation (Araújo *et al.* 2005a), new promising modelling approaches such as the Maximum Entropy methodology (Maxent) have been proposed and assessed (Phillips *et al.* 2006, Hernandez *et al.* 2006, Pearson *et al.* 2007, Peterson *et al.* 2007) and many important reviews on the subject have been published since then (Guisan & Thuiller 2005, Araújo & Guisan 2006, Guisan *et al.* 2006, Elith *et al.* 2006, Austin 2007). There is currently a much greater knowledge and consciousness among ecologists on the advantages and weaknesses of several available modelling techniques and that no such thing as the "best method" actually exists: predictions that represent a compromise among outputs of different methods are always preferable (for a review see Araújo & New 2006).

6.2 Is the effect of spatial autocorrelation contingent to the goals and the kind of modelling approaches?

A reasoning that has seldom been mentioned in the literature is that the effect of features such as ecological traits, prevalence and range sizes on the accuracy of distribution models may partially be attributed to the effect of spatial autocorrelation of species' occurrences. Macroecological attributes of species such as those considered in Chapter 2 may determine the spatial dependencies of their occurrence, e.g. species with higher marginality and lower tolerance to climatic variables (or areas with simultaneously lower extent of occurrence and area of occupancy) tend to show a clumped distribution and therefore to exhibit higher spatial autocorrelation of its occurrence locations.

In fact, in Chater 3 it is shown that the ability of variables to explain and predict species distribution can almost be attributed to spatial autocorrelation alone. According to the results, spatial autocorrelation strongly inflates type I errors of models of species distribution, independently of the main goal and the kind of modelling technique. The effect of spatial autocorrelation did not vary significantly between models' explanatory and predictive powers. Although there were some variations in the effect of spatial autocorrelation on the inflation of type I errors of different kinds of modelling approaches - decreasing effect from totally parametric to totally non-parametric models - these were very subtle. Therefore, the answer to the question raised above in this section would be that spatial autocorrelation has a strong effect regardless of the goal and the statistical approach used to model species distributions.

6.3 How to deal with spatial autocorrelation?

This question was also addressed in chapter 3. One main problem of inferring species-environment relationships is that the relative effect of space and environment on species distribution is often entangled due to spatial autocorrelation of both species occurrences and environmental variables. In other words, there is in most cases a large fraction of shared variation between the effect of space and the environment (Bocard *et al.* 1992). The effect of space alone is due to different endogenous factors such as dispersal, home-range sizes and biotic interactions, while the effect of the environment is due to exogenous factors such as climate, soil type and land use, that typically exhibit some degree of spatial autocorrelation (Dormann 2007). Although it has been argued that models showing an autocorrelated structure in their residuals may indicate that important ecological (or endogenous) variables are missing and that incorporating a term for autocorrelation would mimic biological effects (e.g. dispersion), one cannot be sure if there are other important environmental factors that are being ignored instead (Lichstein *et al.* 2002, Guisan & Thuiller 2005). More complex experimental frameworks are required to discern whether the influential factors that are inducing spatial autocorrelation into species distribution are actually endogenous or exogenous. In fact this is a crucial issue that needs urgently further research if more realism, and therefore more predictive power, is to be conferred to models.

Nevertheless, as shown in chapter 3, the incorporation of a term for spatial autocorrelation (e.g. contagion) in the regression equation may be sufficient to compensate the effect of spatial autocorrelation. Furthermore, according to the results presented in chapter 2, the inclusion of the contagion on both Generalized Linear Models and Generalized Additive Models resulted in overall improvement of model performances. This was particularly evident for widespread species, especially because it prevented absences to be misclassified as presences. In fact, a recent review based on 21 ecological studies (among them that of chapter 2) concluded that, when spatial autocorrelation was incorporated into models of species distribution, model accuracies were improved across all groups of organisms and all spatial scales (Dormann 2007).

The incorporation of spatial autocorrelation may lead to a different set of environmental variables being selected in the final model and it can even completely invert the effect of some variables (Kühn 2007). However, when autocorrelation terms or filters are incorporated into a model, they will necessarily capture part of the explained variation that was shared with environmental variables. A possible consequence is that part of that captured variation can be actually causative and hence influential environmental variables can be rejected during the variable selection process. As a result, by avoiding type I errors there is always the possibility of type II errors to be inflated by our correction. Furthermore, in chapter 3 it was shown that the inclusion of contagion in the model may inverse the effect of autocorrelation, i.e., originate a decrease of variable significance with increasing spatial autocorrelation, or inflating type II errors. Another problem of incorporating spatial autocorrelation arises when the available data do not represent the actual species distribution and consequently model predictions underestimate the true distribution range. Therefore the effect of the inclusion of spatial filters should be carefully examined before deciding to incorporate them into models.

6.4 Is the potential contribution of biotic interactions on species distributions contingent to spatial scale of analysis?

The study presented in Chapter 4 showed that examining patterns of coexistence at different spatial extents and resolutions might lead to opposing conclusions about the nature of biotic interactions among species. At the macroscale a broad spatial overlap between the two studied species was found and, moreover, both species tended to respond identically to environmental variables. At the local scale species tended to show complementary distributions and to respond at opposing directions to most environmental variables. This means that patterns of co-occurrence at local scales - as the result of biotic interactions or other ecological processes - are not necessarily reflected in the observed patterns of co-occurrence at broader spatial scales.

Although this result is not surprising, as ecological factors are expected to be more influential at smaller spatial scales (Brändle & Brandl 2001, Pearson & Dawson 2003), it has seldom been quantified empirically (Guisan & Thuiller 2005). The contribution of ecological processes, such as biotic interactions, on species occurrence should therefore be more relevant for studies with higher resolutions (Pearson & Dawson 2003, Guisan & Thuiller 2005). On the other hand, at macro-scales the effect of environmental factors on species occurrence becomes dominant and therefore ignoring ecological processes at broader extents and lower resolutions is more acceptable (Pearson & Dawson 2003). This is a relevant issue because recent claims for the need to incorporate biological processes on predictive models might be less justifiable for studies undertaken at very broad and coarse spatial scales.

As mentioned earlier in this chapter (section 6.3) the explicit incorporation of spatial autocorrelation into models of species distribution leads to model improvement across all spatial scales (Dormann 2007). This could mean that the role of biological factors is relevant at all spatial scales and should be considered in the modelling process. However, it is probable that the spatial structure at broader spatial scales will be at a greater extent the result of geographic barriers and historical constraints, rather than of ecological processes such as biotic interactions. Furthermore, it is possible that models frequently miss important environmental variables, which would explain the observed improvements after accounting for spatial structure (Dormann 2007).

6.5 How to deal with small sample sizes?

The main problem of studying patterns of occurrence of rare or very cryptic species, apart from the fact that only small sample sizes are typically available, is that absence data are less reliable (Engler *et al.* 2004, MacKenzie *et al.* 2005, Pearson *et al.* 2007) and therefore there is a greater uncertainty whether the available records, which often represent ad-hoc observations, represent the true set of biotic and abiotic requirements of species (Guisan *et al.* 2006). A possible and simple approach to address this problem is to compare species' observations with independent random samples of observations with the same size taken from the whole available set of biotic or abiotic conditions, i.e., to test the null hypothesis that species records are random expectations from the whole "population" of conditions. The problem of this approach is common to the difficult issue of achieving a proper null space to obtain realistic random expectations of observations in typical null model analyses (Gotelli & Graves 1996). A further problem is that rare species show typically less aggregated occurrences (Gaston 2003) and are more often absent from locations with suitable conditions.

