U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

.

Environment and biodiversity

A multidisciplinary approach to dynamic patterns in the Iberian Peninsula

José Pedro Rodrigues Tarroso Gomes

Supervisors:
Paulo Célio Alves
José Carlos Brito
Rachid Cheddadi

Porto
2013

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

.

Environment and biodiversity

A multidisciplinary approach to dynamic patterns in the Iberian Peninsula

José Pedro Rodrigues Tarroso Gomes

Tese apresentada à Faculdade de Ciências da
Universidade do Porto para a obtenção do grau
de Doutor em Biodiversidade, Genética e Evolução

Porto
2013

Este trabalho foi apoiado pela Fundação
para a Ciência e Tecnologia (FCT) através da atribuição
da bolsa de doutoramento (SFRH/BD/42480/2007)

Nota Prévia

Na elaboração desta tese, e nos termos do número 2 do Artigo 4º do Regulamento Geral dos Terceiros Ciclos de Estudos da Universidade do Porto e do Artigo 31º do D.L. 74/2006, de 24 de Março, com a nova redação introduzida pelo D.L. 230/2009, de 14 de Setembro, foi efetuado o aproveitamento total de um conjunto coerente de trabalhos de investigação já publicados ou submetidos para publicação em revistas internacionais indexadas e com arbitragem científica, os quais integram alguns dos capítulos da presente tese. Tendo em conta que os referidos trabalhos foram realizados com a colaboração de outros autores, o candidato esclarece que, em todos eles, participou ativamente na sua conceção, na obtenção, análise e discussão de resultados, bem como na elaboração da sua forma publicada.

To the memory of my mother.
To Anna and Inês.

Acknowledgements

I fell back asleep some time later on
And I dreamed the perfect song
It held all the answers, like hands laid on
I woke halfway and scribbled it down
And in the morning what I wrote I read
It was hard to read at first but here's what it said

Eid ma clack shaw
Zupoven del ba
Mertepy ven seinur
Cofally ragdah

— BILL CALLAHAN, *Eid ma clack shaw*

I am able to say now that I understand the pathological psychosis inherent to the process of writing a thesis. References are a very important part of the scientific method, providing to the reader the source of information and the path to find it in the massive network of scientific works. Looking for references is, thus, a game to play most likely during all the scientific career. However, in the process of writing this thesis, the quest for the reference has expanded from the scientific sphere to the "real" world. In fact, I felt all the climate events during the process of starting, building and writing a PhD as psychological oscillations.

This thesis had a last glacial maximum ($\sim$4 yr before thesis, hereafter BT) when I didn't knew exactly how to start. A trigger event was required, soon followed by a warming trend, sun shining on a wide extent of the brain, ice gone and, thus, less albedo and more grey matter. Dating of this event was not easy, but around 3.5 yr BT I went to Montpellier to start working with Rachid Cheddadi. I am deeply grateful to him for showing me the joy of doing science, even if it was with fossil pollen! He is enthusiastic about science and that is contagious. I was able to deal with the complex pollen database and climate reconstruction methods until the onset of the Younger Dryas. It was not easy. Things can go wrong but persistence in a cryptic refugium can

be helpful, and allow future expansion of ideas. It was soon followed by an increasing trend of the psychological availability to accept science as it is, with its uncertainties, eurekas, and errors. It didn't avoid, however, the increasing aridity. An opportunity of doing field work in the Sahara desert (Mauritania, c. 3 yr BT) with José Carlos Brito made me think about many things. Also about science and the ongoing PhD. I am especially thankful to him for the supervision over these four long years, and, of course, the friendship that came along.

The transition to the Holocene wasn't a easy one, but it was followed by a warming trend, with expansion processes promoting secondary contacts with the previous work. This include periods even before the beginning of the PhD, when I met Paulo Célio and start this long collaboration that created many opportunities for me. I am thankful for his confidence in me. After some cold events, the climate reconstructions were done, several ideas got written and the thesis suddenly appeared. All these processes took place in CIBIO where I spent most of the last four years, but also at ISEM in Montpellier. I am grateful for the hosting and for collaborating with the wonderful teams that are in both places, but especially thankful to Sara and Sandra who evaporated many administrative nightmares. Also, I thank FCT for providing financial support with a PhD grant. After all these analyses, only one who has gone through all this process can know what is the feeling of the excessive warming to current levels, consequence of the increasing concentration of greenhouse gases in the cranium.

Nine months of expectation and a abrupt event of stress, changed my view of the world: my beautiful daughter was born! And, if she learned to stand up and walk always with a smile and a laugh, how could I not write a simple thesis? I didn't even had to lift my body from the chair! This puts all things in a different perspective. And Anna was tremendously helpful with everything. If I could transpose her work capacity (taking care of the baby, working in her own projects, helping others, helping me, etc) to this thesis, you would be reading this several months earlier! I miss you both every minute. Family is, thus, important to survive this process of building a PhD and chance did play in my favour here. My father provided me full support to my work at all levels, as always, and I don't know how to compensate him for that. I am thankful also to my brother and sister for tolerating my absence when I should be there helping them. A special thanks for Helena and her children who were very kind during this time.

Many people were important during the last four years. Collaborating makes the work easier, more interesting and it also makes the interdisciplinary analysis possible. Fernando is a great friend since a long time and he made the work about a viper's contact zone possible with his extensive knowledge on the species and the massive

field work. With Ricardo I had amazing conversations about the theory of the hybrid zone and, due to the different time-zones, most of those happened from sunset to the same sunset! Raquel was responsible for all the molecular methods, and was really supportive and enthusiastic about this work.

Of course, this thesis benefited immensely from the high quality reviews from Antigoni and Xico. Thank you very much for the friendship, the food, the fun, the doubts, the answers, the rides, the support, the conversations, the help, the ideas, the trust... I hope you got the idea. It is good to work in a place where making good friends is so easy. Thank you Raquel, James, Catarina, Miguel and Daniele for your support and friendship over these years. The work would feel just like work if it weren't you all! Our reunions are always very fun! I also must thank to my first great friend in CIBIO, Catarina Ferreira, who always believed in me, and to Sílvia who trusted me to be part of her own project. With Neftalí I shared many sand grains in Mauritania and it seems that I can learn something else with him every time we met. Fernando and Nuno have always been very supportive since I first entered in 'their' lab. To my long time friends Pedro, Adriano, Ivan and the amazing sisters Elin, Sofia and Sara I owe you so much... Thank you.

I have learnt that the joy of life is not doing a thesis: is having and maintaining the friends I have while doing a thesis! I sincerely dedicate this effort to all of you. I am aware that each page of this document has some of your contribute and I don't forget it.

Thank you all!

Sumário

A análise detalhada dos padrões de biodiversidade é apenas possível usando uma abordagem multidisciplinar. Neste estudo integramos dados resultantes de diferentes áreas de investigação, incluindo ecologia espacial, evolução molecular, paleoecologia e modelização climática, demonstrando o potencial destas análises abrangentes. A Península Ibérica oferece condições excelentes para o desenvolvimento destes estudos uma vez que possui níveis altos de endemismo, uma variabilidade climática excecional e uma abundante disponibilidade de dados provenientes das mais diversas áreas. A abordagem utilizada neste trabalho recorreu ao uso de diferentes escalas temporais e espaciais, abrangendo um extenso período desde o fim do Quaternário (entre há 15 000 até 3 000 anos atrás) até ao século XXI, para responder ainda a questões relativas aos processos biológicos que ocorrem a macro e micro-escala.

Neste contexto foram desenvolvidos dois novos softwares (capítulo 3): 1) o E-Clic (secção 3.1) converte previsões do clima futuro para o formato mais comum em estudos de modelização ecológica; e 2) o Simapse (secção 3.2) é uma ferramenta de modelização ecológica que implementa um algoritmo de redes neuronais artificiais para a criação de modelos usando informação da distribuição de espécies. Foi ainda adaptado e melhorado um método para a reconstrução de variáveis climáticas, resultando em três variáveis que representam treze períodos diferentes do fim do Quaternário (capítulo 4, secção 4.1).

As ferramentas computacionais e os dados climáticos produzidos foram aplicados a duas escalas espaciais (capítulo 5). Num contexto de macro-escala, realizou-se um estudo da dinâmica da riqueza de espécies de répteis e anfíbios no passado e da prevista para o século XXI, considerando a atual tendência de aquecimento climático (secção 5.1). A análise a micro-escala incidiu sobre a divergência ecológica numa zona de contacto entre três espécies de víboras (*Vipera latastei*, *V. aspis* e *V. seoanei*), usando a modelização de nicho ecológico e múltiplos caracteres (nucleares, mitocondriais e morfológicos; secção 5.2).

Durante o fim do Quaternário, o clima Ibérico caracterizou-se por uma tendência

geral de aquecimento, pontuada por eventos abruptos com amplas consequências na organização espacial das variáveis climáticas. A análise da evolução do clima permitiu classificar a área da Península Ibérica em quatro grupos distintos que partilham uma dinâmica similar de temperatura e precipitação. Os resultados a uma macro-escala revelaram que os padrões de biodiversidade foram fortemente afetados pelos processos de evolução climática, gerando áreas de alta velocidade de alterações da composição de espécies, mas também áreas de baixa velocidade relacionadas com a permanência das espécies. Os modelos ecológicos elaborados para o século atual preveem enormes alterações nos padrões de riqueza de espécies, impondo grandes desafios de conservação.

A aplicação da ferramenta Simapse no estudo à micro-escala da zona de contacto entre as três espécies de víboras estudadas permitiu uma abordagem inovadora para o estudo da divergência ecológica no limite da distribuição destas três espécies. Os resultados possibilitaram a descrição genética e ecológica da zona de contacto, caracterizando a estrutura populacional. Descreveram-se as necessidades ecológicas particulares de cada espécie, e um ecótono de transição onde foram identificados indivíduos híbridos.

A abordagem multidisciplinar aqui aplicada foi essencial para descrever exaustivamente os processos climáticos e biológicos que ocorrem à macro e micro-escala na Península Ibérica, confirmando assim o potencial da integração de múltiplas áreas científicas. É esperado que os resultados aqui obtidos possam fortalecer a conservação da herpetofauna Ibérica, e também produzir um impacto em estudos futuros de zonas híbridas num contexto ecológico. As análises efectuadas não esgotam o potencial das ferramentas desenvolvidas, sendo esperada também a ampliação das áreas de aplicação a outros domínos relacionados com a biodiversidade.

Abstract

The multidisciplinary approach to the study of biodiversity patterns offers detailed analyses that are often impossible to achieve under single-discipline studies. This is demonstrated here with an integrative approach merging data from different research fields, including landscape ecology, molecular evolution, paleoecology and climate modelling. The Iberian Peninsula offers exceptional conditions for the development of these studies: there are high levels of endemism, climate variability and abundance of data available stemming from different fields. The integrative approach presented here is extended to a different time and spatial scales, covering a wide period from the late-Quaternary (15,000 to 3,000 years before present) to the current century, with different questions addressed at macro- and micro-scales.

In this scope, two new software applications were developed (chapter 3): 1) E-Clic (section 3.1) is a converter of future climate prediction data to commonly used formats in landscape ecology; and 2) Simapse (section 3.2) is an ecological niche modelling tool that implements artificial neural networks to model species' distribution data. Moreover, thirteen spatial layers for three variables representing the late-Quaternary climate were built. This was achieved with an improved method for the reconstruction of past climate using fossil pollen data as a proxy(chapter 4, section 4.1).

Integration of all data and methods was done at two different scales (chapter 5). In a macro-scale context, it was achieved by the analysis of the dynamics of Iberian herpetofauna species composition in the past and predicted for the current century under the current trend of climate warming (section 5.1). A micro-scale analysis was performed for the study of ecological divergence in a contact zone between three viper species (*Vipera latastei*, *V. aspis* and *V. seoanei*) using ecological niche modelling and multi-trait analysis (nuclear, mitochondrial and morphological; section 5.2).

The Iberian climate in the late-Quaternary was characterized by a general warming trend with abrupt transitions that had wide impacts in the spatial organization of each climate variable. The area of the Iberian Peninsula was clustered into four distinct groups exhibiting similar patterns of climate evolution. The results of the macro-scale

analysis revealed that biodiversity patterns were largely affected by these processes, with areas of high velocities of species composition change, but also areas with low velocities related to long species persistence. By the end of the current century, major changes on species richness patterns are predicted that will pose many conservation challenges.

The application of Simapse to the study of the hybrid zone at micro-scale resulted in an innovative approach to the study of ecological divergence in the range limits of the three vipers. The results allowed a full description of the contact zone, both at genetic and environmental levels, with a characterization of the population structure. Different ecological requirements from each species were found and a transitional ecotone was suggested where the hybrids were identified.

The integrative approach followed here provided exhaustive examples describing macro- and micro-scale processes occurring in Iberian Peninsula, confirming the potential of merging results from different research fields. The results presented here are expected to support the conservation effort on Iberian herpetofauna and to have impact on future studies of hybrid zones within an ecological context. However, the potential of the developed tools is not limited to the analyses carried here, and it is also expected the expansion of the applicability to other domains related to biodiversity.

Résumé

L'approche multidisciplinaire pour l'étude des modèles de biodiversité offre des analyses détaillées qui sont souvent difficiles à atteindre dans le cadre d'études monodisciplinaires. Ceci est démontré ici avec une approche intégrative combinant des données issues de différents domaines de recherche incluant l'écologie du paysage, l'évolution moléculaire, la paléoécologie et la modélisation du climat. La péninsule ibérique offre d'excellentes conditions pour développer ces études car il y a un fort taux d'endémisme, une variabilité du climat exceptionnel et une abondance de données disponibles issues de différents domaines. L'approche intégrative présentée ici est étendue avec différentes échelles temporelles et spatiales, couvrant une large période de la fin du Quaternaire, (15.000 à 3.000 ans avant le présent) ainsi que le siècle actuel avec différentes questions abordées à des échelles différents (macro et micro).

Dans le cadre de ce sujet de recherche, deux nouveaux logiciels ont été développés (chapitre 3): 1) E-Clic (section 3.1) est un convertisseur de données de prédiction climatiques futures à des formats usuels utilisés dans l'écologie du paysage et 2) Simapse (section 3.2) est un outil de modélisation de la niche écologique qui met en œuvre une méthode statistique basée sur les réseaux de neurones artificiels pour modéliser la distribution des espèces. La méthode de reconstruction de variables climatiques à partir de données polliniques fossiles a également été adaptée et améliorée pour la quantification de treize variables représentant trois périodes de temps différentes de la fin du Quaternaire (chapitre 4, section 4.1).

L'intégration de toutes les données et les méthodes a été réalisée à deux échelles différentes (chapitre 5). Dans un contexte macro-échelle, j'ai analysé la dynamique de la composition spécifique de l'herpétofaune ibérique dans le passé et j'ai effectué une simulation prédictive pour le siècle actuel sous la tendance d'un réchauffement climatique (section 5.1). Une analyse micro-échelle a été réalisée avec l'étude de la divergence écologique dans une zone de contact entre les espèces de trois vipères (*Vipera latastei*, *V. aspis* et *V. seoanei*) en utilisant la modélisation de la niche

écologique et l'analyse multi-traits (nucléaire, mitochondrial et morphologique; section 5.2).

Le climat ibérique à la fin du Quaternaire a été caractérisé par une tendance générale au réchauffement avec des transitions brusques et des conséquences importantes sur l'organisation spatiale des variables climatiques. L'analyse de l'évolution du climat a permis de classerla péninsule ibérique en quatre zones distinctes qui partagent une même dynamique climatique (température et précipitations). Les résultats de l'analyse macro-échelle a révélé que les patrons de biodiversité ont été largement affectés par ces processus climatiques, avec des zones ayant des vitesses élevées de changement de la composition des espèces, mais aussi des zones avec des vitesses faibles liées à la longue persistance des espèces. À la fin de ce siècle, des changements importants dans les patrons de la richesse spécifique sont prédits et ils posent de nombreux problèmes de conservation.

L'utilisation de Simapse dans l'étude micro-échelle de la zone de contact entre les trois espèces de vipères a permis une nouvelle approche pour aborder la divergence écologique dans les limites de leur distribution géographique. Les résultats ont permis la description génétique et écologique de la zone de contact et la caractérisation de la structure de la population. Des exigences écologiques différentes de chaque espèce ont été mises en évidence et un écotone où les hybrides ont été identifiés a été suggéré.

L'approche intégrative suivie ici a fourni des exemples décrivant les processus macro-et micro-échelle qui se produisent dans la péninsule ibérique, ce qui confirme le potentiel de la combinaison des résultats issus de différents domaines de la recherche. Les résultats présentés ici ont pour but de soutenir l'effort de conservation de l'herpétofaune dans la péninsule ibérique et d'avoir un impact sur les futures études de zones hybrides dans un contexte écologique.

Le potentiel des outils développés dans cette thèse ne se limite pas aux analyses effectuées ici, et une extension est également prévue de son applicabilité à d'autres domaines liés à la biodiversité.

Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Networks
AP	Annual Precipitation
AUC	Area Under the Curve
BA	Bølling-Allerød
BP	Before Present
BPN	Back-Propagation Learning
CMP	Cluster Membership Probabilities
D-O	Dansgaard-Oeschger events
DBDT	Distance to Broadleaved Deciduous Thicket
DHA	Distance to Humid Areas
DNA	Deoxyribonucleic Acid
DNEF	Distance Evergreen Forests
E-Clic	Easy Climate Data Converter
ENFA	Ecological Niche Factor Analysis
ENM	Ecological Niche Modelling
ENV	Ecological Niche Variables
far	false absence rate
fAUC	final AUC
FN	False Negatives
FP	False Positives
fPCA	Functional Principal Component Analysis
fpr	false presence rate
GCM	Global Climate Models
GIS	Geographic Information Systems
GLM	Generalized Linear Model
GPL	GNU Public License
GVS	Generalist Virtual Species
H1	Heinrich event 1

HTO	Holocene Thermal Optimum
IDW	Inverse Distance Weighting
IPCC	Intergovernmental Panel on Climate Change
ka	Thousand years
LGM	Last Glacial Maximum
LR	Learning Rate
MTCM	Minimum Temperature of the Coldest Month
mtDNA	Mitochondrial DNA
MTWM	Maximum Temperature of the Warmest Month
NC	Niche Conservatism
OD	Oldest Dryas
PaD	Partial Derivatives Algorithm
PDF	Probability Density Function
PDM	Precipitation of the Driest Month
Pmin	Minimum Annual Precipitation
PR	Precision-Recall
rAUC	real AUC
ROC	Receiver Operating Characteristic curve
RVS	Random Virtual Species
Simapse	Simulation Maps for Ecological Niche Modelling
SVS	Specialist Virtual Species
TAR	Temperature Annual Range
teAUC	test AUC
Tjan	January Minimum Temperature
Tjul	July Maximum Temperature
trAUC	train AUC
Vm	Velocity of Pixel Value Maintenance
VS	Virtual Species
Vs	Velocity of Richness Shifts
YD	Younger Dryas

Chapter 1

Introduction

> There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved.
>
> — CHARLES DARWIN, *The Origin of Species*

> If artificial selection makes such changes in only a few thousand years, what must natural selection working for billions of years be capable of? The answer is all the beauty and diversity in the biological world.
>
> — CARL SAGAN, *Cosmos*

1.1 Climate evolution

1.1.1 General trends in climate oscillations

The climate on Earth is continuously changing. At a time scale that humans can easily perceive, climate is dynamic with a known periodicity, that translates to the succession of seasons. This effect is known to be dependent on interactions between the Earth and the Sun. However, this relation also causes a longer periodicity that is promoting greater shifts on Earth, ranging from warm trends with ice-free polar caps to glacial ages, known as the succession of glacial and interglacial cycles. The Milankovitch theory summarises three relevant periodic effects of Earth's orbital components (Fig. 1.1;Hays *et al.*, 1976; Zachos *et al.*, 2001): 1) eccentricity which measures the shape of Earth's orbit around the sun, ranging from circular to elliptical, with a

period of 400 ka and 100 ka cycle; 2) obliquity or axial tilt is the angle between the Earth's rotational axis and the plane formed by its orbit and changes with a periodicity of 41 ka; and 3) axial precession, with a period of 23 ka, is related to the orientation of the axis. The interaction between these orbital components determines the insolation on Earth which is the amount of energy that the planet receives. Climate evidences from the geological record are in close agreement with the predicted pace of climate (Hays *et al.*, 1976; Zachos *et al.*, 2001; Cheddadi *et al.*, 2005). Other responses of climate have been documented and attributed to Earth's geographical and topological features and phenomena, as the concentration of atmospheric greenhouse gases (Pielke *et al.*, 1998; Zachos *et al.*, 2001; Clark *et al.*, 2012).

As seen, these large-scale factors dominate the periodic nature of glacial and interglacial cycles. Instead of a smooth transition from cold to warm conditions, the period after a glaciation event also inherits this oscillating nature with abrupt climate transitions (Alley *et al.*, 2003; Cheddadi *et al.*, 2005; Clark *et al.*, 2012). This has been particularly evident since the last post-glacial process until the present, when a chain

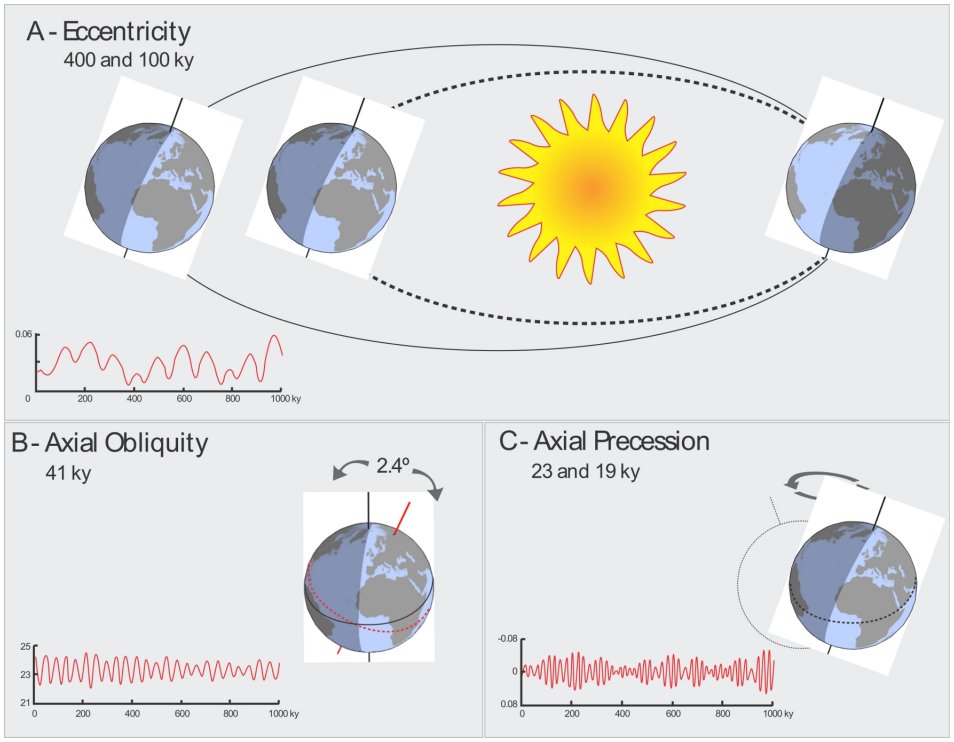


Fig. 1.1 – Major orbital perturbations defining long term climate cycles by changing the amount of radiation reaching the planet, as predicted by the Milankovitch theory. Orbital eccentricity refers to the shape of Earth's obit around the Sun, responsible for long cycles of 400 and 100 ka. The axial obliquity is the angle formed by Earth axis and it orbit and has a medium period of 41 ka. The axial precession refers to the orientation of the axis and controls shorter cycles of 23 and 19 ka. Combinations of these periods are responsible for major insolation changes on Earth. Adapted from Zachos *et al.* (2001).

of cold and warm events took place (Bond *et al.*, 1993, 1997; Dansgaard *et al.*, 1993; von Grafenstein *et al.*, 1999; Renssen & Isarin, 2001; Clark *et al.*, 2012). During the glacial period, the Dansgaard-Oeschger events (D-O) were responsible for warming trends with a rhythm of 1,470 years that were followed by a longer colder period (Bond *et al.*, 1993; Dansgaard *et al.*, 1993; von Grafenstein *et al.*, 1999). These events are generally followed by oceanic Heinrich events, resulting in fast cold-to-warm transitions with massive discharges of icebergs in the North Atlantic (Bond *et al.*, 1993). The Heinrich event 1 (H1), simultaneous to the cold period of the Oldest Dryas (OD; 18.0 - 14.6 ka), ended with the Bølling-Allerød warm interstadial. The last D-O event includes a period of cold climate between ∼12.9 to 11.6 ka known as the Younger Dryas (YD), and considered as a Heinrich event by some authors (Bond *et al.*, 1993), marking the beginning of the Holocene.

The Holocene, although more stable, had also a ∼1470 year cycle responsible for the major shifts to cold during this epoch, including the smaller 8.2 ka event (Bond *et al.*, 1997; von Grafenstein *et al.*, 1999; Heiri *et al.*, 2004). However, this period of time is clearly marked by the Holocene thermal optimum (HTO; ∼11 to 5 ka) with increasing summer temperatures resulting by the combination of orbital forcing and feedback coming from the melting of the Laurentide ice sheet (Seppä & Birks, 2001; Renssen *et al.*, 2009).

1.1.2 European climate

Several reconstructions depicted most climate oscillations in Europe, exposing a divergent spatial pattern of climate-related variables (Birks & Ammann, 2000; Lotter *et al.*, 2000; Renssen & Isarin, 2001; Seppä & Birks, 2001; Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009). At local and regional scales, reconstructions of temperature and precipitation values are usually based on fossil pollen data (e.g. Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009) or fossil remains of invertebrates (e.g. Tóth *et al.*, 2012). Using fossil pollen, Davis *et al.* (2003) found different responses in Europe for mean annual temperature: north and central western Europe have a ∼4ºC anomaly between 12 ka and the present, while southern western and eastern Europe were generally warmer, with anomalies peaking at -3 and 3ºC, respectively. Major differences in Europe were found in winter temperature, being the summer temperatures very stable. Similar results were found by Cheddadi & Bar-Hen (2009), describing a gradient between colder north-western and warmer south-eastern Europe.

The succession of events after the LGM had vast repercussions in Europe. The retraction of the ice after the LGM exposed a great area of land (Fig. 1.2), counter-

balanced by the increase of sea level (Peltier, 1994). The cold events that took place in this period had a wide impact on the summer temperatures. During the OD there was an expansion of the glacial vegetation indicating lower temperatures in southern Europe (Naughton *et al.*, 2007) and colder summers in the east (Tóth *et al.*, 2012). During the YD there was a fast decline to ∼9 ^{o}C (Birks & Ammann, 2000; Lotter *et al.*, 2000; Renssen & Isarin, 2001) and also precipitation suffered a drastic decrease (Dormoy *et al.*, 2009). The HTO has a strong latitudinal variation. Evidences of warmer and more humid summers were found in northern Europe (Seppä & Birks, 2001; Davis *et al.*, 2003), but the reconstructions for southwestern Europe point toward cooler summers. At 8.2 ka, during a brief temporal period (∼600 years), the temperature suddenly decreased about ∼1 ^{o}C (Heiri *et al.*, 2004).

During the past decades, the scientific community converged efforts on the as-

Fig. 1.2 – Extension of the ice sheet during the last glacial maximum (∼21 ka). The grey area corresponds to the ice cover during the LGM, when the Weichselian ice sheet covered a great extent of northern Europe, and the Laurentide ice sheet covered most of northern North America. Light green depicts the land area exposed during this period due to the drop of sea level. Based on data from Peltier (2004).

sessment of other factors shaping global climate, mainly those related to the impact of human activities. Climate changes are driven by changes in the insolation, which, by its turn, is controlled by orbital parameters and Earth features, such as topography, plate tectonics, oceanic circulation, surface albedo and concentration of greenhouse gases (Pielke *et al.*, 1998; Zachos *et al.*, 2001; Clark *et al.*, 2012). The interactions between atmosphere and terrestrial ecosystems are also an important factor modelling Earth's climate. This relation is largely controlled by feedbacks processes triggered by water evaporation from plant transpiration, carbon cycle and human activities shaping the landscape (Pielke *et al.*, 1998). Larger-scale anthropogenic influence on the European landscape has its roots in the development of agriculture, especially during the late-Quaternary (Berglund, 2003; Jalut *et al.*, 2009). More recently, the escalating concentration of greenhouse gases in the atmosphere has already been reflected in an increase of temperatures with an anthropogenic origin (IPCC, 2007b). The human climate forcing has a predicted increase of severity for the next century with wide physical and biological consequences (IPCC, 2007a; Rosenzweig *et al.*, 2008).

In summary, the global climate since the LGM was a succession of cold and warm events. We have seen that some were very drastic and intense, whereas others persisted for longer periods. What is the spatial impact of the velocity of climate change? This question arises frequently due to the wide impact on biodiversity, and answers point towards a simple explanation: complex topographies are more prone to refrain climate change while flat areas have higher velocities (Loarie *et al.*, 2009; Sandel *et al.*, 2011). Which consequences may interactions between climate and topography have on the general patterns of biodiversity?

1.2 Species sensitivity to climate change

1.2.1 Consequences of past climate change

The major biological patterns are largely controlled by the climate (Parmesan & Yohe, 2003; Araújo & Rahbek, 2006; Araújo *et al.*, 2008; Willis & MacDonald, 2011) and the velocity of climate change is a major force behind the global patterns of biodiversity. It has already acted in the past, shaping the current distribution of endemic species (Sandel *et al.*, 2011) and it is predicted to be an extremely severe shaping force during the next century, given the scenarios of climate change (Loarie *et al.*, 2009; Burrows *et al.*, 2011). The velocity of climate change since the LGM is noticeably higher at northern latitudes, reaching a peak of $\sim$170 m/yr on Canada (Sandel *et al.*, 2011). This corresponds to the area that was mostly covered by the Laurentide ice sheet

during the last glacial period (see Fig. 1.2). Despite the relatively lower velocities, Europe has a south-to-north gradient of increasing velocities (Sandel *et al.*, 2011), highlighting the prominent role of biodiversity refugia of the Mediterranean basin. The milder climate of southern European latitudes allowed several species to persist during the harsh conditions of the last glaciation (Weiss & Ferrand, 2007). Sandel *et al.* (2011) related high levels of endemism (in the author's view, small-ranged species of amphibians, mammals and birds) with lower velocities of climate change, which creates an interesting connection to the world hotspots of biodiversity, like the above case of the Mediterranean basin (Myers *et al.*, 2000).

The velocity of climate change is known to have a strong impact on biodiversity, however, the mechanisms behind Quaternary extinctions are not fully understood. The interactions between Quaternary climate change and the sprout of human populations are, nevertheless, related to most extinctions during the last glaciations (Burney & Flannery, 2005; Koch & Barnosky, 2006; Wroe *et al.*, 2006; Brook *et al.*, 2008; Nogués-Bravo *et al.*, 2008, 2010; Barnosky *et al.*, 2010). Although there are some discrepancies between authors on which factor is more influential in this interaction (see, for instance, Burney & Flannery, 2005; Wroe *et al.*, 2006), there is a consensus that the combined action of climate warming and human activities probably led many species to extinction. On one side, temperature increase since the LGM restricted the area where some cold-adapted species were able to thrive (Nogués-Bravo *et al.*, 2008, 2010). Additionally, humans, either directly by active hunting or indirectly by inducing major landscape changes and introducing predators, may have caused drastic reductions of population sizes (Burney & Flannery, 2005; Koch & Barnosky, 2006). For instance, through the analysis of potential niche of the woolly mammoth, *Mammuthus primigenius*, in different time-frames, Nogués-Bravo *et al.* (2008) suggested that the survival of this species would have been possible in very small patches and, thus, attributed the extinction to the synergistic effects of climate change and human activities. In Eurasia, extinction processes in the Quaternary are responsible for the loss of 36% of the mega-fauna genera (Barnosky *et al.*, 2010) and are divided in two stages: between 46 ka and 22 ka, with the loss of warm-adapted species; and between 12 ka and 8 ka, during the Pleistocene-Holocene transition, with the extinction of cold-adapted species (Barnosky *et al.*, 2010).

The impact of climate change in species extinctions is an important factor modulating biodiversity during the Quaternary. Nevertheless, the migration response or resilience of extant species to climate change defines the current biodiversity patterns (Svenning & Skov, 2007; Araújo *et al.*, 2008; Sandel *et al.*, 2011). During the climate

oscillations of the Quaternary, species experienced several contractions, expansions and extirpation of their ranges to track suitable conditions (Taberlet *et al.*, 1998; Hewitt, 2000; Davis & Shaw, 2001; Hewitt, 2004; Taberlet & Cheddadi, 2002). In general, migration paths in Europe follow a south-to-north direction, corresponding to the spread of species from the Mediterranean area, which broadly served as refugia, despite the recent ramifications of this term (Ashcroft, 2010). The analogous climate areas between present and the LGM in Europe revealed interesting relations with the broad pattern of refugia areas (Ohlemüller *et al.*, 2012). These authors divided Europe in source and sink areas depending on whether they had currently wide areas of analogous or non-analogous climate. The southern Europe, mostly below 45^{o} of latitude, revealed an extremely high source potential, along with some northern areas in France and Germany. Sink potential was found to the east of 10^{o}, covering most of Eastern Europe. Despite the importance of the latitudinal movement of the species when tracking the changing climate, the altitudinal gradient was equally important (Davis & Shaw, 2001).

Evidences of plants migrations during the late Quaternary are abundant from the paleorecord. In Europe, plant assemblages during the LGM exhibited a gradient from southern steppe to northern tundra that evolved mainly to forest-like biomes during the middle Holocene (Elenga *et al.*, 2000; Prentice *et al.*, 2000; Cheddadi & Bar-Hen, 2009; Clark *et al.*, 2012). Herbaceous taxa dominated the pollen record during the LGM and higher densities of tree pollen appear in southern Europe at 13 ka and increased pronouncedly all over Europe until 5 ka (Cheddadi & Bar-Hen, 2009). Migrations routes drawn from fossil records and phylogenetic analysis of species from the genus *Pinus* and for *Fagus sylvatica* located several potential refugia in southern Europe (Cheddadi *et al.*, 2006; Magri *et al.*, 2006). Along with the southern European peninsulas, cryptic refugia was also detected for tree species in central and northern Europe (Cheddadi *et al.*, 2006; Magri *et al.*, 2006; Magri, 2010; Svenning *et al.*, 2008). The same pattern of species' persistence and migration routes is found on animals (Taberlet *et al.*, 1998; Hewitt, 2000, 2004). The paradigmatic cases of the grasshopper (*Chorthippus parallelusi*), the hedgehog (*Erinaceus europeus*) and the bear (*Ursus arctos*) reveal expansion from southern peninsulas, with differences on the extent of expansion from the source in each case (Hewitt, 2000). Others, like small mammals and vipers, revealed cryptic refugia north to the classic Mediterranean area (Ursenbacher *et al.*, 2006; Provan & Bennett, 2008; Fløjgaard *et al.*, 2009; Vega *et al.*, 2010).

The oscillations in species' ranges have consequences for interspecific genetic

structure (Hewitt, 1999, 2000, 2004; Petit *et al.*, 2003). The harsh conditions during the glacial period that forced temperate species to reduce their ranges to the southern Europe, also promoted vicariance effects, reducing gene flow and increasing the differentiation between lineages (Taberlet *et al.*, 1998). The horizontal colonization pattern is expressed through wider areas than the vertical, which is restricted to mountain ranges and have less available area for expansion processes. The fast expansion of a population creates a gradient of high to low genetic diversity from source areas to the edge of the distribution (Hewitt, 1999, 2000, 2004). Although species respond differently to climate change, most populations suffered strong bottlenecks and founder effects, recovering from a small number of individuals with a rapid expansion (Taberlet *et al.*, 1998; Hewitt, 1999, 2000, 2004). This resulted in a loss of genetic diversity at the leading edge of the range expansion (Hewitt, 1999). Moreover, the establishment of new populations may hamper the introduction of new individuals and, thus, prevent the increase of diversity (Hewitt, 1999). In plants, due to the dispersal mechanisms by seeds and pollen, adaptation was found important both in the edges of expansion and in the core of the population distribution (Davis & Shaw, 2001). As a result of the expansions that took place after the LGM, several suture zones are found in Europe, which correspond to hybrid zones where previously isolated lineages met in secondary contact (Taberlet *et al.*, 1998; Hewitt, 1999, 2000, 2011). Important suture zones in Europe are coincident with major topographic features like the Alps and the Pyrenees. Populations dispersing from the Iberian and the Balkan Peninsulas, met in a generally flat area corresponding to France and Germany. It is interesting to notice that climate change velocity is higher in flat areas (Loarie *et al.*, 2009; Sandel *et al.*, 2011), probably assisting longer dispersal events. A last common hybrid zone in Europe resides in the Scandinavian Peninsula, where species from several refugia met. The exact location of suture zones depends on the interactions between intervening organisms, and tend to float more on flat areas (Taberlet *et al.*, 1998).

1.2.2 Predicted impacts of future changes

During the late Quaternary, species responded rapidly to climate oscillations. Will biodiversity be able to cope with future climate change velocity? The predicted climate change for the next century raises several conservation issues (Skov & Svenning, 2004; Thomas *et al.*, 2004; Araújo & Rahbek, 2006; Araújo *et al.*, 2006; IPCC, 2007a; Petit *et al.*, 2008; Loarie *et al.*, 2009; Willis & Bhagwat, 2009; Carvalho *et al.*, 2010a,b; Dawson *et al.*, 2011; Hof *et al.*, 2011; Schloss *et al.*, 2012). The impact of recent climate change (over the last century) is already noticeable on several species (Root

et al., 2003; Parmesan, 2006; Moritz *et al.*, 2008; Tingley *et al.*, 2009) and is predicted to result in multiple extinctions (Parmesan, 2006) and, thus, raise the need for conservation measures. There are multiple operators involved on global extinctions besides climate change (Brook *et al.*, 2008; Willis *et al.*, 2010), and lessons from the past, through the study of fossil evidence and climatic reconstructions, show high resilience of biodiversity to past changes, giving a new breath to biodiversity conservation (Willis *et al.*, 2010). However, the authors advert that future changes have a very different nature, given the higher velocity predicted. In fact, Sandel *et al.* (2011) predicts a difference of more than two orders of magnitude between past (since LGM) and future climate change velocities. Thus, dispersal ability will be one of the most influential processes defining species' success in the next century (Schloss *et al.*, 2012), as it was in the past. Biodiversity conservation under such scenarios of rapid climate change is difficult but several avenues of actions have been proposed for conservation at a regional level (Carvalho *et al.*, 2010a), either by bearing in mind the balance between cost and effectiveness (Carvalho *et al.*, 2010b) or by incorporating evolutionary processes into conservation design (Klein *et al.*, 2009).