In Chapter 5 two working examples were presented in which a simple methodology based on the above-mentioned approach was proposed. The purpose of both studies was to test whether the two freshwater turtle species occurring in Portugal shared common ecological requirements. As was mentioned in the introductory chapter, while one species (*Emys orbicularis*) is very rare, the other (*Mauremys leprosa*) is very common. This has been especially suitable for the purpose of comparing ecological and spatial requirements of these particular species because *E. orbicularis* observations could be confronted to random samples of *M. leprosa* with the same size, i.e., to test the null hypothesis that the sample of *E. orbicularis* observations was a random subsample of the *M. leprosa* "population" of observations. In this particular case the issue of finding an adequate null space was partially resolved because both species share the same chief ecological requirements and therefore a record of *M. leprosa* is a potential record for *E. orbicularis*.

This method was successfully used to test for differences in the use of space and environment by the two species at several spatial scales (Chapter 5, working examples 1 and 2), as well as to assess the marginality of *E. orbicularis* distribution along univariate environmental gradients in relation to that of *M. leprosa* (Chapter 5, working example 2).

6.6 Concluding remarks

As evidenced in the last sections, all sources of model uncertainty that were dealt in this thesis are in some way related to each other. For example, the effect of species distribution and environmental use on the performance of different niche-modelling approaches is most probably related to the effect of spatial autocorrelation on model accuracies. In turn, species' rarity or detectability affects negatively the amount of spatial autocorrelation among occurrences and, moreover, induces a great amount of uncertainty on the estimation of the kind of species distribution and environmental use. On the other hand, spatial autocorrelation of species occurrences and the relative importance of its causal factors (e.g. environmental autocorrelation, dispersal and biotic interactions) are influenced by spatial scale. Such interplay of different sources of model uncertainty has seldom been addressed using a single framework. At this point, in the particular context of niche modelling, an important future challenge would be to provide more comprehensive assessments of the interplay between the different factors that contribute to model uncertainty in order to produce less wrong and more useful ecological models.

Literature cited

Akaike, H., 1974. A new look at statistical model identification. *IEEE Transactions on Automated Control*, 19: 716-722.

Amarasekare, P., 2003. Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters*, 6: 1109-1122.

Anadón, J.D., Gimenez, A., Martinez, M., Martinez, J., Perez, I. & Esteve, M.A., 2006. Factors determining the distribution of the spur-thighed tortoise *Testudo graeca* in south-east spain: a hierarchical approach. *Ecography*, 29: 339-346.

Anderson, R.P., Peterson A.T. & Gómez-Laverde, M., 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 98: 3–16.

Apps, C.D., McLellan, B.N & Woods, J.G., 2006. Landscape partitioning and spatial inferences of competition between black and grizzly bears. *Ecography*, 29: 561-572.

Araújo, M.B. & Guisan, A., 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33(10): 1677-1688.

Araújo, M. B. & New, M., 2006. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22: 42-47.

Araújo, M.B. & Rahbek, C., 2006. How does climate change affect biodiversity? *Science*, 313: 1396–1397.

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

Araújo, M.B. & Williams, P.H., 2001. The bias of complementarity hotspots toward marginal populations. *Conservation Biology*, 15: 1710-1720.

Araújo, M.B., Williams, P.H. & Fuller, R.J., 2002. Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society London,* B, 269: 1971-1980.

Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L. & Williams, P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology*, 10: 1618–1626.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M., 2005a. Validation of species-climate impact models under climate change. *Global Change Biology*, 11: 1504-1513.

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

Araújo, M.B., Thuiller, W. & Pearson, R.G., 2006. Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography*, 33: 1365–2699.

Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M., 2005c. Reducing uncertainty in projections of extinction risk from climate change. Global Ecology and Biogeography, 14: 529-538.

Araújo, P., Segurado, P. & Santos, N., 1997. *Bases para a Conservação das tartarugas de água doce, Emys orbicularis e Mauremys leprosa. Estudos de Biologia e Conservação da Natureza*, 24, Instituto de Conservação da Natureza, Lisboa.

Arntzen, J.W., 2006. From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles. *Frontiers in Zoology*, 3: 8 http://www.frontiersinzoology.com/content/3/1/8

Arriaga, L., Castellanos, A.E., Moreno, E. & Alarcón, J., 2004. Potential ecological distribution of alien invasive species and risk assessment: a case study of buffel grass in arid regions of Mexico. *Conservation Biology*, 18: 1504–1514.

Augustin, N.H., Cummings, R.P. & French, D.D. 2001. Exploring spatial vegetation dynamics using logistic regression and a multinomial logit model. *Journal of Applied Ecology*, 38: 991–1006.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T., 1996. An autologistic model for spatial distribution of wildlife. *Journal of Applied Ecology*, 33: 339-47.

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

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

Ayres, C.F & Cordero, A.R. 2000. A new subespecies of *Emys orbicularis* in the Iberian Peninsula? *Chelonii*, 2: 20-22.

Ayres C.F. & Cordero A.R., 2001. Sexual dimorphism and morphologoical differentiation in European Pond Turtle (*Emys orbicularis*) populations from Northwestern Spain. *Chelonian Conservation Biology*, 4: 100-106.

Bailey, T.C. & Gatrell, A.C., 1995. *Interactive spatial data analysis*. Longman, England.

Berg, A, Gärdenfors, U. & von Proschwitz, T., 2004. Logistic regression models for predicting occurrence of terrestrial mollusks in southern Sweden - importance of environmental data quality and model complexity. *Ecography*, 27: 83-93.

Beutel, T.S., Beeton, R.J.S. & Baxter, G.S., 1999. Building better wildlife-habitat models. *Ecography*, 22: 219-223.

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

Bodie, J.R. & Semlitsch, R.D., 2000. Spatial and temporal use of floodplain habitats by lentic and lotic species of aquatic turtles. *Oecologia*, 122: 138-146.

Bodie, J.R., Semlitsch, R.D. & Renken R.B., 2000. Diversity and structure of turtle assemblages: associations with wetland characters across a floodplain landscape. *Ecography*, 23: 444–456.

Bomhard, B., Richardson, D.M., Donaldson, J.S., Hughes, G.O., Midgley, G.F., Raimondo, D.C., Rebelo, A.G., Rouget, M. & Thuiller, W., 2005. Potential impacts of future land use and climate change on the Red List status of the Proteaceae in the Cape Floristic Region, South Africa. *Global Change Biology*, 11: 1452–68.

Bonin, F., Devaux, B. & Dupré, A., 1998. *Toutes les tortues du monde*. Delachaux et Niestlé, Paris, 2ème Édition.

Bons, J. & Geniez, P., 1996. *Amphibiens et reptiles du Maroc (Sahara Occidental compris) Atlas Biogéographique*. Asociación Herpetológica Española, Barcelona.

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

Bossard, M., Feranec, J. & Otahel, J., 2000. *CORINE Land Cover Technical Guide – Addendum 2000*. Technical report 40, EEA.

Botkin, D.B., Saxe, H. Araújo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P., Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert, D.W., Loehle, C., Margules, C., New, M., Sobel, M.J. & Stockwell, D.R.B., 2007. Forecasting the effects of global warming on biodiversity. BioScience, 57(3): 227-236.