1.3 Evolutionary patterns in the Iberian Peninsula

1.3.1 Characteristics of Iberia

The climate dynamics patterns observed in Europe are also seen at lower spatial scales in the Iberian Peninsula. Although it is considered part of a southern European refugium with a high potential of being a source area for species spreading (Ohlemüller *et al.*, 2012), climate was not stable in Iberia. High amplitude climatic oscillations were also frequent in Iberia during the late Quaternary, following the general warming trend from harsh conditions and promoting shifts of the species' ranges within the Iberian Peninsula. All major climatic events of the Quaternary are discernible in the Iberian margin fossil record with wide impact on land biodiversity (Sánchez-Goñi *et al.*, 2000; Roucoux *et al.*, 2005; Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008; Fletcher *et al.*, 2010).

The Iberian Peninsula has peculiarities that make it a unique area in Europe. Located in the south-western part of Europe, the peninsula is partially isolated from the rest of Europe with a strong geographical barrier of the Pyrenees mountains (Fig. 1.3). Similarly, to the south, it is bathed by the Mediterranean Sea with the relatively small Strait of Gibraltar ($\sim$14km wide) separating it from the African continent (Fig. 1.3). Such isolation features hampers the dispersal of several species, particularly non-

flying animals or short-dispersal plants, though the strait and the Pyrenees mountains are known to have been permeable to species dispersal several times in the past (e. g. Petit *et al.*, 2003; Carranza *et al.*, 2006; Arroyo *et al.*, 2007). The high level of endemism in Iberia is comparable to those exhibited by the other peninsulas of the Mediterranean Basin and was recognized as part of a global biodiversity hotspot (Myers *et al.*, 2000). The Iberian topographic features include two plateaus that dominate the landscape north and south of the Central Mountain System (Fig. 1.3). Other mountain systems include the northern Cantabric mountains, the southern Baetic system, the Sierra Morena mountains, and the eastern Iberian system (Fig. 1.3). The peninsula exhibits two major bioclimatic zones: the Mediterranean that covers most of the peninsula, except northern areas, and the mountain ranges where Atlantic bioclimate dominates.

1.3.2 Climate change and landscape dynamics

During the LGM, the Iberian Peninsula was cold and dry (Naughton *et al.*, 2007; Fletcher *et al.*, 2010), however, these conditions were extreme nearly after the LGM, at the OD (Roucoux *et al.*, 2005; Naughton *et al.*, 2007; Fletcher *et al.*, 2010). During this period, Iberian landscapes were dominated by steppe and tundra vegetation (Sánchez-Goñi *et al.*, 2000; Carrión *et al.*, 2010a), with constrained forests in the northwest and south (Carrión, 2002; Muñoz-Sobrino *et al.*, 2006; Carrión *et al.*,

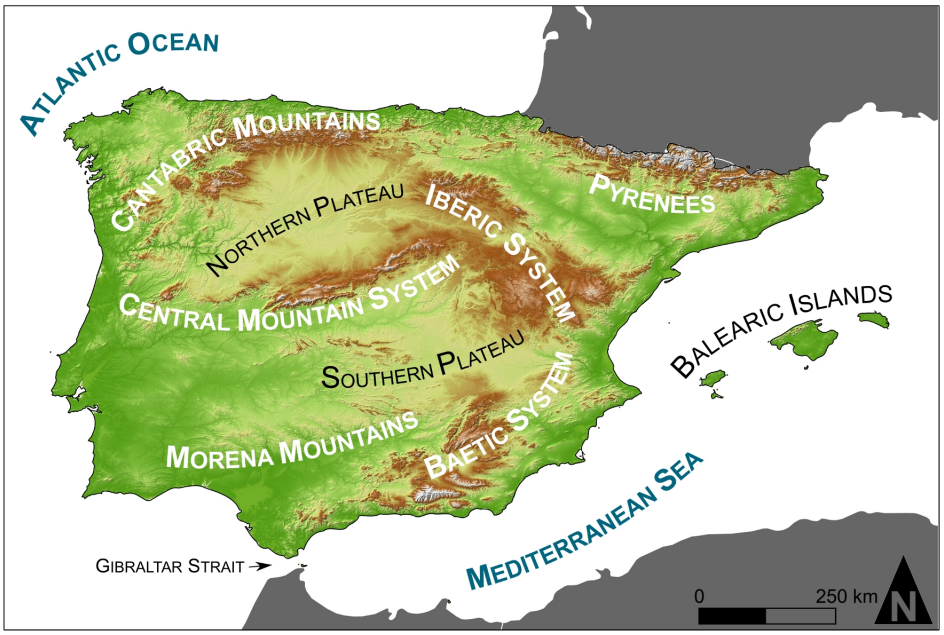


Fig. 1.3 – The Iberian Peninsula and Balearic island with major topographic features.

2010a), though the LGM had enough humidity to support the development of shrub vegetation (Fletcher & Sánchez-Goñi, 2008). The late OD witnessed the expansion of pioneer trees like the case of *Betula* in the north (Muñoz-Sobrino *et al.*, 2006). The BA is marked by the increasing temperature and precipitation, during which warm and moist conditions were propitious for the advance of forests over steppe landscapes (Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008; Carrión *et al.*, 2010a). The expansion of pine and oak forests is very intense during this period (Fletcher *et al.*, 2007; Naughton *et al.*, 2007; Carrión *et al.*, 2010a), at least until the onset of the YD, when abrupt drops of temperature and precipitation took place.

Temperature reconstructions show that southwest Europe had a severe transition from the YD to Holocene, similarly to northern and central Europe, but much colder than the southeast (Davis *et al.*, 2003). This pattern was confirmed by different temperature reconstruction methods from fossil pollen (Cheddadi & Bar-Hen, 2009). The severe decrease of temperatures on the YD is not noticeable in all Iberian pollen sites, though some reveal extreme sensitivity to the climatic event (Carrión *et al.*, 2010a). Marine pollen records show that the Iberian Peninsula experienced a strong retraction of temperate forests during this period, particularly oak, with the return to steppe landscape dominated by herbaceous plants (Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008; Carrión *et al.*, 2010a). Nevertheless, *Betula* trees expanded slightly during this stage (Muñoz-Sobrino *et al.*, 2006; Naughton *et al.*, 2007).

The transition to the Holocene has similarities to the BA period. Cold-to-warm changes marked the end of the YD and the beginning of the new period. A striking pattern observed in temperature reconstructions is the absence of the HTO in Iberia, while it was evident in other European areas (Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009). However, the expansion of *Quercus*, *Corylus* and *Alnus* in northern Iberia (Naughton *et al.*, 2007) and of other thermophile vegetation, including species from the genera *Quercus* and *Olea*, in central and southern Iberia, lead the general increase of forest cover (Dorado-Valiño *et al.*, 2002; Fletcher *et al.*, 2007), indicating an rise of temperature and moisture availability. The substitution of well established pine by oak forests is often regarded as a result of climate change and/or a modification of the fire regime (Carrión *et al.*, 2010a). After 5 ka there was a general expansion of green shrublands, with a decrease of *Quercus* and *Pinus* forests (Dorado-Valiño *et al.*, 2002; Fletcher *et al.*, 2007; Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008; Carrión *et al.*, 2010a). The cold events of low amplitude at 10.1, 9.3, 8.2 and 7.4 ka are reflected in forest declines (Fletcher *et al.*, 2010).

Despite the impact of climate dynamics in the Iberian landscape, the anthro-

pogenic source of disturbance was also vast, however, very difficult to date. It may have started after the Neolithic with the advent of agriculture in Iberia (Zapata *et al.*, 2004). Nevertheless, human impact on the landscape remained at the minimum levels and centred in high altitude grasslands (Carrión *et al.*, 2010a,b). Signals of agricultural presence (e.g. *Olea*, *Castanea* and *Cerealia*) in north-western Iberia are dated after 2,000 yr BP (Allen *et al.*, 1996), but evidences show earlier agriculture activities in the northeast (Allen *et al.*, 1996; Zapata *et al.*, 2004). Forest decline reached its peak during the Roman invasion of the peninsula, around 2,000 yr BP, following a period of increasing resources exploitation by humans that, combined with increasing aridity, contributed to the spread of xerophytic vegetation (Carrión *et al.*, 2010b).

1.3.3 Cryptic refugia in a glacial refugium

The climate oscillations in Iberia had a wide impact on species, by directly affecting them, due to the harsh conditions out of their niche range, or indirectly due to landscape changes as the vegetation composition, as seen above. In general, a dynamic pattern of suitable habitats emerged and species responded with contractions and expansions of their ranges within Iberia, following the hypothesis of small glacial refugia inside a general European refugium (Gómez & Lunt, 2007; Médail & Diadema, 2009). Fossil pollen evidence for multiple refugia in the Iberian Peninsula (e.g. Cheddadi *et al.*, 2006) is well supported by molecular analysis showing that species of the genus *Quercus* had different refugia (Olalde *et al.*, 2002; López de Heredia *et al.*, 2007), and also by potential distribution models for species of the genus *Abies* (Alba-Sánchez *et al.*, 2010) and *Pinus* (Benito Garzón *et al.*, 2007). Moreover, evidence from multiple refugia also arrives from Iberian fauna. Similar patterns were found in mammals (Branco *et al.*, 2002; Jaarola & Searle, 2004; Rebelo *et al.*, 2012), reptiles (Godinho *et al.*, 2006; Miraldo *et al.*, 2011), invertebrates (Vila *et al.*, 2005; Muñoz *et al.*, 2008) and rotifers (Campillo *et al.*, 2011).

What consequences do these patterns of refugia and colonization routes have? In Iberia, the colonization routes after several vicariance events led to a complex network of hybrid zones. The expansion of species' ranges with the advance of milder climates allowed the lineages previously isolated in small patches of suitable habitat to meet and promote the gene flow. This is observed both at intraspecific level (e.g. Sequeira *et al.*, 2005; Godinho *et al.*, 2006), at interspecific level (e.g. López de Heredia *et al.*, 2007; Geraldes *et al.*, 2008), and also between Iberian species and others expanding from refugia elsewhere (e.g. Martínez-Freiría *et al.*, 2006; Melo-Ferreira *et al.*, 2007).

As a study case, the Iberian vipers are very appealing because they exemplify the aforementioned impacts of retraction into refugia and posterior colonization. Three parapatric species of vipers are currently described for the Iberian Peninsula: 1) *Vipera latastei* with a wide distribution covering most of Iberia except the north and also northwest Africa; 2) *Vipera aspis*, occurring in northeast Iberia and extending to the east, covering most of France and Italy; and 3) *Vipera seoanei* with a distribution restricted to the northwest of the peninsula. While the first two are closely related, *V. seoanei* is phylogenetically distant from the others (Garrigues *et al.*, 2005). The distribution of the three species intersects in northern Iberia, in a transition area between the Atlantic and Mediterranean bioclimates (Martínez-Freiría *et al.*, 2006, 2008, 2009), where *V. aspis* and *V. latastei* apparently coexist (Martínez-Freiría *et al.*, 2006). This crossroad between Atlantic and Mediterranean climates with colonization routes arriving from Europe, is prone to maintain hybrid zones for several species (Bridle *et al.*, 2001; Melo-Ferreira *et al.*, 2007), mainly due to the ecotone resulting from the bioclimatic transitions.

The particular characteristics of the Iberian Peninsula described above, renders it as a unique place to study evolutionary processes including speciation events, impact of isolation episodes, and hybridization phenomena. Moreover, due to the edaphologic compositions of the Iberian Peninsula, areas with adequate conditions to preserve pollen over several millennia are abundant, as are also the studied pollen records. These allow to analyse in detail the evolution of the landscape, showing flora migrations and refugia, and to reconstruct climate serving as a biological proxy.

1.4 Reconstructing the past, predicting the future

1.4.1 Pollen as a proxy of the past

The impact of climate change on biodiversity is observed on the distribution of species. The biological material is deposited in layers through time accordingly to the climatic and geological changes operating, constituting a well preserved record of past changes. Although there are a multitude of biological proxies available to study past environments (Anderson *et al.*, 2006; Birks & Birks, 2006; Willis *et al.*, 2010), a very common proxy is the fossil pollen. An immense quantity of information is preserved in fossilized pollen, including the flora present at the range period covered in the record and its relative abundance. The multidimensional information about the landscape, together with the ubiquity of sites with conditions for preserving pollen through ages on land (e.g. Allen *et al.*, 1996; Carrión, 2002; Cheddadi *et al.*, 2005; Roucoux *et al.*,

2005; Fletcher *et al.*, 2007) and oceans (e.g. Sánchez-Goñi *et al.*, 2000; Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008), renders this proxy as one of the most used to study past environments. The analysis of a single site may provide enough information for a detailed description of the surroundings areas of the pollen site. However, the comparative analysis of nearby sites contributes to a better understanding of the spatial evolution of the area under study. Several levels of spatial information may be retrieved in this way, including the distribution of a particular taxon (e.g Cheddadi *et al.*, 2006), or wider combinations of taxa like trees (e.g. Cheddadi & Bar-Hen, 2009), and of more complex plant assemblages as biomes (e.g. Prentice *et al.*, 1996; Cheddadi *et al.*, 1997; Prentice *et al.*, 2000).

Fossil pollen presents several advantages over other climate indicators: 1) it offers a multitude of data of past landscapes that are convertible in quantitative climate reconstructions; 2) it is virtually ubiquitous since plants usually produce large quantities of pollen and it is also used at long-dispersal events; 3) it provides large quantities of samples per depth for the same reasons as before; and 4) pollen grains are relatively easy to identify due to well preserved structures and large databases of reference collections (Traverse, 2007). Some of the disadvantages of using this proxy are related to the specific medium conditions for pollen preservation and to the poor identification resolution for some taxa. The small size of the pollen grains may require also expensive laboratory material as electronic microscopes (Traverse, 2007).

Due to the presence of biological material, the radiocarbon (C^{14}) is used frequently for dating the cores covering the late Quaternary. This method is based on measurements of the ratio of carbon C^{12} to C^{14} content in the sample and on the comparison with a living sample. This allows to calculate the age based on the half-life of C^{14}, producing very accurate results. However, since the radiocarbon content in the atmosphere has not been stable through time, a calibration is needed to accurately convert C^{14} ages to calendar ages (Reimer *et al.*, 2009).

The past distribution of plants is a direct evidence from fossil pollen assemblages. A complete analysis of the dynamics of past ecosystems is possible using this source of information (Sánchez-Goñi *et al.*, 2000; Naughton *et al.*, 2007; Fletcher & Sánchez-Goñi, 2008; Carrión *et al.*, 2010a), as well as the analysis of the evolution of vegetation types (Prentice *et al.*, 1996, 2000; Cheddadi *et al.*, 1997; Williams *et al.*, 2004). The time-series contained in pollen records allows identifying refugia and inferring migration routes and rates, particularly when complemented with molecular analysis (Davis & Shaw, 2001; Cheddadi *et al.*, 2006; Magri *et al.*, 2006). One interesting case resulting from the direct analysis of pollen is the Reid's paradox, which

describes the discrepancies between the higher rates of trees' past migration inferred from fossil pollen and the lower migration estimates based on current data (Clark *et al.*, 1998). This fast response of vegetation to climate change is evident on the sequential analysis of biomes during the late Quaternary (Williams *et al.*, 2004).

1.4.2 Quantitative reconstructions of climate

The last two decades have witnessed an increasing use of biostratigraphical data to produce quantitative estimates of past climate change instead of purely descriptive analyses (Birks, 1998). The quantitative reconstruction of past climates relies on the record of vegetation compositions present in the pollen assemblage, and on the assumption that plant distributions are in equilibrium with climate (Webb III, 1986), which is supported by the bidirectional relation of influence between ecosystems and climate (Pielke *et al.*, 1998). Although this assumption is rarely limiting, reconstructions are only possible if enough is known about the ecological niche of the taxa being used. Moreover, the reconstruction methods can be further improved to provide past climate values with less uncertainty (Anderson *et al.*, 2006). Birks *et al.* (2010) summarized other assumptions for climate reconstruction: 1) the climate variables to be reconstructed should be ecologically meaningful for the taxa used; 2) fossil data and recent data (training data) should include the same biological entities; 3) the algorithms being used should have sufficient power to derive reconstructions with low uncertainties; 4) other environmental variables should have negligible effects on the taxa envelope and exclude other vegetation assemblages; 5) the validation dataset must be independent from the training set. A final assumption, more general to pollen analyses, is the linear relationship between pollen percentages and plant density (Fig. 1.4).

Different reconstruction algorithms have different degrees of tolerance to these assumptions and are classified under three major categories (Birks & Birks, 2006; Birks *et al.*, 2010). The first is the multivariate calibration-function approach, also known as transfer function, which relies on surface sediments to establish a relation, usually by means of linear or non-linear regression, with an environmental variable. This transfer function is used to project the relation with climate variables to past pollen assemblages and, thus, generates a reconstruction of past climate. Although accurate reconstructions have been built with this method (e.g. Birks & Ammann, 2000), one of its main disadvantages is its dependence on surface sediments, which may have been altered by human activities and are likely to result in incorrect transfer functions. The second family of methods is the assemblage approach that includes the widely used modern analogue technique (e.g. Cheddadi *et al.*, 1997; Davis *et al.*, 2003). Modern

pollen assemblages are compared to fossil pollen assemblages using a dissimilarity metric (usually chord distances) to find the most likely combination acting in the past and, thus, producing a climate reconstruction using the climate values of the equivalent (less dissimilar) modern combination. Disadvantages usually attributed to this approach are the unknown number of analogues to include in the reconstructions and the possible absence of past compositions in the modern distributions of taxa (Birks *et al.*, 2010). The third approach, like the previous one, is based on modern distribution of taxa and includes several algorithms under the name of indicator-species approach. The technique used in this approach relates to the climate envelope for each taxon under study. Current plant distributions are related with an ecological pertinent modern climate variable to build the climate range where the taxon thrives. This range is used to calculate an anomaly to current climate based on the presence of the taxon in the past. The method was further extended to support multiple taxa (Birks *et al.*, 2010). Kühl *et al.* (2002) developed the probability density functions (PDF) method which is based on parametric density functions of the taxa' presence in relation to a climate variable. The density functions of the taxa present in the past are intersected, after weighting by a measure of dispersion of the modern climate values, generating a climate reconstruction value with an associated uncertainty range. This class of methods is usually based on presence-only or presence/absence of taxa, discarding a source of information related to taxa abundance. Additionally, the PDF method is usually based on the normal density function that is not always observed in plant distributions in relation to climate variables, and forcing the removal of those taxa from the reconstruction (Kühl *et al.*, 2002).

1.4.3 Predicting future climate

The evidence for past climate comes from various sources, with different algorithms providing accurate reconstructions of climate oscillations. Thus, is it possible to apply such knowledge to predict future climates? During the last decades, the scientific community has put attention on the anthropogenic influence on climate change, as seen by the increasing alarm of its potential impact in a near future (IPCC, 2007a; Rosenzweig *et al.*, 2008) and by the creation of international entities, like the Intergovernmental Panel on Climate Change (IPCC), to assess those changes. The continuous weather measurements available since the 19^{th} century and the discovery of greenhouse gases produced by human activities, especially after the industrial revolution, has stimulated the research on predictions of future climate change and its impacts. The orbital forces promote long cycles of climate change, but evidence

for abrupt changes related to Earth internal variability are common, and, putting all together, these mechanisms generate feedbacks and amplifications of the change (Zachos *et al.*, 2001; Alley *et al.*, 2003). Examples of these feedbacks include the decrease of moisture related to the increasing dryness and its impact on flora; or the decrease of solar reflectance as consequence of the retreating ice cover promoting increasing temperatures (Pielke *et al.*, 1998; Alley *et al.*, 2003; IPCC, 2007a). Mathematical models merging these factors to forecast weather are generally known as climate models. The dawn of climate modelling was based on very simple models, with increasing complexity over time, fueled by the increasing computational power and the accumulated knowledge about the factors promoting climate change (IPCC, 2007a). By the end of the 1960s, an important work triggered the research on global climate models (GCM) integrating models of atmospheric and ocean fluid dynamics (also known as general circulation models) to generate forecasts of climate (Manabe & Bryan, 1969). New terms representing measured physical processes, tested to

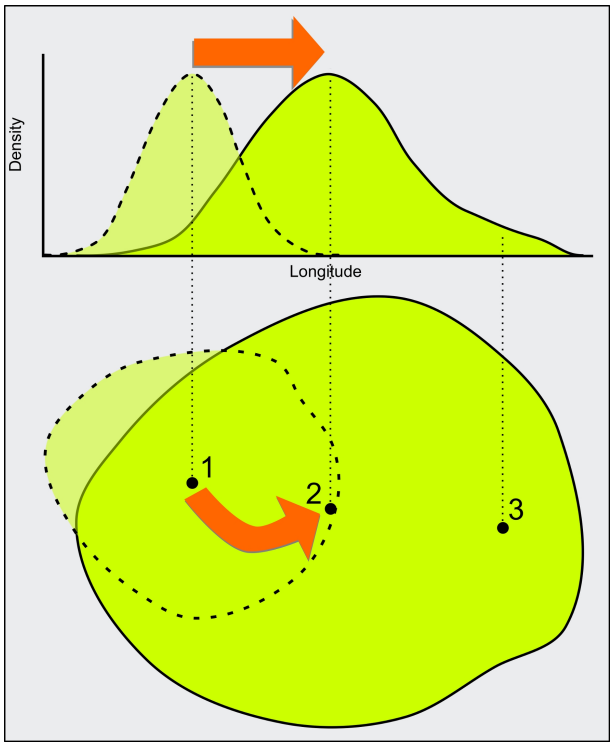


Fig. 1.4 – Relationship between pollen percentages and taxa density. The pollen sample sites (numbered from 1 to 3) report different pollen abundances. The restrained distribution during the harsh conditions of the last glacial maximum (distribution with dashed line) expands to the current wider distribution (solid line). The centre of the distribution is shifted accordingly to the arrows. Site 1 reports a near 100% pollen percentage for the pertinent taxa during the LGM, corresponding to highest densities, but lower at the present. Site 2 will report near 0% at the LGM as consequence of being at the edge of the distribution where densities are very low, but near 100% at the present. The detection of the pollen taxa at site 3 will not be possible at the LGM and near present will report a similar value to site 1.

quantify the impact on climate (e.g. clouds, land heterogeneity, greenhouse gases, ice cover, aerosols, vegetation and others), have been sequentially added to the modelling equations, contributing to the increase of their complexity, more accurate results, and more processing time (IPCC, 2007a).

To face the uncertainties of global economic trends and the effects of human activities on greenhouse gases concentrations, the IPCC has built emission scenarios (SRES; IPCC, 2000) with four categories. Such scenarios rely on driving forces of greenhouse emissions, such as demographic evolution and socio-economic features (Fig. 1.5; Table 1.1). The underling storyline behind the A1 scenario family describes a homogeneous world with reduction of major per capita income differences between regions, and intensive economic growth, and an increasing population until mid-century, followed by a decrease. This scenario is further divided in three groups accordingly to major energy source trends: 1) A1T, intensive use of fossil fuels; 2) A2T, use of

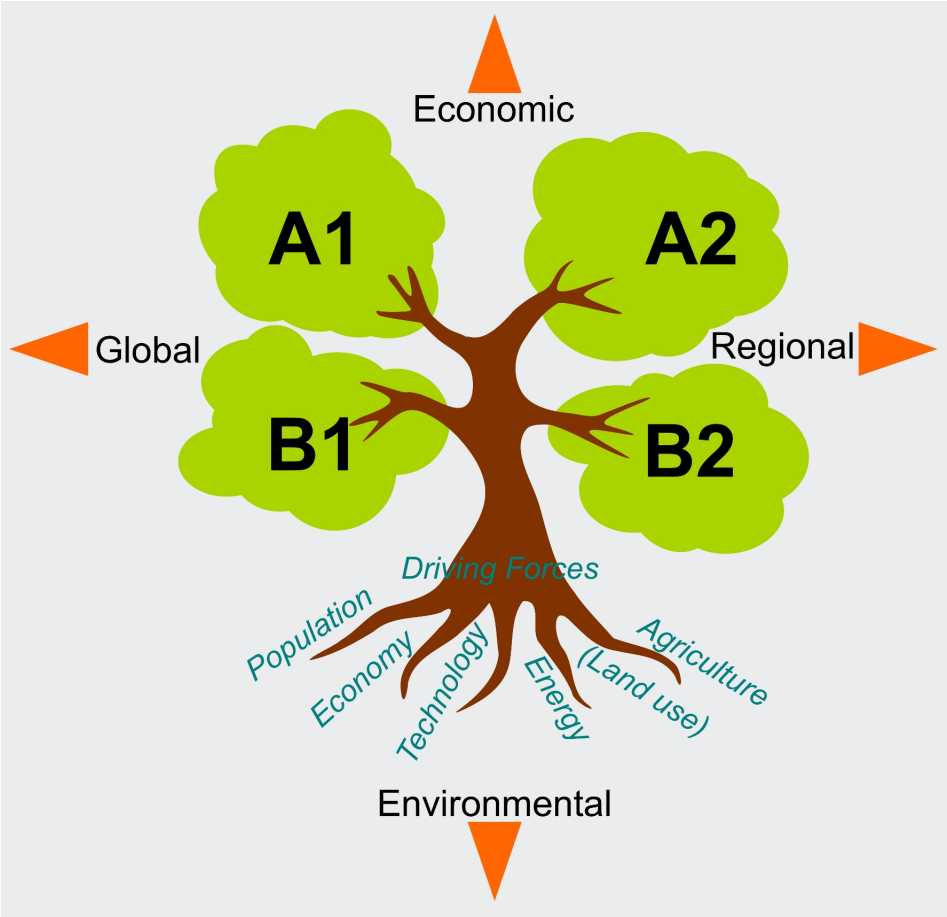


Fig. 1.5 – Families of emissions scenarios accordingly to IPCC (adapted from IPCC, 2000). Vertical axis represents trends from environmental-enemy to environmental-friendly economies. Horizontal axis denotes gradients from globalised to regionalised worlds.

Table 1.1 – Predicted temperatures for 2100 from different emission scenarios based on the IPCC 4^{th} assessment report (IPCC, 2007b). Values given in Celsius degrees (ºC) represent the multi-model global averages of surface warming in relation to average historical climate (1980-1999) and the $\pm$1 standard deviation range of individual model annual averages (between squared brackets).

	Economic based			Environmental based
Increasing globalization	**A1B**	**A1T**	**A1FI**	**B1**
	2.8 ºC	2.2ºC	4.0 ºC	1.9 ºC
	[1.7, 4.4]	[1.4, 3.8]	[2.4, 6.4]	[1.1, 2.9]
Increasing regionalization	**A2**			**B2**
	3.4 ºC			2.2 ºC
	[2.0, 5.4]			[1.4, 3.8.]

non-fossil energy; 3) A1B, a balanced energy source. The B1 scenario family is also described by a storyline with high emphasis on a homogeneous world with population trends as described previously, but including an option for clean energies, less materialistic societies and economies focused on services and information. The A2 and B2 families of scenarios describe an extremely heterogeneous world with continuously growing population, with lower rates in B2. Both scenarios draw regionally controlled economies, but A2 has a greater fragmentation of per capita income and technology change than other storyline. On the other hand, B2 scenario is more environmentally friendly, with local solutions to sustainability, resulting in a more diversified technology than the A2 scenario.

The spatial resolution of GCMs has been increasing, following also the same trend of computational power due to more demanding calculations (IPCC, 2007a). The resolution has increased almost 5 times since the first IPCC's assessment report models during 1990 ($\sim$500 km resolution). Better resolutions allow depicting regional trends, but biodiversity patterns operate at smaller scales, demanding even higher resolutions (e.g. Araújo *et al.*, 2005; Peterson *et al.*, 2007; Waltari *et al.*, 2007; Brito *et al.*, 2009; Carvalho *et al.*, 2010a,b; McKelvey *et al.*, 2011). The increase of the spatial resolution, or downscaling, is a common process for current weather data, with several interpolating algorithms available. The thin plate smoothing splines algorithm with latitude, longitude and altitude as independent variables is known to produce good results with climate data (Hijmans *et al.*, 2005; Kriticos *et al.*, 2011). However, past and future climate estimates have less resolution to depict relationships between climate and independent variables, especially elevation. To overcome this problem, a simple method known as statistical downscaling has been used to produce higher resolution maps from past and future climate layers (Waltari *et al.*, 2007; Tabor & Williams, 2010; McKelvey *et al.*, 2011). Since a linear relationship with elevation is

difficult to achieve, anomalies to the present are calculated at coarse resolutions and interpolated to a finer spatial scale, usually with splines, and reprocessed to full climate values by adding the present data at the same finer scale (Tabor & Williams, 2010). The anomalies have wider relations with space and, thus, are easier to interpolate. Moreover, the relationship with elevation is usually well established with the available data from weather stations. This method allows building spatial layers with enough resolution to establish spatial relationships with biodiversity distribution, although with an increase of uncertainty (Tabor & Williams, 2010).

1.5 Ecological niche modelling as a tool to study niche dynamics

1.5.1 Niche concept and traits

The niche concept is transversal to many subjects discussed above. For instance, building past environmental layers of climate using plant taxa is based on the concept of climate niche and its stability over time. Also, species' distributions are largely maintained by niche preferences and their expansion and contraction are also niche driven. Additionally, niche can play an important role in evolutionary processes like speciation (Holt, 2003; Wiens & Graham, 2005). Many species' related processes are based on the concept of niche and such relation raises broad questions about its definition and the number of niche types a definition may hold.

The term niche was first applied by Grinnell (1917) to define the set of environmental conditions where a species is able to thrive and reproduce. Later, Elton (1927), in his description of animal communities, defined the niche as the interacting factors that act as forces maintaining the distribution. This is a more functional view of the niche, where species relations with competitors and resources play a more prominent role. These two type of niches, currently known as Grinnellian and Eltonial niches, respectively (Wiens & Graham, 2005; Soberón, 2007; Wiens *et al.*, 2009), operate on different spatial scales due to the nature of the variables acting in each concept (Soberón, 2007). The variables used to construct the Grinnellian niche, the scenopoetic variables (Hutchinson, 1978; Soberón, 2007; Wiens *et al.*, 2009), are measured at large scales (climate data, for instance); whereas the bionomic variables are related to the Eltonian niche (Hutchinson, 1978; Soberón, 2007; Wiens *et al.*, 2009), and due to their biological nature (e.g. competition for resources), operate mostly at local scales (Soberón, 2007).

Three decades after the Eltonian niche definition, Hutchinson (1957) summarized the niche concept in the currently most commonly used form: a hypervolume

of n-dimensions where a species is able to persist indefinitely. The dimensions are the niche-related variables meaningful for the species. The Hutchinson niche concept is further divided in the fundamental and realized niches. The first is, in fact, the n-dimensional hypervolume that describes the ecological requirements of the species, whereas the second is a section of that hypervolume where the species exists *de facto* due to other biological pressures such as the presence of competitors and resources availability (Hutchinson, 1957; Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005). These niche concepts are similar but they rarely completely overlap (Fig. 1.6; Soberón & Peterson, 2005). The complete picture of the niche should, however, include a time component. This was later introduced by incorporating the source-sink theory to the niche concept (Pulliam, 2000), and also the range of dispersal ability (Holt, 2003). These concepts add a new interacting sphere to the niche concept representing the area where species may disseminate given their dispersal ability (Fig. 1.6).

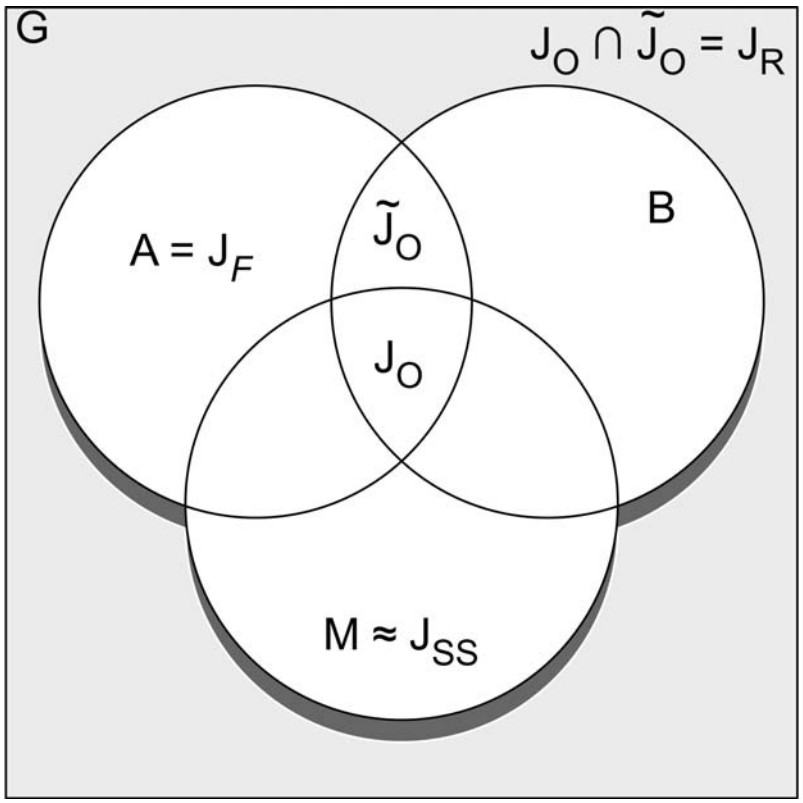


Fig. 1.6 – Biotic-Abiotic-Movement diagram representing driving forces of species' distributions. The study area G represents the scenopoetic and bionomic space containing the species' niche and accessible areas. A and J_F represent the fundamental niche of the species with positive growth rates and composed by Grinnellian variables; B is the space within Eltonian variables where species can compete and coexist with others; and M denotes the area accessible by the species. J_R is the sum of the area where species is found (J_O) plus the area where species has potential to be found ($\tilde{J}_O$). The area where species is found, irrespective of growth rate is J_{SS}, and assumed to be similar to M. Adapted from Soberón (2007), using author's niche notation for simplicity.

The tendency to maintain the fundamental and/or realized niche over time is called niche conservatism (NC; Wiens & Graham, 2005; Pearman *et al.*, 2008; Peterson, 2011). The conservation of the niche is central to the analysis of biodiversity patterns due to its broad impact on evolution and, consequently, on species' distributions. For instance, NC is a major force acting in allopatric speciation (Wiens, 2004, 2008; Wiens & Graham, 2005; Pearman *et al.*, 2008; Glor & Warren, 2011). Vicariance events are related to NC over time, by forcing population splits under conditions that are outside the original niche of the population (Wiens, 2004; Wiens & Graham, 2005; Wiens *et al.*, 2009). In its basic definition, a barrier to species dispersal is a patch of unsuitable or suboptimal habitat dividing a population (Wiens & Graham, 2005). Thus, species dispersal is also controlled by NC positively, by allowing individuals to migrate within their niche, and negatively by creating a resistance to the movement in areas outside the species' niche range.

Time is an influential factor of niche stability (Pearman *et al.*, 2008) and evidences of NC come from very different time-scales (Peterson, 2011). On one hand, short to medium term (10^0 to 10^6 years) phenomena including species invasions and geographical shifts orchestrated during the Quaternary, show a trend to preserve the niche of the species (Peterson, 2011). On the other hand, evolutionary scale events between sister species or distant species occur at higher time-scales and involve phylogenetic hypotheses to be tested (Peterson, 2011). Several studies have demonstrated the existence of labile niches while others have found strong evidence of NC between species (for a review on NC see Pearman *et al.*, 2008). Additionally, cryptic NC was found between evolutionary lineages of lizards with stable fundamental niches hidden within different realized niches (Schulte *et al.*, 2012). Detecting NC may be also hampered by the choice of niche predictors used in each study due to different degrees of stability for different variables (Rödder & Lötters, 2009). Nevertheless, NC is an assumption for ecological niche modelling and should be addressed especially with model transference to other temporal and spatial scales (Wiens & Graham, 2005; Wiens *et al.*, 2009; Petitpierre *et al.*, 2012; Warren, 2012).

1.5.2 Ecological niche-based modelling

Ecological niche-based modelling (ENM) is the exercise of capturing a species' niche based on the locations of presence or presence and absence data and a set of environmental variables. It is based on correlative approaches to predict the distribution of a species and it is often compared to a process-based approach (Guisan & Zimmermann, 2000; Kearney, 2006; Soberón & Peterson, 2005), where the relation between

the species and environmental factors is determined mechanistically to find the range where fitness is maximised. Kearney (2006) argued that only mechanistic models can help defining the niche of the species, whereas ENMs should use the term 'distribution model' preferably to 'niche model'. However, the purpose of ENMs is in fact the assessment of the niche (Warren, 2012). Nevertheless, the presence data of a species is restricted to the realized niche, and there is a lack of consensus in the scientific community about which niche, fundamental or realized, is the result of an ENM (Araújo & Guisan, 2006). A simple answer to this problematic may reside in the complexity of the models being built (Sillero, 2011), that is related to the chosen algorithm (from climate envelope based to machine learning) and to the amount of ENVs used. Simple models, particularly those with few predictors, tend to capture the part of the niche that is conserved (Peterson, 2011) which corresponds to the fundamental niche, because conservatism applies to the Grinnelian space (Soberón, 2007). On the other hand, very complex models tend to over-fit the realized niche. However, there is no estimate available on the correct number of variables to be used in ENM (Peterson, 2011).

Ecological niche modelling is a modern tool and its usage is transverse to research fields of landscape ecology and genetics. It is a major tool in conservation planning (Brito *et al.*, 2009; Torres *et al.*, 2010), especially under scenarios of climate change (Araújo *et al.*, 2006; Carvalho *et al.*, 2010a,b), but also fundamental in studies of niche evolution (Peterson *et al.*, 1999; Rödder & Lötters, 2009; Tingley *et al.*, 2009; McCormack *et al.*, 2010; Petitpierre *et al.*, 2012; Schulte *et al.*, 2012). Finding and quantifying relationships between species presence data and environmental data requires an algorithm capable of discovering patterns. Several algorithms are available to this purpose, ranging from very simplistic envelope models to regression-based analyses, and the more complex machine-learning techniques (Guisan & Zimmermann, 2000; Elith *et al.*, 2006). Algorithms for modelling are grouped based on whether they use only species presence data or they require data on localities of both species presence and absence. Due to the ambiguous nature of absences (Lobo *et al.*, 2010), these are often substituted by artificially generated data (pseudo-absences; Zaniewski *et al.*, 2002; Pearce & Boyce, 2005). Envelope models require presence-only data to build the hypervolume niche based on the locations where the species is present (Carpenter *et al.*, 1993; Elith *et al.*, 2006; Elith & Leathwick, 2009). Niche estimations with presence-only data can be also achieved using a modified principal component analysis, a technique implemented in the Ecological-Niche Factor Analysis (ENFA; Hirzel *et al.*, 2002).