Bozhansky, A.T. & Orlova, V.F., 1998. Conservation status of the European pond turtle, *Emys orbicularis* (Linnaeus, 1758), in European Russia. *Mertensiella*, 10: 41-46.

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

Brändle, M. & brandl, R., 2001. Distribution, abundance and niche breadth of birds: scale matters. Global Ecology & Biogeography, 10: 173–177.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J., 1984. *Classification and regression trees.* Wadsworth International Group, California.

Brito, J.C., Brito e Abreu, F., Paulo, O.S., Rosa H.D. & Crespo E.G., 1996. Distribution of schreiber's green lizard (*Lacerta schreiberi*) in Portugal: a predictive model. *Herpetological Journal*, 6: 43-47.

Brito, J.C., Crespo, E.G. & Paulo, O.S., 1999. Modelling wildlife distributions: Logistic multiple regression vs overlap analysis. *Ecography,* 22: 251–260.

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H., 2004. Presence-absence versus presence-only habitat suitability models: the role of species ecology and prevalence. *Ecography*, 27: 165–72.

Brown, J.H., 1995. *Macroecology.* The University of Chicago Press, Chicago.

Buckland S.T. & Elston D.A., 1993. Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, 30: 478–495.

Buckland, S.T., Elston, D.A. & Beaney, S.J., 1996. Predicting distributional change, with application to bird distributions in northeast Scotland. *Global Ecology and Biogeography Letters*, 5: 66-84.

Buckley, L.B. & Roughgarden, J., 2006. Climate, competition, and the coexistence of island lizards. *Functional Ecology*, 20(2): 315-322.

Bullock, J.M., Edwards, R.J., Carey, P.D. & Rose, R.J., 2000. Geographical separation of two Ulex species at three spatial scales: does competition limit species' ranges? *Ecography,* 23: 257–271.

Burke, V.J. Lovich, J.E. & Gibbons J.W., 2000. Conservation of freshwater turtles. *Turtle Conservation.* (ed. by M.W. Klemens) , pp. 156-179. Smithsonian Institution Press, Washington.

Burnham, K.P. & Anderson, D.R., 2002. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach.* 2nd ed., Springer-Verlag, New York.

Bury, R.B., Germano, D.J. & David, J., 2003. Differences in Habitat Use by Blanding's Turtles, *Emydoidea blandingii*, and Painted Turtles, *Chysemys picta*, in the Nebraska Sandhills. *American Midland Naturalist*, 149: 241-244.

Buskirk J. & Servan J. (Ed.), 2000. Proceedings of the 2nd International Symposium on *Emys orbicularis*. *Chelonii*, 2.

C.N.A., 1983. *Atlas do Ambiente.* Direcção Geral do Ambiente. Ministério do Ambiente e dos Recursos Naturais, Lisbon.

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

Cabral, M.J., Almeida, J., Almeida, P.R., Dellinger, T., Ferrand de Almeida, N., Oliveira, M.E., Palmeirim, J.M., Queiroz, A.I., Rogado, L. & Santos-Reis, M. (Eds.), 2005. *Livro Vermelho dos Vertebrados de Portugal.* Instituto da Conservação da Natureza, Lisboa.

Cadena, C.D. & Loiselle, B.A., 2007 Limits to elevational distributions in two species of emberizine finches: disentangling the role of interspecific competition, autoecology, and geographic variation in the environment. *Ecography*, 30: 491-504.

Cadi, A. & Joly, P., 2003. Competition for basking places between the endangered European pond turtle (*Emys orbicularis galloitalica*) and the introduced red-eared slider (*Trachemys scripta elegans*). *Canadian Journal of Zoology*, 81: 1392-1398.

Cadi, A. & Joly, P., 2004. Impact of the introduction of the red-eared slider (*Trachemys scripta elegans*) on survival rates of the European pond turtle (*Emys orbicularis*). *Biodiversity Conservation*, 13 : 2511 – 2518.

Cadi, A. Nemoz, M., Thienpont, S. & Joly, P., 2004. Home range, movements, and habitat use of the European pond turtle (*Emys orbicularis*) in the Rhone-Alpes region, France. *Biologia*, 59 (Suppl. 14): 89-94.

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.

Catry, P., Campos A., Segurado, P., Silva, M. & Strange I., 2003. Population census and nesting habitat selection of thin-billed prion *Pachyptila belcheri* on New Island, Falkland Islands. *Polar Biology*, 26: 202-207.

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

Chesson, P., 2000. General Theory of competitive coexistence in spatially - varying environments. *Theoretical Population Biololgy*, 58: 211 237.

Clark, L.A. & Pregibon, D., 1992.. Tree-based Models. *Statistical Models in S.* (ed. by J.M. Chambers & T.J. Hastie), pp. 377-419. Chapman and Hall, New York.

Clarke, K.R. & Gorley, R.N., 2001. *PRIMER v5: Users Manual/Tutorial.* PRIMER-E Ltd, Plymouth, UK.

Cliff, A.D. & Ord, J.K., 1973. *Spatial autocorrelation.* Pion Limited, London.

Cocks, K.D. & Baird, I.A., 1989. Using mathematical programming to address the multiple reserve selection problem: an example from Eyre Peninsula, South Australia. *Biological Conservation*, 49: 113-130.

Colwell, R.K. & Lees, D.C., 2000. The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution*, 15: 70–76.

Connor, E.F. & Bowers, M.A., 1987. The spatial consequences of interspecific competition. *Annales Zoologici Fennici*, 24 : 213-226.

Connor, E.F. & Simberloff, D., 1979. The assembly of species communities: chance or competition? *Ecology*, 60, 1132–1140.

Cordero, A.R. & Ayres, C.F., 2004. A management plan for the European pond turtle (*Emys orbicularis*) populations of the Louro river basin (Northwest Spain). *Biologia*, 59 (Suppl. 14): 161-171.

Corsi, F., de Leeuw, J. & Skidmore, A., 2000. Modeling Species Distribution with GIS. *Research Techniques in Animal Ecology. Controversies and Consequences.* (ed. by L. Boitani & T.K. Fuller), pp 389-434. Columbia University Press, New York.

Cumming, G.S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. *Journal of Biogeography*, 27: 441–455.

Cunningham, R.B. & Lindenmayer, D.B., 2005. Modeling count data of rare species: some statistical issues. *Ecology*, 86: 1135–1142.

Dale, M.R.T. & Fortin, M.-J., 2002. Spatial autocorrelation and statistical tests in ecology. *Écoscience*, 9: 162-167.

De'Ath, G. & Fabricius, K.E., 2000. Classification and Regression Trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81: 3178-3192.

Derksen, S. & Keselman, H.J., 1992. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. *British Journal of Mathematical & Statistical Psychology*, 45: 265-282.

Dettmers, R., Buehler, D.A. & Bartlett, J.G., 2002. A test and comparison of wildlife-habitat modelling techniques for predicting bird occurrence at a regional scale. *Predicting Species Occurrences: Issues of Accuracy and Scale.* (ed. by J.M. Scott, M.L. Morrison & P.J. Heglund), pp. 607-615. Island Press, Covelo, CA.