Regression-based methods are extensively used to model presence and ab-

sence data. The aim of regression models is to find relationships between one or more predictors and the presence and absence data. An important advantage over the envelope methods is the quantification of predictor importance. These methods include linear regression, which tries to find linear solutions with maximum likelihood explaining the species data, but may generate, in turn, unreal values (negative probabilities or higher than 1.0; Guisan & Zimmermann, 2000). More commonly used is the extension of linear models to other possible statistical distributions found in the generalized linear models by means of a link function (Guisan & Zimmermann, 2000). The logit function is widely applied link function in the literature to model data that approximate a binomial distribution, as is the case of binary species presence/absence data. Generalized additive models are extensions of these models to fit non-parametric data and better describe non-linear relationships between the environmental predictors and species data (Guisan & Zimmermann, 2000; Guisan *et al.*, 2006; Elith & Leathwick, 2009).

With the increasing availability of larger datasets of both species occurrence and environmental data, the relationships between species and environment become more complex to model and demand efficient and powerful techniques. Machine learning methods are very efficient at pattern-detection, but also, highly demanding in terms of computer processing. Several algorithms are described for the purpose of finding species distribution's patterns, including genetic algorithms, classification and regression trees, maximum entropy, and artificial neural networks. The latter has been proved efficient in detecting very complex patterns that are usually found in ecological relations. It offers a multitude of advantages, however, it application to ecological modelling is still lacking dedicated tools as other commonly used algorithms.

1.5.3 Artificial neural networks

Artificial neural networks (ANN) have several advantages over other approaches. First, it is a simple method with broad literature available. The extensive application in ecological research, perhaps inspired by the biological motivation behind the artificial neuron, generated a diverse and broad literature explaining the ANNs in great detail, resulting in an easy to understand methodology. Secondly, ANNs allow different settings of the learning parameters and reshaping of the network (explained in detail below), creating a vast array of learning options that the user may explore within an ensemble forecasting framework (Araújo & New, 2007; Olden *et al.*, 2008). Thirdly, the train and test error assessment allow reducing model overfitting and thus control the generalization ability of the trained network (Dimopoulos *et al.*, 1999; Özesmi *et al.*,

2006b). Fourthly, ANNs accept continuous or discrete data, both for dependent and independent variables, thus, models may be made using binary or continuous presence data (e.g. abundance) and benefiting from the panoply of possible ecological niche variables available. Other features usually attributed to ANNs include the ability to process complex, non-linear relations and non-parametric data, found regularly in ecological datasets, with a high prediction success even in the presence of noise (Lek *et al.*, 1996; Dimopoulos *et al.*, 1999; Lek & Guégan, 1999; Spitz & Lek, 1999; Pearson *et al.*, 2002; Özesmi *et al.*, 2006a,b; Olden *et al.*, 2008).

The development of ANNs was motivated by the efficient pattern interpretation capabilities of the brain and the introduction of the mathematical model of a simplified neuron by McCulloch & Pitts (1943). During the last decade, a relevant number of scientific papers reviewed the application of ANNs in ecology and the analysis of ANNs' internal behaviour (Olden & Jackson, 2002; Gevrey *et al.*, 2003, 2006a,b; Özesmi *et al.*, 2006a; Park & Chon, 2007), revealing a growing interest on ANNs, despite the restricted use by a small circle of ecologists with a background in informatics (Olden *et al.*, 2008). The interest of the biological scientific community in ANNs, with pioneering applications using species' data during the 1990s (Lek *et al.*, 1996; Tan & Smeins, 1996; Mastrorillo *et al.*, 1997), triggered new developments of the method in an ecological context, especially with sensitivity analyses of the model output (Lek & Guégan, 1999; Olden & Jackson, 2002; Gevrey *et al.*, 2006b; Özesmi *et al.*, 2006a). The ecological application of ANNs is still widening, encompassing multiple research fields such as predictive modelling (Özesmi & Özesmi, 1999; Özesmi *et al.*, 2006b), plant ecology (Hilbert & Muyzenberg, 1999), impact assessment (Spitz & Lek, 1999), and potential effects of climate warming (Pearson *et al.*, 2002; Araújo *et al.*, 2006; Xavier *et al.*, 2010), supporting the extreme potential of this method. With the increase of computational power, the demand for spatially explicit predictions has also increased.

The ANNs spectrum is constituted of various types of networks, varying in structure and learning methods. One common type of ANN is the feed-forward neural network with back-propagation learning (BPN; Lek & Guégan, 1999). This network has a layered structure of neurons connecting the inputs to an output through one or several hidden layers (Fig. 1.7). The BPN has been used to model ecological systems due to its efficient learning ability with complex relationships between dependent and independent variables and also due to its simple nature, which makes it easy to understand (Lek & Guégan, 1999; Özesmi *et al.*, 2006a). The artificial neuron constitutes

the unit of the BPN. The output of a neuron with n inputs is defined as:

$$(1)\ \ Output = f(net) = f(\sum_{i=1}^{n} w_i x_i)$$

where $f(net)$ represents the activation function and w_i represents the weight connecting to input x_i. The activation function, usually a linear or a sigmoid function (2), squashes the sum of the products of all connecting weights and the respective neuron output in the previous layer to a value that is passed to the next layer of neurons by the connecting weights.

$$(2)\ \ f(x) = \frac{1}{1+e^{-x}}$$

The learning of the network occurs during the training stage where vectors of variables for each target (presence or absence in the ecological data) are presented to the network, in a process called supervised learning. The values of the connecting weights are initialized to random values and the input vectors of targets and respective variables vectors are propagated through the network producing an output. The error of the output in relation to the target value of the input (i.e. the value that corresponds to the set of inputs of the training data) is assessed and back-propagated in the network. Usually, the error processed by the network is half the sum of squared errors given by the formula:

$$(3)\ \ E = \frac{1}{2}\sum_{d \in D}(t_d - o_d)^2$$

where D is the training set, t_d the target value and o_d the output of the network for the d sample in the training data. The learning in BPN occurs by gradient descent when the weight's change is updated accordingly to a set of learning rules. In a BPN, the learning rules are the back-propagation algorithm:

$$(4)\ \ \triangle w_{ij} = -\eta \frac{\sigma E}{\sigma w_{ij}} + \alpha \triangle w_{ij}^{t-1}$$

where w_{ij} is the weight connecting a neuron in layer i to neuron in layer j, η represents the learning rate and α the momentum. The last two parameters are commonly used to control the learning ability. The learning rate affects the convergence of the network by defining the allowed quantity of weight's change (Özesmi & Özesmi, 1999; Olden & Jackson, 2002). While small learning rates will slowly converge to a global minimum, high values will oscillate in the error surface, and may fail in finding the global minimum. This parameter highly interacts with momentum which defines how much of the last weight change is incorporated in the new weight change, thus giving a direction of progress on the error surface (Lek *et al.*, 1996; Spitz & Lek, 1999; Olden & Jackson, 2002). The process of feeding the network with training data is repeated to minimise

the error until the network is fully trained (Lek & Guégan, 1999). Once a network is trained with the combination of input data, independent variables and targets, it can be used to predict to other locations where the same set of variables is available.

1.5.4 Ensemble modelling and model evaluation

The use of different models allows a better assessment of uncertainty due to the variability in the outputs (Pearson *et al.*, 2002; Araújo & New, 2007). Stacking several models allows defining a consensus prediction by averaging and to compute a standard deviation as a measure of prediction uncertainty. The ensembling can be done at several levels: across replicates of the same algorithm with resampling methods or across algorithms. This method reduces the over fitting trend of very complex models and increases accuracy (Breiman, 1996).

Assessment of model fitting is done by calculating the rates of correct classification. Most common metrics of algorithm performance (e.g. the receiver operating characteristic curve and its area) are based on the confusion matrix where real values are

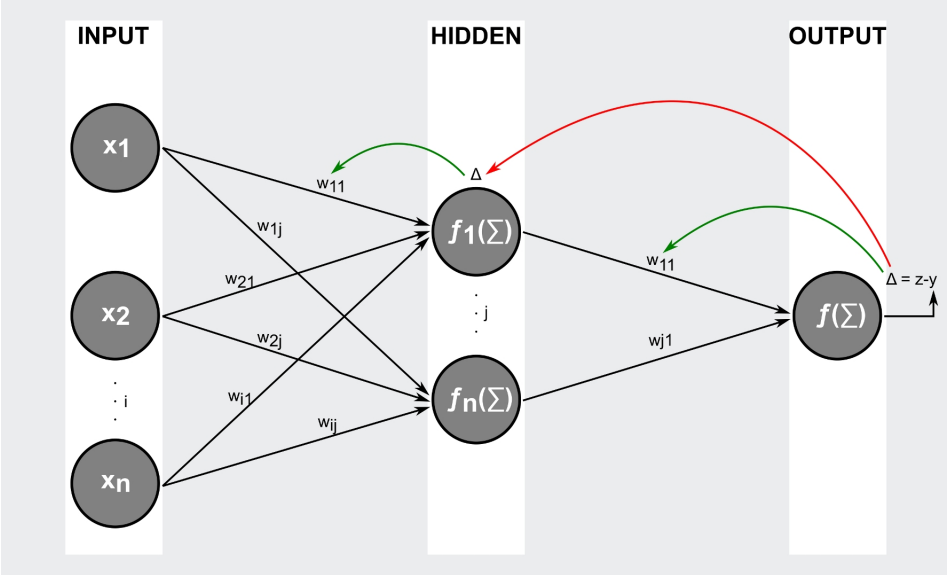


Fig. 1.7 – Typical structure of a simple artificial neural network with backpropagation algorithm and one hidden layer. The artificial neural network is a structure of layers with neurons (dark grey circles) connected by weights (black arrows). The most basic structure has three layers: the input layer where values from different variables are fed to the network; the hidden layer where neurons sum the information from the previous layer and apply an activation function, sending the result to the neurons on the next layer; and the output layer where the results of the network are returned to the user. The first stage of training the network is to propagate the input data to the output neuron (black arrows) and calculate the error at the output. The second stage is to backpropagate the error to each neuron in the previous hidden layers (red arrow). The third stage consists in adapt the weights to new values based on the calculated error with a learning algorithm (green arrow). After this process, an artificial neural network is considered trained and is able to predict to different combination of values of the same set of variables used in input.

confronted with values predicted by the model, producing counts of true positives and negatives, when estimated values are in agreement with real values; and false positives and negatives, in the case of prediction failure. Such estimates pose problems for presence-only methods, especially when based on pseudo-absences/background data (Lobo *et al.*, 2007). An output that is commonly of interest for researchers is the binary model, depicting the presence and absence area, that is obtained by applying a threshold to the probabilistic model output. In this case, a count of true positives and false negatives can be reported by intersecting the binary model with the species data, providing an indication of algorithm success (Lobo *et al.*, 2007).

ENMs are an essential tool for modern conservation and reserve design under scenarios of climate change and also for landscape studies of evolution. Due to the wide availability of high amounts of data and methods, and like in other research areas, models may deceive the researcher to an illusion of accuracy. These *good* models are false positives resulting from looking to a small fraction of the reality being modelled or from of a lack of post analyses (Macarthur, 2012). Accurate models are time-consuming due to the computing requirements of modern techniques, the available data at increasing resolutions and the complexity of the hypotheses being tested. But, above all, because of all the assumptions that obligate to careful steps, including those from niche modelling and those particular to each algorithm.

Niche definition is central to all major aspects discussed here. Despite the increasing availability of data, most are stored under formats difficult to manage, hidden in very complex websites or representing series with such quantity of information that summarizing is still computing and time demanding. In the same line, algorithms for ENM are available in advanced software that often requires programming abilities. Given their importance in many research lines and to generate conservation policies, such tools and data should be readily available and easy to use. User friendly software incorporating state-of-art modelling algorithms, such as ANNs, is especially needed. On the other hand, increasing computing power and data availability offer the possibility of multi-species and multi-disciplinary studies that aim to strengthen the knowledge of impacts of climate change on biodiversity, to define areas where mitigation measures are urgent and to study evolutionary process under a spatial and ecological scenario.

1.6 References

Alba-Sánchez, F., López-Sáez, J.a., Pando, B.B.d., Linares, J.C., Nieto-Lugilde, D. & López-Merino, L. (2010) Past and present potential distribution of the Iberian *Abies* species: a phytogeographic approach using fossil pollen data and species distribution models. *Diversity and Distributions* **16**, 214–228.

Allen, J.R.M., Huntley, B. & Watts, W.A. (1996) The vegetation and climate of northwest Iberia over the last 14 000 yr. *Journal of Quaternary Science* **11**, 125–147.

Alley, R.B., Marotzke, J., Nordhaus, W.D., Overpeck, J.T., Peteet, D.M., Pielke, R.a., Pierrehumbert, R.T., Rhines, P.B., Stocker, T.F., Talley, L.D. & Wallace, J.M. (2003) Abrupt climate change. *Science* **299**, 2005–2010.

Anderson, N.J., Bugmann, H., Dearing, J.A. & Gaillard, M.J. (2006) Linking palaeoenvironmental data and models to understand the past and to predict the future. *Trends in Ecology & Evolution* **21**, 696–704.

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**, 1712–1728.

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

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

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* **22**, 42–47.

Araújo, M.B., Nogués-Bravo, D., Diniz-Filho, J.A.F., Haywood, A.M., Valdes, P.J. & Rahbek, C. (2008) Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31**, 8–15.

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

Arroyo, J., Aparicio, A., Albaladejo, R.G., Muñoz, J. & Braza, R. (2007) Genetic structure and population differentiation of the Mediterranean pioneer spiny broom *Calicotome villosa* across the Strait of Gibraltar. *Biological Journal of the Linnean Society* **93**, 39–51.

Ashcroft, M.B. (2010) Identifying refugia from climate change. *Journal of Biogeography* **37**, 1407–1413.

Barnosky, A.D., Koch, P.L., Feranec, R.S., Wing, S.L. & Shabel, A.B. (2010) Assessing the Causes of Late Pleistocene Extinctions on the Continents. *Science* **306**, 70–75.

Benito Garzón, M., Sánchez de Dios, R. & Sáinz Ollero, H. (2007) Predictive modelling of tree species distributions on the Iberian Peninsula during the Last Glacial Maximum and Mid-Holocene. *Ecography* **30**, 120–134.

Berglund, B.E. (2003) Human impact and climate changes—synchronous events and a causal link? *Quaternary International* **105**, 7–12.

Birks, H.H. & Ammann, B. (2000) Two terrestrial records of rapid climatic change during the glacial-Holocene transition (14,000- 9,000 calendar years B.P.) from Europe. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 1390–1394.

Birks, H.J.B. (1998) Numerical tools in palaeolimnology - Progress, potentialities, and problems. *Journal of Paleolimnology* **20**, 307–332.

Birks, H.H. & Birks, H.J.B. (2006) Multi-proxy studies in palaeolimnology. *Vegetation History and Archaeobotany* **15**, 235–251.

Birks, H., Heiri, O., Seppä, H. & Bjune, A. (2010) Strengths and weaknesses of quantitative climate reconstructions based on Late-Quaternary biological proxies. *The Open Ecology Journal* **3**, 68–110.

Bond, G., Broecker, W., Johnsen, S., McManus, J., Labeyrie, L., Jouzel, J. & Bonani, G. (1993) Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147.

Bond, G., Showers, W., Cheseby, M., Lotti, R., Almasi, P., DeMenocal, P., Priore, P., Cullen, H., Hajdas, I. & Bonani, G. (1997) A pervasive millennial-scale cycle in North Atlantic Holocene and glacial climates. *Science* **278**, 1257–1266.

Branco, M., Monnerot, M., Ferrand, N. & Templeton, A.R. (2002) Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* **56**, 792–803.

Breiman, L. (1996) Bagging Predictors. *Machine Learning* **140**, 123–140.

Bridle, J.R., Baird, S.J. & Butlin, R.K. (2001) Spatial structure and habitat variation in a grasshopper hybrid zone. *Evolution* **55**, 1832–1843.

Brito, J.C., Acosta, A.L., Álvares, F. & Cuzin, F. (2009) Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African canids. *Biological Conservation* **142**, 3020–3029.

Brook, B.W., Sodhi, N.S. & Bradshaw, C.J.A. (2008) Synergies among extinction drivers under global change. *Trends in Ecology & Evolution* **23**, 453–460.

Burney, D.A. & Flannery, T.F. (2005) Fifty millennia of catastrophic extinctions after human contact. *Trends in ecology & evolution* **20**, 395–401.

Burrows, M.T., Schoeman, D.S., Buckley, L.B., Moore, P., Poloczanska, E.S., Brander, K.M., Brown, C., Bruno, J.F., Duarte, C.M., Halpern, B.S., Holding, J., Kappel, C.V., Kiessling, W., O'Connor, M.I., Pandolfi, J.M., Parmesan, C., Schwing, F.B., Sydeman, W.J. & Richardson, A.J. (2011) The pace of shifting climate in marine and terrestrial ecosystems. *Science* **334**, 652–655.

Campillo, S., Serra, M., Carmona, M.J. & Gómez, A. (2011) Widespread secondary contact and new glacial refugia in the halophilic rotifer *Brachionus plicatilis* in the Iberian Peninsula. *PloS one* **6**, e20986.

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

Carranza, S., Arnold, E.N. & Pleguezuelos, J.M. (2006) Phylogeny, biogeography, and evolution of two Mediterranean snakes, *Malpolon monspessulanus* and *Hemorrhois hippocrepis* (Squamata, Colubridae), using mtDNA sequences. *Molecular Phylogenetics and Evolution* **40**, 532–546.

Carrión, J.S. (2002) Patterns and processes of Late Quaternary environmental change in a montane region of southwestern Europe. *Quaternary Science Reviews* **21**, 2047–2066.

Carrión, J.S., Fernández, S., González-Sampériz, P., Gil-Romera, G., Badal, E., Carrión-Marco, Y., López-Merino, L., López-Sáez, J.a., Fierro, E. & Burjachs, F. (2010a) Expected trends and surprises in the Lateglacial and Holocene vegetation history of the Iberian Peninsula and Balearic Islands. *Review of Palaeobotany and Palynology* **162**, 458–475.

Carrión, J., Fernández, S., Jiménez-Moreno, G., Fauquette, S., Gil-Romera, G., González-Sampériz, P. & Finlayson, C. (2010b) The historical origins of aridity and vegetation degradation in southeastern Spain. *Journal of Arid Environments* **74**, 731–736.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2010a) From climate change predictions to actions - conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology* **16**, 3257–3270.

Carvalho, S.B., Brito, J.C., Pressey, R.L., Crespo, E. & Possingham, H.P. (2010b) Simulating the effects of using different types of species distribution data in reserve selection. *Biological Conservation* **143**, 426–438.

Cheddadi, R., Yu, G., Guiot, J., Harrison, S. & Prentice, I.C. (1997) The climate of Europe 6000 years ago. *Climate Dynamics* **13**, 1–9.

Cheddadi, R. & Bar-Hen, A. (2009) Spatial gradient of temperature and potential vegetation feedback across Europe during the late Quaternary. *Climate Dynamics* **32**, 371–379.

Cheddadi, R., de Beaulieu, J.L., Jouzel, J., Andrieu-Ponel, V., Laurent, J.M., Reille, M., Raynaud, D. & Bar-Hen, A. (2005) Similarity of vegetation dynamics during interglacial periods. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13939–13943.

Cheddadi, R., Vendramin, G.G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J.M., de Beaulieu, J.L., Sadori, L., Jost, A. & Lunt, D. (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography* **15**, 271–282.

Clark, J.S., Fastie, C., Hurtt, G., Jackson, S.T., Johnson, C., King, G.a., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E.W., Webb, T. & Wyckoff, P. (1998) Reid's Paradox of Rapid Plant Migration. *BioScience* **48**, 13.

Clark, P.U., Shakun, J.D., Baker, P.a., Bartlein, P.J., Brewer, S., Brook, E., Carlson, A.E., Cheng, H., Kaufman, D.S., Liu, Z., Marchitto, T.M., Mix, A.C., Morrill, C., Otto-Bliesner, B.L., Pahnke, K., Russell, J.M., Whitlock, C., Adkins, J.F., Blois, J.L., Clark, J., Colman, S.M., Curry, W.B., Flower, B.P., He, F., Johnson, T.C., Lynch-Stieglitz, J., Markgraf, V., McManus, J., Mitrovica, J.X., Moreno, P.I. & Williams, J.W. (2012) Global climate evolution during the last deglaciation. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E1134–E1142.

Dansgaard, W., Johnsen, S., Clausen, H., Dahl-Jensen, D., Gundestrup, N., Hammer, C., Hvidberg, C., Steffensen, J., Sveinbjörnsdottir, A., Jouzel, J. & Bond, G. (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220.

Davis, B.A.S., Brewer, S., Stevenson, A.C., Guiot, J. & Contributors, D. (2003) The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews* **22**, 1701–1716.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science* **292**, 673–679.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011) Beyond predictions: biodiversity conservation in a changing climate. *Science* **332**, 53–8.

Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A. & Lek, S. (1999) Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling* **120**, 157–165.

Dorado-Valiño, M., Rodríguez, A.V., Zapata, M.B.R., García, M.J.G. & Gutiérrez, I.D.B. (2002) Climatic changes since the Late-glacial/Holocene transition in La Mancha Plain (South-central Iberian Peninsula, Spain) and their incidence on Las Tablas de Daimiel marshlands. *Quaternary International* **93-94**, 73–84.

Dormoy, I., Peyron, O., Combourieu Nebout, N., Goring, S., Kotthoff, U., Magny, M. & Pross, J. (2009) Terrestrial climate variability and seasonality changes in the Mediterranean region between 15 000 and 4 000 years BP deduced from marine pollen records. *Climate of the Past* **5**, 615–632.

Elenga, H., Peyron, O., Bonnefille, R., Jolly, D., Cheddadi, R., Guiot, J., Andrieu, V., Bottema, S., Buchet, G., De Beaulieu, J., Hamilton, A.C., Maley, J., Marchant, R., Perez-Obiol, R., Reille, M., Riollet, G., Scott, L., Straka, H., Taylor, D., Campo, E.V., Vincens, A., Laarif, F. & Jonson, H. (2000) Pollen-based biome reconstruction for southern Europe and Africa 18,000 yr bp. *Journal of Biogeography* **27**, 621–634.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., 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.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697.

Elton, C.S. (1927) *Animal ecology*. The Macmillan Company, London.

Fletcher, W.J., Boski, T. & Moura, D. (2007) Palynological evidence for environmental and climatic change in the lower Guadiana valley, Portugal, during the last 13 000 years. *The Holocene* **17**, 481–494.

Fletcher, W.J., Sanchez-Goñi, M.F., Peyron, O. & Dormoy, I. (2010) Abrupt climate changes of the last deglaciation detected in a Western Mediterranean forest record. *Climate of the Past* **6**, 245–264.

Fletcher, W.J. & Sánchez-Goñi, M.F. (2008) Orbital- and sub-orbital-scale climate impacts on vegetation of the western Mediterranean basin over the last 48,000 yr. *Quaternary Research* **70**, 451–464.

Fløjgaard, C., Normand, S., Skov, F. & Svenning, J.C. (2009) Ice age distributions of European small mammals: insights from species distribution modelling. *Journal of Biogeography* **36**, 1152–1163.

Garrigues, T., Dauga, C., Ferquel, E., Choumet, V. & Failloux, A.B. (2005) Molecular phylogeny of *Vipera* Laurenti, 1768 and the related genera *Macrovipera* (Reuss, 1927) and *Daboia* (Gray, 1842), with comments about neurotoxic *Vipera aspis aspis* populations. *Molecular Phylogenetics and Evolution* **35**, 35–47.

Geraldes, A., Carneiro, M., Delibes-Mateos, M., Villafuerte, R., Nachman, M.W. & Ferrand, N. (2008) Reduced introgression of the Y chromosome between subspecies of the European rabbit (*Oryctolagus cuniculus*) in the Iberian Peninsula. *Molecular Ecology* **17**, 4489–4499.

Gevrey, M., Dimopoulos, I. & Lek, S. (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* **160**, 249–264.

Gevrey, M., Dimopoulos, I. & Lek, S. (2006a) Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecological Modelling* **195**, 43–50.

Gevrey, M., Lek, S. & Oberdorff, T. (2006b) Utility of sensitivity analysis by artificial neural network models to study patterns of endemic fish species. *Ecological Informatics: Scope, techniques and applications* (ed. F. Recknagel), chap. 14, pp. 293–306, Springer, Berlin, 2^{nd} edn.

Glor, R.E. & Warren, D. (2011) Testing ecological explanations for biogeographic boundaries. *Evolution* **65**, 673–683.

Godinho, R., Mendonça, B., Crespo, E.G. & Ferrand, N. (2006) Genealogy of the nuclear beta-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity* **96**, 454–463.

Gómez, A. & Lunt, D. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* (eds. S. Weiss & N. Ferrand), pp. 155–188, Springer Netherlands.

Grinnell, J. (1917) The niche-relationships of the California Thrasher. *The Auk* **34**, 427–433.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* **43**, 386–392.

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

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

Hays, J., Imbrie, J. & Shackleton, N. (1976) Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132.

Heiri, O., Tinner, W. & Lotter, A.F. (2004) Evidence for cooler European summers during periods of changing meltwater flux to the North Atlantic. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15285–15288.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–13.

Hewitt, G.M. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87–112.

Hewitt, G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* **359**, 183–195.

Hewitt, G.M. (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617–638.

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

Hilbert, D.W. & Muyzenberg, J.V.D. (1999) Using an artificial neural network to characterize the relative suitability of environments for forest types in a complex tropical vegetation mosaic. *Diversity and Distributions* **5**, 263–274.

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

Hof, C., Levinsky, I., Araújo, M.B. & Rahbek, C. (2011) Rethinking species' ability to cope with rapid climate change. *Global Change Biology* **17**, 2987–2990.

Holt, R. (2003) On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research* **5**, 159–178.

Hutchinson, G. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 42, pp. 415–427.

Hutchinson, G. (1978) *An Introduction to Population Ecology*. Yale University Press, New Haven, 1^{st} edn.

IPCC (2000) *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

IPCC (2007a) *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

IPCC (2007b) *Climate Change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Jaarola, M. & Searle, J.B. (2004) A highly divergent mitochondrial DNA lineage of *Microtus agrestis* in southern Europe. *Heredity* **92**, 228–234.

Jalut, G., Dedoubat, J.J., Fontugne, M. & Otto, T. (2009) Holocene circum-Mediterranean vegetation changes: Climate forcing and human impact. *Quaternary International* **200**, 4–18.

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

Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., Smith, M.S., Mackey, B. & Possingham, H. (2009) Incorporating ecological and evolutionary processes into continental-scale conservation planning. *Ecological applications : a publication of the Ecological Society of America* **19**, 206–217.

Koch, P.L. & Barnosky, A.D. (2006) Late Quaternary Extinctions : State of the Debate. *Annual Review of Ecology, Evolution, and Systematics* **37**, 215–252.

Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. & Scott, J.K. (2011) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* **3**, 53–64.

Kühl, N., Gebhardt, C., Litt, T. & Hense, A. (2002) Probability Density Functions as Botanical-Climatological Transfer Functions for Climate Reconstruction. *Quaternary Research* **58**, 381–392.

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.

Lek, S. & Guégan, J.F. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* **120**, 65–73.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature* **462**, 1052–1055.

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

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

López de Heredia, U., Carrión, J.S., Jiménez, P., Collada, C. & Gil, L. (2007) Molecular and palaeoecological evidence for multiple glacial refugia for evergreen oaks on the Iberian Peninsula. *Journal of Biogeography* **34**, 1505–1517.

Lotter, A., Birks, H., Eicher, U., Hofmann, W., Schwander, J. & Wick, L. (2000) Younger Dryas and Allerød summer temperatures at Gerzensee (Switzerland) inferred from fossil pollen and cladoceran assemblages. *Palaeogeography, Palaeoclimatology, Palaeoecology* **159**, 349–361.

Macarthur, D. (2012) Face up to false positives. *Nature* **487**, 427–428.

Magri, D. (2010) Persistence of tree taxa in Europe and Quaternary climate changes. *Quaternary International* **219**, 145–151.

Magri, D., Vendramin, G.G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., Latałowa, M., Litt, T., Paule, L., Roure, J.M., Tantau, I., van der Knaap, W.O., Petit, R.J. & de Beaulieu, J.L. (2006) A new scenario for the quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *The New Phytologist* **171**, 199–221.

Manabe, S. & Bryan, K. (1969) Climate calculations with a combined ocean-atmosphere model. *Journal of the Atmospheric Sciences* **26**, 786–789.

Martínez-Freiría, F., Santos, X., Pleguezuelos, J.M., Lizana, M. & Brito, J.C. (2009) Geographical patterns of morphological variation and environmental correlates in contact zones: a multi-scale approach using two Mediterranean vipers (Serpentes). *Journal of Zoological Systematics and Evolutionary Research* **47**, 357–367.

Martínez-Freiría, F., Sillero, N., Lizana, M. & Brito, J.C. (2008) GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. *Diversity and Distributions* **14**, 452–461.

Martínez-Freiría, F., Brito, J.C. & Avia, M.L. (2006) Intermediate forms and syntopy among vipers (*Vipera aspis* and *V. latastei*) in Northern Iberian Peninsula. *Herpetological Bulletin* **97**, 14–18.

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

McCormack, J.E., Zellmer, A.J. & Knowles, L.L. (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation? Insights from tests with niche models. *Evolution* **64**, 1231–1244.

McCulloch, W. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology* **5**, 115–133.

McKelvey, K.S., Copeland, J.P., Schwartz, M.K., Littell, J.S., Aubry, K.B., Squires, J.R., Parks, S.A., Elsner, M.M. & Mauger, G.S. (2011) Climate change predicted to shift wolverine distributions, connectivity, and dispersal corridors. *Ecological Applications* **21**, 2882–2897.

Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36**, 1333–1345.

Melo-Ferreira, J., Boursot, P., Randi, E., A. Kryukov, Suchentrunk, F., Ferrand, N. & Alves, P.C. (2007) The rise and fall of the mountain hare. *Molecular Ecology* **16**, 605– 618.

Miraldo, A., Hewitt, G.M., Paulo, O.S. & Emerson, B.C. (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology* **11**, 170.

Moritz, C., Patton, J.L., Conroy, C.J., Parra, J.L., White, G.C. & Beissinger, S.R. (2008) Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science* **322**, 261–264.

Muñoz, J., Gómez, A., Green, A.J., Figuerola, J., Amat, F. & Rico, C. (2008) Phylogeography and local endemism of the native Mediterranean brine shrimp *Artemia salina* (Branchiopoda: Anostraca). *Molecular ecology* **17**, 3160–77.

Muñoz-Sobrino, C., Ramil-Rego, P. & Gómez-Orellana, L. (2006) Late Würm and early Holocene in the mountains of northwest Iberia: biostratigraphy, chronology and tree colonization. *Vegetation History and Archaeobotany* **16**, 223–240.

Myers, N., Mittermeier, R., Mittermeier, C., Da Fonseca, G. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

Naughton, F., Sanchez-Goñi, M., Desprat, S., Turon, J.L., Duprat, J., Malaizé, B., Joli, C., Cortijo, E., Drago, T. & Freitas, M. (2007) Present-day and past (last 25000 years) marine pollen signal off western Iberia. *Marine Micropaleontology* **62**, 91–114.

Nogués-Bravo, D., Rodríguez, J., Hortal, J., Hortal, J., Batra, P. & Araújo, M.B. (2008) Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biology* **6**.

Nogués-Bravo, D., Ohlemüller, R., Batra, P. & Araújo, M.B. (2010) Climate predictors of late quaternary extinctions. *Evolution* **64**, 2442–2449.

Ohlemüller, R., Huntley, B., Normand, S. & Svenning, J.C. (2012) Potential source and sink locations for climate-driven species range shifts in Europe since the Last Glacial Maximum. *Global Ecology and Biogeography* **21**, 152–163.

Olalde, M., Herrán, A., Espinel, S. & Goicoechea, P.G. (2002) White oaks phylogeography in the Iberian Peninsula. *Forest Ecology and Management* **156**, 89–102.

Olden, J.D. & Jackson, D.A. (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling* **154**, 135–150.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology* **83**, 171–193.

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

Özesmi, S., Tan, C. & Özesmi, U. (2006a) Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling* **195**, 83–93.

Özesmi, U., Tan, C., Özesmi, S. & Robertson, R. (2006b) Generalizability of artificial neural network models in ecological applications: Predicting nest occurrence and breeding success of the red-winged blackbird *Agelaius phoeniceus*. *Ecological Modelling* **195**, 94–104.

Park, Y. & Chon, T. (2007) Biologically-inspired machine learning implemented to ecological informatics. *Ecological Modelling* **203**, 1–7.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42.

Parmesan, C. (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* **37**, 637–669.

Pearce, J. & Boyce, M. (2005) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* **43**, 405–412.

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

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

Peltier, W. (1994) Ice age paleotopography. *Science* **265**, 195–201.

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

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

Peterson, A.T., Papeş, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**, 550–560.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565.

Petit, R.J., Hu, F.S. & Dick, C.W. (2008) Forests of the past: a window to future changes. *Science* **320**, 1450–2.

Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C. & Guisan, a. (2012) Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science* **335**, 1344–1348.

Pielke, R.A., Avissar, R., Raupach, M., Dolman, A.J., Zeng, X. & Denning, A.S. (1998) Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. *Global Change Biology* **4**, 461–475.

Prentice, C., Guiot, J., Huntley, B., Jolly, D. & Cheddadi, R. (1996) Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics* **12**, 185–194.

Prentice, I., Jolly, D. & BIOME 6000 Participants (2000) Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography* **27**, 507–519.

Provan, J. & Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* **23**, 564–71.

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

Rebelo, H., Froufe, E., Brito, J.C., Russo, D., Cistrone, L., Ferrand, N. & Jones, G. (2012) Postglacial colonization of Europe by the barbastelle bat: agreement between molecular data and past predictive modelling. *Molecular ecology* **21**, 2761–74.

Reimer, P.J., Baillie, M.G.L., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Burr, G.S., Edwards, R.L., Friedrich, M., Grootes, P.M., Guilderson, T.P., Hajdas, I., Heaton, T.J., Hogg, A.G., Hughen, K.A., Kaiser, K.F., Kromer, B., McCormac, F.G., Manning, S.W., Reimer, R.W., Richards, D.A., Southon, J.R., Talamo, Turney, C.S.M., van der Plicht, J. & Weyhenmeyer, C.E. (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150.

Renssen, H. & Isarin, R.F.B. (2001) The two major warming phases of the last deglaciation at 14.7 and 11.5 ka cal BP in Europe: climate reconstructions and AGCM experiments. *Global and Planetary Change* **30**, 117–153.

Renssen, H., Seppä, H., Heiri, O., Roche, D.M., Goosse, H. & Fichefet, T. (2009) The spatial and temporal complexity of the Holocene thermal maximum. *Nature Geoscience* **2**, 411–414.

Rödder, D. & Lötters, S. (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography* **18**, 674–687.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C. & Pounds, J.A. (2003) Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60.

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G., Menzel, A., Root, T.L., Estrella, N., Seguin, B., Tryjanowski, P., Liu, C., Rawlins, S. & Imeson, A.

(2008) Attributing physical and biological impacts to anthropogenic climate change. *Nature* **453**, 353–357.

Roucoux, K., Abreu, L.D., Shackleton, N.J. & Tzedakis, P.C. (2005) The response of NW Iberian vegetation to North Atlantic climate oscillations during the last 65kyr. *Quaternary Science Reviews* **24**, 1637–1653.

Sánchez-Goñi, M.F., Turon, J.L., Eynaud, F. & Gendreau, S. (2000) European Climatic Response to Millennial-Scale Changes in the Atmosphere–Ocean System during the Last Glacial Period. *Quaternary Research* **54**, 394–403.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J. & Svenning, J.C. (2011) The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664.

Schloss, C.A., Nunez, T.A. & Lawler, J.J. (2012) Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8606–8611.

Schulte, U., Hochkirch, A., Lötters, S., Rödder, D., Schweiger, S., Weimann, T. & Veith, M. (2012) Cryptic niche conservatism among evolutionary lineages of an invasive lizard. *Global Ecology and Biogeography* **21**, 198–211.

Seppä, H. & Birks, H. (2001) July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: pollen-based climate reconstructions. *The Holocene* **11**, 527–539.

Sequeira, F., Alexandrino, J., Rocha, S., Arntzen, J..W. & Ferrand, N. (2005) Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Molecular ecology* **14**, 245–254.

Sillero, N. (2011) What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling* **222**, 1343–1346.

Skov, F. & Svenning, J. (2004) Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography* **27**, 366–380.

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

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology letters* **10**, 1115–1123.

Spitz, F. & Lek, S. (1999) Environmental impact prediction using neural network modelling. An example in wildlife damage. *Journal of Applied Ecology* **36**, 317–326.

Svenning, J. & Skov, F. (2007) Ice age legacies in the geographical distribution of tree species richness in Europe. *Global Ecology and Biogeography* **16**, 234–245.

Svenning, J.C., Normand, S. & Kageyama, M. (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology* **96**, 1117–1127.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453–464.

Taberlet, P. & Cheddadi, R. (2002) Quaternary refugia and persistence of biodiversity. *Science* **297**, 2009–2010.

Tabor, K. & Williams, J.W. (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications* **20**, 554–565.

Tan, S.S. & Smeins, F.E. (1996) Predicting grassland community changes with an artificial neural network model. *Ecological modelling* **84**, 91–97.

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

Tingley, M.W., Monahan, W.B., Beissinger, S.R. & Moritz, C. (2009) Birds track their Grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences of the United States of America* **106 Suppl**, 19637–19643.

Torres, J., Brito, J., Vasconcelos, M., Catarino, L., Gonçalves, J. & Honrado, J. (2010) Ensemble models of habitat suitability relate chimpanzee (*Pan troglodytes*) conservation to forest and landscape dynamics in Western Africa. *Biological Conservation* **143**, 416–425.

Tóth, M., Magyari, E.K., Brooks, S.J., Braun, M., Buczkó, K., Bálint, M. & Heiri, O. (2012) A chironomid-based reconstruction of late glacial summer temperatures in the southern Carpathians (Romania). *Quaternary Research* **77**, 122–131.

Traverse, A. (2007) *Paleopalynology*. Springer, Dordrecht, The Netherlands, 2nd edn.

Ursenbacher, S., Carlsson, M., Helfer, V., Tegelström, H. & Fumagalli, L. (2006) Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. *Molecular ecology* **15**, 3425–3437.

Vega, R., Fløjgaard, C., Lira-Noriega, A., Nakazawa, Y., Svenning, J.C. & Searle, J.B. (2010) Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography* **33**, 260–271.

Vila, M., Vidal-Romaní, J.R. & Björklund, M. (2005) The importance of time scale and multiple refugia: incipient speciation and admixture of lineages in the butterfly *Erebia triaria* (Nymphalidae). *Molecular Phylogenetics and Evolution* **36**, 249–260.

von Grafenstein, U., Erlenkeuser, H., Brauer, A., Jouzel, J. & Johnsen, S.J. (1999) A Mid-European Decadal Isotope-Climate Record from 15,500 to 5000 Years B.P. *Science* **284**, 1654–1657.

Waltari, E., Hijmans, R.J., Peterson, a.T., Nyári, A.S., Perkins, S.L. & Guralnick, R.P. (2007) Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PloS one* **2**, e563.

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

Webb III, T. (1986) Is vegetation in equilibrium with climate? How to interpret late-Quaternary pollen data. *Vegetatio* **67**, 75–91.

Weiss, S. & Ferrand, N. (eds.) (2007) *Phylogeography of Southern European Refugia*. Springer, Dordrecht.

Wiens, J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* **58**, 193–197.

Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A. & Snyder, M.A. (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19729–19736.

Wiens, J.J. (2008) Commentary on Losos (2008): niche conservatism déjà vu. *Ecology letters* **11**, 1004–1005.

Wiens, J.J. & Graham, C.H. (2005) Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics* **36**, 519–539.

Williams, J.W., Shuman, B.N., Webb III, T., Bartlein, P.J. & Leduc, P.L. (2004) Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. *Ecological Monographs* **74**, 309–334.

Willis, K.J., Bennett, K.D., Bhagwat, S.a. & Birks, H.J.B. (2010) 4°C and beyond: what did this mean for biodiversity in the past? *Systematics and Biodiversity* **8**, 3–9.

Willis, K.J. & Bhagwat, S.A. (2009) Biodiversity and climate change. *Science* **326**, 806–807.

Willis, K. & MacDonald, G. (2011) Long-Term Ecological Records and Their Relevance to Climate Change Predictions for a Warmer World. *Annual Review of Ecology, Evolution, and Systematics* **42**, 267–287.

Wroe, S., Field, J. & Grayson, D.K. (2006) Megafaunal extinction: climate, humans and assumptions. *Trends in Ecology & Evolution* **21**, 61–62.

Xavier, R., Lima, F.P. & Santos, A.M. (2010) Forecasting the poleward range expansion of an intertidal species driven by climate alterations. *Scientia Marina* **74**, 669–676.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693.

Zaniewski, A.E., Lehmann, A. & Overton, J.M.C. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* **157**, 261–280.

Zapata, L., Peña Chocarro, L., Pérez-Jordá, G. & Stika, H.P. (2004) Early Neolithic Agriculture in the Iberian Peninsula. *Journal of World Prehistory* **18**, 283–325.

Chapter 2

Objectives

> - Com estas ferramentas que problemas posso resolver?
> (Esta é a pergunta tonta.)
> - Com estes problemas que ferramentas preciso?
> (Esta pergunta é melhor.)
> - Com estes problemas que ferramentas tenho de aprender a utlizar?
> (Esta pergunta é ainda melhor: pressupõe vontade e um plano de acção.)
>
> — GONÇALO M. TAVARES, *Breves notas sobre ciência*

2.1 General objectives

The main objective of this work is to make an exhaustive use of new tools and innovating methods to study macro- and micro-scale processes, using the amphibians and reptiles of the Iberian Peninsula as study models. The main objective is branched into four relevant goals: 1) the production of new tools on the scope of landscape ecology and genetics research lines; 2) the reconstruction of spatial layers of past climates; 3) the analysis of the impacts of past and future climate change on species richness at a macro-scale; and 4) the analysis of environmental factors in the dynamics of a contact zone at the micro-scale.

2.2 Detailed objectives and structure of the thesis

The contents of this dissertation are organized in seven chapters. In chapter 1, I make a general introduction to the multidisciplinary subject of study, focusing on the general patterns of present climate and its known dynamics in the past and predictions

of future change. I provide further details for the case of the Iberian Peninsula that is the main study area used for most works in this document. In this chapter, I also introduce the potential impacts of climate change on biodiversity over broad scales and in the Iberian Peninsula in particular. This introduction covers the likely impacts of the warming trend and abrupt climate events that occurred since the last glacial maximum on diverse taxa, and details the cases of reptiles, in particular of the Iberian vipers. I also provide the theoretical background on climate reconstruction using fossil pollen data and on future climate prediction using global climate models. As it is one main subject of this work, I provide an introduction to the species niche concept, its broad implications, and tools to study its dynamics. Although I present some commonly used algorithms, I focus on the artificial neural networks due to the ability to fit non-linear relations often find in ecological datasets.

In the current chapter 2, I expose the main objectives and detailed questions addressed in this dissertation, and I provide a description of its organization.

In chapter 3, we present two new applications to the study of species distributions by means of ecological niche modelling and an exhaustive analysis of a new software application. This chapter is constituted by two published scientific articles in SCI journals and a third manuscript submitted to an international listed in the science-citation index:

- Tarroso, P. & Rebelo, H. (2010) E-Clic - Easy climate data converter. *Ecography*, **33**, 617–620.
- Tarroso, P., Carvalho, S. & Brito, J.C. (2012) Simapse - Simulation Maps for Ecological Niche Modelling. *Methods in Ecology and Evolution*, **3**, 787-791
- Tarroso, P. & Brito, J.C. Evaluating ecological niche models with virtual and real species. *submitted.*

In the first article we put forward a new tool (E-Clic) for the conversion of climate layers in a readily usable format within an ecological niche model framework and in most GIS packages. The source climate data are freely available and holds an historical data set with nine variables and climate predictions (from year 2001 to 2100) for five variables with combinations of five climate models and four emission scenarios. In the second article we propose an application to build complete ecological niche models from species data (presence and absence, presence only and continuous data) and environmental predictors using the artificial neural networks algorithm (Simapse). This application provides full control of network parameters, learning ability and subsampling strategies and provides extensive analysis of the built model. We later use

both applications to provide detailed analyses of past and future biodiversity shifts resulting from predicted climate changes.

In the third manuscript of this chapter we make an exhaustive analysis of Simapse's ability to build accurate models. We use a virtual species to derive positive and negative controls of the built models and further test them with real data from three reptile species of the Iberian Peninsula with different ecological requirements.

In chapter 4 we study climate dynamics in the Iberian Peninsula and Balearic Islands in the past, covering a period from 15,000 to 3,000 years BP. This study is presently submitted to an international journal listed in the science-citation index:

- Tarroso, P., Carrión, J., Dorado-Valiño M., Queiroz, P., Santos, L., Valdeolmillos-Rodríguez, A., Alves, P. C., Brito, J. C., & Cheddadi, R. Spatial climate dynamics in the Iberian Peninsula since 15 000 Yr BP. *submitted.*

The main objective of this work is to analyse the climate evolution of the Iberian Peninsula, depicting areas that shared similar climates during the studied period. We present new additions to methods of climate reconstruction using fossil pollen and spatial interpolation of the obtained values. From this process we obtain several spatial layers for three climate variables that later are used for analysing temporal patterns in the distribution of species richness of amphibians and reptiles.

In chapter 5 we use the applications presented in chapter 3 and the climate variables derived in chapter 4 to provide an exhaustive analysis of dynamic patterns of biodiversity change (using amphibians and reptiles as models) and evolutionary processes occurring in the Iberian Peninsula. In this chapter we present two manuscripts:

- Tarroso, P., Alves, P. C., Cheddadi, R., Brito, J. C. Velocity of biodiversity change: a case study in a global hotspot. *in prep.*
- Tarroso, P., Pereira, R., Martínez-Freiria, F., Godinho, R., Brito, J.C. Hybridization at an ecotone: ecological and genetic barriers between three Iberian vipers. *submitted.*

In the first manuscript of this chapter we expose relationships between species richness of amphibians and reptiles and climate dynamics in the Iberian Peninsula, using Simapse (chapter 3) with reconstructed past climate variables (chapter 4) and a set of future climates derived from E-Clic (chapter 3). We address two main questions: What is the dynamic of change in past species composition and its relation to present diversity?; and What is predicted to occur in terms of species compositional changes for the near future? To answer these questions we provide two new velocity indices

measuring species compositional change. The comparative analysis of past and future changes in species richness provides new insights to the spatial location of glacial refugia for amphibians and reptiles, and identifies areas facing potential dramatic impacts of future climate change. In the second manuscript of this chapter we analyse evolutionary processes occurring at a micro-scale in the Iberian Peninsula. A contact zone between three viper species is analysed using Simapse (chapter 3) to assess the role of environmental factors in shaping the dynamics and genetic structure of the populations. We address three main questions: What is the genetic structure of the focal taxa?; Is such genetic structure associated to ecological segregation between taxa?; and Is the ecological transition associated with barriers at the several species traits? To answer these questions we develop new methods in the scope of landscape genetics that provide an exhaustive picture on how environmental factors influence the maintenance of the contact zone.

In chapter 6 I provide a general discussion on the subjects addressed in the previous chapters, emphasizing the general achievements and suggesting future research lines. Finally, the chapter 7 summarises the major conclusions of this dissertation.

Chapter 3

Analyzing species' distributions with ecological niche modelling

> I call our world Flatland, not because we call it so, but to make its nature clearer to you, my happy readers, who are privileged to live in Space.
>
> — A SQUARE [EDWIN A. ABBOT], *Flatland*

3.1 E-Clic - Easy climate data converter[1]

3.1.1 Abstract

There are an increasing number of studies that are now focusing on the influence of climate change on species' distributions. However, access to predictive climatic datasets for future scenarios is difficult due to their specific formats and/or the need to be geographically downscaled. The TYN dataset is freely available to users and provides a synthetic format with several climatic models and IPCC future climate scenarios. Moreover, the CRU historical dataset (1901 – 2000) is also available which allows users to create baseline models for current climatic variables. E-Clic is a free, user-friendly software package that offers three different ways to convert these two datasets into a spatially explicit raster format which is compatible with the most common geographic information systems and usable on different platforms.

[1] This section refers to the published article: Tarroso, P. & Rebelo, H. (2010) E-Clic - Easy climate data converter. *Ecography*, **33**, 617–620.

3.1.2 Easy climate data converter

Climatic variables provide key input parameters when modelling species' distributions (Guisan & Zimmermann, 2000). Despite the range of different variables that are frequently used in ecological modelling, climate data is commonly used due to its continuous influence in shaping and limiting species' distributions, and also because of its correlation with other biologically meaningful variables (e.g. land cover) (Huntley & Webb III, 1989; Thomas *et al.*, 2004; Thuiller *et al.*, 2004). The changing dynamics of climatic forces are highly correlated with the shift of species' distributions, as shown in the past by the repeated moving patterns of populations during the climate oscillations of the glacial and interglacial periods (Huntley & Webb III, 1989). The current forecast of rapid climate change will force a shift in species' distributions and, in some cases, will increase their risk of extinction (Parmesan & Yohe, 2003; Thomas *et al.*, 2004). Predicting changes in species' distributions and the likelihood of their extinction is an important measure in understanding the possible impacts of climate change on biodiversity (Hannah *et al.*, 2002). Recently, several studies have been carried out on a number of species, including plants, birds, marine animals, amphibians and reptiles (Hawkes *et al.*, 2007; Lemoine *et al.*, 2007; Thuiller *et al.*, 2005; Araújo *et al.*, 2006). Consequently, achieving accurate distribution models using present day climatic variables is the first step in projecting these models for future predicted climate scenarios. This will allow for more informed decisions regarding the possible loss of biodiversity.

As one of the greatest political, social and scientific concerns of our time, research on climate change has generated a multidisciplinary interest which has produced a wide range of models that combine different social-economic scenarios to predict several climatic variables in the future (Nakicenovic & Swart, 2000). Climatic models are available at different geographic scales but usually have a coarse resolution and can be stored in a difficult format that is not easily integrated into the most commonly used geographic information systems (GIS) (e.g. ArcMap) and ecological modelling software (e.g. MaxEnt). The TYN dataset (Mitchell *et al.*, 2004) provides several downscaled models that are frequently used to study the impact of climate change. Specifically, it consists of monthly data broken down into five variables (precipitation, daily mean temperature, diurnal temperature range, vapour pressure, cloud cover) for five global circulation models (CGCM2, CSIRO mk 2, DOE PCM, HadCM3, ECHam4) over 100 year intervals (from 2001 to 2100). It also includes four special reports on emission scenarios (A1FI, A2, B1 and B2) resulting from the Intergovernmental Panel on Climate Change (IPCC) meeting (Nakicenovic & Swart, 2000). These high-resolution datasets are freely available from the Climate Research Unit web page

(`http://www.cru.uea.ac.uk/cru/data/hrg.htm`) at two different scales: the TYN SC 1.0 with European data downscaled to 10' resolution and the TYN SC 2.0 with global coverage at a 0.5º resolution. Although new future climate models and datasets occasionally emerge, the TYN dataset remains as a stable and complete source of future climate predictions for ecological studies. In addition to the TYN dataset, the CRU TS 1.0 dataset (Mitchell *et al.*, 2004) is also freely available to researchers and comprises of historical climate data ranging from 1901 to 2000 for the same variables, coverage and resolution as TYN SC 1.0. This data is highly useful because it provides researchers the opportunity to widen the time frame of their studies and to test the accuracy of their models if a species range has been well documented over the last century. The equivalent historical dataset for the TYN SC 2.0, the CRU TS 2.10 (Mitchell & Jones, 2005) is also available, including four new variables (monthly average daily maximum temperature, monthly average daily minimum temperature, wet day frequency, frost day frequency).

Here we present a new software program to convert public climatic datasets (TYN SC and CRU TS) into formats that are more commonly used and therefore can be directly utilized in GIS. 'Easy climate data converter' (E-Clic) (`http://purl.oclc.org/eclic/`) is a free, open-source application which is written in Python and can be used in different operating systems such as Windows, Mac OS or Linux (Fig. 3.1). It can read both TYN SC and CRU TS datasets and is able to convert them to ASCII raster maps - a format easily readable in most GIS and ecological modelling software (e.g. Arcgis, Idrisi, Maxent).

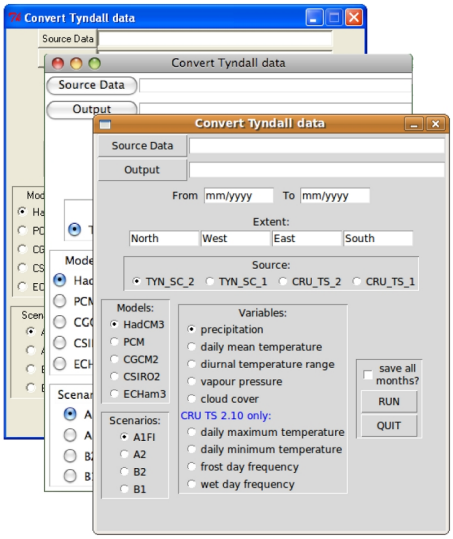


Fig. 3.1 – E-Clic built-in graphical user interface with all the available options to convert from TYN or CRU data to a raster format, running in (from front to back) Linux, Mac OS and Windows.

Prior to data conversion, users should download and decompress the TYN and/or CRU datasets to a chosen folder. The CRU dataset is available in two different forms; one single file that includes all the CRU TS range data, and as separate files containing data for each decade. The single file that includes all data must be used. Unpacking of the data from the files done by E-Clic follows the procedure described online with the dataset (see dataset Internet site for more details). In short, it consists of searching the data for the defined period of time, converting it into real units given by the model/scenario, and then testing it with the allowed minimum and maximum limits for the chosen variable. The user can choose one model, scenario and variable from all of the TYN supporting data. The output is the average model for the chosen time period and, optionally, a raster for each month within the same period.

The output created by E-Clic is easily integrated into most GIS packages and modelling software. The ability to query a spatial and temporal dataset and create raster data from them avoids, in most cases, the need to post-process the raster datasets. Although there are other software packages available to convert TYN SC and CRU TS data into different formats (see the dataset web page for more software and Solymosi *et al.* 2008), they lack some of the features presented in E-Clic such as a user-friendly interface, ability to be used on different platforms and/or the option to export data into a raster format. As E-Clic is written in Python, it benefits from the multi-platform availability of this programming language (Fig. 3.1). Moreover, it is not dependent upon external python modules and can be run directly after python installation (`http://www.python.org`), which is already available as default on some platforms (e.g. most of Linux distributions and Mac OS). In addition, this software also offers three different interfaces to convert data, depending on the user's preferences.

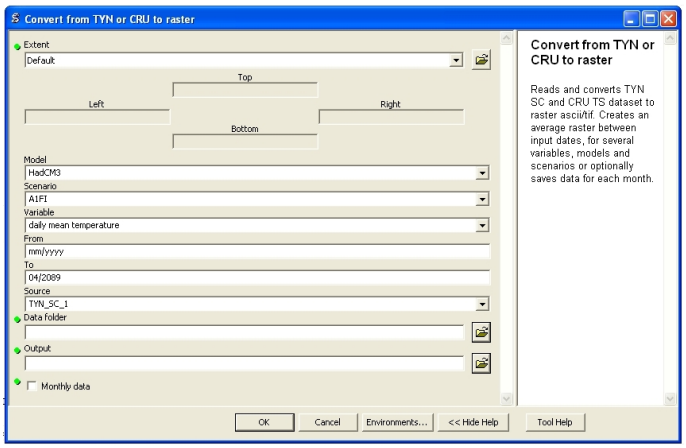


Fig. 3.2 – E-Clic interface using ArcGIS software as a toolbox. All options are available and the extent may be automatically defined.

E-Clic may be used at the command line with data inputted in a strict sequence format (see the online instructions in E-Clic website for more details on how to use E-Clic at the command line). The command line feature may be used to build quick batches to produce large quantities of data. The built-in graphical user interface (Fig. 3.1) is displayed when the user runs E-Clic without any additional command, the interface is launched displaying all the possible options making the process of data conversion

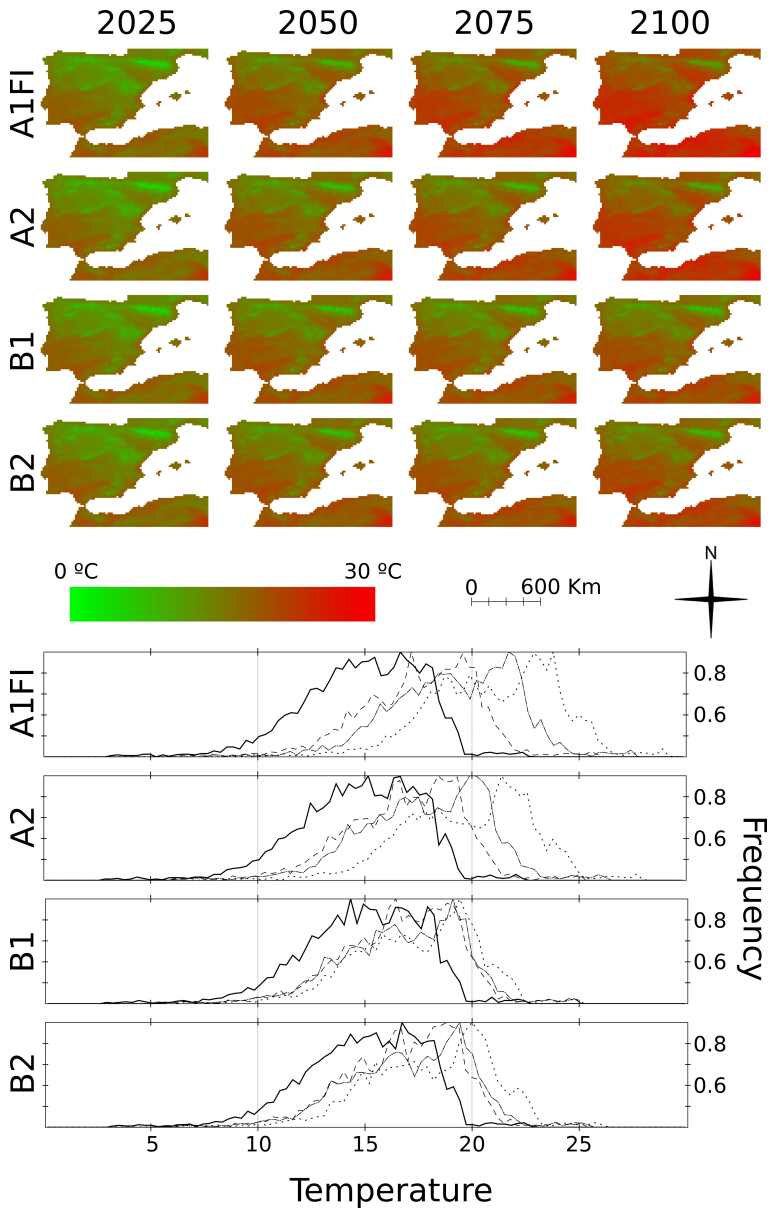


Fig. 3.3 – Average daily temperature forecasts for the years 2025, 2050, 2075 and 2100 for four available scenarios (A1FI, A2, B1 and B2). The plots represent the frequency of pixels with the same value of temperature. The year 2025 is plotted with a solid line (—) 2050 with a dashed line (- - -), 2075 with a dash and dot line (- . -) and 2100 with a dotted line (. . .).

between formats relatively simple. Finally, it may also be integrated into ArcGIS (ESRI, Redlands, CA, USA) as a toolbox (Fig. 3.2) where the graphical interface provided by the GIS package can be used to access all of E-Clic's functions. When it is used within the GIS or when it detects a valid license, it will also output the raster in a geoTIF format. All result files are saved to the chosen output directory and the names are explicit to the data they contain.

We predict that the main application of the output raster maps created by E-Clic will be in species distribution modelling for future climate scenarios. We present here a simple example of the output maps for an area covering the Iberian Peninsula and part of North Africa (Fig. 3.3). With TYN_SC_1.06 data, we have built annual mean values for four daily mean temperature scenarios (ranging from the more extreme A1FI, A2, B2 to the less severe B1) for the years 2025, 2050, 2075 and 2100. We synthesized the data using histograms where the frequency of map cells with the same value is plotted against the temperature. The results show that all the scenarios predict an increasing number of cells with higher temperatures with time and, as expected, the A1FI scenario has the greatest increase in temperature. The increase of the maximum values of the predicted mean temperatures for this area between 2025 and 2100 reaches the highest value of 6.7^{o}C with the A1FI scenario, whereas the lowest is 2.7^{o}C for the B1 scenario.

Even though there are a limited number of climatic data conversion software packages available, the majority of them require some knowledge of computer programming. Since the majority of users will not be familiar with this type of programming, especially those who come from different scientific fields such as ecology, accessing climatic data can be difficult. Here we present a user-friendly application that allows anyone with the most basic GIS knowledge to easily access climatic data for future scenarios. Given that climate change studies are becoming more important, we hope that this software will help more researchers to have access to the core datasets for their studies.

3.1.3 Acknowledgments

The authors are grateful to Jon Flanders for his extensive review of the manuscript that greatly improved its quality. The authors are also grateful to James Harris, Anna Perera and two anonymous referees for their comments and suggestions. PT and HR were funded by the 'Fundação para a Ciência e Tecnologia' doctoral grants SFRH/BD/42480/2007 and SFRH/BD/17755/2004, respectively.

3.1.4 References

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**, 1712–1728.

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

Hannah, L., Midgley, G. & Millar, D. (2002) Climate change-integrated conservation strategies. *Global Ecology and Biogeography* **11**, 485–495.

Hawkes, L.a., Broderick, a.C., Godfrey, M.H. & Godley, B.J. (2007) Investigating the potential impacts of climate change on a marine turtle population. *Global Change Biology* **13**, 923–932.

Huntley, B. & Webb III, T. (1989) Migration: species' response to climatic variations caused by changes in the earth's orbit. *Journal of Biogeography* **16**, 5–19.

Lemoine, N., Schaefer, H.C. & Böhning-Gaese, K. (2007) Species richness of migratory birds is influenced by global climate change. *Global Ecology and Biogeography* **16**, 55–64.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M. & New, M. (2004) A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100).

Mitchell, T.D. & Jones, P.D. (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* **25**, 693–712.

Nakicenovic, N. & Swart, R. (2000) *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42.

Solymosi, N., Kern, A., Maróti-Agóts, A., Horváth, L. & Erdélyi, K. (2008) TETYN: An easy to use tool for extracting climatic parameters from Tyndall data sets. *Environmental Modelling & Software* **23**, 948–949.

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

Thuiller, W., Araújo, M. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography* **31**, 353–361.

Thuiller, W., Lavorel, S., Araújo, M. & Sykes, M. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences* **102**, 8245–8250.

3.2 Simapse - Simulation Maps for Ecological Niche Modelling[2]

3.2.1 Abstract

1. Artificial neural networks (ANN) are known for their powerful predictive power in the analysis of both linear and non-linear relationships. They have been successfully applied to several fields including ecological modelling and predictive species' distributions.
2. Here we present Simapse – Simulation Maps for Ecological Niche Modelling, a free and open-source application written in Python and available to the most common platforms. It uses ANNs with back-propagation to build spatially explicit distribution models from species data (presence/absence, presence-only, and abundance).
3. The main features include the automatic production of replicates with different subsampling methods and total control of ANN structure and learning parameters.
4. Simapse uses common text formats as main input and output and provides assessment of variable importance and behaviour and measurement of model fitness.

Keywords: artificial neural networks, ecological niche modelling, species distribution, Python

3.2.2 Introduction

Artificial Neural Networks (ANNs) have been used in scientific fields where pattern recognition is a primary need and its application to biological systems had a considerable growth during the past decades (Lek *et al.*, 1996; Lek & Guégan, 1999; Özesmi *et al.*, 2006a). As a machine learning algorithm, it is with no surprise that ANNs are increasingly being used in ecological studies. These algorithms are usually seen as more powerful to deal with complex ecological datasets than other methods (Brosse & Lek, 2000; Olden *et al.*, 2008; Özesmi *et al.*, 2006b; Pearson *et al.*, 2002).

A common type of ANN is the feed-forward neural network with back-propagation learning (BPN; Lek & Guégan, 1999). This network has a layered structure of neurons connecting the inputs to an output through one or several hidden layers (see supplementary material A.1 for more details). The BPN has been used to model ecological systems due to its efficient learning ability and to its simple nature that makes it easy to understand (Lek & Guégan, 1999; Özesmi *et al.*, 2006a). The unit of the BPN is an artificial neuron with an activation function, usually linear or sigmoid. It squashes the

[2]This section refers to the published article: Tarroso, P., Carvalho, S. & Brito, J.C. (2012) Simapse - Simulation Maps for Ecological Niche Modelling. *Methods in Ecology and Evolution*, **3**, 787-791

sum of the products of all connecting weights and the respective neuron output in the previous layer to a value to be passed to the next layer of neurons by the connecting weights.

3.2.3 Simapse

Simulation Maps for Ecological Niche Modelling (Simapse - `http://purl.oclc.org/simapse`) is an open source and multi-platform application released under GNU Public License (GPL) and written in Python (`http://www.python.org`) that applies the pattern recognition power of BPN to ecological data within a spatially explicit framework (Fig. 3.4). Although Simapse's dependence on a few external python modules for graphing purposes, a complete model can be built only with Python core installation. The process of creating potential distribution maps with Simapse is straightforward and benefits from the graphical user interface, the strong spatial component, input and output with common text formats and options to full control the sub-sampling and learning ability (table 3.1). This application automates the process of building several models with different sub-sampling methods and the creation of a final averaged prediction, assuring robust results by taking into account the independent information of the individual models (Araújo & New, 2007) and a description of uncertainty between individual models. Simapse is also able to project models to a different set of the same variables, including distinct spatial or temporal extents.

3.2.4 Input data and general options

The graphical interface of Simapse has five main areas (Fig. 3.4): 1) the input/output definitions; 2) the sub-sampling methods; 3) the BPN proprieties and general options; 4) the buttons area; and 5) the text box. Simapse uses target data (i.e. species presence) and independent variables as inputs to build a model The independent variables are ascii raster files and should be placed inside a directory that is given to Simapse. These variables are automatically standardised to z-scores (Lek *et al.*, 1996; Özesmi & Özesmi, 1999; Özesmi *et al.*, 2006a). All standardised variables are saved as ascii rasters in a directory inside the provided raster directory.

The targets may be presence/absence, presence-only, or abundance data in a text file formatted with a header (e.g. target;longitude;latitude) and samples by row and fields separated by semicolon. The presence and absence data are defined in the text file by 1 and 0, respectively. When using presence-only data, the user may define a ratio of presence and pseudo-absences. When using presence/absence or presence-only data, the evaluation of the final model is made by a Receiver Operating

Characteristic (ROC) and Precision-Recall (PR) curves with the respective Area Under the Curve (AUC) values. These methods are based on the confusion matrix of real and predicted values but whereas ROC uses the full table by comparing the sensitivity and 1-specificity, the PR avoids the use of the true negative values by using the precision and recall values (Davis & Goadrich, 2006). These measures have a high discrimination power but should be analysed with care, especially with pseudo-absences or when comparing different algorithms (Lobo *et al.*, 2007; Peterson *et al.*, 2007). Since all results are outputted as text files, model's performance may be evaluated by means other than those available natively in the application.

Modelling abundance data with Simapse is possible but it requires continuous data between zero and one. The input species data must be previously scaled to this range. To evaluate the performance of the model, Simapse builds a cross-validation plot where each abundance value is plotted against the output predicted value of the model.

To construct several models replicates there are three options of sub-sampling the target data: 1) random repetitions, where the user defines the number of model repetitions and a percentage of the data randomly chosen to test each repetition; 2) k-fold cross-validation, where data is divided in k folds and each fold is tested against

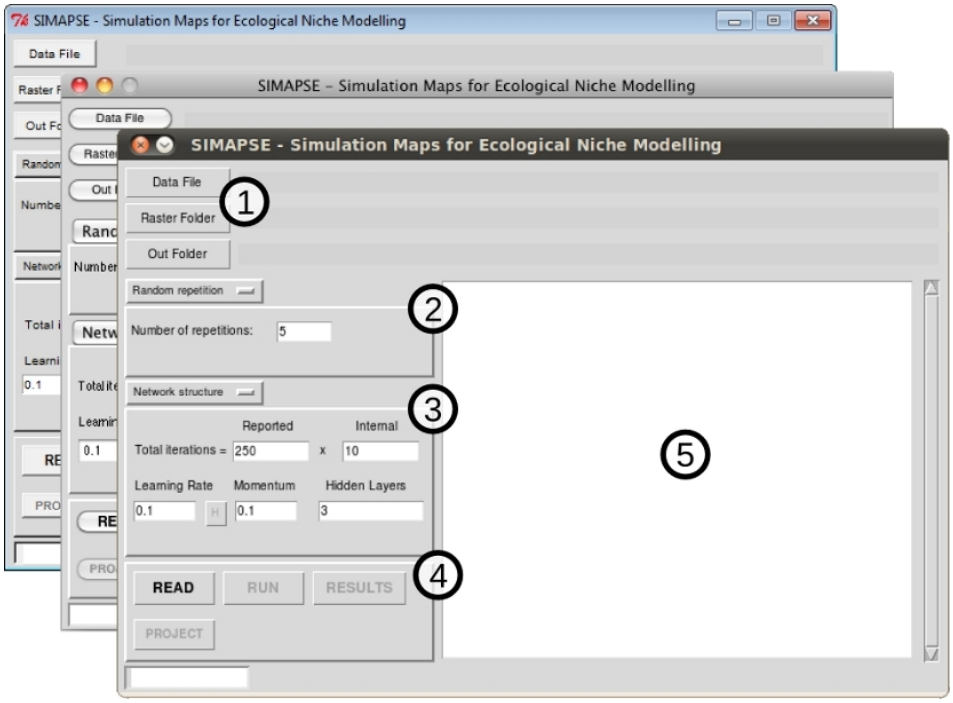


Fig. 3.4 – Graphical user interface from Simapse in different operating systems. Simapse has a simple layout of options divided by five main areas: 1) the input/output; 2) the sub-sampling methods; 3) the network proprieties and general options; 4) the buttons; and 5) the text box.

Table 3.1 – Overview of Simapse's general options.

Sub-sampling	
Random repetition	A percentage of data is randomly set aside of the sample records for evaluating the error or the network
K-fold cross-validation	The dataset is divided in k folds, and each model is trained with k-1 folds and tested with the remaining fold
Bootstrapping	Each sub-sample of user-defined size is obtained by random sampling with replacement from the dataset
Network structure	
Iterations	The iterations are divided in internal (number of times that the data are passed through the network to minimise the error until the report) and reported (quantity of reports that should be made to choose the best network)
Learning rate	Defines the learning amount (an indication of the value is given by the hint button)
Momentum	Defines the inertia of the learning, i.e. the influence of the previous weight change in the current change
Hidden layers	Network's hidden layers architecture defined by the user: neurons per layer separated by comma (e.g. '3, 2, 4' creates three hidden layers with three, two and four neurons)
Options	
Test percentage	The percentage of the sub-sampled dataset (by random repetition or bootstrap) that will be used to test the network
Pseudo-absences ratio	With presence-only datasets, defines the proportion of pseudo-absences to be created in relation to the number of presences
Burn-in iteration	Number of beginning iterations to achieve minimum learning
AUC filter	If active, defines the AUC threshold to accept a reported network

the remaining fold, resulting in a total of k models; 3) bootstrapping, where a new dataset is created, based on a percentage defined by the user of the original dataset, by randomly sampling the data with replacement for both training and test datasets.

The options for the BPN assembly are divided by two main groups: the network structure and the learning options. The user has to define the structure of the hidden layers, since the input layer is based on the number of detected variables in the rasters directory and a single output is always used. The hidden structure is defined with the number of neurons per layer separated with commas. Although Simapse allows several hidden layers, one is usually enough to solve complex iterations between the dependent and the independent variables. A simple structure usually is more prone to generalization, thus avoiding overfitting, and requires less computing power and time (Özesmi & Özesmi, 1999). Usually, the choice of the hidden architecture is made by trial-and-error (Dimopoulos *et al.*, 1999; Özesmi *et al.*, 2006a). An indication of the learning rate value is obtained using the hint button, which tests several learning rates

values with the internal iterations, momentum and hidden structure defined by the user and presents the values classified by the amount of learning they can produce.

The learning options available refer to the number of iterations, learning rate and momentum. The final number of iterations is the product of internal iterations (i.e. the number of times each sequence of targets is passed through the network during the training phase) and the reported iterations, where Simapse reports the error and, optionally, AUC value of the training process. During the training stage, Simapse saves each reported network to the output folder. After the training process, only the best network is preserved and the choosing algorithm acts by selecting the network that presents the lowest sum of training and test errors. When an AUC threshold is defined, all networks that did not meet the threshold are removed previously to the test. This process results in a model that is representative of the training data relationships, avoiding possible overfitting of the BPN by testing each network (i.e. training) against a second dataset (i.e. test). This procedure allows the achievement of a good generalization (Dimopoulos *et al.*, 1999).

3.2.5 Output results

After running the model, the user defined output folder contains all the results produced by Simapse, saved in text and image formats. The successful built models are saved in the output folder as rasters and are averaged to a single consensus model. Simapse also produces rasters of prediction uncertainties by calculating the spatial standard deviation of all models.

Although ANNs are still seen as “black boxes”, there are several processes to disentangle the effect of predictors in the model (Fu & Chen, 1993; Dimopoulos *et al.*, 1995; Lek *et al.*, 1996; Olden & Jackson, 2002; Gevrey *et al.*, 2003, 2006a,b; Özesmi *et al.*, 2006a). Simapse incorporates sensitivity techniques that provide reliable results in identifying the variables general contribution and response (Gevrey *et al.*, 2003). The partial derivatives algorithm (PaD) measures the sensitivity of the network in respect to the input data. Two outputs are given by Simapse using PaD: 1) the variable contribution to the model and 2) the individual partial derivatives that measure the sensitivity throughout each variable range. The profile algorithm acts by setting all variables to zero except one, for which it depicts the predictive behaviour throughout its range of values. We added a third method, the variable surface, that is similar to the profile method and is based on Lek’s algorithm (Lek *et al.*, 1996). It plots the prediction surface of a variable throughout the range of all other variables.

In addition to the plots, Simapse also outputs text files with the results data for

all sensitivity analyses and model building stages. To easily retrieve any of those files in the output folder, Simapse also creates a report of the model with a full summary and links to all images and respective text files. All spatial results are output as ascii raster files and are ready to import to most GIS packages.

3.2.6 Example

To better illustrate the work flow with Simapse and its outputs, we created a simple VS widespread throughout Europe based on five real environmental variables (Fig. 3.5). The original variables data were downloaded from Worldclim (`http://www.worldclim.org/`) with 10' resolution and further processed to create the maximum and minimum precipitations and temperatures, plus the altitude data. The presence area of the VS was obtained by averaging the Gaussian or logistic functions applied to the variables (Fig. 3.5; see supplementary material A.2). A dataset with 100 presence locations chosen randomly from the presence area of the VS were used as input to the model (this dataset is included with the download of Simapse).

We used a single hidden layer network with five neurons and set the learning rate to 0.1 after the hint given by the application. The sub-sampling method was set to 50 random repetitions. Each replicate was trained with 1000 iterations and filtered with the AUC value (0.9 for train and 0.8 for test). All other parameters were set to the application's default. The same set of variables used to construct the VS was used as predictors to build the model.

After running Simapse, five replicates were discarded from the consensus model due to not meeting the AUC threshold. The consensus model matched the area of

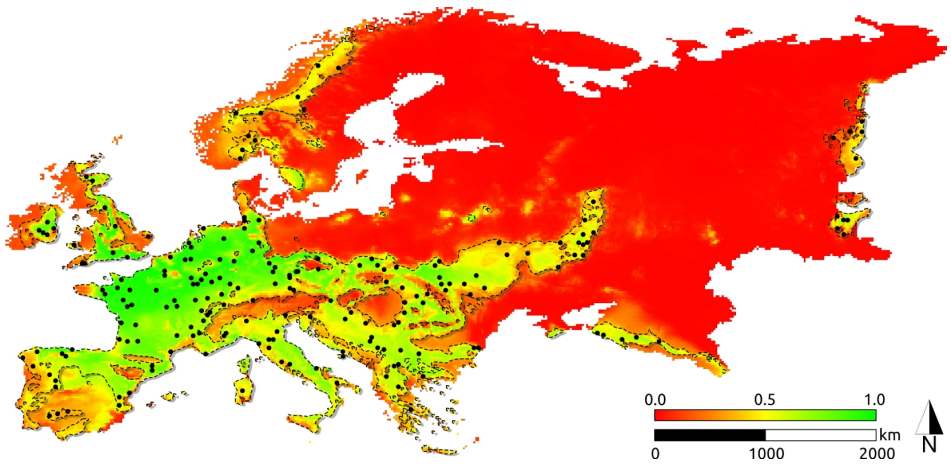


Fig. 3.5 – Consensus model of the virtual species presence. The gradient describes the probability of presence. The black dots are the locations of the 100 presences used to model that were randomly selected from the distribution area of the virtual species delimited by the dashed line.

presence of the VS (Fig. 3.5) from the presences data, though locations in the border of the presence area showed higher standard deviation (Fig. 3.5; supplementary material A.3).