Devaux, B. & Bley, S., 1998. *Emys orbicularis galloitalica* in Provence: an example of a small threatened population. *Mertensiella*, 10: 83-88.

Diniz-Filho, J.A.F. & Bini, L.M., 2005. Modelling geographical patterns in species richness using eigenvector-based spatial filters. *Global Ecology and Biogeography*, 14: 177–186.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A., 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, 12: 53–64.

Donald, P.F. & Greenwood, J.J.D., 2001. Spatial patterns of range contraction in British breeding birds. *IBIS*, 143: 593-601.

DonnerWright, D.M., Bozek, M.A., Probst, J.R. & Anderson, E.M., 1999. Responses of turtle assemblage to environmental gradients in the St. Croix River in Minnesota and Wisconsin, U.S.A. *Canadian Journal of Zoology*, 77: 989-1000.

Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, 16: 129–138.

Dutilleul, P., 1993. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49: 305-314.

Efron, B. & Tibshirani, R.J., 1993. *An Introduction to the Bootstrap.* Chapman & Hall, San Francisco, USA.

Elith, J., 2000. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative Methods for Conservation Biology*, (ed. by S. Ferson & M. Burgman), pp. 39–58. Springer-Verlag, New York.

Elith, J & Burgman, M., 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. *Predicting Species Occurrences: Issues of Accuracy and Scale*, (ed. by J.M. Scott, M.L. Morrison & P.J. Heglund), pp. 303-313. Island Press, Covelo, CA.

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

Elkie, P.C., Rempel, R.S. & Angus, P.C. 1999. *Patch Analyst User's Manual.* Natur. Resour. Northwest Sci & Technol., Ontario.

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

Ernst, C.H. & Barbour, R.W., 1989. *Turtles of the World.* Smithsonian Institution Press, Washington, D.C.

ESRI., 1999. *ArcView GIS Help (v 3.2).* Environmental Systems Research Institute, Inc. New York, USA.

Fattizzo, T., 2004. Distribution and conservational problems of *Emys orbicularis* in Salento (South Apulia, Italy). *Biologia*, 59 (Suppl. 14): 13-18.

Feria, T.P. & Peterson, A.T., 2002. Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. *Diversity & Distributions*, 8: 49-56.

Fertig, W. & Reiners, W.A., 2002. Predicting presence/absence of plant species for range mapping: a case study from Wyoming. *Predicting Species Occurrences: Issues of Accuracy and Scale.* (ed. by J.M. Scott, M.L. Morrison & P.J. Heglund), pp. 483-489. Island Press, Covelo, CA.

Ficetola, G.F., Padoa-Schioppa, E., Monti, A., Massa, R., De Bernardi, F. & Bottoni L., 2004. The importance of aquatic and terrestrial habitat for the European pond turtle (*Emys orbicularis*): implications for conservation planning and management. *Canadian Journal of Zoology*, 82: 1704-1712.

Ficetola, G.F. & De Bernardi, F., 2006. Is the European "pond" turtle *Emys orbicularis* strictly aquatic and carnivore? *Amphibia-Reptilia*, 27: 445-447.

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.

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

Fischer, J., Lindenmayer, D.B., Nix, H.A., Stein, J.L. & Stein, J.A., 2001. Climate and animal distribution: a climatic analysis of the Australian marsupial *Trichosurus caninus*. *Journal of Biogeography*, 28: 293–304.

Forman, R.T.T. 1995. *Land Mosaics. The ecology of landscapes and regions.* Cambridge University Press, Cambridge.

Fortin, M.-J. & Jacquez, G.M., 2000. Randomization tests and spatially autocorrelated data. *Bulletin of the ESA*, 81: 201-206.

Fortin, M.-J. & Payette, S., 2002. How to test the significance of the relation between spatially autocorrelated data at the landscape scale: A case study using fire and forest maps. *Écoscience*, 9 : 213-218.

Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, 9: 733-748.

Fritz, U., 1998. Introduction to zoogeography and subspecific differentiation in *Emys orbicularis* (Linnaeus, 1758). *Mertensiella*, 10: 1-27.

Fritz, U. & Havas, P. (Ed.), 2004. Proceedings of the 3rd International Symposium on *Emys orbicularis*. *Biologia*, 59 (Suppl. 14).

Fritz, U., Barata, M., Busack, S., Fritzsch, G. & Castilho, R., 2006. Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, Mauremys leprosa (Reptilia, Testudines, Geoemydidae). *Zoologica Scripta*, 35(1): 97-108.

Fritz, U., Joger, U., Podloucky, R. & Servan, J. (Ed.), 1998. Proceedings of the *Emys* Symposium Dresden 96. *Mertensiella*, 10.

Fuselier, L. & Edds, D., 1994. Habitat partitioning among three sympatric species of map turtles, genus *Graptemys*. *Journal of Herpetology*, 28: 154-158.

Gasith, A. & Resh, V.H., 1999. Streams in Mediterranean climate regions: Abiotic influences and biotic responses to predictable seasonal events. *Annual Review Of Ecology Evolution And Systematics*, 30: 51–81.

Gaston, K.J., 1994. *Rarity.* Chapman & Hall, New York.

Gaston, K.J., 1996. Species-range-size distributions: patterns, mechanisms and applications. *Trends in Ecology and Evolution*, 11: 197-201.

Gaston, K.J., 2003. *The structure and dynamics of geographic ranges.* Oxford University Press, Oxford.

Gates, S. & Donald, P.F., 2000. Local extinction of British farmland birds and the prediction of further loss. *Journal of Applied Ecology*, 37: 806-820.

Gates, S., Gibbons, D.W., Lack, P.C. & Fuller, R.J., 1994. Declining farmland bird species: modelling geographical patterns of abundance in Britain. *Large-scale ecology and conservation biology.* (ed. by P.J. Edwards, R.M. May & N.R. Webb), pp. 153-177. Blackwell Science, London.

Godinho, R., Teixeira, J., Rebelo, R., Segurado, P., Loureiro, A., Alvares, F., Gomes, N., Cardoso, P., Camilo-Alves, C. & Brito, J.C., 1999. Atlas of the continental Portuguese herpetofauna: an assemblage of published and new data. *Revista Española de Herpetologia*, 13: 61-82.

Gomez-Mestre, I. & Keller, C., 2003. Experimental Assessment of Turtle Predation on Larval Anurans. *Copeia*, 2003, 349–356.

Gómez, A. & Lizana, M. 1998. Distribución y uso del habitat de los galápagos (*Mauremys leprosa* e *Emys orbicularis*) en la provincia de Salamanca. *Bol. Asoc. Herpetol. Esp.*, 11: 4-8.

Gotelli, N., 2000. Null model analysis of species co-occurrence patterns. *Ecology*, 81: 2606–2621.

Gotelli, N.J., 2001. Research frontiers in null model analysis. *Global Ecology & Biogeography*, 10: 337–343.

Gotelli, N.J. & Entsminger, G.L., 2006. *EcoSim: Null models software for ecology. Version 7.* Acquired Intelligence Inc. and Kesey-Bear. Jericho, VT 05465. http://garyentsminger.com/ecosim.htm.

Gotelli, N.J. & Graves, G.R., 1996. *Null models in ecology.* Smithsonian Institute Press, Washington, D.C.

Gravel, D., Canham, C.D., Beaudet, M. & Messier, C., 2006. Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9: 399-409.

Guisan, A. & Hofer, U., 2003. Predicting reptile distributions at mesoscale: relation to climate and topography. *Journal of Biogeography*, 30, 1233–1243.

Guisan, A. & Theurillat, J.P., 2000. Assessing alpine plant vulnerability to climate change: a modelling perspective. *Integrated Assessment*, 1: 307–320.

Guisan, A. & Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135: 47-186.

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

Guisan, A., Lehmann, A., Ferrier, S., Aspinall, R., Overton, R., Austin, M. & Hastie, T., 2006. Making better biogeographic predictions of species distribution. *Journal of Applied Ecology*, 43: 386–392.

Gutiérrez, D., Fernández, P., Seymour, A. S. & Jordano, D., 2005. Habitat distribution models: are mutualist distributions good predictors of their associates? *Ecological applications*, 15: 3-18.

Hampe, A., 2004. Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, 13: 469-471.

Hanski, I. & Gilpin, M., 1997. *Metapopulation Biology: Ecology, Genetics and Evolution*. Academic Press, London.

Harrel, F.E. Jr, Lee K.L., Califf, R.M., Pryor D.B. & Rosati R.A., 1984. Regression modeling strategies for improved prognostic prediction. *Statistics in Medicine*, 3: 143-152.

Harrison, S., 1997. How natural habitat patchiness affects the distribution of diversity in Californian serpentine chaparral. *Ecology*, 78: 1898-1906.

Haslett, J.R., 1990. Geographic information systems: a newapproach to habitat definition and the study of distributions. *Trends in Ecology and Evolution*, 5: 214–218.

Hastie, T. & Tibshirani, R., 1990. *Generalised additive models.* Chapman and Hall, London.

Heikkinen, R.K., Luoto, M., Kuussaari, M. & Pöyry, J., 2005. New insights to butterfly–environment relationships with partitioning methods. *Proceedings of the Royal Society London Series B, Biological Sciences*, 272: 2203–2210.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R. Thuiller, W. & Sykes, M.T. 2007. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography*, 30: 751–777.

Hepinstall, J.A, Krohn, B. & Sader, S.A., 2002. Effects of niche width on the performance and agreement of avian habitat models. *Predicting Species Occurrences: Issues of Accuracy and Scale*. (ed. by J.M. Scott, M.L. Morrison & P.J. Heglund), pp. 593-606. Island Press, Covelo, CA.

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

Hirzel, A., 2001. *When GIS come to life. Linking landscape- and population ecology for large population management modelling: the case of Ibex (Capra ibex) in Switzerland.* PhD thesis. University of Lausanne.

Hirzel, A., Hausser, J. & Perrin, N., 2001. *Biomapper 1.0.* Lausanne, Lab. for Conservation Biology. URL: http://www.unil.ch/biomapper

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

Hirzel, A.H., Le Lay, G., Helfer, V.,Randin, C. & Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199: 142–152.

Hofer, U., Bersier, L.-F. & Borcard, D., 2004. Relating niche and spatial overlap at the community level. *Oikos*, 106: 366-376.

Holt, R.D., 2006. Emergent neutrality. *Trends in Ecology and Evolution*, 21(10): 531-532.

Holyoak, M., Loreau, M. & Strong, D., 2006. Neutral community ecology. *Ecology*, 87: 1368–1369.

Hosmer Jr., D.W. & Lemeshow, S., 1989. *Applied Logistic Regression*. John Wiley & Sons, New York.

Hubbell, S.P., 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton.

Huston, M.A. 2002. Introductory essay: critical issues for improving predictions. *Predicting species occurrences: issues of accuracy and scale* (ed. by J. M. Scott, P. J. Heglund, M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall & F. B. Samson), pp. 7-21. Island Press, Covelo, California.

Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22: 145-159.

Illoldi-Rangel, P., Sanchez-Cordero, V. & Peterson, A.T., 2004. Predicting distributions of Mexican mammals using ecological niche modeling. *Journal of Mammalogy*, 85: 658–662.

Jetz, W. & Rahbek, C., 2002. Geographic range size and determinants of avian species richness. *Science*, 297: 1548–1551.

Joyal, L.A., McCollough, M. & Hunter, M.L., 2001. Landscape ecology approaches to wetland species conservation: a case study of two turtle species in southern Maine. *Conservation Biology*, 15:1755-1762.

Karl, J.W., Bomar, L.K., Heglund, P.J., Wright, N.M. & Scott, M., 2002. Species commonness and the accuracy of habitat-relationship models. *Predicting Species Occurrences: Issues of Accuracy and Scale*. (ed. by J.M. Scott, M.L. Morrison & P.J. Heglund), pp. 573-580. Island Press, Covelo, CA.

Kearney, M., 2006. Habitat, environment and niche: what are we modelling? *Oikos* 115(1): 186-191.

Kearney, M. & Porter, W.P., 2004. Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology*, 85: 3119–3131.

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

Keller, C., 1997. *Ecología de poblaciones de Mauremys leprosa e Emys orbicularis en el Parque Nacional de Doñana*. PhD Thesis, Universidad de Sevilla, Sevilla.

Keller, C. & Andreu, A.C., 2002. *Emys Orbicularis. Atlas y Libro Rojo de los Anfibios y Reptiles de España* (ed. by J.M. Pleguezelos, R. Marquez and M. Lizana), pp. 137-140. Dirección General de la Conservación de la Naturaleza, Madrid.

Keller, C. & Busack, S.D., 2001. *Mauremys leprosa* (Schweigger, 1812) — Maurische Bachschildkröte. Handbuch der Reptilien und Amphibien Europas. (ed. by U. Fritz), pp. 57-88. Band 3/IIIA: Schildkröten I. Wiebelsheim, Aula-Verlag, Wiesbaden.

Keller, C., Andreu, A.C. & Ramo, C., 1998. Aspects of the population structure of *Emys orbicularis hispanica* from southwestern Spain. *Mertensiella*, 10: 147-158.

Keller, C., Díaz-Paniagua, C., Andreu, A.C. & Bravo, M.A., 1995. Distribution pattern of freshwater turtles in the Doñana National Park (SW Spain). *I International Congress of Chelonian Conservation, 1995, Gonfaron*, pp192-195. Proceedings of the I International Congress of Chelonian Conservation. Gonfaron: SOPTOM.

Klockars, A. & Sax, G., 1986. *Multiple Comparisons*. Quantitative Applications in the Social Sciences series Sage Publications, Thousand Oaks, CA.

Koenig, W.D. & Knops, M.H., 1998. Testing for spatial autocorrelation in ecological studies. *Ecography*, 21: 423-429.

Kofron, C.P.& Schreiber, A.A., 1985. Ecology of two endangered aquatic turtles in Missouri: *Kinisternon flavescens* and *Emidoidea blandigii*. *Journal of Herpetology*, 19: 27-40.

Kotenko, T., 2004. Distribution, habitats, abundance and problems of conservation of the European pond turtle (*Emys orbicularis*) in the Crimea (Ukraine): first results. *Biologia*, 59 (Suppl. 14): 33-46.

Kühn, I., 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions*, 13: 66-69.

Lacomba, J.L. & Sancho, V., 2000. The European pond turtle conservation plan in the Valencian Community: a proposal. *Chelonii*, 2: 130-132.

Lacomba, I.A. & Sancho, V.A., 2004. Advances in the action plan for *Emys orbicularis* in the Valencia region, Spain. *Biologia*, 59 (Suppl. 14): 173-176.