Simapse provides images of the model and exhaustive analysis (see supplementary material A.3-A.4). For each variable, it is produced a series of partial derivatives, profiles and variable surfaces plots (see supplementary material A.5-A.7).

Simapse exhibited a good learning ability depicting the distribution of the species with presence-only data. The learning process was able to detect the general trend of the variable use with the presence-only data, as shown by the comparison of the real use of the variable by the VS and the sensitivity analysis results (see supplementary material A.8), found in the text files in output folder. The user benefits from the availability of results in this format to easily produce plots to fit particular purposes.

3.2.7 Discussion

Simapse provides a spatially explicit framework to model species' distributions with ANNs with sensitivity analysis for studying the influence of each explanatory variable. The example shown here illustrates the work flow with a VS that has a very simple relation with the descriptors. Despite the extensive use of ANNs with ecological data, testing Simapse with more complex models and real case examples and different sampling strategies is needed. Nevertheless, we expect that the easy learning path of Simapse with native analysis of the results may provide additional advantages over other code demanding approaches, especially under more pragmatic areas like applied conservation. Being an open-source software it also may benefit from the experience of more advanced users to suggest and/or improve it. This system also allows adjustments of the application to fit the specific scope of different user's projects.

The transparent framework, with all data outputted as text, is expected to result in ample understanding of the built models, allowing completing the descriptive analysis by other means then those native to the application. Moreover, the models produced may integrate broader approaches where analyses with different algorithms are used. We expect that the easy pathway provided by this application to predict species distributions along with the efficient pattern finding ability of BPNs will assure the usefulness of Simapse for biodiversity studies.

3.2.8 Acknowledgments

PT, SC and JCB are supported by Fundação para a Ciência e Tecnologia (SFRH/BD/42480/2007, SFRH/BPD/74423/2010 and Programme Ciência 2007, re-

spectively). We thank A. Townsend Peterson and an anonymous reviewer for the helpful comments to a preliminary version of the manuscript.

3.2.9 References

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* **22**, 42–7.

Brosse, S. & Lek, S. (2000) Modelling roach (*Rutilus rutilus*) microhabitat using linear and nonlinear techniques. *Freshwater Biology* **44**, 441–452.

Davis, J. & Goadrich, M. (2006) The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning* pp. 233–240.

Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A. & Lek, S. (1999) Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling* **120**, 157–165.

Dimopoulos, Y., Bourret, P. & Lek, S. (1995) Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters* **2**, 1–4.

Fu, L. & Chen, T. (1993) Sensitivity analysis for input vector in multilayer feed forward neural networks. *IEEE International Conference on Neural Networks*, pp. 215–218.

Gevrey, M., Dimopoulos, I. & Lek, S. (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* **160**, 249–264.

Gevrey, M., Dimopoulos, I. & Lek, S. (2006a) Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecological Modelling* **195**, 43–50.

Gevrey, M., Lek, S. & Oberdorff, T. (2006b) Utility of sensitivity analysis by artificial neural network models to study patterns of endemic fish species. *Ecological Informatics: Scope, techniques and applications* (ed. F. Recknagel), chap. 14, pp. 293–306, Springer, Berlin, 2nd edn.

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.

Lek, S. & Guégan, J.F. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* **120**, 65–73.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

Olden, J.D. & Jackson, D.A. (2002) Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling* **154**, 135–150.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* **83**, 171–193.

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

Özesmi, S., Tan, C. & Özesmi, U. (2006a) Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling* **195**, 83–93.

Özesmi, U., Tan, C., Özesmi, S. & Robertson, R. (2006b) Generalizability of artificial neural network models in ecological applications: Predicting nest occurrence and breeding success of the red-winged blackbird *Agelaius phoeniceus*. *Ecological Modelling* **195**, 94–104.

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.

Peterson, A.T., Papes, M. & Sober, J. (2007) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecology* **3**, 63–72.

3.3 Evaluating ecological niche models with virtual and real species[3]

3.3.1 Abstract

Ecological niche models are widely used to study species' distributions based on presence data collected in the field or gathered from diverse collections. Several algorithms are available and testing an ENM is usually done comparatively between different methods. Since different algorithms tend to answer different questions, this comparison is not trivial to execute.

Here we present a framework to make exhaustive tests of an ENM and to assess its predictive ability under controlled conditions. For that purpose we use a set of virtual species representing different ecological demands and a set of real species based on distributional atlases. Both data sets offer an accurate description of the complete distribution of each species, however, the virtual species allows a full control by providing both presence and absence areas.

To perform the models we used an artificial neural network (ANN) with nine commonly used environmental variables. We test the performance of the ANN with varying network structures and with different sample sizes. Models for the virtual species with pseudo-absences were compared to the control model with the accurate virtual presences and absences. The performance of models for the real species were assessed with the real distribution and compared with the results of the virtual species, under the same conditions.

The main finding is that ANNs are accurate in predicting the distribution of both virtual and real species. The structure of the network has little implications for the final models, but simple networks have generally good results and are preferred. Low sample size produces accurate models and the spatial uncertainty of the resulting models tends to increase with the decreasing number of available samples. The models of the real species were found to correctly predict presence when compared with the full distribution information.

The framework that we have used here can be extended to other ecological modelling algorithms to fully assess their performance. It allows to test the implications of the set of parameters typical of each algorithm, and to assess the performance under different scenarios of data availability.

[3]This study is presently submitted to an international journal in the science-citation index: Tarroso, P. & Brito, J.C. Evaluating ecological niche models with virtual and real species.

3.3.2 Introduction

The spatial context of a species distribution is self-evident and its analysis, either in current time or in other time periods, is a central aspect of biogeography. A research line in this field is the study of the relation between species' locations and the environment. Ecological niche modelling (ENM) has been widely used to study relationships between species' distribution and ecological niche variables (ENV) and to predict occurrence in areas outside the sampling locations (Guisan & Zimmermann, 2000; Ferrier, 2002; Guisan & Thuiller, 2005; Elith & Leathwick, 2009). The wide range of applications of these tools reveals the importance of ENMs in current research. It comprises diverse areas as biodiversity conservation (Brito *et al.*, 2009, 2011; Carvalho *et al.*, 2011a), evolutionary biology (Martínez-Freiría *et al.*, 2009; Tomović *et al.*, 2010), human archeology (Banks *et al.*, 2008), and predictions of climate change effects on biodiversity distribution (Carvalho *et al.*, 2010, 2011b; Rebelo *et al.*, 2010).

To predict distributions, ENMs rely on algorithms that relate species occurrence with the ENVs. Comparisons of different algorithms have suggested that machine learning methods, such as artificial neural networks or maximum entropy, are more robust than other traditional methods (Segurado & Araujo, 2004; Elith *et al.*, 2006). Artificial neural networks (ANN) were defined as a promising method for habitat distribution modelling (Guisan & Zimmermann, 2000), but until recently their use was restricted to a small circle of ecological modellers (Olden *et al.*, 2008). Nonetheless, research with ANN is active (Özesmi *et al.*, 2006; Park & Chon, 2007; Olden *et al.*, 2008; Tarroso *et al.*, 2012) and the method has been applied during the last two decades to study several ecological and environmental processes (Lek & Guégan, 1999; Lek *et al.*, 1996; Araújo *et al.*, 2006; Özesmi *et al.*, 2006; Xavier *et al.*, 2010). An advantage of ANN over other methods is the ability to fit non-linear relationships resulting in an high predictive power which can virtually approximate any continuous function (Gevrey *et al.*, 2003; Olden *et al.*, 2008), and thus rendering it an adequate method to model complex ecological data (Barry & Elith, 2006). ANN have been shown to outperform other methods (Segurado & Araujo, 2004). The unit of ANN is the biologically inspired artificial neuron, which processes an input signal and yields an output accordingly to its internal activation function (usually a sigmoid function). These neurons are fully connected by weights on a network structured in layers. The most simple form is a three layered network: the first layer corresponds to the input neurons (with the same number as ENVs available); the second layer is known as hidden layer with an arbitrary number of neurons; and the third is the output layer with one neuron that returns the final result. As a supervised machine learning method, the error of the network

is assessed with training data and is propagated backwards through the network to adjust the value of the connecting weights (back-propagation learning). The repetition of this process for several iterations produces a trained network with minimal error for the given input that was used to predict the presence of a species in a specific location with the same set of ENVs.

The evaluation of the performance of an ENM is based on the comparative analysis of multiple algorithms with single or multiple species (e.g. Segurado & Araujo, 2004; Elith *et al.*, 2006). This task is not trivial due to the particular design or operation of an algorithm affecting the possible range of commission errors (Peterson *et al.*, 2007), or to the impact of different species strategies (i.e. generalist or specialist), resulting in different occurrence extents that has consequences on the general model performance (Araújo & Williams, 2000; Segurado & Araujo, 2004). Thus, the knowledge of the full distribution of the species would be a valuable asset for a complete analysis of model performance. Although an exhaustive record of distribution data is nearly unattainable with a real species, a virtual species provides an efficient method to overcome this difficulty (Hirzel *et al.*, 2001; Guisan *et al.*, 2006; Zurell *et al.*, 2010). Virtual species enlighten model performance under different circumstances, as the researcher is aware of the complete virtual reality. However, caution is needed when analysing virtual species results. Despite the immense data it potentially provides, the success of ENMs in the virtual world does not guarantee the same success in the real world (Zurell *et al.*, 2010). As such, modelling species distribution with distribution data from atlases constitutes a complementary alternative to evaluate ENM performance. These presence data compendiums usually provide details of sampling effort and are based on a uniform strategy over a defined geographical area and time period constituting an important resource for biodiversity conservation (Robertson *et al.*, 2010) and extensive species data for modelling purposes. A framework to assess the performance of a given algorithm for ENM combining results from virtual species and atlases data are lacking in the scientific literature, especially in the case of ANN.

Species' distribution datasets are usually composed of presence-only data. Absences are associated to high uncertainty (Lobo *et al.*, 2010), therefore most studies are limited to the presence locations. Simulating absences, generally referred as pseudo-absences, is a common method to overcome this limitation. Nevertheless, indulgent usage of artificially generated absence data may have negative impacts on the final model (Chefaoui & Lobo, 2008; Vanderwal *et al.*, 2009; Lobo *et al.*, 2010).

Much has been done to evaluate ENMs strategies; however, a framework that provides positive and negative controls to analyse the performance of an ENM using

Table 3.2 – Ecological niche variables (ENV) general information and usage by the generalist virtual species (GVS), specialist virtual species (SVS), and the random virtual species (RVS).

				GVS (n=4612)		SVS (n=778)		RVS (n=3764)
ENV description	*Code*	*Units*	*Range*	*Function*	*Mean±sd*	*Function*	*Mean±sd*	*Mean±sd*
Elevation	elv	m	0–2989	logistic	590.7±363.6	–	300±280.2	645±416.3
Annual precipitation	ap	mm	227–1755	logistic	586.2±199.9	–	1075.2±216.5	653.7±256.7
Precipitation of the driest month	pdm	mm	0–110	normal	13.1±9.9	–	32.6±14.3	18.1±15.9
Precipitation seasonality	ps		11–76	logistic	42±13.4	–	38.5±11.9	39.6±14.1
Precipitation of the wettest month	pwm	mm	32–270	logistic	78.8±31	–	145.9±34.3	86.1±35.9
Slope	slp	º	0–68.2	logistic	4.7±5.1	–	6.8±7.1	6.1±7.1
Temperature annual range	tar	ºC	13.5–34.1	logistic	27.5±3.8	logistic	18.6±2.2	26.7±4.1
Minimum temperature of the warmest month	mtwm	ºC	11.4–36.5	logistic	29.6±3.1	normal	23.2±1.3	28.5±3.9
Minimum temperature of the coldest month	mtcm	ºC	-10.3–8.5	logistic	2.1±2.9	normal	4.6±1.9	1.8±3.0

species with different niche requisites and to test the impact of pseudo-absence data it is still lacking. The combination of virtual species with real data provide an opportunity to test ENM thoroughly with a design that allows comparable models to be built with effective controls of their performance. The objective of this study is to develop a testing framework for evaluation of an ENM. The general objective is divided in three key aspects: 1) comparison between presence-only and presence and absence data; 2) assessment of the impact of algorithm parameters; and 3) effect of sample size. We use presence datasets of virtual species, to test both algorithm parameters and sample size influence, and datasets of real species distribution from atlases to confirm them.

3.3.3 Methods

We developed a framework to extensively test the ANN predicting ability with virtual and real species (Fig. 3.6). The study area included the Iberian Peninsula covering an extent between the coordinates 43.8ºN, 36.1ºN, 9.5ºW and 3.3ºE with a mainland area of 580 000km^2. A set of ENVs including climatic and topographic data (Table 3.2) were used for modelling both real and virtual species. The climatic ENVs were obtained from Worldclim (`www.worldclim.org`; Hijmans *et al.*, 2005) with the exception of slope which was derived from altitude data. All ENVs were resampled to 0.09º to match the resolution of the presence data for real species, resulting in a total of 7686 cells.

Species data

We created three virtual species with different ranges of presence response simulating different niche requisites: 1) a generalist species (GVS) with broad distribution in the Iberian Peninsula; 2) a specialist species (SVS) with restricted and temperature-dependent range; and 3) a random virtual species (RVS) with distribution independent of the used ENVs (appendix B.1). To create both GVS and SVS we used the averaged partial niche coefficient for the chosen variables (Hirzel *et al.*, 2001). Each partial niche coefficient, defining the virtual species response, was extracted from the ENV using either a normal or a logistic function with a range between 0 and 1 (table 3.2). All responses were weighted equally to obtain a distribution with continuous

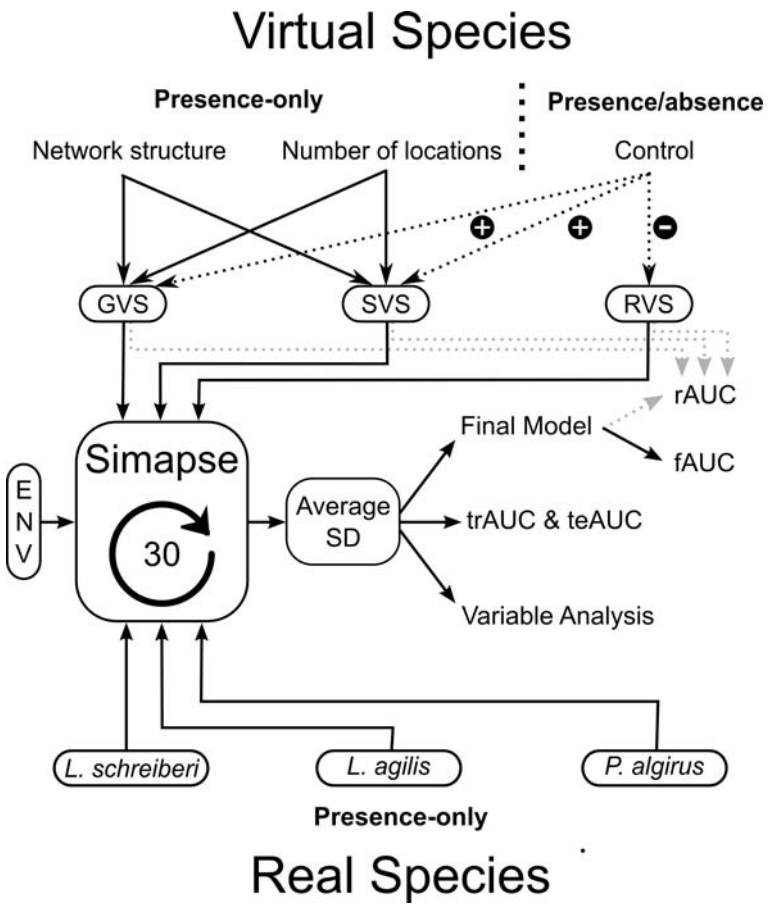


Fig. 3.6 – Framework of the study and general procedures to use artificial neural networks. The input data fed to Simapse was divided in virtual species (top) and real species (bottom). Virtual species provides a context to extensively test the effects of structure of the neural network and sample size. This is compared to presence and absence data which has a dual role (dotted arrows): positive control - it provides full virtual reality to the modelling process; and negative control – it provides a random reality which has no relation with ENVs. The access to full virtual reality allows the calculation of a real AUC (rAUC, with correct virtual absences plus the input presences) as opposed to the final AUC (fAUC, with the pseudo-absences) of the consensus model. Train and test AUC are processed (trAUC and teAUC, respectively). Three real species with presence-only data are used.

value. A threshold was then used to convert the presence probability to a binary presence/absence distribution from which datasets of presence-only and presence-absence locations were randomly extracted. All virtual species were created using R (R Development Core Team, 2012) with the package rgdal (Keitt *et al.*, 2010).

The GVS was created to make a general use of all ENVs range and to have a broad spatial extent in the study area. It had a normal response to the precipitation of the driest month (PDM) and logistic responses to all other ENVs (table 3.2, appendix B.1). An arbitrary threshold of 0.75 was chosen to create a binary distribution for the GVS. We generated a list of 400 random locations equally split between presence and absence datasets. We also generated presence-only datasets with 25, 50, 75, 100, 125, 150, 175 and 200 locations. The SVS was built to be more restricted in its distribution and to have specific thermal requirements. It has a logistic response to the temperature annual range (TAR) and normal response to the maximum temperature of the warmest month (MTWM) and minimum temperature of the coldest month (MTCM; table 3.2, appendix B.1). The binary distribution was created with a threshold of 0.5. The generated datasets include a set of 100 presence and 100 absence locations, plus presence-only datasets with 20, 40, 60, 80 and 100 locations. The RVS was generated by assigning a random number between 0 and 1 to each cell of the study area. The binary distribution of the RVS was produced with a threshold of 0.5 and 400 locations were extracted, equally divided between presences and absences (appendix B.1).

The real species used in this study included three lacertid lizards (*Lacerta schreiberi*, *Lacerta agilis*, and *Psammodromus algirus*) which exhibit different distributions and ecological requirements in Iberia. Both *L. schreiberi* and *L. agilis* are specialists and occupy different spatial extents, the first ranging in larger areas of the humid north-western Iberia (n=1152; 15% of study area) and the latter restricted to a few high altitude cells in the Pyrenees (n=11; 0.14%). *P. algirus* covers wider habitat types, occurring throughout most of the study area (n=4596; 60%). The distribution data for each species were obtained by digitalisation of the atlases of Spain and Portugal (Pleguezuelos *et al.*, 2004; Loureiro *et al.*, 2008).

Modelling procedure

All models were built using the ANN implemented in Simapse (Tarroso *et al.*, 2012). This software uses an ANN with back-propagation learning algorithm to produce potential species' distributions from binary data. All models were ran with random subsampling with 30 repetitions, from which a final averaged consensus model was built

using the standard deviation between models as a measure of uncertainty (Fig. 3.6). Each repetition was trained with 25 burn-in and 2500 training iterations (500 reports after a sequence of 5 internal iterations), and tested with 25% of the data chosen randomly from the input dataset to assure proper generalization by reducing overfitting (Dimopoulos *et al.*, 1995). At each reported iteration, the value of the area under curve (AUC) of the receiver operating characteristic curve (ROC) was saved together with the sum of squared error of the network. A repetition was only accepted in the final consensus model if, at some point of the training process, a network was reached with both train and test AUC$\geq$0.7 with real species, GVS and SVS. For RVS, the network was accepted with AUC$\geq$0.5.

We tested the influence of network structure with the GVS and SVS by running models with different structures, ranging from 1 to 18 neurons (double the available ENVs) in a single layer. These models were fed with the virtual presence-only data: 200 presence locations for the GVS and 100 locations for the SVS. Pseudo-absences were generated in the same number as presences. The learning rate (LR) was optimized heuristically for each species. In the case of GVS, the LR was set to 0.01, for SVS it was set to 0.1. Momentum was kept constant at 0.1 in all models. For each model we calculated the average train and test AUCs, (trAUC and teAUC, respectively) and the final AUC (fAUC) value of the consensus model with the presences and pseudo-absences as given by Simapse. The real AUC (rAUC) of the same model was also processed after the substitution of the pseudo-absences by the correct virtual absences (see Fig. 3.6). To evaluate the influence of the number of available presences on the final model, we produced models for the GVS and SVS presence-only datasets with a single layer of nine neurons (same number as ENVs available). The LR value for the GVS was set to 0.1 for 25 and 50 presence locations and to 0.01 for 75, 100, 125, 150, 175 and 200 locations. The chosen LR value for the SVS was 0.6 for the dataset of 20 presence locations, 0.2 for the 40 presence and 0.1 for 60, 80 and 100 presence locations. The trAUC, teAUC, fAUC and rAUC values were calculated. As a positive control for the models, we produced models with presence-absence datasets from all the virtual species (Fig. 3.6), which allowed evaluating the effect of the pseudo-absences on final model quality. The RVS, with no relation with the environment, acts as a negative control, to assess the ability to fit only noisy data. The models of the real species followed the standard modelling procedures. The locations were input to Simapse with the ENVs to produce an averaged consensus prediction. We have chosen a structure with a single hidden layer with 9 neurons (same number as ENVs). The chosen LR was different for each species to maximize the gradient descent of the

network error. Therefore, for the *L. agilis*, *L. schreiberi* and *P. algirus*, the LR value was set to 0.7, 0.005, and 0.0005, respectively.

3.3.4 Results

Virtual species predictions

As expected, the three virtual species have different niche occupancies (table 3.2) resulting in different spatial extents. The GVS is widespread in southern Iberia (4612 cells), while the SVS occupies the mountainous northern areas (778 cells). The RVS is present throughout all study area (3764 cells) without any visible clustering (appendix B.1).

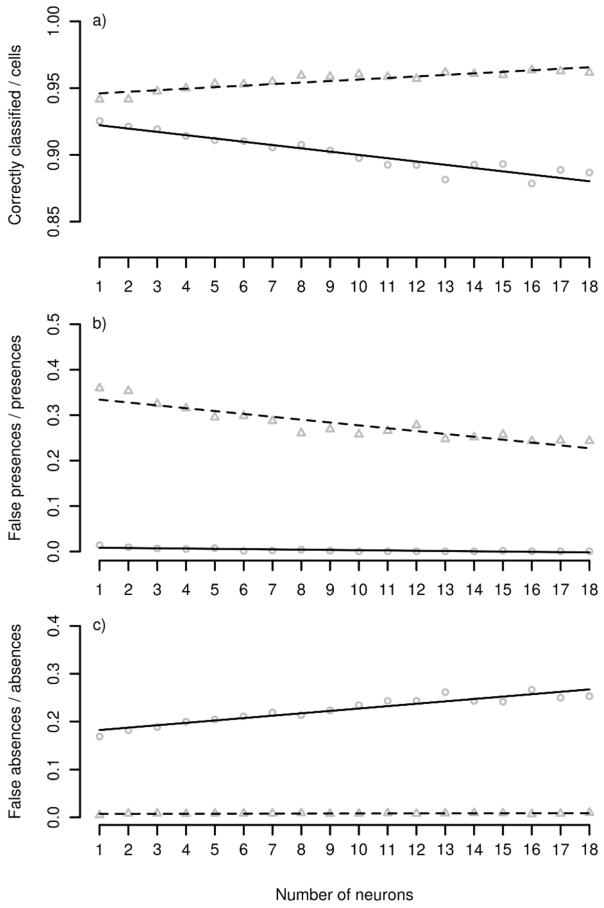


Fig. 3.7 – Effects of the number of neurons in the single hidden layer on network structure. The generalist virtual species (GVS) is represented by solid regression line and circles and the specialist virtual species (SVS) by dashed regression line and triangles. Ratios of cells correctly classified as presence or absence (a), false presences as presences (b), and false absences as absences (c) are given.

Network structure

Models for the GVS and the SVS achieved high correct classification rates (Fig. 3.7, appendix B.2) with the first exhibiting a slight decrease of the correct classification with increasing complexity of the network (slope=-0.002, R^2=0.89), whereas SVS shows a slight increase (slope = 0.001, R^2=0.81). The minimum value of correctly classified cells was 0.88 (GVS with 13 and 16 neurons) and the maximum was 0.96 (SVS with >8 neurons). The rates of false presences and false absences (fpr and far, respectively) also show different patterns between GVS and SVS. The first has low number of false presences among classified presences (fpr_{GVS}<0.015) and an increasing number of false absences with increasing network complexity (far_{GVS} = 0.17 with a single neuron, far_{GVS}=0.25 with 18 neurons; slope=0.005, R^2=0.89). The SVS has a decrease of false presences with the increasing number of neurons in the single layer (fpr_{SVS}=0.36 with single neuron, fpr_{SVS}=0.24 for neurons > 16; slope=-0.006, R^2=0.83) with low number of false absences (far_{SVS}<0.01).

The AUC values are stable with network structure changes. The rAUC had values higher than other AUC values for both GVS and SVS (Fig. 3.8a, 3.8b). The

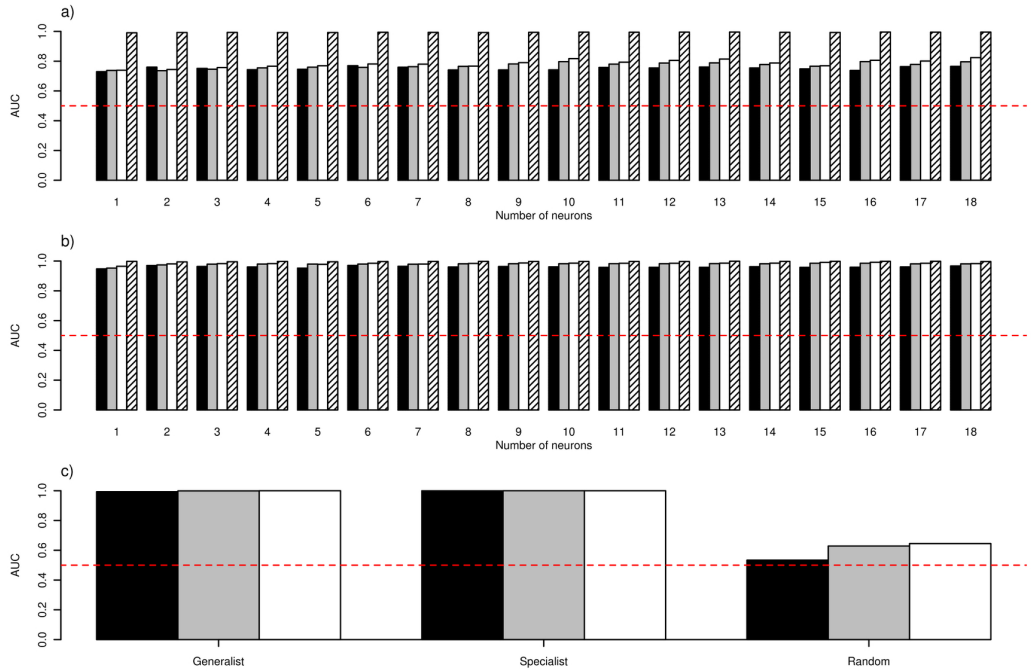


Fig. 3.8 – AUC values resulting from different network structures. Black bars represent the train AUC (trAUC), grey bars represent the test AUC (teAUC), white bars represent the final AUC (fAUC), and dashed bars represent the real AUC (rAUC). The dashed line limits the AUC = 0.5 or random predictions. a) GVS with presence-only data; b) SVS with presence-only data; c) models based on presence-absence data for GVS, SVS and RVS. Note that c) does only have three bars because fAUC = rAUC in the presence-absence models.

SVS achieves higher AUC values in all cases, with the exception of rAUC which was similar between both virtual species models (n=18; rAUC$_{GVS}$=0.994 $\pm$ 0.001; rAUCSVS=0.997 $\pm$ 0.001). The models built with presence-absence datasets revealed high accuracy for GVS and SVS (Fig. 3.8c) and AUC=0.993 was the lowest among all values (fAUC$_{GVS}$=fAUC$_{SVS}$=1.000; fAUC=rAUC due to the usage of real absences). The RVS models achieved AUC values near full randomness (trAUC$_{RVS}$=0.534, teAUC$_{RVS}$=0.639 and fAUC$_{RVS}$=0.645) with no separation between presences and absences in predicted classes of probability (Fig. 3.9).

Sample size

Both GVS and SVS models responded to the increment of available samples with an increase of correctly classified cells, resulting in more accurate average predictions (ccr$_{GVS}$=0.851 $\pm$ 0.084; ccr$_{SVS}$=0.927 $\pm$ 0.067; Fig. 3.10a). The average number of false presences (Fig. 3.10b) is near zero for GVS (fpr$_{GVS}$=0.006 $\pm$ 0.008) for all sample sizes, whereas for SVS is higher (fpr$_{SVS}$=0.356 $\pm$ 0.169) but with a decreasing trend with the increasing number of samples. The opposite pattern was observed for false absences (Fig. 3.10c), where GVS failed to predict a higher number of absences (far$_{GVS}$=0.288 $\pm$ 0.114), however decreasing with the availability of more sample points. SVS has a very low number of false presences irrespective of the sample size (far$_{SVS}$=0.006 $\pm$ 0.002).

The AUC obtained with different numbers of sampled points had high values

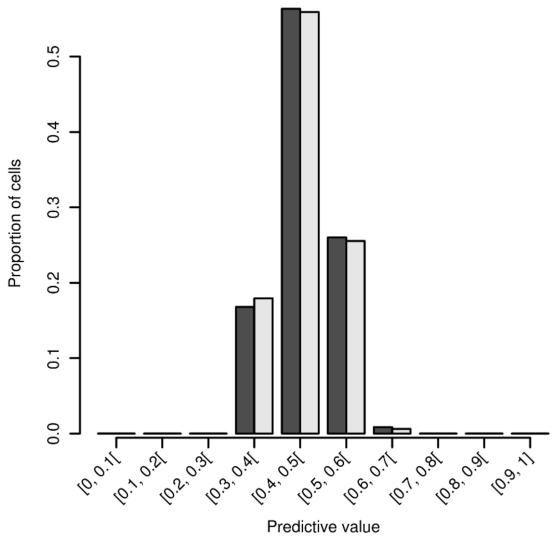


Fig. 3.9 – Proportion of presences and absences per predictive class for the RVS. Dark bars represent the virtual presences and light bars represent the virtual absences.

and exhibited different patterns for GVS and SVS (Fig. 3.11). The values of trAUC, teAUC and fAUC for GVS had a strong drop above 75 sample points, but the minimum values achieved (n=175; trAUC$_{GVS}$ = 0.727; teAUC$_{GVS}$ = 0.722; fAUC$_{GVS}$ = 0.749) were indicating a lack of randomness in model predictions. The SVS presents stable AUC values with a high minimum (n=60; trAUC$_{SVS}$ = 0.962; teAUC$_{SVS}$ = 0.907; fAUC$_{SVS}$ = 0.965). Models for both virtual species achieved very high values for rAUC in all studied number of sampled points (rAUC$_{GVS}$=0.980 $\pm$ 0.020; rAUC$_{SVS}$=0.991 $\pm$ 0.013).

Real species predictions

The models for the three real species accurately predicted the presence locations (Fig. 3.12). All presences fall in the highest prediction class (p>0.9) for the species *L. agilis* (fAUC$_{LA}$=1.000; Fig. 3.12a) and more than 70% of all non-presence cells re-

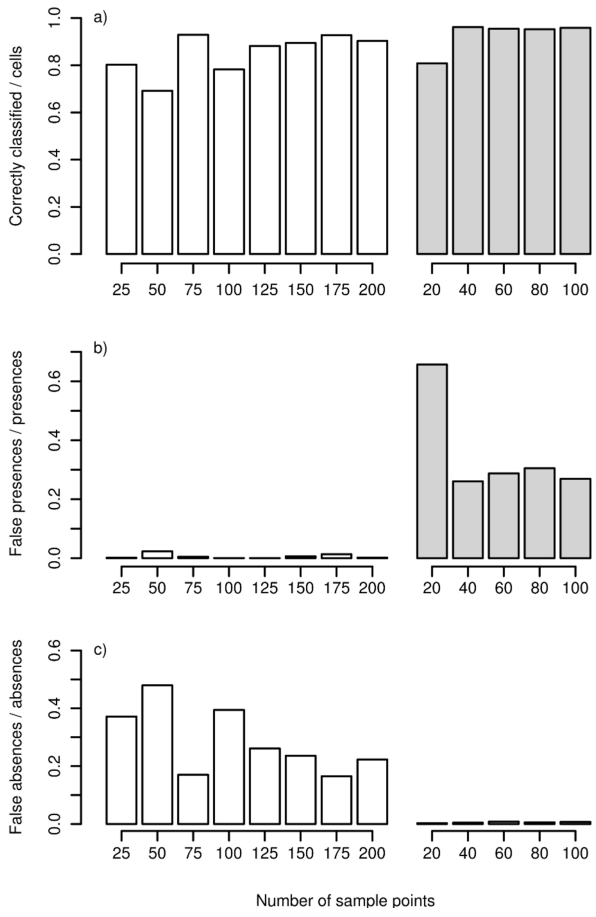


Fig. 3.10 – Sample size effects on model prediction accuracy. GVS is represented by white bars whereas SVS is represented by grey bars. The ratios of cells correctly classified as presence or absence (a), false presences as presences (b), and false absences as absences (c) are given.

mained with low presence probability (p<0.1). A similar pattern was observed in *L. schreiberi* (fAUC$_{LS}$=0.960; Fig. 3.12b), where more than 80% of the presences occurred in high presence probability cells (p>0.7) and 80% of the non-presence cells settled at low presence probability (p<0.3). For *P. algirus* (fAUC$_{PA}$=0.820; Fig. 3.12c), presence and non-presence cells were mixed. The presences were accurately predicted, with 88% laying in classes with p>0.7, and only 27% of the non-presence cells were predicted to have p<0.5.

3.3.5 Discussion

The overall pattern of the results shows a high accuracy of Simapse predictions of species distribution under distinct scenarios of ecological requisites by virtual and real species. This is not surprising due to the high predictive power of the artificial neural networks which make them suitable for the analysis of complex ecological relationships (Dimopoulos *et al.*, 1995; Lek *et al.*, 1996; Gevrey *et al.*, 2003; Segurado &

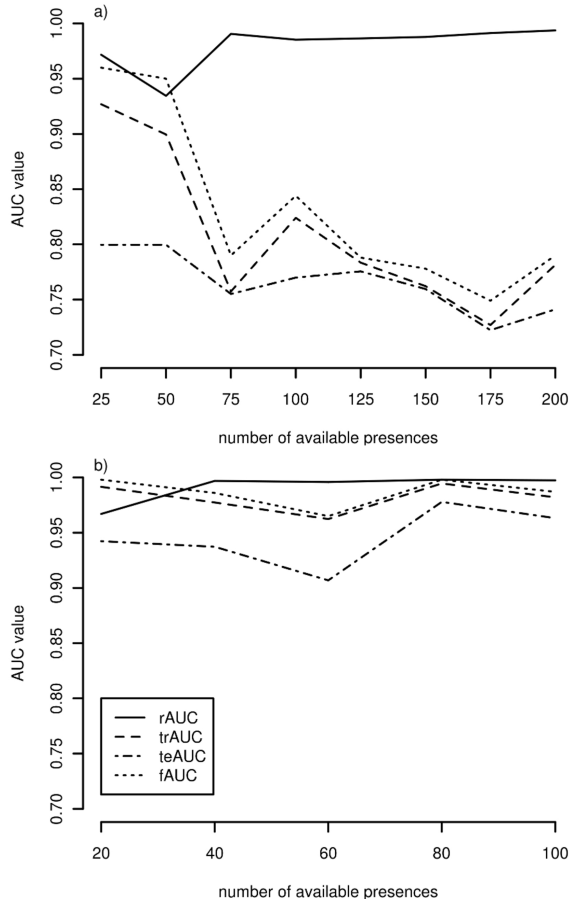


Fig. 3.11 – Sample size effects on AUC estimates for the GVS (a) and SVS (b).

Araujo, 2004). The present study extends the ability to both virtual and real species' data, with presence-absence comparable to presence-only modelling strategies.

General performance of ANN

Modelling with presence and absence data provided accurate models that separated both datasets in the predictive space (Fig. 3.8c and appendix B.1). Thus, AUC values near the maximum value were expected and together demonstrate the good dis-

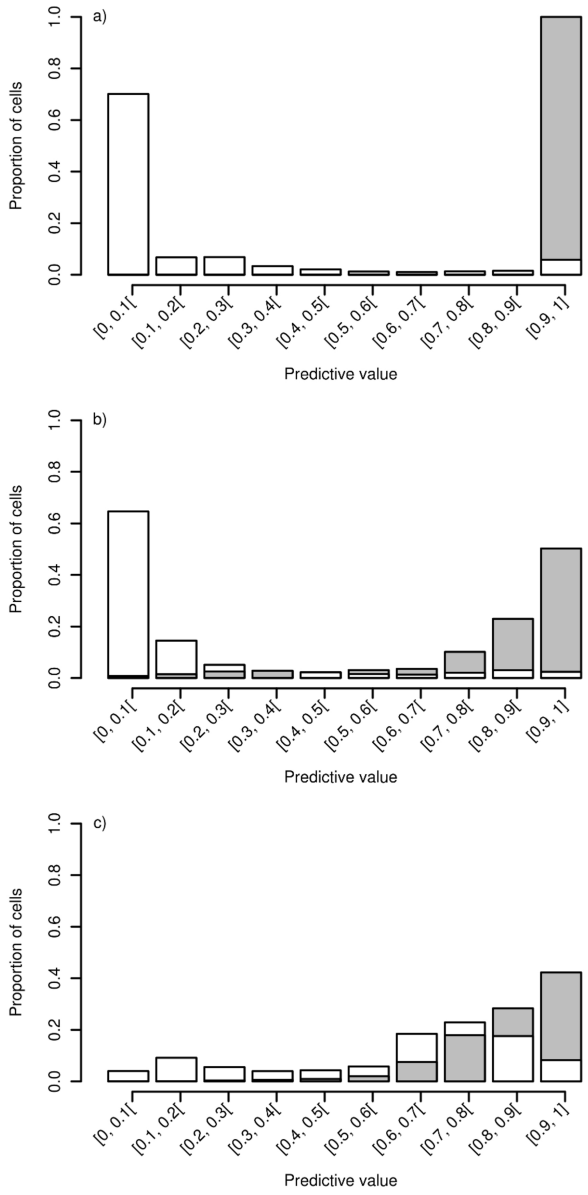


Fig. 3.12 – Proportion of presences and non-presences for the real species *Lacerta agilis* (a), *Lacerta schreiberi* (b), and *Psammodromus algirus* (c). Presences are represented by dark bars and non-presences by light bars.

crimination between presences and absences of the input data. In fact, most of the absences fall in the lowest predictive classes and most of the presences are in the highest predictive classes in both models for GVS and SVS (appendix B.2). The ability to learn complex data by Simapse greatly increases with trustful presences and absences. Here we have absolute confidence in the absences because we have access to the complete distribution of the virtual species. However, in real species, absences are very difficult to sample due to their nature and are usually associated with high uncertainty (Mackenzie & Royle, 2005; Lobo *et al.*, 2010).