Lawton, J.H. & May, R.M., 1995 eds. Extinction Rates. Oxford University Press.

Leathwick, J.R. & Austin, M.P., 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, 82: 2560–2573.

Leathwick, J.R., Rowe, D., Richardson, J., Elith, J. & Hastie, T., 2005. Using multivariate adaptive splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biology*, 50: 2034–2052.

Lebboroni, M. & Chelazzi, G., 1999. Habitat use, reproduction, and conservation of *Emys orbicularis* in a pond system in Central Italy. *Ponds and pond landscapes of Europe: Proceedings of the International Conference of the Pond Life Project*. (ed. by J. Boothby), pp. 169-173. Pond Life Project, Liverpool.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74: 1659–1673.

Legendre, P. & Legendre, L., 1998. *Numerical ecology*. Elsevier, Amsterdam.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Casgrain, P. & Gurevitch, J., 2004. Effects of spatial structures on the results of field experiments. *Ecology*, 85: 3202-3214.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Casgrain, P., Gurevitch, J., Hohn, M. & Myers, K., 2002. The consequence of spatial structure for the design and analysis of ecological field surveys. *Ecography*, 25: 601-615.

Lehmann A., Leathwick J.R. & Overton J.McC., 2002. Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, 11: 2217–2238.

Leibold, M.A. & McPeek, M.A., 2006. Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87: 1399–1410.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling*, 90: 39-52.

Lennon, J.J., 2000. Red-shifts and red herrings in geographical ecology. *Ecography*, 23: 101-113.

Letcher, A.J., Purvis, A., Nee, S. & Harvey, P.H., 1994. Patterns of overlap in the geographic ranges of Palearctic and British mammals. *Journal of Animal Ecology*, 63: 871-879.

Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology*, 73:1943–1967.

Lichstein, J.W., Simons, T.R., Shriner, S.A. & Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72: 445 – 463.

Lindeman, P.V., 1999. Aggressive interactions during basking among four species of emydid turtles. *Journal of Herpetology*, 33: 214–219.

Lindeman, P.V., 2000. Resource use of five sympatric turtle species: Effects of competition, phylogeny, and morphology. *Canadian Journal of Zoology*, 78: 992-1008.

López, P., Marcos, I. & Martín, J., 2005. Effects of habitat-related visability on escape decisions of the Spanish Terrapin *Mauremys leprosa*. *Amphibia-Reptilia*, 26: 557-561.

Luoto, M., Virkkala, R., Heikkinen, R.K. & Rainio, K., 2004. Predicting bird species richness using remote sensing in boreal agricultural forest mosaic. *Ecological Applications*, 14: 1946–1962.

MacArthur, R.H., 1957. On the relative abundance of bird species. *Proceedings of the National Academy of Sciences*, 43: 293-295.

McCauley, S.J. & Bjorndal, K.A., 1999. Response to dietary dilution in an omnivorous freshwater turtle: Implications for ontogenetic dietary shifts. *Physiological and Biochemical Zoology*, 72:101-108.

McCullagh, P. & Nelder, J.A., 1989. *Generalized Linear Models*. Chapman & Hall. London.

Maciantowicz, M. & Najbar, B., 2004. Distribution and active conservation of *Emys orbicularis* in Lubuskie province (West Poland). *Biologia*, 59 (Suppl. 14): 177-183.

MacKenzie, D.I., Bailey, L.L. & Nichols, J.D., 2004. Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, 73: 546–555.

MacKenzie, D.I., Nichols, J.D., Sutton, N., Kawanishi, K. & Bailey, L.L., 2005. Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*, 86: 1101–1113.

Mackey, B.G. & Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography*, 28: 1147–1166.

McPherson, J.M. & Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography*, 30 (online early: doi: 10.1111/j.2006.0906-7590.4823.x).

McPherson, J., Jetz, W. & Rogers, D., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, 41: 811–23.

Magalhães, M.F., Beja, P., Canas, C. & Collares-Pereira, M.J., 2002. Functional heterogeneity of dry-season fish refugia across a Mediterranean catchment: the role of habitat and predation. *Freshwater Biology*, 47: 1919-1934.

Manel, S., Buckton, S.T. & Ormerod, S.J., 2000. Testing large-scale hypotheses using surveys: the effects of land use on the habitats, invertebrates and birds of Himalayan rivers. *Journal of Applied Ecology*, 37: 756 –770.

Manel, D., Dias, J. M., Buckton, S.T. & Ormerod, S.J., 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, 36: 734-747.

Manel, S., Dias, J.M., Buckton, S.T., & Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distribution: a case study with a Himalayan river bird. *Ecological Modelling*, 120: 337–347.

Manel, S., Williams, H.C. & Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38: 921-931

Manly, B.F.J., 1995. A note on the analysis of species co-occurrences. *Ecology*, 76: 1109-1115.

Manly, B.F.J. & Sanderson, J.G., 2002. A note on null models: justifying the methodology. *Ecology*, 83: 580–582.

Marco, A. & Andreu, A.C., 2005. Social interactions among *Emys orbicularis*, redswamp crayfishes, red-eared turtles and *Mauremys leprosa*. *Abstracts of the IV International Symposium on Emys orbicularis*, pp. 5-6. Valencia, Spain.

Margules, C.R. & Nicholls, A.O., 1987. Assessing the conservation value of remnant 'islands': mallee patches on western Eyre Peninsula, South

Australia. *Nature Conservation: The Role of remnants of Native Vegetation* (ed. by D.A. Saunders, G.W. Arnold, A.A. Burbidge & A.J.M. Hopkins), pp. 89-102, Surrey Beatty and Sons, Sidney.

Mascort, R., 1998. Distribution and status of the European pond turtle, *Emys orbicularis*, in Catalonia. *Mertensiella*, 10: 177-186.

Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A., 1997. The use of artificial neural network to predict the presence of small-bodied fish in a river. *Freshwater Biology*, 38: 237–246.

Mau-Crimmins, T., Schussman, H. & Geiger, E., 2006. Can the invaded range of a species be predicted sufficiently using only native-range data? Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modeling*, 193: 736–746.

Maurer, B.A., 1994. *Geographical Population Analysis: Tools for the Analysis of Biodiversity*. Blackwell, Oxford.

Maurer B.A, 1999. *Untangling Ecological Complexity*. The University of Chicago Press, Chicago.

May, R.M. 1994. The effects of spatial scale on ecological questions and answers. *Large-scale ecology and conservation biology*. (ed. by P.J. Edwards, R.M. May, & N.R. Webb), pp. 1-17. Blackwell Scientific Publications, London.

Mazanaeva, L. & Orlova, V., 2004. Distribution and ecology of *Emys orbicularis* in Daghestan, Russia. *Biologia*, 59 (Suppl. 14): 47-53.

Meeske, A.C.M. & Muhlenberg, M., 2004. Space use strategies by a northern population of the European pond turtle, *Emys orbicularis*. *Biologia*, 59 (Suppl. 14): 95-101.

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

Meynard, C.N. & Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *Journal of Biogeography*, 34 (8), 1455–1469.

Midgley, G.F., Hannah, L., Millar, D., Rutherford, M.C. & Powrie, L.W., 2002. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography*, 11: 445-451.