This raises the usefulness of virtual species for primary evaluation of model fitness, with the necessary caution when extrapolating results to the real world (Zurell *et al.*, 2010). While such caution applies to models of virtual species with ecological sense, the same line of reasoning is not applicable to the RVS, where the model is expected to fail due to the lack of relation with the predictors. Despite their good performance with complex data, ANN do need reasonable relationships between species' presences and variables, as shown by the ineptitude to process the RVS presences. Most cells fall in the same predictive space when using a RVS, irrespective of being presence or absence in the original distribution (Fig. 3.9). This may seem contradictory with AUC results for the RVS (Fig. 3.8c) where fAUC > 0.5, pointing towards a modest success in the discrimination of presences and absences of the RVS distribution. However, Simapse chooses the best networks in each repetition by the least test error (i.e., the one that is less prone to overfit the train data), albeit the train results. In the case of RVS, the trAUC value is near the randomness value, indicating absence of learning and demonstrating that one should carefully check the training values for a trace of learning.

Species niche requisites and pseudo-absences

The GVS and SVS allowed to compare different modelling contexts and exhibited an expected dichotomous pattern. Though the GVS was more difficult to predict due to its broader usage of the ENVs resulting in widespread presences (60% of the total area; see table 3.2), the fitness of the model, when measured with real presences and absences, was extremely high. Pseudo-absences are a common surrogate to overcome the sampling difficulties and high uncertainty associated with absences (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Barry & Elith, 2006; Chefaoui & Lobo, 2008; Lobo *et al.*, 2010). Nevertheless, the selection of pseudo-absences locations may limit the position of the model outcome in the potential and realized niche gradient (Chefaoui & Lobo, 2008). Using random locations with a species with such large presence area,

may increase the probability of falling in a non-sampled presence, thus creating difficulties to the learning algorithm and consequently affecting the AUC values (Lobo *et al.*, 2007; Chefaoui & Lobo, 2008). On the other side, modelling with real presences and absences showed a great accuracy with a steep ANN learning path, as is common with presence/absence methods (Zaniewski *et al.*, 2002; Barry & Elith, 2006).

Models for the SVS (10% of the total area; see table 3.2) exhibited high accuracies. The high AUC values when compared to rAUC demonstrate an opposite picture for the GVS. The probability of a pseudo-absence to fall in a location with a real absence was higher, leading to an effective learning and, thus, higher AUC values. In fact, the differences between modelling with real absences or pseudo-absences are unnoticeable for the SVS when analysing the AUC values achieved with the different strategies.

The strategy of comparing models based on the same presence dataset using accurate virtual absences or pseudo-absences allowed to compare AUC values. The reduced uncertainty of the virtual absences was reflected in the stability of the rAUC, but randomly chosen absences resulted in fAUC oscillation. This exposes the relation of AUC values based on pseudo-absences with other factors like the relative area of occurrence of the species and the area where are pseudo-absences are generated. This topic has been subject of intense debate in recent literature (e.g. Engler *et al.*, 2004; Lobo *et al.*, 2007; Peterson *et al.*, 2007; Chefaoui & Lobo, 2008; Vanderwal *et al.*, 2009; Lobo *et al.*, 2010) but further discussion is beyond the objectives of the present research.

ANN parameters

The effect of network structure on the models was analysed by changing the number of neurons available in the single hidden layer. The classification success of the predictive models was very high with all tested network architectures. Interestingly, simple networks, i.e. lower number of neurons, achieve a classification success comparable to more complex networks with the tested datasets. Although the process of selecting network architecture is usually heuristic, modellers should keep the simplest network as it is less prone to overfitting and, thus, is able to predict better in new regions (Dimopoulos *et al.*, 1995; Özesmi & Özesmi, 1999; Özesmi *et al.*, 2006). The problem of overfitting in very complex networks is related to the trend of noise fitting and leading to a lack of predictability (Özesmi & Özesmi, 1999). The slight decrease of predictive success with increasing complexity of the network observed in the GVS reflects this problem: its intricate pattern of presences and pseudo-absences is fitted

along with noise when using a higher number of neurons, inflating the number of false absences. On the other hand, the restricted range and specificity to a few ENVs of the SVS make it less prone to noise presence. Modelling in this case, is very effective with a small increase of predictability accompanying the increasing complexity of the network structure.

Sample size

Modelling the GVS and SVS with different number of presences resulted in high classification rates, even with the lowest number of presences. Even so, additional presences are extremely informative, as shown by the decrease of false presences and false absences in the SVS and GVS, respectively (Fig. 3.10) but also by the decrease of spatial uncertainty (appendix B.2). The increase of the available samples is usually directly proportional to the accuracy of the predictions (Barry & Elith, 2006; Hernandez *et al.*, 2006; Wisz *et al.*, 2008) and the extra information contained in additional presences allows to better depict the relations between species' locations and the predictors. The increasing complexity of these relations are data hungry (Barry & Elith, 2006) as seen by the unstable results of the GVS with low number of samples or, on the other side, the rapid achievement of very high correct classification values with small sample size by the SVS. The increasing number of presences in this study has a parallel increase in number of pseudo-absences and, as seen before, the presence uncertainty related to these extra locations in GVS results in lower AUC values when measured with the generated pseudo-absences (Fig. 3.11), even in the presence of accurate models.

Real species

The final step in the algorithm testing was to submit it to the scrutiny of real data. The predictive models for the three species were very accurate, with well defined presences and absences for *L. agilis* and *L. schreiberi* (Fig. 3.12a, 3.12b). Although the latter has a broader extent, both species have specific niche requirements and the results are similar to those of the SVS. In this case of a very restricted distribution in a larger region of study, generalization of the niche to other areas occurs as a side effect of using pseudo-absences (Engler *et al.*, 2004; Vanderwal *et al.*, 2009; Lobo *et al.*, 2010). This results in very accurate prediction of the presences (all presences were correctly predicted) but areas where the species was not detected also have high values of prediction (i.e. high commission errors).

Regarding *P. algirus*, with a large number of presences widespread throughout

the study area, we achieved less modelling success with an overlap of presence and non-presence area in the predictive space (Fig. 3.12). Most of the non-presence areas resulting from an atlas with intensive and systematic field work can be accepted as a highly probable absence. However, real species distributions are complex, as opposed to the more simple relationships of the virtual world, and despite the importance of the environment as a shaping factor, there are a multitude of other effects, like human and other sources of natural disturbance, limiting or extending the occurrence of species (Barry & Elith, 2006). In this study we do not use a human related predictor or other source of potential disturbance, therefore some of these absences fall in the high predictive space, hampering the detection of those gaps.

3.3.6 Conclusions

The use of virtual and real data allowed testing Simapse with knowledge of the full, though simplistic, virtual reality and the more complex relationships of the real world. The predictive performance of the ANN when modelling both datasets revealed to be extremely high. We demonstrated the efficacy of the algorithm with simple architectures and with an extensive range of sample sizes. When absences are available with high confidence, presence and absence modelling strategies should be prefererred over presence-only data.

The framework we developed here extends previous methods for assessing ENM performance and we expect that it will be used in future algorithm comparisons.

3.3.7 Acknowledgments

PT and JCB are funded by Fundação para a Ciência e Tecnologia: SFRH/BD/42480/2007 and Programa Ciência 2007, respectively.

3.3.8 References

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**, 1712–1728.

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

Banks, W., Derrico, F., Peterson, a., Vanhaeren, M., Kageyama, M., Sepulchre, P., Ramstein, G., Jost, a. & Lunt, D. (2008) Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. *Journal of Archaeological Science* **35**, 481–491.

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

Brito, J., Fahd, S., Geniez, P., Martínez-Freiría, F., Pleguezuelos, J. & Trape, J.F. (2011) Biogeography and conservation of viperids from North-West Africa: An application of ecological niche-based models and GIS. *Journal of Arid Environments* **75**, 1029–1037.

Brito, J.C., Acosta, A.L., Álvares, F. & Cuzin, F. (2009) Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African canids. *Biological Conservation* **142**, 3020–3029.

Carvalho, S.B., Brito, J.C., Crespo, E.G., Watts, M.E. & Possingham, H.P. (2011a) Conservation planning under climate change: Toward accounting for uncertainty in predicted species distributions to increase confidence in conservation investments in space and time. *Biological Conservation* **144**, 2020–2030.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2010) From climate change predictions to actions - conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology* **16**, 3257–3270.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2011b) Incorporating evolutionary processes into conservation planning using species distribution data: a case study with the western Mediterranean herpetofauna. *Diversity and Distributions* **17**, 408–421.

Chefaoui, R.M. & Lobo, J.M. (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* **210**, 478–486.

Dimopoulos, Y., Bourret, P. & Lek, S. (1995) Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters* **2**, 1–4.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., 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.

Elith, J. & Leathwick, J.R. (2009) Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697.

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.

Ferrier, S. (2002) Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology* **51**, 331–63.

Gevrey, M., Dimopoulos, I. & Lek, S. (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* **160**, 249–264.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006) Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* **43**, 386–392.

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

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

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.

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

Hirzel, A., Helfer, V. & Metral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling* **145**, 111–121.

Keitt, T.H., Bivand, R., Pebesma, E. & Rowlingson, B. (2010) rgdal: Bindings for the Geospatial Data Abstraction Library.

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.

Lek, S. & Guégan, J.F. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* **120**, 65–73.

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

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

Loureiro, A., Ferrand, N., Carretero, M.A. & Paulo, O. (2008) *Atlas dos Anfíbios e Répteis de Portugal*. Instituto de Conservação da Natureza e Biodiversidade, Lisboa.

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

Martínez-Freiría, F., Santos, X., Pleguezuelos, J.M., Lizana, M. & Brito, J.C. (2009) Geographical patterns of morphological variation and environmental correlates in contact zones: a multi-scale approach using two Mediterranean vipers (Serpentes). *Journal of Zoological Systematics and Evolutionary Research* **47**, 357–367.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* **83**, 171–193.

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

Özesmi, S., Tan, C. & Özesmi, U. (2006) Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling* **195**, 83–93.

Park, Y. & Chon, T. (2007) Biologically-inspired machine learning implemented to ecological informatics. *Ecological Modelling* **203**, 1–7.

Peterson, A.T., Papeş, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**, 550–560.

Pleguezuelos, J.M., Márquez, R. & Lizana, M. (2004) *Atlas y libro rojo de los anfibios y reptiles de España*. Organismo Autónomo de Parques Nacionales.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rebelo, H., Tarroso, P. & Jones, G. (2010) Predicted impact of climate change on European bats in relation to their biogeographic patterns. *Global Change Biology* **16**, 561–576.

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

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

Tarroso, P., Carvalho, S.B.S. & Brito, J.C. (2012) Simapse - Simulation Maps for Ecological Niche Modelling. *Methods in Ecology and Evolution* **3**, 787–791.

Tomović, L., Crnobrnja-Isailović, J. & Brito, J.C. (2010) The use of geostatistics and GIS for evolutionary history studies: the case of the nose-horned viper (*Vipera ammodytes*) in the Balkan Peninsula. *Biological Journal of the Linnean Society* **101**, 651–666.

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

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

Xavier, R., Lima, F.P. & Santos, A.M. (2010) Forecasting the poleward range expansion of an intertidal species driven by climate alterations. *Scientia Marina* **74**, 669–676.

Zaniewski, A.E., Lehmann, A. & Overton, J.M.C. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* **157**, 261–280.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos* **119**, 622–635.

Chapter 4

Reconstruction of past Iberian climate

> The world, my friend Govinda, is not imperfect or confined at a point somewhere along a gradual pathway toward perfection. No, it is perfect at every moment.
>
> — HERMANN HESSE, *Siddhartha*

4.1 Spatial climate dynamics in the Iberian Peninsula since 15 000 Yr BP. [1]

4.1.1 Abstract

The climate in the Iberian Peninsula has evolved since the last glacial maximum, promoting distributional shifts of most species. This relation between glacial and postglacial climate and species composition is used here to quantify the climate change in the Iberian Peninsula using several fossil pollen records widespread throughout the study area. We have reconstructed spatial layers (1ka interval) of January minimum temperature, July maximum temperature and minimum annual precipitation using a method based on probability density functions and covering the time period between 15ka and 3ka. In order to evaluate the spatial evolution of climate, we used a functional principal component analysis. Using a clustering method we have identified areas that share similar climate evolutions during the studied time period. The spatial reconstructions show a highly dynamic pattern in accordance with climatic events. The four

[1] This study is presently submitted to an international journal listed in the science-citation index: Tarroso, P., Carrión, J., Dorado-Valiño M., Queiroz, P., Santos, L., Valdeolmillos-Rodríguez, A., Alves, P. C., Brito, J. C., & Cheddadi, R. Spatial climate dynamics in the Iberian Peninsula since 15 000 Yr BP.

cluster areas we found exhibit different climate evolution over the studied period. The clustering scheme is compatible with multiple refugial areas.

4.1.2 Introduction

The pattern of present biodiversity is the result of a dynamic process driven at a broad temporal scale by geological events and climatic oscillations. Currently, several threats to biodiversity are identified and the largest implications are distributional shifts (Parmesan & Yohe, 2003) and, more dramatically, species extinction (Hewitt, 2000). The biodiversity hotspots retain high levels of endemism and are considered as the best candidates for preserving species diversity in the future (Myers *et al.*, 2000). The Mediterranean basin was identified as a biodiversity hotspot and palaeoenvironmental studies showed that some areas within this region have played the role of refugia to diverse ecosystems over several hundreds of millenia (Wijmstra, 1969; Wijmstra & Smith, 1976; Van der Wiel & Wijmstra, 1987a,b; Tzedakis *et al.*, 2002). Often, those areas where species have persisted during glacial periods are referred to as glacial refugia (Bennett & Provan, 2008; Carrión *et al.*, 2010a; Hewitt, 2000; Hu *et al.*, 2009; MacDonald *et al.*, 2008; Willis *et al.*, 2010) and the high levels of diversity found at species level is corroborated at molecular level (Hewitt, 2000; Petit *et al.*, 2003). Understanding the past processes that affected biodiversity is fundamental for species conservation decisions facing the future expected global climate changes (Anderson *et al.*, 2006; Willis *et al.*, 2010).

Refugia have been generally defined based on species survival with an evident relationship to climatic components (Hewitt, 2000; Bennett & Provan, 2008; Cheddadi & Bar-Hen, 2009; Médail & Diadema, 2009), nevertheless, the term has been used recently with multiple definitions (Bennett & Provan, 2008; Ashcroft, 2010). The biological definition of refugia is related to the physiological limits of a species that under an increasingly stressing environment are forced to shift to suitable areas. Palaeoenvironmental data and molecular analysis have proven useful to locate species ancestral distributions and migration routes (Petit *et al.*, 2003; Cheddadi *et al.*, 2006). However, the locations and extension range of putative refugia still lack spatial consensus and a quantification of its dynamic nature. Fossil pollen is an appropriate proxy for quantitative reconstruction of past climate variables (Webb *et al.*, 1993; Cheddadi *et al.*, 1997; Guiot, 1997; Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009). Using proxy data to derive a definition of refugia in terms of climate, which has an obvious spatial context, may provide a suitable answer to this problem.

Climate oscillations in Europe during the last 15,000 years exhibited latitudinal

and longitudinal variations (Cheddadi *et al.*, 1997; Davis *et al.*, 2003; Roucoux *et al.*, 2005; Cheddadi & Bar-Hen, 2009; Carrión *et al.*, 2010a). During the last glacial maximum (LGM), several species found refugia in the southern peninsulas (Hewitt, 2000; Tzedakis *et al.*, 2002; Petit *et al.*, 2003; Weiss & Ferrand, 2007; Bennett & Provan, 2008; Hu *et al.*, 2009; Médail & Diadema, 2009; Ohlemüller *et al.*, 2012). The Iberian Peninsula, with a milder climate than the northern European latitudes (Renssen & Isarin, 2001; Carrión *et al.*, 2010a; Pérez-Obiol *et al.*, 2011) served as a refugium to several species that persisted in this area during the LGM. The current patterns of high diversity in the Iberian Peninsula derive partially from this role during harsh glacial conditions and it is the main reason why this peninsula is considered as an important part of the Mediterranean biodiversity hotspot (Médail & Quézel, 1999; Cox *et al.*, 2006). Although it may be seen globally as static refugia, the vegetation and climate dynamics in Iberia reveal a quite complex picture. The abrupt climate changes that occurred after the LGM (Heinrich event 1, Bølling-Allerød interstadial and the Younger Dryas) did not impact uniformly across the peninsula but rather with distinct velocities and patterns as shown by palynologic data (Roucoux *et al.*, 2005; Naughton *et al.*, 2007; Pérez-Obiol *et al.*, 2011) with consequences on biodiversity patterns as described by molecular studies (Branco *et al.*, 2002; Godinho *et al.*, 2006; Weiss & Ferrand, 2007; Miraldo *et al.*, 2011). The Iberian Peninsula is, therefore, a unique area to study the climate dynamics during the late-Quaternary, with immense variation and of paramount importance in terms of biodiversity conservation.

Our main objective in this study is to define areas within the Iberian Peninsula (Balearic Islands included) that share similar climate evolution and susceptible to serve as a potential refugium area to temperate trees. We reconstructed spatially three climate variables and quantified their changes between 15 ka and 3 ka BP, with a 1,000 years interval. Then, using robust statistical methods, we defined geographical areas that have undergone similar climate changes and analysed their spatial dynamics throughout the late Quaternary.

4.1.3 Methods

The area for the spatial reconstruction extends throughout the land area of the Iberian Peninsula and the Balearic islands (Fig. 4.1). The method used to produce past climate grids is based on probability density functions (PDF) and requires both fossil pollen records and full distribution of modern plant taxa. The raw fossil pollen data were gathered from author's contribution and from the European Pollen Database (`www.europeanpollendatabase.net`). We checked each site to fit a quality criteria

regarding the number of radiometric datations and sampling resolution of the pollen records. Using these criteria we selected a total of 31 records which cover different time spans between 15000 and 3000 years BP; table 4.1; Fig. 4.1). The last 3000 have been strongly impacted by human activities in the Iberian peninsula and therefore the fossil data may be biased and may lead to some potential misinterpretations.

Data sources

The current distribution data for 246 taxa was obtained by georeferencing the Atlas of Flora Europaea (Laurent *et al.*, 2004). We gathered additional occurrence data for the Mediterranean flora from the Global Biodiversity Information Facility data portal (`data.gbif.org`; last access 2011-02-01). The georeferenced geographical distributions were then rescaled to the resolution of 30' (∼55 km). The global historic climate data (1950-2000) for January minimum temperature (Tjan), July maximum temperature (Tjul) and monthly precipitation data were obtained from Worldclim database (`www.worldclim.org`; Hijmans *et al.*, 2005) with 5' resolution (∼10km) and values were

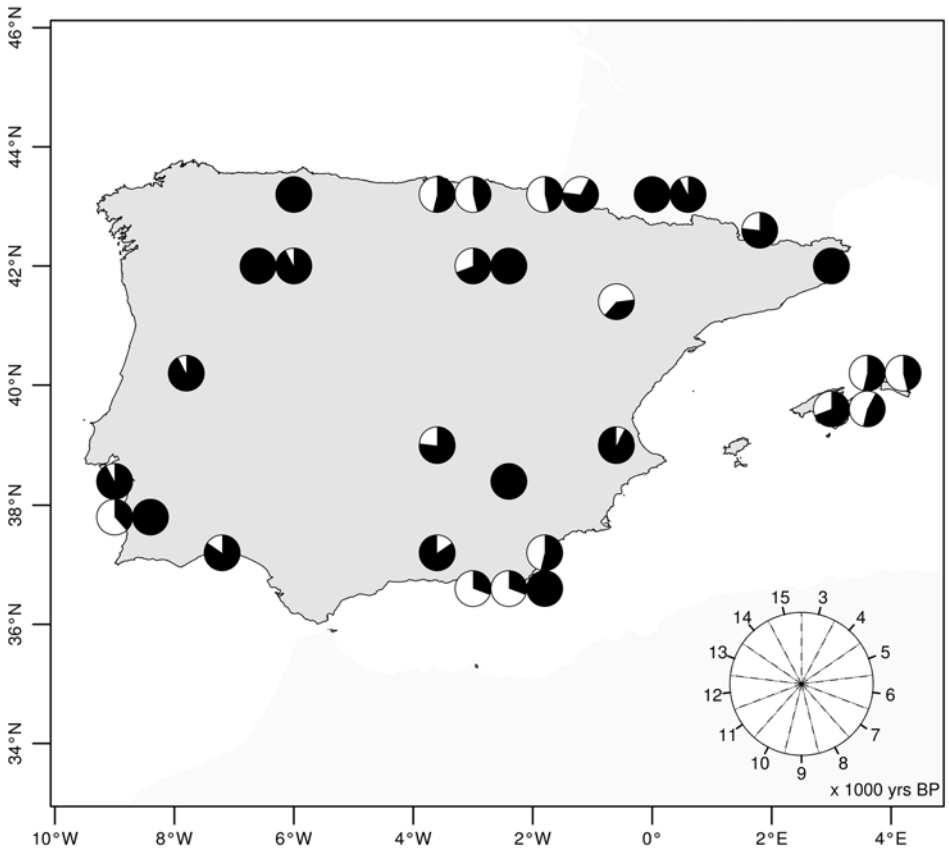


Fig. 4.1 – Study area with sample points. The black area inside each circle represents the ages available in each pollen sequence.

Table 4.1 – Origin and description of the data sources of fossil pollen used to reconstruct the climate in the Iberian Peninsula.

Name	Latitude	Longitude	Source	Reference
Albufera Alcudia	39.793	3.119	EPD	
Algendar	39.941	3.959	EPD	
Antas	37.208	-1.824	EPD	
Barbaroxa	38.071	-8.809	author's contribution	Queiroz 1999
Cala Galdana	39.937	3.965	EPD	
Cala n' Porter	39.871	4.131	EPD	
CC-17	39.08	-3.87	author's contribution	Dorado-Valiño et al. 2002
Charco da Candieira	40.342	-7.576	EPD	
Gádor	36.9	-2.917	author's contribution	Carrión et al. 2003
Golfo	38.561	-9.135	author's contribution	Queiroz 1999
Guadiana	37.267	-7.45	author's contribution	Fletcher 2007
Hoya del Castilho	41.25	-0.5	EPD	
Lago de Ajo	43.05	-6.15	EPD	
Laguna de la Roya	42.217	-6.767	EPD	
Lake Racou	42.554	2.008	EPD	
Las Pardillas Lake	42.03	3.03	author's contribution	Sanchez-Goñi and Hannon 1999
Navarrés 1	39.1	-0.683	EPD	
Puerto de Belate	43.033	-2.05	EPD	
Puerto de Los Tornos	43.15	-3.433	EPD	
Quintanar de la Sierra	42.033	-3.017	EPD	
Roquetas de Mar	36.794	-2.589	EPD	
Saldropo	43.05	-2.717	EPD	
Sanabria Marsh	42.1	-6.733	EPD	
San Rafael	36.773	-2.601	EPD	
Santo André	38.08	-8.78	author's contribution	Santos and Sanchez-Goñi 2003
Siles	38.4	-2.5	author's contribution	Carrión 2002
Padul	37	-3.667		Pons and Reille 1988
Lourdes	43.033	-0.075		Reille and Andrieu 1995
Monge	43.05	-0.017		Reille and Andrieu 1995
Moura	43.45	-1.55		Reille, 1993
Banyoles	42.133	2.75	EPD	

aggregated to the resolution of 30'. Precipitation was further processed to obtain the minimum annual precipitation (Pmin) from the monthly data.

Reconstruction of past climate variables

The climate reconstruction method is based on the PDF of each taxon identified in a pollen assemblage. This approach has already proven successful to reconstruct temperatures from fossil pollen data (Kühl *et al.*, 2002; Cheddadi & Bar-Hen, 2009). This method relies on the PDF intersection of the taxa that are present in a fossil pollen sample to infer the reconstructed climate amplitude. The area of each PDF

available for intersection is defined after applying a threshold limit based on the pollen percentages. A univariate kernel estimator was used to build the single dimension PDF from the modern plant distributions and the Tjan, Tjul and Pmin data. Kernel estimators and parametric estimators provide similar results when fit with a normally distributed variable (Kühl *et al.*, 2002), however, kernel estimators follow closely the real distribution of species with non-normal distributions.

Fluctuations in pollen abundances are related to several factors (Hicks, 2006) including the physiology of the species, but have also a strong climatic component through the influence that climate has on land cover. We used the pollen percentage as a measure of the taxon abundance throughout the time span of the core. Thus, a minimum pollen percentage corresponds to the presence while the maximum defines the highest abundance of the taxon. Using pollen percentages relative to each taxon eliminates the bias of different pollen production by distinct taxa and thus allows quantifying the presence of a species relative to its maximum detection within each site.

The PDF is filtered using the pollen percentage as a threshold. We assume that higher pollen percentages reveal near optimum conditions for the taxon that correspond to the area of the PDF where the taxon is found at higher densities. Restricting the PDF probability area based on the relation of density of taxa and pollen percentage results in narrower range of reconstructed values. The threshold influence on the PDF is controlled by a coefficient: “0” results in no effect of the pollen percentage and uses the full profile for taxa intersection; whereas "1" will likely result in a very narrow area of the PDF around highest densities at maximum pollen detection. We used a value of 0.5 which, at maximum detection of the taxa, constrains the PDF to the area where densities are higher than 50% of the maximum density observed (Fig. 4.2).

The reconstructed climate value is obtained by the weighted mean of the intersection of the PDF area of all taxa identified in a sample. A coefficient with the maximum value of 1 was added to control the number of species in the intersection. For Tjan and Tjul this value was set to 0.95 which means that the intersection area is defined by, at least, 95% of the maximum number of species intersecting, whereas for Pmin this value was set to 0.5. The weights used to find the average reconstruction are the union of the densities found in the intersection space. Since the intersection area takes into account the number of taxa present, the union of the PDFs was preferred to other possible weights because it eliminates the increasing influence of multiple taxa with similar PDFs. The product of the PDFs, in this case, would bias the reconstructed value towards values with similar profiles. This is more striking in the case

of precipitation, which is very challenging to reconstruct because xerophyte taxa are less frequent than taxa supporting average to high precipitation, but their presence is clearly indicative of more arid environments. Moreover, taxa well adapted to dryness in the Mediterranean climate have highest densities with some precipitation, adding a bias towards higher values of precipitation and a misrepresentation of precipitation absence. Increasing the number of available arid taxa in the reconstruction might help detecting lower values, but this will also be limited by the resolution of taxa identification in the pollen analysis and by the scarce distribution data of Mediterranean taxa.

The reconstructed values for each site were fitted with a smoothing spline to produce a continuous time-series.

Spatial analysis of past climate

Thirteen climate grids, ranging from 15 to 3 thousand calendar years BP (hereafter, "ka") with a 1 ka time interval, were obtained for each reconstructed variable by spatial interpolation of the climate anomalies at each available site. The anomalies were

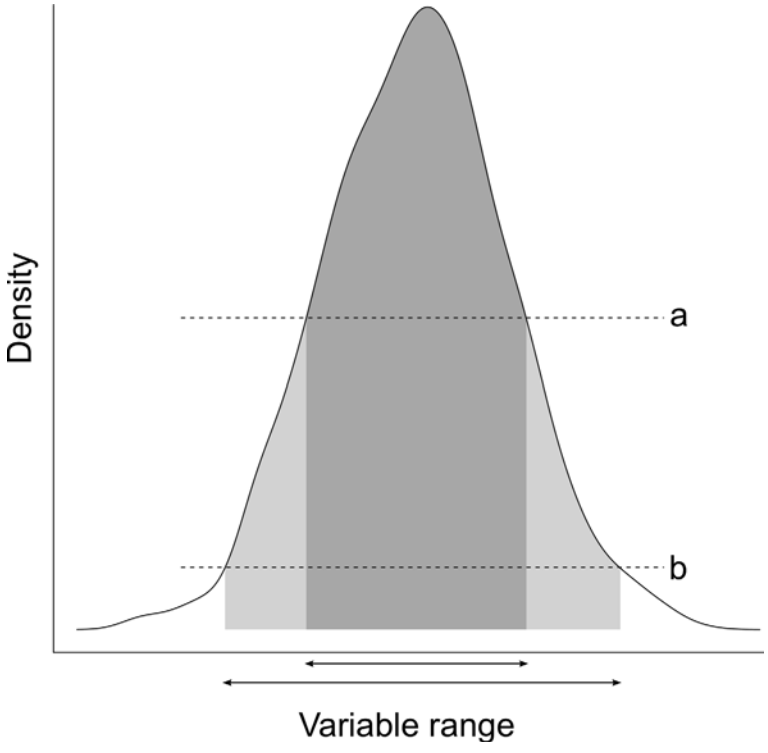


Fig. 4.2 – Example of the influence of pollen abundance on the PDF. The area of the PDF is filtered using the pollen thresholds related to the proximity of core distribution centres: higher pollen percentages (a) are assumed to originate near the distribution core and, thus, have higher densities and lower pollen percentages (b) are originated farther from the core and from optimum conditions. This results in different ranges of the variable that are related to the different densities.

first calculated at 30' resolution by subtracting the current climate variable from the reconstructed one for each time slice and each site. We interpolated these values onto a 5' resolution grid using thin-plate smoothing splines.

To further summarize the spatial and time variability of the data we applied a functional principal component analysis (fPCA). This method extends the exploratory data analysis of the principal components analysis to functional data (Ramsay & Silverman, 2005), depicting both spatial and time patterns on the original data. Cheddadi & Bar-Hen (2009) applied a fPCA in nearly the same timescale as the present study to depict January temperature patterns in European pollen sites. Here we have broadened the approach to each climate time-series available in each grid cell to produce gridded spatial components. The functional data was built by combining B-spline basis functions to fit the time-series. We have retained the components that explain more than 90% of the variance and rescaled the range from -1 to 1. We used hierarchical cluster analysis over the produced first components grids of each variable to identify areas in the Iberian Peninsula that shared similar climate evolution over the past 15 ka.

All analysis were performed using The R Project for Statistical Computing (R Development Core Team, 2012) with packages fields (Furrer *et al.*, 2012), rgdal (Keitt *et al.*, 2012), gstat (Pebesma, 2004) and fda (Ramsay *et al.*, 2012).

4.1.4 Results

The reconstruction of three climate variables exhibited high spatial variability over the period between 15 ka and 3 ka (Fig. 4.3, appendix C.1). The Iberian Peninsula had extensive areas with extremely low Tjan that gradually increased to warmer values, and markedly after 10 ka. Conversely, the pattern of Tjul over the same time span remained quite stable, with extended areas of lower values at 12, 11 and 5 ka. Areas with the highest reconstructed temperatures were found between 15 and 13 ka at the southern plateau. Between 15 and 7 ka the areas with increasing values of Pmin have expanded, with the exception between 12 and 11 ka, when extensive areas with very low precipitation values were observed. After 6 ka, a general trend towards aridification of the Iberian Peninsula was observed, with extended areas with lower Pmin than the early Holocene.

The clustering of the first fPCA component of the three reconstructed variables identified areas with congruent spatial structure (Fig. 4.4), and allowed summarizing the evolution of these three climate variables in the Iberian Peninsula (Fig. 4.5, appendix C.2). The first two components of each variable explained more than 90% of

the variation (Tjan: 75.0%, 13.6%; Tjul: 91.2%, 6.4%; Pmin: 87.6%, 7.2%, for the first and second components, respectively). The cluster C1 (22% of the total area) is located mostly on north-eastern Iberia and includes the Iberian mountain ranges but also areas with low altitude. It is the coldest area in the late-Quaternary, with Tjan

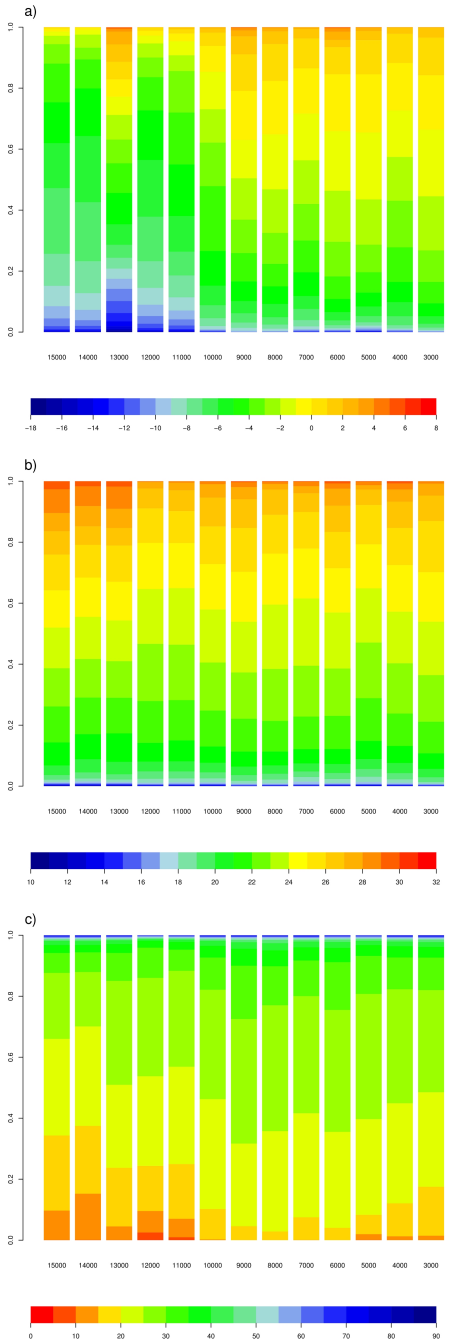


Fig. 4.3 – Distribution of the reconstructed climate variables in the Iberian Peninsula and Balearic Islands in the last 15 ka. Colours show the proportion of area covered with each class of a) minimum temperature of January; b) maximum temperature of July and c) minimum annual precipitation.

from -10.6°C to -4.3°C, Tjul ranging between 20.3°C and 21.2°C and the wettest with Pmin between 32mm and 41mm. The cluster C2 (26% of the total area) occupies most of the southern Iberian Plateau. It holds the warmest values for Tjan varying between -4.8°C and 1.0°C and Tjul between 26.1°C and 27.7°C. The dissimilarities between clusters C3 and C4 (each 26% of the total area) occur in Tjan, having the former higher temperatures (-4.9°C to -0.1°C) than the latter (-6.6°C to -1.5°C). Pmin is similar within the C2 and C4 (19 to 25mm and 19 to 26mm, respectively) while C3 had higher values (23 to 28mm).

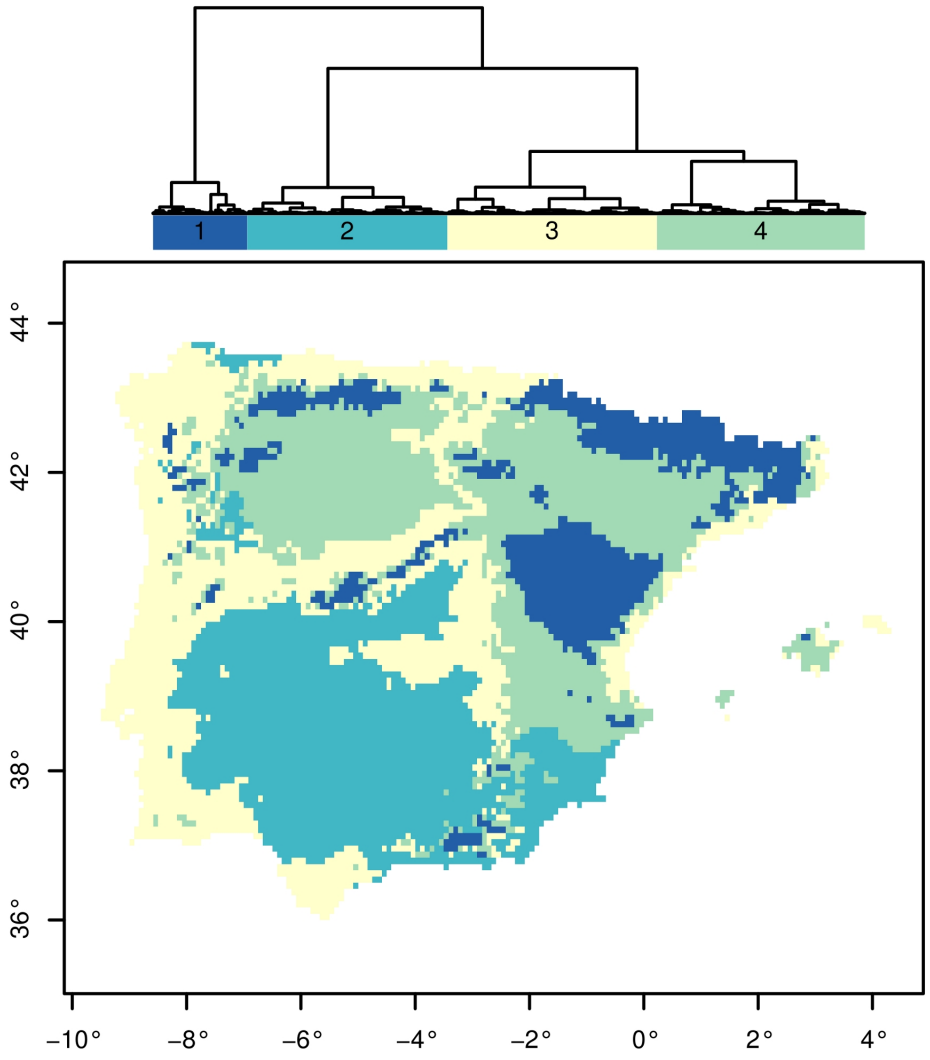


Fig. 4.4 – Hierarchical cluster analysis of the functional PCA components of Tjan, Tjul and Pmin in the last 15ka found in the study area. The top dendrogram represents the size of the clusters of similar climate evolution and the relations between them. Numbers correspond to the cluster.

4.1.5 Discussion

The climate of the last 15 ka in the Iberian Peninsula was dynamic, with oscillations of temperature and precipitation that had a major impact on the location, extent and evolution of the glacial refugia during the post-glacial period. Nonetheless, the reconstructed overall trend is a noticeable warming after the 15 ka that results from the increase of the summer insolation in the northern hemisphere (Berger, 1978). An evident pattern that strikes from the results presented here is the homogeneous Tjan throughout the Iberian Peninsula at 15ka, with 2°C difference between clusters. This pattern evolved towards a widening of the temperature range during the Holocene. Despite the stability within each cluster, Tjul and Pmin markedly divide the peninsula since 15ka (Fig. 4.5, appendix C.1).

Fossil pollen data provide a record of vegetation changes which constitutes a valuable proxy for reconstructing past climate changes. The major constrain we found

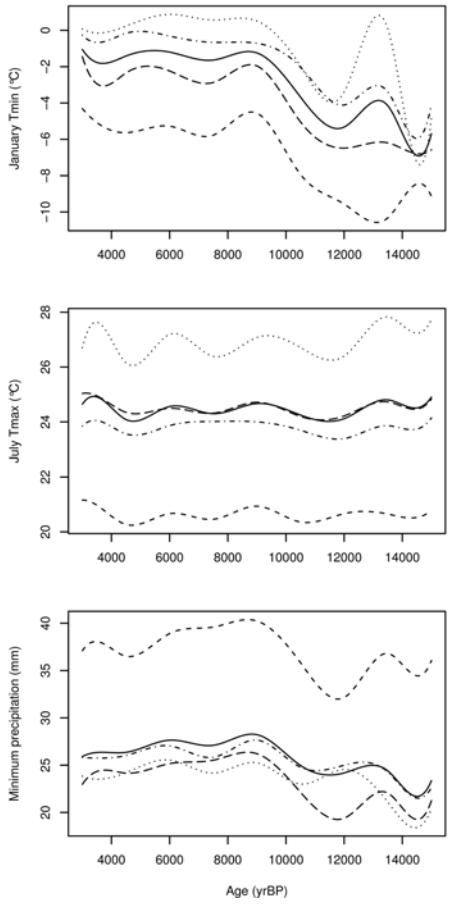


Fig. 4.5 – Minimum and maximum temperatures of January and July, respectively, and minimum annual precipitation during the last 15 ka. The solid line represents the average climate in the study area. The remaining lines are the average of each cluster found: C1: short dash line; C2: dotted line; C3: dash-dot line and C4: long dashed line.

in the Iberian Peninsula was the low number of sequences available according to our quality criteria for a robust spatial climate reconstruction, both in terms of sampling resolution and number of C^{14} dates. Nevertheless, the resulting climate reconstruction is consistent with other studies (Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009), placing the Iberian Peninsula in a particular case in Europe in terms of sensitivity to climatic stadials and interstadials like the Oldest Dryas (OD), Bølling-Allerød interstadial (BA) and Younger Dryas (YD).

This study shows that Tjan and Tjul had different patterns over the last 15ka. We identify areas where climate had different patterns and also depict regions of similar climate conditions throughout the last 15 ka. Although we have reconstructed climate with a 1 ka period, the climate stadials and interstadials have a spatial imprint that is visible in the spatial time-series. The OD ($\sim$18 to 14.7 ka) is characterized in Iberia by a vegetation changes compatible with cold and humid conditions followed by a warming trend (Naughton *et al.*, 2007). The OD is followed by the warmer BA ($\sim$14.7 to 12.9 ka). Our results show a similar pattern, with colder condition between 15 ka and 14 ka, followed by a warming trend until 13 ka, with different clusters exhibiting different sensibility to these fluctuations (Fig. 4.5), and showing extreme contrasts for Tjan (Figs. 4.3, 4.5). Precipitation values are low between 15 ka and 14 ka except in mountainous areas, comprised mostly in the first cluster, which remain relatively more humid than the other areas (Figs. 4.3, 4.5).

As described earlier in Europe (Renssen & Isarin, 2001; Heiri *et al.*, 2004), Tjan shows more abrupt changes than Tjul. The cold to warm transitions that occurred at $\sim$14.7 and 11.5 ka (Renssen & Isarin, 2001; von Grafenstein *et al.*, 1999) in Europe had a spatial impact that is well depicted in Tjan. The increasing variability of Tjan after $\sim$14 ka is related to the expansion of trees from glacial refugia which have modified the albedo (Cheddadi & Bar-Hen, 2009).

The BA is followed by the cold YD ($\sim$12.9 to 11.6 ka) during which we observe a reduction of the warmer areas at 12 and 11 ka and an expansion of the more arid ones (Fig. 4.5). When Tjan exhibited a fast increase at 9 ka, Tjul remained stable. This is in agreement with Davis *et al.* (2003), who showed a similar Summer and Winter temperature evolution for South-western Europe. Nevertheless, a closer look at the scale of Iberia allows to observe that some areas had different vegetation compositions (Carrión *et al.*, 2010a) and vegetation is known to respond to both long term climate trends as well as to abrupt changes (Roucoux *et al.*, 2005).

The Holocene warm period (approximately between 8.2 and 5.6 ka, depending on where in Europe) is characterized by increasing summer temperatures (Seppä &

Birks, 2001), being more evident in Northern Europe and the Alps and simultaneous with a cooling at lower latitudes (Davis *et al.*, 2003). Our results show this dichotomy within the Iberian Peninsula where two areas (clusters 2 and 3) exhibit a warming between 9 and 5 ka while the other areas record a cooling pattern. This cooling at 8 and 7 ka is likely to be related to the impact of the ∼8.2 ka cold event (Heiri *et al.*, 2004; von Grafenstein *et al.*, 1999). Concerning the precipitation, there is evidence of a wetter climate between 9 ka and 6 ka which confirms what was previously known for the southern European lowlands (Cheddadi *et al.*, 1997).

The behaviour of the reconstructed variables after 5ka is likely to be influenced by non-natural ecosystem changes due to human activities such as the forest degradation that begun in lowlands and later in mountainous areas (Carrión *et al.*, 2010b), increasing the difficulty of its interpretation. These human impacts add confounding effects in the fossil pollen record and may lead to reconstructed lower temperatures at 5 ka. On the other hand, human impact at larger scales, capable of leaving noticeable imprints on landscape were likely to happen later (Carrión *et al.*, 2010b) and, furthermore, there are evidences of a cooling and drier stage after 5 ka, marking the end of the Holocene warm period in Europe (Seppä & Birks, 2001), and particularly in the Iberian Peninsula (Dorado-Valiño *et al.*, 2002).

The northern areas of the Iberian Peninsula had high values of precipitation during the last 15 ka. An interesting pattern arises at 9 ka, in the mid and lower latitudes, where we observe a shift of precipitation intensity. Mid-latitudes registered the highest differences in precipitation, with a beginning of Holocene increasingly humid. On the other hand, the southernmost latitudes became increasingly arid, with the exception of the 3 ka time slice. This could be due to the increasing human activity with a greater impact on the ecosystems (Carrión *et al.*, 2010b) and their feedback on climate (Pielke *et al.*, 1998).

The variability of the climate change during the last 15ka in the Iberian Peninsula had an obvious impact on the presence of putative refugia, migrating pathways of species from these refugia and on the overall recolonisation processes during the postglacial period within the Iberian Peninsula. During this period, climate favoured migrations and expansion processes that culminated in secondary contacts for several lineages previously isolated in patches of suitable habitat (Branco *et al.*, 2002; Godinho *et al.*, 2006; Weiss & Ferrand, 2007; Miraldo *et al.*, 2011). The clustering scheme (Fig. 4.4) is consistent with the molecular evidence of a network of putative refugia within Iberia (Weiss & Ferrand, 2007). Refugia have been associated with climate and habitat stability, with both playing complementary roles (Ashcroft, 2010).

However, as shown by large scale landscape analysis (Carrión *et al.*, 2010a,b) and climate reconstructions (Davis *et al.*, 2003; Cheddadi & Bar-Hen, 2009), both have a strong dynamic nature in the Iberian Peninsula, and promoted the formation of patches of suitable habitat during harsh conditions. The highly structured population that many species exhibit in the Iberian Peninsula have contributed decisively to the idea of refugial diversity (Hewitt, 2000; Weiss & Ferrand, 2007). Overall, the information included in the multidimensional climate data allowed us to define putative refugia as areas that shared a similar climate evolution during the late-Quaternary and where temperature and precipitations are suitable to support the survival of temperate trees. The range of climate changes over the past 15,000 years has largely affected the recolonization process of species. The defined clusters can be associated with potential isolation or dispersal events of species throughout the studied time span. Particularly, the third cluster (Fig. 4.4) includes areas that have already been described as glacial refugia for several animal and plant species (Weiss & Ferrand, 2007, see chapter 5 for a review of refugia in Iberian Peninsula). In the area represented by this cluster, the climate evolved with a lower amplitude ($\sim$4^{o}C) and was less sensitive to extreme fluctuations than the other clusters, which is compatible with the persistence of species in these areas. The southern plateau, mostly comprised in the second cluster (Fig. 4.4), recorded also mild conditions which are often associated with southern refugia but a rapid feedback to late-Quaternary events may have prevented persistence or recolonisation processes.

4.1.6 Conclusions

The reconstruction of past climates using biological data is an invaluable resource for the study of dynamics of glacial refugial areas. Although there is a limited number of available sites and time range coverage, the spatial combination of fossil pollen data provides a continuous record with a climate signal that can be translated into spatially explicit analysis of climate dynamics.

The reconstructed climate variables for the post-glacial period show different patterns of evolution but clearly marked by the lasting impact of episodic climatic events. The Iberian Peninsula had areas that shared similar climate evolution during the late-Quaternary. Some areas that we have identified as potential glacial refugia are consistent with those areas where genetic diversity is highest and which are often considered as refugial areas for several animal and plant species.

The analysis of these areas and the related climate provides new insights about the dynamics of refugia through time and space which helps a better understanding of

the evolution of biodiversity hotspots both at the species and the intraspecific levels. Thus, such study on the Iberian Peninsula has an obvious interest for conservation issues, especially under the expected future climate change.

4.1.7 Acknowledgments

PT was funded with a PhD grant (SFRH/BD/42480/2007) and JCB has a contract (Programme Ciência 2007), both from Fundação para a Ciência e Tecnologia. JC contribution was funded by the project Paleoflora y Paleovegetación ibérica, Plan Nacional de I+D+i, Ref. CGL-2009-06988/BOS. LS acknowledges the contribution of M. C. Freitas and C. Andrade (University of Lisbon) who provide the cores. The authors would like to acknowledge all contributors of the European Pollen Database and the Global Biodiversity Information Facility for making their datasets publicly available to the scientific community. We are very grateful to Basil Davis, for his kind support and comments. We also thank William Fletcher and Maria Sanchez-Goñi for data contribution and comments, and also Penélope González-Sampériz contributions. Their contributions greatly improved the quality of the manuscript. This is an ISEM-contribution nº xx-xxxx

4.1.8 References

Anderson, N.J., Bugmann, H., Dearing, J.A. & Gaillard, M.J. (2006) Linking palaeoenvironmental data and models to understand the past and to predict the future. *Trends in Ecology & Evolution* **21**, 696–704.

Ashcroft, M.B. (2010) Identifying refugia from climate change. *Journal of Biogeography* **37**, 1407–1413.

Bennett, K. & Provan, J. (2008) What do we mean by 'refugia'? *Quaternary Science Reviews* **27**, 2449–2455.

Berger, A. (1978) Long-term variations of caloric insolation resulting from the Earth's orbital elements. *Quaternary Research* **9**, 139–167.

Branco, M., Monnerot, M., Ferrand, N. & Templeton, A.R. (2002) Postglacial dispersal of the European rabbit (*Oryctolagus cuniculus*) on the Iberian Peninsula reconstructed from nested clade and mismatch analyses of mitochondrial DNA genetic variation. *Evolution* **56**, 792–803.

Carrión, J.S., Fernández, S., González-Sampériz, P., Gil-Romera, G., Badal, E., Carrión-Marco, Y., López-Merino, L., López-Sáez, J.A., Fierro, E. & Burjachs, F.

(2010a) Expected trends and surprises in the Lateglacial and Holocene vegetation history of the Iberian Peninsula and Balearic Islands. *Review of Palaeobotany and Palynology* **162**, 458–475.

Carrión, J., Fernández, S., Jiménez-Moreno, G., Fauquette, S., Gil-Romera, G., González-Sampériz, P. & Finlayson, C. (2010b) The historical origins of aridity and vegetation degradation in southeastern Spain. *Journal of Arid Environments* **74**, 731–736.

Cheddadi, R., Yu, G., Guiot, J., Harrison, S. & Prentice, I.C. (1997) The climate of Europe 6000 years ago. *Climate Dynamics* **13**, 1–9.

Cheddadi, R. & Bar-Hen, A. (2009) Spatial gradient of temperature and potential vegetation feedback across Europe during the late Quaternary. *Climate Dynamics* **32**, 371–379.

Cheddadi, R., Vendramin, G.G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J.M., de Beaulieu, J.L., Sadori, L., Jost, A. & Lunt, D. (2006) Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography* **15**, 271–282.

Cox, N., Chanson, J. & Stuart, S. (2006) *The status and distribution of reptiles and amphibians of the Mediterranean Basin*. IUCN, Gland, Switzerland and Cambridge, U.K.

Davis, B.A.S., Brewer, S., Stevenson, A.C., Guiot, J. & Data Contributors (2003) The temperature of Europe during the Holocene reconstructed from pollen data. *Quaternary Science Reviews* **22**, 1701–1716.

Dorado-Valiño, M., Rodríguez, A.V., Zapata, M.B.R., García, M.J.G. & Gutiérrez, I.D.B. (2002) Climatic changes since the Late-glacial/Holocene transition in La Mancha Plain (South-central Iberian Peninsula, Spain) and their incidence on Las Tablas de Daimiel marshlands. *Quaternary International* **93-94**, 73–84.

Furrer, R., Nychka, D. & Sain, S. (2012) *fields: Tools for spatial data*. URL `http://CRAN.R-project.org/package=fields`.

Godinho, R., Mendonça, B., Crespo, E.G. & Ferrand, N. (2006) Genealogy of the nuclear beta-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity* **96**, 454–463.

Guiot, J. (1997) Palaeoclimatology: Back at the last interglacial. *Nature* **388**, 25–27.

Heiri, O., Tinner, W. & Lotter, A.F. (2004) Evidence for cooler European summers during periods of changing meltwater flux to the North Atlantic. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15285–15288.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.

Hicks, S. (2006) When no pollen does not mean no trees. *Vegetation History and Archaeobotany* **15**, 253–261.

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

Hu, F.S., Hampe, A. & Petit, R.J. (2009) Paleoecology meets genetics: deciphering past vegetational dynamics. *Frontiers in Ecology and the Environment* **7**, 371–379.

Keitt, T.H., Bivand, R., Pebesma, E. & Rowlingson, B. (2012) rgdal: Bindings for the Geospatial Data Abstraction Library. URL `http://CRAN.R-project.org/package=rgdal`.

Kühl, N., Gebhardt, C., Litt, T. & Hense, A. (2002) Probability Density Functions as Botanical-Climatological Transfer Functions for Climate Reconstruction. *Quaternary Research* **58**, 381–392.

Laurent, J.M., Bar-Hen, A., François, L., Ghislain, M. & Cheddadi, R. (2004) Refining vegetation simulation models: from plant functional types to bioclimatic affinity groups of plants. *Journal of Vegetation Science* **15**, 739–746.

MacDonald, G., Bennett, K., Jackson, S., Parducci, L., Smith, F., Smol, J. & Willis, K. (2008) Impacts of climate change on species, populations and communities: palaeobiogeographical insights and frontiers. *Progress in Physical Geography* **32**, 139–172.

Médail, F. & Diadema, K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography* **36**, 1333–1345.

Médail, F. & Quézel, P. (1999) Biodiversity hotspots in the Mediterranean Basin: setting global conservation priorities. *Conservation Biology* **13**, 1510–1513.

Miraldo, A., Hewitt, G.M., Paulo, O.S. & Emerson, B.C. (2011) Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology* **11**, 170.

Myers, N., Mittermeier, R., Mittermeier, C., Da Fonseca, G. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

Naughton, F., Sanchez Goñi, M., Desprat, S., Turon, J.L., Duprat, J., Malaizé, B., Joli, C., Cortijo, E., Drago, T. & Freitas, M. (2007) Present-day and past (last 25000 years) marine pollen signal off western Iberia. *Marine Micropaleontology* **62**, 91–114.

Ohlemüller, R., Huntley, B., Normand, S. & Svenning, J.C. (2012) Potential source and sink locations for climate-driven species range shifts in Europe since the Last Glacial Maximum. *Global Ecology and Biogeography* **21**, 152–163.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42.

Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**, 683–691.

Pérez-Obiol, R., Jalut, G., Julia, R., Pelachs, A., Iriarte, M.J., Otto, T. & Hernandez-Beloqui, B. (2011) Mid-Holocene vegetation and climatic history of the Iberian Peninsula. *The Holocene* **21**, 75–93.

Petit, R.J., Aguinagalde, I., de Beaulieu, J.L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565.

Pielke, R.A., Avissar, R., Raupach, M., Dolman, A.J., Zeng, X. & Denning, A.S. (1998) Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. *Global Change Biology* **4**, 461–475.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J.O., Wickham, H., Graves, S. & Hooker, G. (2012) *fda: Functional Data Analysis*.

Ramsay, J. & Silverman, B. (2005) *Functional data analysis*. Springer, New York, 2nd edn.

Renssen, H. & Isarin, R.F.B. (2001) The two major warming phases of the last deglaciation at 14.7 and 11.5 ka cal BP in Europe: climate reconstructions and AGCM experiments. *Global and Planetary Change* **30**, 117–153.

Roucoux, K., Abreu, L.D., Shackleton, N.J. & Tzedakis, P.C. (2005) The response of NW Iberian vegetation to North Atlantic climate oscillations during the last 65kyr. *Quaternary Science Reviews* **24**, 1637–1653.

Seppä, H. & Birks, H. (2001) July mean temperature and annual precipitation trends during the Holocene in the Fennoscandian tree-line area: pollen-based climate reconstructions. *The Holocene* **11**, 527–539.

Tzedakis, P.C., Lawson, I.T., Frogley, M.R., Hewitt, G.M. & Preece, R.C. (2002) Buffered tree population changes in a quaternary refugium: evolutionary implications. *Science* **297**, 2044–2047.

Van der Wiel, A.M. & Wijmstra, T.A. (1987a) Palynology of the 112.8-197.8 m interval of the core Tenaghi Philippon III, middle Pleistocene of Macedonia. *Review of Palaeobotany and Palynology* **52**, 89–108.

Van der Wiel, A.M. & Wijmstra, T.A. (1987b) Palynology of the lower part (78-120 m) of the core Tenaghi Philippon II, Middle Pleistocene of Macedonia, Greece. *Review of Palaeobotany and Palynology* **52**, 73–88.

von Grafenstein, U., Erlenkeuser, H., Brauer, A., Jouzel, J., Johnsen, S.J. & von Grafenstein, U. (1999) A Mid-European Decadal Isotope-Climate Record from 15,500 to 5000 Years B.P. *Science* **284**, 1654–1657.

Webb, R.S., Anderson, K.H., Webb, T. & Others (1993) Pollen response-surface estimates of late-Quaternary changes in the moisture balance of the northeastern United States. *Quaternary Research* **40**, 213–227.

Weiss, S. & Ferrand, N. (eds.) (2007) *Phylogeography of Southern European Refugia*. Springer, Dordrecht.

Wijmstra, T.A. & Smith, A. (1976) Palynology of the middle part (30–78 metres) of the 120 m deep section in northern Greece (Macedonia). *Acta Botanica Neerlandica* **25**, 297–312.

Wijmstra, T. (1969) Palynology of the first 30 m. of a 120m. deep section in northern Greece. *Acta Botanica Neerlandica* **18**, 511–527.

Willis, K.J., Bailey, R.M., Bhagwat, S.a. & Birks, H.J.B. (2010) Biodiversity baselines, thresholds and resilience: testing predictions and assumptions using palaeoecological data. *Trends in Ecology & Evolution* **25**, 583–591.

Chapter 5

Spatial dynamic patterns in the Iberian Peninsula at two scales

> Our intelligence and our technology have given us the power to affect the climate. How will we use this power? Are we willing to tolerate ignorance and complacency in matters that affect the entire human family? Do we value short-term advantages above the welfare of the Earth? Or will we think on longer time scales, with concern for our children and our grandchildren, to understand and protect the complex life-support systems of our planet? The Earth is a tiny and fragile world. It needs to be cherished.
>
> — CARL SAGAN, *Cosmos*

5.1 Velocity of biodiversity change: a case study in a global hotspot[1]

5.1.1 Abstract

The velocity of climate change has been assessed globally for the past and predicted for the current century. High velocities are predicted for the near future, exceeding those estimated since the last glacial maximum. The past biodiversity patterns reveal a succession of contraction and expansion stages, strongly linked to climate change. The projection of these changes to the predicted future velocities exposes a dramatic scenario of biodiversity shifts, and, in some cases, extinctions. However, the relation between velocity of climate change and species impact is mediated by the niche breath of the species.

[1] This work is in preparation for publication.

Here we provide velocities estimates of species richness change using Iberian herpetofauna as model species. We developed two velocity indexes to describe the patterns of past changes in Iberian Peninsula using 13 layers of climate variables, covering a period from 15,000 to 3,000 years BP. We also predict future impacts using two climate models and four emission scenarios. We show that species compositional change had lower velocities in the past when compared to those predicted for this century. Niche-based velocities provided reliable estimates of the past environments dynamics and the same reliability is expected to future projections.

5.1.2 Introduction

Climate is a major shaping force of biodiversity distribution (Hewitt, 2000). Fluctuations of past climates were responsible for distributional shifts that resulted in the current patterns of diversity, both at inter and intraspecific levels, as observed by species richness (Araújo *et al.*, 2006) and genetic diversity (Hewitt, 2000). The harsh climate of the Last Glacial Maximum (LGM), especially for warm dependent species in temperate regions, forced a contraction of their ranges to areas classified as glacial refugia, located at southern latitudes (Hewitt, 2000; Gómez & Lunt, 2007; Ohlemüller *et al.*, 2012). Though a warming trend is evident since the LGM, the evolution of the climate was not stable, often punctuated by abrupt events with a dramatic range of variation (Alley *et al.*, 2003, but see also Section 4.1). These recurring events interacted with biodiversity patterns, resulting in expansions and contractions of species' ranges driven by the climatic oscillations (Carrión *et al.*, 2010).

Future climate poses a similar situation for biodiversity, with the aggravating factor of the shorter time period when changes with equivalent amplitude are predicted to occur (Willis *et al.*, 2010). Such changes might lead to migrations to northern latitudes (e.g. Rebelo *et al.*, 2010), extreme contractions of distributions (e.g. Araújo *et al.*, 2006; Carvalho *et al.*, 2010) or, in the worst case scenario, a profusion of species extinctions (Thomas *et al.*, 2004). However, climate predictions for the current century differ in magnitude of changes between models and emissions scenarios (IPCC, 2000), increasing the difficulty to accurately predict potential impacts for biodiversity (Knutti *et al.*, 2010). Effects of the climate change are already discernible in some taxa, with adaptation to new conditions occurring at multiple levels, including range shifts, phenological and species interaction changes (Parmesan, 2006). Dispersal is a strong limiting factor due to the imposition of a boundary to the species' ability to follow the pace of climate change (Araújo *et al.*, 2006; Schloss *et al.*, 2012). If a population is able to move as fast as climate, than the probability of survival is higher.

Including accurate dispersal estimates in the study of climate change impacts on biodiversity is difficult, and usually limited to extreme scenarios of dispersal (e.g. Thomas *et al.*, 2004; Araújo *et al.*, 2006; Carvalho *et al.*, 2010). An estimation of how fast the species need to move is indirectly given by the climate change velocity (Loarie *et al.*, 2009; Sandel *et al.*, 2011). Velocities of temperature change were shown to have an altitudinal gradient, with mountainous areas more resistant to change and flat lowland areas more prone to climate sweeps, promoting differential impacts on species (Loarie *et al.*, 2009). The translation of the velocity of climate change to the distributional shifts to track the change is dependent on the tolerance of each species (Parmesan, 2006; Loarie *et al.*, 2009). Species' niche assumes a paramount importance in the analysis of the impact of climate change by defining the breadth of tolerance of a species and also the ability to conquer novel habitats. This renders the ecological niche modelling (ENM) as an appropriate tool to study the effects of climate change on biodiversity patterns.

The species' presence data, however, reflects the biotic and abiotic interactions limiting the distribution. The niche represented by both interactions is the realized niche, which is a restriction of the fundamental niche in the Grinnellian space (Soberón, 2007). Niche transference, i.e. projection, into other time frames should be achieved using the fundamental niche, or at least the most proximate solution, because it holds the ability of a species to occupy novel habitats where there is less physiological stress and, thus, providing also higher degrees of confidence to the predicted range shifts (Peterson *et al.*, 2007b). Modelling the fundamental niche depends on factors such as the number of variables in a model. Although there is no optimum solution for the number of variables to include, using many will tend to result in an overfitted model, whereas a small number will tend to build very simple models and to capture the most conserved part of the niche (Peterson, 2011; Rödder & Lötters, 2009). This will be a part of the fundamental niche, however, with reduced multidimensionality. Moreover, the succession of modelling approaches, from mechanistic to correlative with presence-only data and to correlative with presences and absences, show a general trend of capturing the fundamental niche to the realized niche (Sillero, 2011). While different modelling algorithms have distinct degrees of transferability (Randin *et al.*, 2006; Peterson *et al.*, 2007b; Phillips, 2008; Wenger & Olden, 2012), artificial neural networks were shown to have a good transferability potential with simple network schemes (Heikkinen *et al.*, 2012; Wenger & Olden, 2012).

The herpetofauna of the Iberian Peninsula are an interesting model group to study the effects of climate change. Firstly, they are ectotherms, thus, highly depen-

dent on external sources of heat, which is an advantage in studies under the Grinnellian niche space. Secondly, there are detailed information of the Iberian distribution of herpetofauna. Thirdly, there are more than 30 species of herptiles endemic or nearly endemic to the Iberian Peninsula, allowing the development of a multi-species studies with wide coverage. Fourthly, due to their sensibility to climate change, the awareness for herpetofauna conservation is increasing (Carvalho *et al.*, 2010). Finally, the Iberian Peninsula is part of the Mediterranean biodiversity hotspot due to the high levels of endemism and threats to biodiversity (Myers *et al.*, 2000).

The objective of this study is to elucidate relationships between reptile and amphibians species richness with the climate change in the Iberian Peninsula. To accomplish this we address two questions: What is the dynamic of change in past species composition and its relation to present diversity?; What is predicted to occur in terms of species compositional changes for the near future? We propose new velocity indices to analyse these topics. Answering these questions will offer an comprehensive understanding of the biodiversity patterns in the Iberian Peninsula, contributing to a better knowledge on the dynamics of the shaping processes acting in the area with potential implications for other biodiversity hotspots.

5.1.3 Methods

Species data

We gathered distributional data for 11 amphibians and 19 reptiles species from the Spanish and Portuguese herpetological atlases (Pleguezuelos *et al.*, 2004; Loureiro *et al.*, 2008) and complemented with the data available at SIARE (`http://siare.herpetologica.es/`, last access on June 2012). The original data was digitized with the UTM 10km squared grid, and converted to a geographical grid with 5 arc-minutes resolution ($\sim$9km) with square pixels, avoiding the use of different UTM zones. This conversion was based on the intersection of the original data and the geographical grid, only retaining presences in the final grid where the intersecting area was greater than 25% of the original UTM square.

We followed the most updated taxonomy (Carretero *et al.*, 2011) from the Spanish taxonomic list (Montori *et al.*, 2005). Species with taxonomic conflicts between the two atlases (*Triturus* sp. and *Pelodytes* sp.) and those with less than five observations in the grid were removed from the analysis. To reach the final dataset, we selected the species with 75% or more of their distribution included in the Iberian Peninsula, resulting in the 30 taxa selected with a high degree of Iberian endemism.

Climate data

We downloaded current monthly climate data from WorldClim (www.worldclim.org;Hijmans *et al.*, 2005) with a resolution of 5 arc-minutes. These data were obtained by spatial interpolation of climate weather data covering the world. We clipped the obtained data to the extent of the Iberian Peninsula and prepared three ecological niche variables (ENV): 1) minimum temperature of January (Tjan); 2) maximum temperature of July (Tjul); and 3) minimum annual precipitation (Pmin). These ENVs represent extreme climate and are assumed to be distribution limiting for ectoterms and for water dependent animals (Costa *et al.*, 2008). Temperatures are weakly correlated (Pearson's r=0.31) while precipitation is moderately and inversely correlated with Tjan and Tjul (r=-0.53 and r=-0.83, respectively).

We obtained past climate data with a period of 1,000 years within the time period between 15,000 and 3,000 years BP (see Section 4.1). These layers of climate were reconstructed using fossil pollen as a climate proxy and spatially interpolated to the extent of the Iberian Peninsula with a pixel size of 5 arc-minutes resolution ($\sim$9km).

We used the high-resolution gridded dataset TYN SC 2.0 (http://www.cru.uea.ac.uk/cru/data/hrg/;Mitchell *et al.*, 2004) and converted to raster format using E-Clic (Tarroso & Rebelo, 2010). We extracted two models (CSIRO2 and HadCM3) with four emissions scenarios (A1FI, A2, B1 and B2). These scenarios represent different economic directions for future societies, varying broad parameters as globalization intensity and environmental friendliness, and generating different estimations of climate (IPCC, 2000). These datasets offer two main advantages over other datasets (e.g. WorldClim future climate datasets): 1) availability of five variables for five climate models; and 2) continuous monthly data for the period between 2001-2100. We extracted monthly data between 2001 and 2100 for three variables (daily mean temperature, diurnal temperature range and precipitation) at the resolution available (10 arc-minutes). The Tjan and Tjul were built from subtracting and adding, respectively, the temperature range to the mean temperature. We downscaled the data to the study resolution using statistical downscaling (Waltari *et al.*, 2007; Svenning *et al.*, 2008; Tabor & Williams, 2010): anomalies were calculated to the present (WorldClim data) at 10 arc-minute resolution and interpolated using splines algorithm to 5 arc-minute. The correct values of the variables were then rebuilt by adding the current climate to the anomalies at higher resolution. We averaged the variables by decade and did a linear regression between each decade and current climate to estimate linear equations. First decade (2001-2010) was used to generate calibrating values for the slope and intercept for temperature variables. For precipitation data only slope was calibrated

and intercepted was forced at zero. This procedure allows to adequate climate projections to current climate when using different sources, maintaining a similar degree of variance. The obtained calibration values were used to adjust all successive decades, generating more reliable estimates of the variables.

Ecological niche modelling

We used species and climate data as input variables to assess the niche of reptiles and amphibians in the Iberian Peninsula by means of ecological niche modelling. The chosen algorithm was the Artificial Neural Networks (ANN) implemented in the software Simapse v1.1 (Tarroso *et al.*, 2012). The ANN are a machine-learning algorithm known to be very efficient in detecting patterns maintained by non-linear relationships between variables, even in the presence of some noise, as is often the case of ecological models (Olden *et al.*, 2008; Tarroso *et al.*, 2012). ANN attractive features are reflected in the extensive application in ecology, ranging from simple ecological models (Lek *et al.*, 1996), to algorithm ensemble approaches in conservation studies (Carvalho *et al.*, 2010), hybrid zone analysis (see Section 5.2) and climate reconstruction (Cheddadi *et al.*, 1997). It consists on a layered structure of interconnected processing nodes (i.e. artificial neurons) that is able to learn patterns during the training stage, by adjusting the connections between nodes with the purpose of minimizing the output error after each iteration. When fed with species data and ENVs, it will be able to predict presence probability to other locations with the same set of ENVs.

We used the current climate ENVs to assess the niche of the 30 selected species. Parameters were optimized heuristically and maintained transversally to all studied taxa. These include 5,000 iterations (with cycles of 1 internal iteration) after 15 burn-in iterations and a structure of the hidden layer with 6 nodes. Learning rates were adapted to presence data complexity (Table D.1): difficulty to model increases with data clustering and smaller learning rates are used (Wilson & Martinez, 2001). The number of pseudo-absences was calculated based on 20% of the available locations without presence data. We built a consensus model after running Simapse for 10 replicates using random sub-sampling. All other values were maintained at Simapse's default.

The final probability model was converted to binary using the 10^{th} percentile value of the model predictions for the presences. Although this threshold classifies 10% of presences as false negatives, it provides accurate potential distributions (Brito *et al.*, 2009). Models for each species were evaluated with receiver operating characteristic curve (ROC) and the respective area under the curve (AUC). The niche models

using current climate data were projected to the 13 layers of past climate and to 10 decades of predicted future climate, based on 8 combinations of models and scenarios. The projected models of the 30 species, for each projection layer were added to obtain projections of species richness per time layer.

Velocity calculation

Current climate models were used to assess the species niches and velocities were calculated based on species richness of projected models. Two different velocities measures (Fig. 5.1) were processed for past and future projections independently: 1) the velocity of pixel value maintenance (Vm) sums the minimum distance that the richness value of each pixel has to travel from one layer to the next layer in time and is divided by the total time; and 2) the velocity of richness shifts (Vs) sums the minimum distance a value of richness has to travel in each layer, arranged by chronological order, and divided by the total time. Both velocities have different purposes: while the Vm measures how far a value of richness has to travel after reaching a location, the

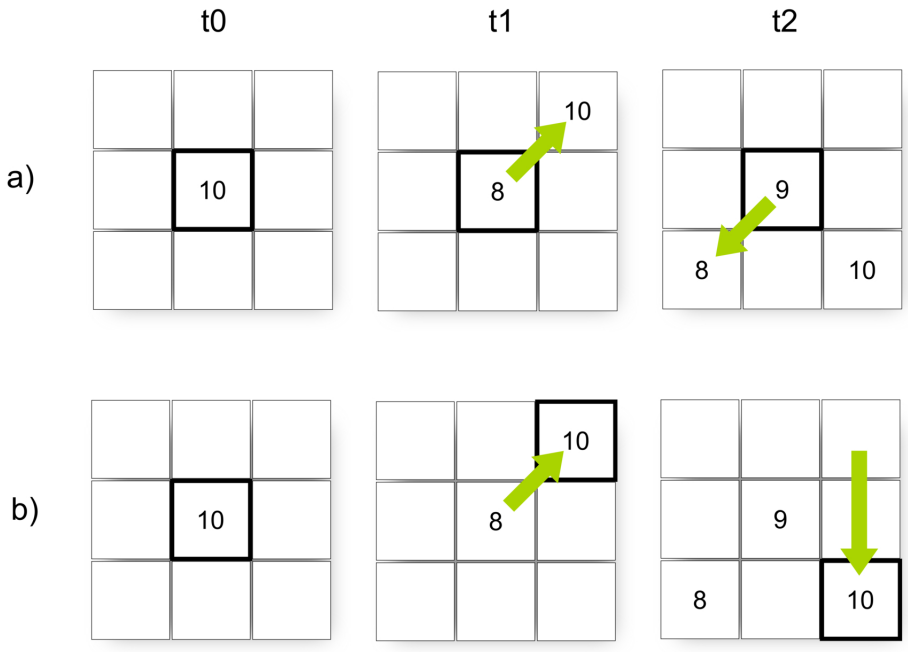


Fig. 5.1 – Illustration of the velocities metrics. The velocity of pixel maintenance (Vm) corresponds to the cell value maintenance throughout time (a). It is measured the shortest distance between one cell at time t and the nearest cell at the next time layer (t+1) holding an equivalent value of species richness. The velocity of richness shift (Vs), corresponds to the shift of pixel value in space (b). The distance here is the tracking of the species richness value at time t in all subsequent time layers (t+n), irrespective of the original cell location. Distances measure in km are summed and divided by years to calculate the velocities in km.yr-1 units. Species values are tracked with a tolerance value of 1 species.

Vs describes how far did each value of richness travelled from its original location. The tracking of species richness has a tolerance of one species, meaning that the distance is found by locating the value with the richness plus or minus one. This procedure allows the possibility of positive infinite velocities: if a value being tracked is not found in the next time layer, an infinite distance is attributed. Geographical coordinates were transformed using a gnomonic projection, adjusted to the centroid of the Iberian Peninsula area, for calculation of velocities in standard units (km.yr^{-1}).

Since different future climate scenarios have different parametrizations of physical climate properties and impacts of society behaviours, multi-model averaging may result in deviations and physically impossible scenarios (Knutti *et al.*, 2010). Thus, we projected each combination of model and scenario individually and summarized the results by calculating the minimum and the maximum predicted by the four scenarios, creating a best and worst case scenario for each climate model.

All analysis were performed and automated with R language (R Development Core Team, 2012) using the packages 'rgdal' (Keitt *et al.*, 2012) with the Geospatial Data Abstraction Library (GDAL Development Team, 2011), 'ROCR' (Sing *et al.*, 2009) and 'RSQLite' (James, 2011) to interact with an SQLite (`www.sqlite.org`) database storing all the data. Additional processing and automation were done using python (`www.python.org`) and shell scripts. Means are reported with one standard deviation.

5.1.4 Results

The models produced for the present climatic conditions had a high average AUC (0.850 $\pm$ 0.114, appendix Table D.1 and Fig. D.1), indicating high accuracy. However, some species, particularly those with wide distributions, reported low AUC values. There is a general increasing trend between achieved AUC and the proportion of clustered absences (i.e. non-presence locations; appendix D.2), relating low values of AUC with higher scattering of absences' distribution. The threshold chosen for each model (appendix D.1) classifies some presences located at the edges of the predicted distributions as false negative (appendix D.3).

The calibration of the future climate datasets with the present data resulted in less drastic changes of climate (appendix D.4). For 2100, calibrated predictions points to a Tjul $\sim$0.7ºC lower than original models, in average. For Tjan is shown an increase of $\sim$0.5ºC, while for Pmin, calibrated models indicate more $\sim$5.6mm of precipitation, for average values.

Both calculated velocities generated similar spatial patterns (Fig. 5.2). The Vm have slightly higher average values than Vs (Vm = 0.0098 $\pm$ 0.0090 km.yr^{-1} and Vs

= 0.0093 $\pm$ 0.0086 km.yr^{-1}), indicating that pixel values travel longer distances every temporal layer than the original value has to shift during the 12,000 years. Major differences between velocities are located at the western part of the Iberian Peninsula where the values of each pixel are preserved through time, but original species richness (at 15,000 years BP) travelled long distances. Central area of the Peninsula reveal high velocities of change, with pockets of low velocities within higher altitudes.