Moilanen, A., 2005. Reserve Selection Using Nonlinear Species Distribution Models. *The American Naturalist*, 165: 695–706.

Mönkkönen, M., Forsman, J. T. & Thomson, R.L., 2004. Qualitative geographical variation in interspecific interactions. *Ecography*, 27: 112-118.

Monserud, R.A. & Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, 62: 275-293.

Moran, M.D., 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, 22:403-405.

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.

Myers, J.C., 1997. *Geostatistical error management: quantifying uncertainty for environmental sampling and mapping*. Van Nostrand Reinhold, New York.

Neuhaüser, M., 2004. Testing whether any of the significant tests within a table are indeed significant. *Oikos*, 106: 409-410.

New, M., Hulme, M. & Jones, P.D., 2000. Representing twentieth century spacetime climate variability. Part 2: development of 1901-96 monthly grids of terrestrial surface climate. *Journal of Climate*, 13: 2217-2238.

Nogués-Bravo, D. & Araújo, M.B., 2006. Species richness, area and climate correlates. *Global Ecology and Biogeography*, 15: 452–460.

Olden, J.D. & Jackson, D.A., 2002. A comparison of statistical approaches for modeling fish species distributions. *Freshwater Biology*, 47:1976-1995.

Olden, J.D., Jackson, D. A. & Peres-Neto, P.R., 2002. Predictive Models of Fish Species Distributions: A Note on Proper Validation and Chance Predictions. *Transactions of the American Fisheries Society*, 131: 329-336.

Osborne, P.E. & Suarez-Seoane, S. 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling*, 157: 249 259.

Osborne, P.E., Alonso, J.C. & Bryant, R.G., 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*, 38: 458-471.

Ottonello, D., Salvidio, S. & Rosecchi, E., 2005. Feeding habits of the European pond terrapin *Emys orbicularis* in Camargue (Rhône delta, Southern France). *Amphibia- Reptilia*, 26: 562-565.

Özesmi, S.L. & Özesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, 116: 15–31.

Palmer, M.W. & Van der Maarel, E., 1995. Variance in species richness, species association, and niche limitation. *Oikos*, 73: 203–213.

Paul, R. & Andreas, B., 1998. Migration and home renge of female European pond turtles (*Emys o. orbicularis*) in Brandenburgh (NE Germany), first results. *Mertensiella*, 10: 41-46.

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, Ferrier, S. & Scotts, D., 2001. An evaluation of the predictive performance distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management*, 62: 171-184.

Pearson, R.G. & Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12: 361-371.

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

Pearson, R.G., Dawson, T.E. & Liu, C., 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27: 285 –298.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martínez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C., 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography*, 33: 1704–1711.

Pearson, R.G., Raxworthy, C.J., Nakamura, M. & Peterson, A.T., 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography,* 34: 102-117.

Peek, M.S. Leffler, A.J., Flint, S.D. & Ryel, R.J., 2003. How much variance is explained by ecologists? Additional perspectives. *Oecologia*,137: 161-170.

Pereira, P.M. & Fonseca, M. P., 2003. Nature vs. nurture: the making of the montado ecosystem. *Conservation Ecology*, 7(3): 7. [online] URL: http://www.consecol.org/vol7/iss3/art7

Pereira, J.M.C. & Itami, R.M., 1991. GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham Red Squirrel. *Photogrammetric Engineering and Remote Sensing*, 57: 1475–1486.

Peres-Neto, P.R. 2004. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*, 140: 352–360.

Peres-Neto, P.R., Olden, J.D. & Jackson, D.A., 2001. Environmentally constrained null models: site suitability as occupancy criterion. *Oikos*, 93: 110-120.

Perrin, N., 1984. *Contribution à l'écologie du genre Cepaea (Gastropoda): Approche descriptive et expérimentale de l'habitat et de la niche écologique.* PhD thesis. University of Lausanne.

Perry, J.N. & Dixon, P.M., 2002. A new method to measure spatial association for ecological count data. *Ecoscience*, 9: 133-141.

Peterson, T. Soberón, J. & Sánchez-Cordero, V., 1999. Conservatism of Ecological Niches in Evolutionary Time. *Science*, 285: 1265-1267.

Peterson, A.T., Papes, M. & Kluza, D.A., 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science*, 51: 863–868.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H. & Stockwell, D.R.B., 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature*, 416: 626–629.

Peterson, A.T., Papes, M. & Eaton, M., 2007. Transferability and model evaluation in ecological nichemodeling: a comparison of GARP and Maxent. *Ecography*, 30 (online early: doi: 10.1111/j.2007.0906-7590.05102.x)

Petren, K. & Case, T.J., 1998. Habitat structure determines competition intensity and invasion success in gecko lizards. *Proc. Natl. Acad. Sci. USA.*, 95(20):11739–11744.

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

Pleguezuelos, J.M., Márquez, R. & Lizana, M. (eds.), 2002. *Atlas y Libro Rojo de los Anfibios y Reptiles de España.* Dirección General de Conservación de la Naturaleza-Asociación Herpetologica Española (2ª impresión), Madrid.

Polasky, S. & Solow, A. R., 2001. The value of information in reserve site selection. *Biodiversity and Conservation*, 10: 1051-1058.

Pontin, A.J., 1982. *Competition and coexistence of species.* Pitman, London.

Puky, M., Gemesi, D. & Schad, P., 2004. Distribution of *Emys orbicularis* in Hungary with notes on related conservational and environmental education activities. *Biologia*, 59 (Suppl. 14): 55-60.

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

Quesada, R., 2000. La cistude d'Europe dans le Departement de l'Isere. *Chelonii*, 2 : 33-36.

Quinn, G.P. & Keough, M.J., 2002. *Experimental design and data analysis for biologists.* Cambridge University Press, Cambridge, UK.

Randin, C., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A., 2006. Are species distribution models transferable in space? *Journal of Biogeography*, 33: 1689–1703.

Rejmánek, M. & Leps, J., 1996. Negative associations can reveal interspecific competition and reversal of competitive hierarchies during succession. *Oikos*, 76: 161-168.

Robertson, M.P., Peter, C.I., Villet, M.H. & Ripley, B.S., 2003. Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, 64: 153-167.

Rodrigues, A.S.L. & Gaston, K.J., 2002. Rarity and conservation planning across geopolitical units. *Conservation Biology*, 16: 674–682.

Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. & Possingham, P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology Letters*, 9: 1136–1145.

Rowlingson, B. & Diggle, P.J., 1993. Splancs: spatial point pattern analysis code in S-PLUS. *Computers and Geosciences*, 19: 627-655.

Roxburgh, S.H. & Chesson, P., 1998. A new method for detecting species associations with spatially autocorrelated data. *Ecology*, 79: 2180–2192.

Roxburgh, S.H. & Matsuki, M., 1999. The statistical validation of null models used in spatial association analyses. *Oikos*, 85: 68-78.

Rushton, S.P., Ormerod, S.J. & Kerby, G., 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41, 193 –200.

Sanderson, J.G., 2000. Testing ecological patterns. *American Scientist*, 88: 332-339.

Sanderson, J.G., Moulton, M.P. & Selfridge, R.A., 1998. Null models and the analysis of species co-occurrences. *Oecologia*, 116: 275-283.

Schneeweiss, N., 1998. Status and protection of the European pond turtle (*Emys o. orbicularis*) in Brandenburg, Northeast Germany. *Mertensiella*, 10: 219-226.

Segurado, P. & Araújo, A.P.R., In review. Population structure of *Emys orbicularis* in syntopy and allotopy with *Mauremys leprosa*. *Revista Española de Herpetologia.*

Segurado, P. & Araújo, A.P.R., 2004a. Coexistence of *Emys orbicularis* and *Mauremys leprosa* in Portugal at two spatial scales: is there evidence of spatial segregation? *Biologia*, 59 (Suppl. 14): 61-72. (Chaper 5, section 5.2)

Segurado, P. & Araujo, M.B., 2004b. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31: 1555–1568. (Chapter 2)

Segurado, P., Araújo, M.B. & Kunin, W.E. 2006. Consequences of spatial autocorrelation on niche-based models. *Journal of Applied Ecolology*, 43: 433-444. (Chapter 3)

Segurado, P. & Figueiredo, D., 2007. Coexistence of two freshwater turtle species along a Mediterranean stream: The role of spatial and temporal heterogeneity. *Acta Oecologica*, 37: 134-144. (Chapter 5, section 5.1)

Sfenthourakis, S, Tzanatos, E. & Giokas, S. 2005. Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. *Global Ecology and Biogeography*, 15: 39–49.

Siegel, R.A. & Dodd, C.K. Jr., 2000. Manipulation of turtle populations for conservation: halfway technologies or viable options? Turtle Conservation. (ed. by M.W. Klemens), pp. 218-238. Smithsonian Institution Press, Washington.

Sidis, I. & Gasith, A., 1985. Food habits of the Caspian terrapin (*Mauremys caspica rivulata*) in unpolluted and polluted habitats in Israel. *Journal of Herpetology*, 19: 108-115.

da Silva, E., 2002. *Mauremys leprosa* (Schweiger, 1812). *Atlas y Libro Rojo de los Anfibios y Reptiles de España* (ed. by J.M. Pleguezelos, R. Marquez and M. Lizana), pp. 143-146. Dirección General de la Conservación de la Naturaleza, Madrid.

Silva, T., Reino, L.M. & Borralho, R., 2002. A model for range expansion of an introduced species: the common waxbill *Estrilda astrild* in Portugal. *Diversity & Distributions*, 8: 319–326.

Snyder, R.E. & Chesson, P., 2004 How the spatial scales of dispersal, competition, environmental heterogeneity interact to affect coexistence. *The American Naturalist*, 164(5): 634-650.

Smith, P.A., 1994. Autocorrelation in logistic regression modelling of species distributions. *Global Ecology and Biogeography Letters*, 4: 47-61.

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

Sokal, R.R. & Rohlf, F.J., 1995. Biometry: *The principles and practice of statistics in biological research*. W. H. Freeman, New York.

Sprent, P. & Smeeton, N. C., 2001. *Applied nonparametric statistical methods. Texts in Statistical Science*. Chapman & Hall/CRC, Boca Raton.

Statistical Sciences, 1999. *S-Plus, version 2000 for Windows*. Mathsoft, Inc., Seattle, Washington, USA.

Stephens, P.A., Buskirk, S.W., Hayward, G.D. & Martínez del Rio, C., 2005. Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology*, 42: 4-12.

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

Stockwell, D.R.B. & Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148: 1-13.

Storch, D., Konvièka, M., Beneš, J., Martinková, J. & Gaston, K.J., 2003. Distribution patterns in butterflies and birds of the Czech Republic: separating effects of habitat and geographic position. *Journal of Biogeography*, 30: 1195-1205.

Swets, K.A., 1988. Measuring the accuracy of diagnostic systems. *Science*, 240: 1285-1293.

Szczerbak, N., 1998. The European pond turtle (*Emys orbicularis*) in Ukraine. *Mertensiella*, 10: 259-266.

Taskavak, E. & Reimann M.J., 1998. The present status of *Emys orbicularis* (Linnaeus, 1758) in southern Central Anatolia. *Mertensiella*, 10: 267-278.

Teixeira, J., Ferrand, N. & Arntzen, J.W., 2001. Biogeography of the golden-striped salamander *Chioglossa lusitanica*: a field survey and spatial modelling approach. *Ecography*, 24: 618–624.

Therneau, T.M. & Atkinson, E.J., 1997. An introduction to recursive partitioning using the RPART routines. URL: http://www.mayo.edu/hsr/techrpt/61.pdf

Therneau, T.M. & Atkinson, E.J., 2000. An introduction to recursive partitioning using the RPART routines. URL: http://www.mayo.edu/hsr/techrpt /rpartmini.pdf

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., Ferreira de Siqueira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E., 2004. Extinction risk from climate change. *Nature*, 427: 145–148.

Thuiller, W., 2003. BIOMOD – Optimising predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, 9: 1353-1362.

Thuiller, W., 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10: 2220-2227.

Thuiller, W., Araújo, M.B. & Lavorel, S., 2003. Generalized models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, 14: 669–680.

Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L. & Lavorel, S., 2004. Uncertainty in predictions of extinction risk. *Nature*, 430: 33.

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, USA.*, 102: 8245-8250.

Thuiller, W., Lavorel, S., Sykes, M.T. & Araújo, M.B., 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions*, 12: 49–60.

Tomlin, C.D., 1990. *Geographical Information Systems and Cartographic Modeling*. Prentice Hall, New York.

Vayssières, M.P., Plant, R.E. & Allen-Diaz, B.H., 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 11: 679-694.

Vázquez, D.P., Melián, C.J., Williams, N.M., Blüthgen, N., Krasnov, B.R. & Poulin, R. 2007. Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116: 1120-1127.

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

Venables, W.N. & Ripley, B.D., 2002. *Modern Applied Statistics with S-PLUS*. Springer-Verlag, New York.

Verhoven, K.J.F., Simonsen K.L. & McIntyre L.M., 2005. Implementig false discovery rate control: increasing your power. *Oikos*, 108: 643-647.

Vetaas, O.R., 2000. Comparing species temperature response curves: population density versus second-hand data. *Journal of Vegetation Sciences*, 11: 659-666.

Walker, P.A., 1990. Modelling wildlife distributions using a geographic information system: 22 Kangaroos in relation to climate. *Journal of Biogeography*, 17: 279-289.

Walker, P.A. & Cocks, K.D., 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters*, 1: 108–118.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O. & Bairlein, F., 2002. Ecological responses to recent climate change. *Nature*, 416: 389–95.

Watkins, A.J. & Wilson, J.B., 1992. Fine-scale community structure of lawns. *Journal of Ecology*, 80: 15-24.

Weiher, E. & Keddy, P., 1995. Assembly rules, null models, and trait dispersion: new questions from old patterns. *Oikos*, 74: 159-164.

Williams, P.H. & Araújo, M.B., 2000. Using probability of persistence to identify important areas for biodiversity conservation. *Proceedings of the Royal Society London, Biological Series*, 267: 1959–1966.

Williams, P.H. & Araújo, M.B., 2002. Apples, oranges, and probabilities: integrating multiple factors into biodiversity conservation with consistency. *Environmental Modeling and Assessment*, 7: 139–151.

Zaykin, D.V., Zhivotovsky, L.A., Westfall, P.H. & Weir B.S., 2002. Truncated product method for combining P-values. *Genetic Epidemiology*, 22, 170-185.