Higher velocities and average richness are generaly located between 500 and 1,000 m altitude (Fig. 5.3), however, with high variance in species richness through the studied time range. The mountain areas higher than 1,000m shown a trend to decrease to near zero velocities and with low changes in species richness. Areas with higher altitudes (>1,000m) had a low to average richness during the studied period, with stable values of species number (appendix D.5). Most diversity was found below 1250 m and, particularly, below 500 m, with high variance in the number of species. The average richness had an absolute decrease from 15,000 (17.3 $\pm$ 3.5) to 3,000 years BP (14.7 $\pm$ 3.9). However, this process had an oscillating nature and the minimum was registered at 9,000 years BP (13.7 $\pm$ 3.0; Fig. 5.4). Locations with highest variance in species number through the studied time period were those that registered highest number of species at the oldest age (Fig. 5.4).

The general result of the velocities based on predicted future climate layers is an increase of two orders of magnitude in relation to those of the past (Fig. 5.5). The average Vs is higher than the Vm, however HadCM3 model (minima: Vm = 0.6891

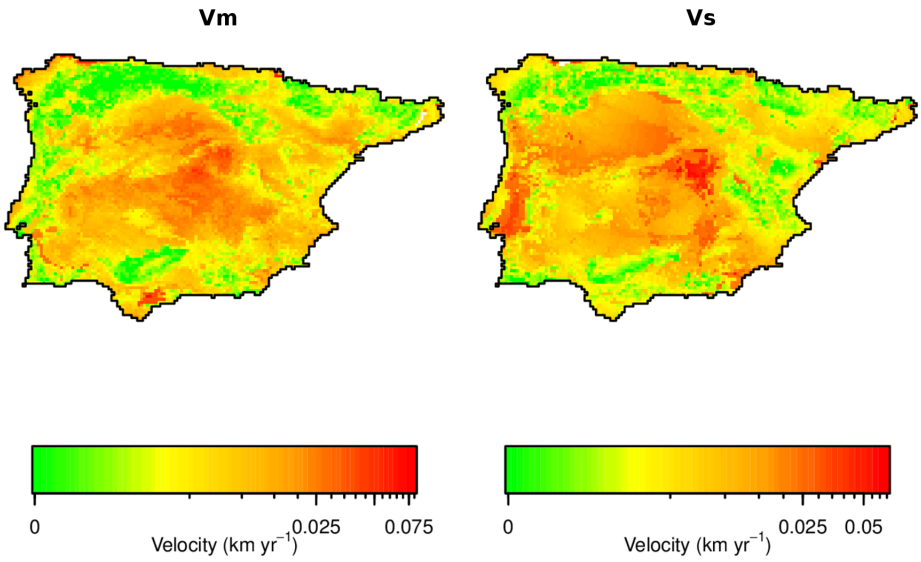


Fig. 5.2 – Velocities of species composition change: Velocity of pixel maintenance (Vm) and velocity of richness shift (Vs) estimated for the period between 15,000 and 3,000 years BP.

$\pm$ 0.9566 km.yr^{-1}< Vs = 1.371 $\pm$ 1.4446 km.yr^{-1}; maxima: Vm = 1.5479 $\pm$ 1.6452 km.yr^{-1} < Vs 2.0168 $\pm$ 2.4781 km.yr^{-1}) exhibit higher differences than CSIRO2 (minima: Vm = 0.5084 $\pm$ 0.6249 km.yr^{-1} < Vs = 0.5632 $\pm$ 0.6408 km.yr^{-1}; maxima: Vm = 1.1754 $\pm$ 0.9742 km.yr^{-1} < Vs = 1.392 $\pm$ 1.2041 km.yr^{-1}). The climate model HadCM3 also shows a higher number of cells with infinite values (Fig. 5.5), and, minimum values between scenarios also display infinite velocities, predicting that values of richness in these locations will be lost in all available scenarios.

The value of richness is predicted to be most unstable at elevations between 750 and 1000 m and near sea level for the next 100 years, and also values at these locations will have to move longer distances (Fig. 5.6). Nevertheless, the two climate models provide different effects of both velocities calculated per altitude. The impact of HadCM3 model is seen more generally at altitudes bellow 1750 m, with many locations reaching high velocities. The highest values of average species richness for the next century are restricted to middle altitudes, and with high levels of variance, indicating changes in number of species at these locations (appendix D.6). The average values of species richness were predicted to decrease in all combinations of models and scenarios (Fig. 5.7).

The scenario A1FI showed the highest differences (CSIRO2: 12.7 $\pm$ 4.0 km.yr^{-1}, 9.4 $\pm$ 2.6 km.yr^{-1}; HadCM3: 12.8 $\pm$ 4.0 km.yr^{-1} , 6.2 $\pm$ 1.4 km.yr^{-1}, for 2010 and 2100 values, respectively for each model), followed by the A2 (CSIRO2: 13.0 $\pm$ 4.2 km.yr^{-1}, 10.0 $\pm$ 3.0 km.yr^{-1}; HadCM3: 12.8 $\pm$ 4.0 km.yr^{-1}, 7.4 $\pm$ 1.6 km.yr^{-1}). The scenarios B1 (CSIRO: 12.9 $\pm$ 4.2 km.yr^{-1}, 11.6 $\pm$ 3.2 km.yr^{-1}; HadCM3: 12.8

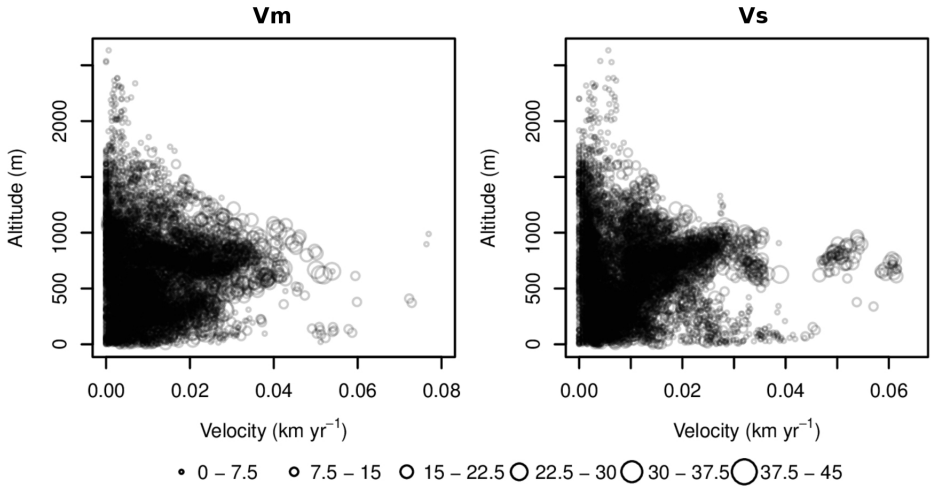


Fig. 5.3 – Relation between velocities of pixel maintenance (Vm) and richness shift (Vs) with altitude in the Iberian Peninsula. Circle size corresponds to the variance of species richness (n. of species/cell) registered for each cell in all time layers.

$\pm$ 4.0 km.yr^{-1}, 9.6 $\pm$ 2.7km.yr^{-1}) and B2 (CSIRO: 12.8 $\pm$ 4.2 km.yr^{-1}, 11.0 $\pm$ 3.0 km.yr^{-1}; HadCM3: 12.8 $\pm$ 4.1 km.yr^{-1}, 9.1 $\pm$ 2.5 km.yr^{-1}) were very similar, but with differences in severity. The locations predicted to have the highest variance in number of species for the next century are those with a high number of species currently.

5.1.5 Discussion

The issue of model transferability has been extensively discussed on scientific literature (Randin *et al.*, 2006; Peterson *et al.*, 2007b; Phillips, 2008; Wenger & Olden, 2012; Heikkinen *et al.*, 2012). Although the question raised by Thuiller *et al.* (2008) about the relation of model complexity and accuracy is still unanswered, there are arguments pointing to both directions (Stockwell, 2006; Peterson, 2007). Simple models, however, are more prone to capture a well conserved part of the niche, despite being less effective capturing the realized niche, due to the high generalization resulting from a low dimensional niche assessment. The overfitted models were shown to have a negative effect on their transferability (Randin *et al.*, 2006). The framework used here seems appropriate for the study of niche extrapolation to other time frames: 1) the Artificial Neural Networks (ANN) were shown before as an appropriate algorithm to study niche transference (Wenger & Olden, 2012; Heikkinen *et al.*, 2012); 2) although the over parametrization of ANN may require time to fine tune models, it allowed a good control of model overfitting; 3) using few niche limiting variables allowed to build a more relaxed model, supporting the conclusions by Rödder & Lötters (2009); and

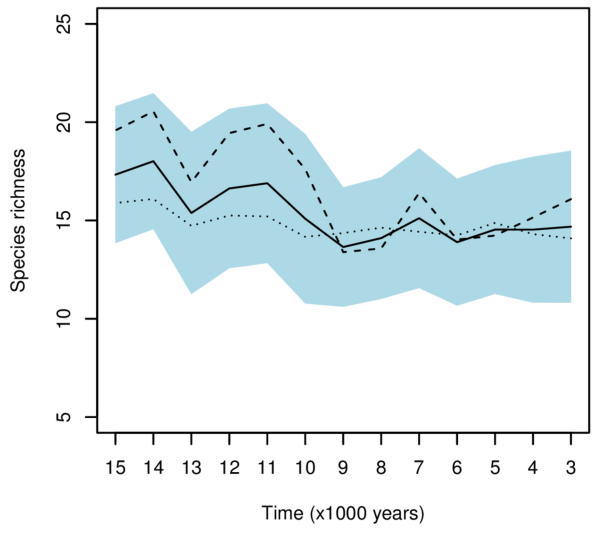


Fig. 5.4 – Evolution of species richness for the period between 15,000 and 3,000 years BP. Solid line represents the average species richness in the study area and blue band is one standard deviation. Dotted and dashed lines corresponds to the average values of locations in the first and forth quartiles of variance.

4) using 30 species allowed to balance the species-dependent transferability (Randin *et al.*, 2006).

By using an efficient machine-learning algorithm with a low number of predictors we were able to build accurate models of the species' distributions, with low number of either false positives (FP) and negatives (FN). The number of FN is obviously dependent on the threshold chosen for each model. The exclusion of some presences of the model is based on two assumptions: 1) species niche is also dispersal dependent (Soberón, 2007) and individuals may be found outside their fundamental niche; and, at less extent, 2) translocation of animals, either directly or indirectly, has occurred as result of human activities in Iberia (Pleguezuelos *et al.*, 2004). These individuals are, therefore, treated as outliers for the transferable niche and are mainly located at the borders of the distribution (appendix D.3). On the other hand, the number of FP, although also dependent on the threshold, is related to the predictive ability of the model, finding potential presences outside the original locations, i.e., generalization of the model. As observed by the predictive distribution area, some species have higher tendency to be generalists, with predicted ranges extending the area where species were found (appendix D.3). Distributional restricting forces like biotic factors (e.g. presence of competitors; Soberón, 2007), may be among the underlying causes of the high difference between predicted areas and realized niche, as given by pres-

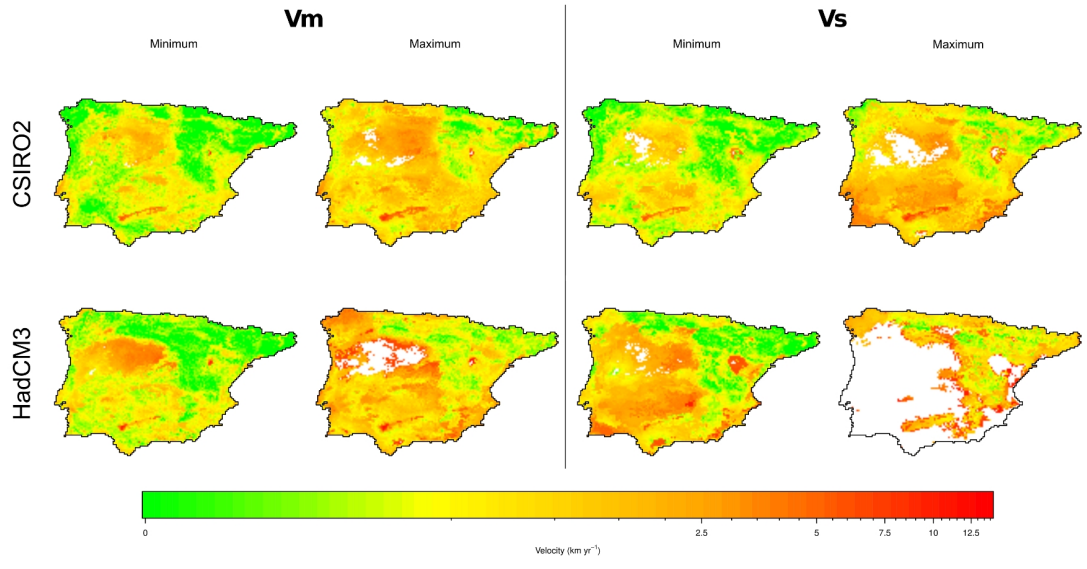


Fig. 5.5 – Velocities for predicted scenarios of future climates. The velocity of pixel maintenance (Vm) and velocity of richness shift (Vs) for climate models CSIRO2 and HadCM3 are summarized with minimum and maximum values achieved for four emissions scenarios (A1FI, A2, B1 and B2). White area represents location where infinite velocities were found.

ence's locations. However, due to the simplicity of the predictors used in this study, we cannot assess if species with high generalization would be limited by an abiotic factor other than those used. On other cases, especially on species with wider distributions (e.g. *Podarcis hispanica*, *Rhinechis scalaris* and others), FP is inflated by scattered absences. Several factors may be creating these isolated absences, including human disturbance in the habitat or evasive species which are difficult to sample. The existence of many isolated absences was shown to be an influential factor lowering the AUC values of the models (appendix Table D.1 and Fig. D.2). This metric can provide misleading results with presence-only modelling (Lobo *et al.*, 2007; Peterson *et al.*, 2007a) but it is reliable when analysed with care, being a widely used metric in ecological studies (Randin *et al.*, 2006; Peterson *et al.*, 2007a; Svenning *et al.*, 2008; Brito *et al.*, 2009; Carvalho *et al.*, 2010; Heikkinen *et al.*, 2012). However, the extensive fieldwork supporting the atlases, which are the basis of the present work, assures a higher degree of confidence on absences data, especially when clustered, and, thus, AUC is expected to show a higher accuracy.

The calculation of two velocities measuring richness maintenance and shift, have provided insightful results of the past dynamics and predicted future change in species assemblages. Other studies analysing the velocity of climate change, have based their calculations on linear temporal gradients of temperature and precipitation

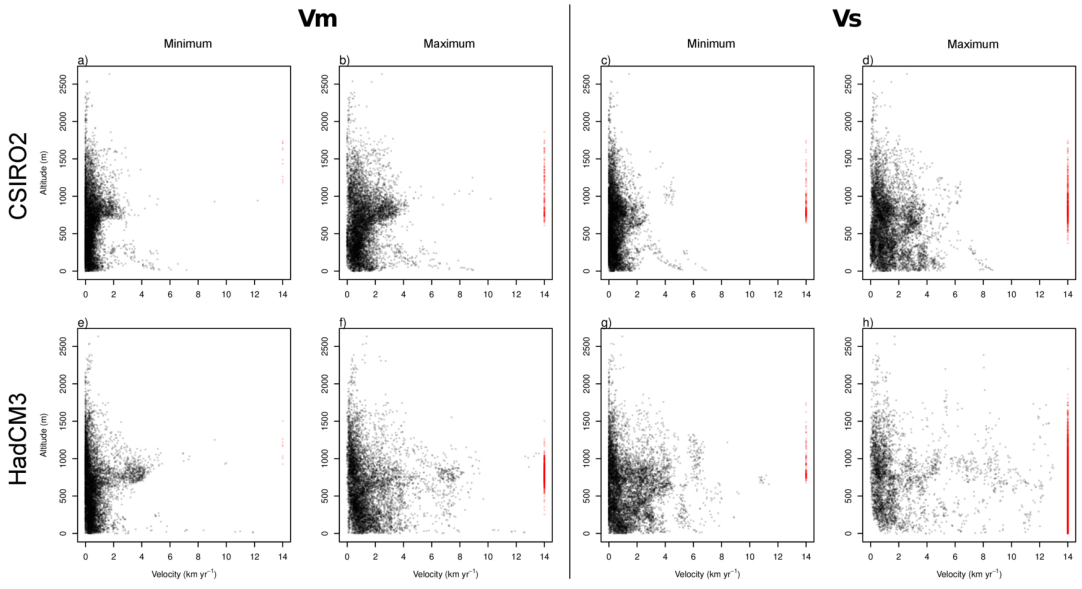


Fig. 5.6 – Relation between velocities and altitude for future climate predictions. Emission scenarios for CSIRO and HadCM3 climate models are summarized with minimum and maximum values achieved for velocity of pixel maintenance (Vm) and velocity of richness shift (Vs). Red dots at the right of each plot correspond to estimated infinite velocities.

changes (Loarie *et al.*, 2009; Sandel *et al.*, 2011). Predictions of temperature and precipitation for the next 100 years have a linear increase (Loarie *et al.*, 2009) and, due to the short time scale, no major fluctuations are expected. Past changes, on the contrary, are usually analysed for longer time periods, capturing intense oscillations within longer trends (Zachos *et al.*, 2001). This is the case of the warming trend after the last glacial maximum until the present days, that registered major cold and warm periods (see Section 4.1; Willis & MacDonald, 2011). Excluding these events from the analyses forces lower predictions of the climate change velocity, related to a spurious linear warming trend since the end of the glacial period (Sandel *et al.*, 2011). The oscillating nature of the past climate warming is, therefore, eliminated by linear estimations of the temporal component of velocities. Our dataset is not possible to analyse with such velocity calculation: species richness oscillates with great variance during the 13,000 years BP studied for the past climate (Fig. 5.3, appendix D.5). Fitting a line to these data would cancel the oscillations in species richness and provide erroneous estimates of velocity. In fact, in extreme cases, predicted expansions and contractions of species ranges would generate such variation of species richness through time that a linear fitting results in a flat slope, indicating no change in species composition.

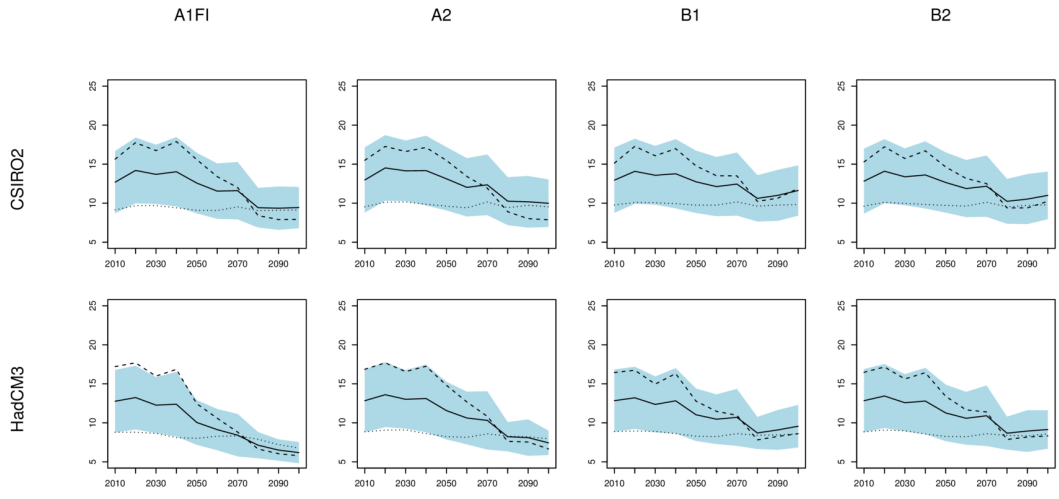


Fig. 5.7 – Predicted evolution of species richness for the current century under four different emissions scenarios. Solid line corresponds to the average species richness in the Iberian Peninsula with one standard deviation (blue band). Dotted and dashed lines represent the average value of species richness for locations in the first and forth quartiles of variance.

Velocity of changes in the past

The past changes on species numbers in the Iberian Peninsula were processed at different velocities. The period between 15,000 and 3,000 years BP has experienced dramatic climate oscillations, as part of the process of glacial to interglacial transition (see Section 4.1; Willis & MacDonald, 2011). The Iberian Peninsula underwent through these changes with a pronounced spatial dynamics (see Section 4.1) that are reflected in the different velocities at which species compositions operate. Most of Central Iberia underwent dramatic changes in this context, and areas of low velocity are scattered throughout the study area. Sandel *et al.* (2011) points to a maximum climate change velocity of 0.0148 km.yr^{-1} in the Iberian Peninsula which is faster than the average velocity of species composition change estimated in this study. This may indicate that species assemblages and individual species, since each one was modelled independently, had some degree of resilience to climate change in the past.

The climate in the Quaternary conducted species distributional shifts, finally shaping the current diversity pattern of reptiles and amphibians (Araújo *et al.*, 2008). The combination of two velocities provides insightful results: areas that preserve quite effectively the species richness are those retaining the original value (measured by Vs) nearby its location (measured by Vm). Areas that share these conditions are found mostly scattered around the coastal line, with few exceptions in the Central System mountain range. Ohlemüller *et al.* (2012), based on analogous climate locations between the current climate and at the LGM, classified most of southern Iberian Peninsula as high potential source of species for European colonizations. However, as also seen by the different velocities, the climate is a general proxy for species, indicating areas with suitable conditions that may hold species. Here, we follow a nearly opposite reasoning, by using species data and by assuming that areas with slower changes in species number are indicative of climate conditions that favoured species persistence and, thus, preserved overall diversity. Our results support not only the idea of southern Iberian Peninsula as a source area (i.e. with low velocities), but we also add the northern area, with lower velocities, as potential species source for expansion and recolonization processes.

The interest of the ecological niche modelling community to derive past distributions of species has gradually increased in the last few years (Waltari *et al.*, 2007; Svenning *et al.*, 2008; Graham *et al.*, 2010). These tools, especially when integrated with molecular analysis, have potential to provide spatial insights of putative refugia for species (Kozak *et al.*, 2008; Hickerson *et al.*, 2010) under a static (e.g. Waltari *et al.*, 2007; Svenning *et al.*, 2008) or dynamic perspective (e.g. Graham *et al.*, 2010). The

notion of velocity used here converges with the concept of species' refugia. Areas with low velocities provide locations where species numbers are maintained with few alterations, while those with higher velocities indicate places where species number fluctuates, and are compatible with expansion processes. The existence of cryptic refugia for reptiles and amphibians within a more broad European refugia (Gómez & Lunt, 2007) is supported by the velocities of species composition change. The contractions of the suitable area during harsh conditions of the past, favoured different species assemblages from those of the present (Hewitt, 2000; Svenning & Skov, 2007; Svenning *et al.*, 2008), probably contributing for higher richness of some areas. We show here that species richness has been oscillating with climate fluctuations, but with a general trend to a gradual decrease in the number of species sharing the same location, being particularly sharp in areas with higher species richness in the past (Fig. 5.4). The areas with high variance of richness are located mainly at low altitudes (below 500 m) which was expected due to harsh glacial conditions at higher altitudes. We believe that expansion processes may be also responsible for diluting species richness to current patterns. Historical climate has been shown to exert some control on the current tree species richness patterns (Svenning & Skov, 2007) and also in reptiles and amphibians in Europe (Araújo *et al.*, 2008), especially in those taxa with short-range dispersal abilities or narrow ranges. Our results show that these processes are dynamically controlled, with species richness patterns of reptiles and amphibians responding very rapidly to an oscillating climate. Major fluctuations in species number seem to be related to major climate events in the late-Quaternary. Rapid declines of species richness occurred during the warm Bølling-Allerød interstadial ($\sim$14,700 to 12,900 years BP), with a fast increase in the Younger Dryas ($\sim$12,900 to 11,600 years BP). The following warming period witnessed the most dramatic decrease in species number, which propagated through most of the Holocene, including the Holocene climatic optimum ($\sim$6,000 years BP). This reduction was only interrupted between 8,000 and 7,000 years BP, coinciding with the abrupt cold event at 8,200 years BP. Nevertheless, the past climate layers used in this study do not detected these events precisely, due to the 1,000 years time interval between them. However, they do depict the general spatial impact of temperature and precipitation abrupt changes may had in the Iberian Peninsula (see Section 4.1). Therefore, these events may have caused greater impacts in species richness than those predicted here. The described dynamic pattern follows the same reasoning raised by Nogués-Bravo *et al.* (2010), showing that evolutionary mechanisms (e.g. speciation, extinction and dispersal) driven by climate, shaped the actual richness patterns, although current mechanisms (e.g. human intervention, cur-

rent climate, land cover) should also be acknowledged with an important role in the current patterns of species richness (Moreno-Rueda & Pizarro, 2008; Sillero *et al.*, 2009).

Velocities for an uncertain future

The most evident pattern from predicted velocities of changes in species assemblages is the difference between models and scenarios. Such difference result from the high uncertainty related to future predictions, which vary greatly between combinations of models and scenarios. Therefore, necessary caution must be used when interpreting these data (Buisson *et al.*, 2010; Carvalho *et al.*, 2010).

Our models predict that much of southern Iberia will have higher velocities compared to the northern part. The CSIRO2 model, although predicted severe impacts, is less alarming than the predictions obtained when using the HadCM3 (Fig. 5.5). The combination of both velocities show that, although some southern locations will maintain the species richness nearby in the subsequent decades, long distances towards north will have to be travelled to find similar values to the current richness. Our results are coherent to those of other authors working with herptiles (Araújo *et al.*, 2006; Carvalho *et al.*, 2010), especially at the north-eastern part of the Iberian Peninsula where few changes are predicted to occur, both in species gain or loss. The southern part of Iberia, however, exhibit most differences, with reduced area with low velocities, encircled by an extensive area with high velocities with a trend of richness drop. These differences are likely to arise due to the distinct methods based on alternative predictions of future climate. Also the choice of niche modelling algorithm may affect the outcome (Pearson *et al.*, 2006).

The estimated velocities show a second order magnitude of difference from those predicted for the past. This implies that species in the next decades will have to move faster than they did during the Holocene to follow the pace of climate change in the following century. High differences were also predicted between future (Loarie *et al.*, 2009) and past (Sandel *et al.*, 2011) climate change velocity. Loarie *et al.* (2009) predicted lower global velocities for future climates with an intermediate scenario than those presented here for species composition change. This discrepancies may be due to different methods of velocity calculation and different combinations of models and scenarios, but also to the effect of using the species' climatic niche. Future climate forecasts predict major changes in a few decades, which may obligate to larger displacements, contributing to higher richness velocities. This same effect may be responsible for the opposite trend in the past, by providing some resilience to

climate change that operated on longer time scales. Nevertheless, species composition seems to adapt quite rapidly to climate oscillations, as seen from the relation of predicted species richness in the past. This opens a window of opportunity for amphibians and reptiles during the next century, but habitat destruction a high fragmentation levels might hamper dispersal and potentiates the impacts of climate change. Even in the least dramatic scenario, changes might involve loss of species due to extreme area reduction of potential niche. These dramatic results seem ubiquitous to studies analysing herptiles under future climate predictions with extreme dispersal scenarios (e.g. Araújo *et al.*, 2006; Carvalho *et al.*, 2010). The same predictions occur among other taxa with a possible widespread tree richness lost in Europe (Thuiller *et al.*, 2005) and restricted dispersal ability, hampering mammals to track climate change (Schloss *et al.*, 2012). Even long-dispersal mammals, as bats, showed that predicted ranges of species with strong affinities to Mediterranean climate might shift towards northern latitudes in Europe (Rebelo *et al.*, 2010).

The areas of low velocity are related to species refugia in the past. This raises a question: will these areas constitute refugia to a drastic warming in the next century? These areas are predicted to maintain species richness with few changes, and, thus, be more resilient to climate change. However, conservation should not be only focused in these regions. There is a general trend among scenarios and models to a drastic reduction of herptile richness in Iberia (Fig. 5.7). High velocities in most of the study area and the infinite values of Vs in the worst scenario, show that average to high values of richness will be lost in the next century, likely resulting in species lost. If the aim of conservation is to preserve most species, then these areas are indicative of where less effort may be needed to preserve a reduced number of species that are more resilient to climate warming. We stress that we provide broad hints of species extinctions or extreme contractions of distributions because the study area is limiting potential shifts to northern latitudes. Additionally, the absence of dispersal scenarios in the modelling process might overestimate distributional shifts. Also, metrics such as gain and loss of species should be assessed for conservation purposes (Thuiller *et al.*, 2005; Araújo *et al.*, 2006; Carvalho *et al.*, 2010).

5.1.6 Final remarks

We have shown here that using niche modelling algorithms with relatively few variables and two different measures of velocities generate good predictions of past evolution of species composition and, based on those, expect to have reliable results for the next century. Variables common for past and future periods are difficult to obtain with such

periodicity as we used here (every thousand years for the past and every decade for this century). These results should, however, be further tested with more predictors and algorithms. Also, considering the quantification of velocities along with species gain, lost and turn-over should provide detailed results to conservation decision making.

Velocities form the last 15,000 years BP and those predicted for the future century are difficult to compare due to the very different time periods they represent. Wider shifts are expected within each 1,000 years layers we used here to model past distributions, as response to climate dynamics at sub-millennial time scales (Hof *et al.*, 2011). However, the landscape has changed extensively over the last millenia, and the velocity that those changes operated was slower when compared to predicted future changes. A direct comparison of both velocities may be difficult to achieve, but it provides a clear hint of the magnitude of changes to come.

5.1.7 References

Alley, R.B., Marotzke, J., Nordhaus, W.D., Overpeck, J.T., Peteet, D.M., Pielke, R.a., Pierrehumbert, R.T., Rhines, P.B., Stocker, T.F., Talley, L.D. & Wallace, J.M. (2003) Abrupt climate change. *Science* **299**, 2005–2010.

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**, 1712–1728.

Araújo, M.B., Nogués-Bravo, D., Diniz-Filho, J.A.F., Haywood, A.M., Valdes, P.J. & Rahbek, C. (2008) Quaternary climate changes explain diversity among reptiles and amphibians. *Ecography* **31**, 8–15.

Brito, J.C., Acosta, A.L., Álvares, F. & Cuzin, F. (2009) Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African canids. *Biological Conservation* **142**, 3020–3029.

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

Carretero, M.A., Ayllón, E. & Llorente, G. (2011) *Lista patrón de los anfibios y reptiles de España*. Asociación Herpetológica Española, Madrid.

Carrión, J.S., Fernández, S., González-Sampériz, P., Gil-Romera, G., Badal, E., Carrión-Marco, Y., López-Merino, L., López-Sáez, J.a., Fierro, E. & Burjachs, F.

(2010) Expected trends and surprises in the Lateglacial and Holocene vegetation history of the Iberian Peninsula and Balearic Islands. *Review of Palaeobotany and Palynology* **162**, 458–475.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2010) From climate change predictions to actions - conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology* **16**, 3257–3270.

Cheddadi, R., Yu, G., Guiot, J., Harrison, S. & Prentice, I.C. (1997) The climate of Europe 6000 years ago. *Climate Dynamics* **13**, 1–9.

Costa, G.C., Wolfe, C., Shepard, D.B., Caldwell, J.P. & Vitt, L.J. (2008) Detecting the influence of climatic variables on species distributions: a test using GIS niche-based models along a steep longitudinal environmental gradient. *Journal of Biogeography* **35**, 637–646.

GDAL Development Team (2011) *GDAL - Geospatial Data Abstraction Library, Version 1.8.0*. Open Source Geospatial Foundation.

Gómez, A. & Lunt, D. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* (eds. S. Weiss & N. Ferrand), pp. 155–188, Springer Netherlands.

Graham, C.H., VanDerWal, J., Phillips, S.J., Moritz, C. & Williams, S.E. (2010) Dynamic refugia and species persistence: tracking spatial shifts in habitat through time. *Ecography* **33**, 1062–1069.

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

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.

Hickerson, M.J., Carstens, B.C., Cavender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F. & Yoder, A.D. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* **54**, 291–301.

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

Hof, C., Levinsky, I., Araújo, M.B. & Rahbek, C. (2011) Rethinking species' ability to cope with rapid climate change. *Global Change Biology* **17**, 2987–2990.

IPCC (2000) *Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

James, D.A. (2011) *RSQLite: SQLite interface for R*. URL `http://CRAN.R-project.org/package=RSQLite`.

Keitt, T.H., Bivand, R., Pebesma, E. & Rowlingson, B. (2012) rgdal: Bindings for the Geospatial Data Abstraction Library. URL `http://CRAN.R-project.org/package=rgdal`.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G.a. (2010) Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate* **23**, 2739–2758.

Kozak, K.H., Graham, C.H. & Wiens, J.J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23**, 141–148.

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.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature* **462**, 1052–1055.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

Loureiro, A., Ferrand, N., Carretero, M.A. & Paulo, O. (2008) *Atlas dos Anfíbios e Répteis de Portugal*. Instituto de Conservação da Natureza e Biodiversidade, Lisboa.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M. & New, M. (2004) A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100).

Montori, A., Llorente, G.A., Alonso-Zarazaga, M.A., Arribas, O., Ayllón, E., Bosch, J., Carranza, S., Carretero, M.A., Galán, P., García-París, M., Harris, D.J., Lluch, J.,

Márquez, R., Mateo, J.A., Navarro, P., Ortiz, M., Mellado, V.P., Pleguezuelos, J.M., Roca, V., Santos, X. & Tejedo, M. (2005) *Lista patrón actualizada de la herpetofauna española: Conclusiones de nomenclatura de anfibios y reptiles de España*. Aasociación Herpetológica Española, Barcelona.

Moreno-Rueda, G. & Pizarro, M. (2008) Relative influence of habitat heterogeneity, climate, human disturbance, and spatial structure on vertebrate species richness in Spain. *Ecological Research* **24**, 335–344.

Myers, N., Mittermeier, R., Mittermeier, C., Da Fonseca, G. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858.

Nogués-Bravo, D., Ohlemüller, R., Batra, P. & Araújo, M.B. (2010) Climate predictors of late quaternary extinctions. *Evolution* **64**, 2442–2449.

Ohlemüller, R., Huntley, B., Normand, S. & Svenning, J.C. (2012) Potential source and sink locations for climate-driven species range shifts in Europe since the Last Glacial Maximum. *Global Ecology and Biogeography* **21**, 152–163.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* **83**, 171–193.

Parmesan, C. (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* **37**, 637–669.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-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.

Peterson, A.T. (2007) Why not WhyWhere: The need for more complex models of simpler environmental spaces. *Ecological Modelling* **203**, 527–530.

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

Peterson, A.T., Papes, M. & Soberón, J. (2007a) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecology* **3**, 63–72.

Peterson, A.T., Papeş, M. & Eaton, M. (2007b) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**, 550–560.

Phillips, S.J. (2008) Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography* **31**, 272–278.

Pleguezuelos, J.M., Márquez, R. & Lizana, M. (2004) *Atlas y libro rojo de los anfibios y reptiles de España*. Organismo Autónomo de Parques Nacionales.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

Rebelo, H., Tarroso, P. & Jones, G. (2010) Predicted impact of climate change on European bats in relation to their biogeographic patterns. *Global Change Biology* **16**, 561–576.

Rödder, D. & Lötters, S. (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography* **18**, 674–687.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J. & Svenning, J.C. (2011) The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664.

Schloss, C.A., Nunez, T.A. & Lawler, J.J. (2012) Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8606–8611.

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

Sillero, N. (2011) What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling* **222**, 1343–1346.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2009) *ROCR: Visualizing the performance of scoring classifiers*. URL `http://CRAN.R-project.org/package=ROCR`.

Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology letters* **10**, 1115–1123.

Stockwell, D.R. (2006) Improving ecological niche models by data mining large environmental datasets for surrogate models. *Ecological Modelling* **192**, 188–196.

Svenning, J. & Skov, F. (2007) Ice age legacies in the geographical distribution of tree species richness in Europe. *Global Ecology and Biogeography* **16**, 234–245.

Svenning, J.C., Normand, S. & Kageyama, M. (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology* **96**, 1117–1127.

Tabor, K. & Williams, J.W. (2010) Globally downscaled climate projections for assessing the conservation impacts of climate change. *Ecological Applications* **20**, 554–65.

Tarroso, P., Carvalho, S.B.S. & Brito, J.C. (2012) Simapse - Simulation Maps for Ecological Niche Modelling. *Methods in Ecology and Evolution* **3**, 787–791.

Tarroso, P. & Rebelo, H. (2010) E-Clic - Easy climate data converter. *Ecography* **33**, 617–620.

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

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

Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* **9**, 137–152.

Waltari, E., Hijmans, R.J., Peterson, a.T., Nyári, A.S., Perkins, S.L. & Guralnick, R.P. (2007) Locating pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PloS one* **2**, e563.

Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* **3**, 260–267.

Willis, K.J., Bennett, K.D., Bhagwat, S.a. & Birks, H.J.B. (2010) 4 ^{o}C and beyond: what did this mean for biodiversity in the past? *Systematics and Biodiversity* **8**, 3–9.

Willis, K. & MacDonald, G. (2011) Long-Term Ecological Records and Their Relevance to Climate Change Predictions for a Warmer World. *Annual Review of Ecology, Evolution, and Systematics* **42**, 267–287.

Wilson, D. & Martinez, T. (2001) The need for small learning rates on large problems. *Proceedings of the International Joint Conference on Neural Networks*, pp. 115–119, IEEE.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693.

5.2 Hybridization at an ecotone: ecological and genetic barriers between three Iberian vipers[2]

5.2.1 Abstract

The formation of stable genetic boundaries between emerging species is often diagnosed by reduced hybrid fitness relative to parental taxa, due to endogenous or exogenous barriers to gene flow. Although the latter mechanism is difficult to detect in nature, the integration of molecular and ecological niche modelling tools can shed light on the relative contribution of ecological divergence. Here, we focus on a three-way contact zone between viper species (*Vipera aspis*, *V. latastei* and *V. seoanei*) that likely diverged through allopatry and have maintained secondary contacts over several climatic cycles. In a cross-disciplinary approach we examine 218 vipers at the High Ebro in Northern Spain, scoring morphological traits, genotyping ten microsatellite loci plus the mitochondrion, and analysing their georeferenced environments in a spatially explicit framework. The study area is recognized as a steep ecotonal region between Atlantic and Mediterranean biogeographic provinces, where the three taxa overlap and hybridize. Despite a high frequency of hybrids between the species pair (*V. aspis* / *V. latastei*), parental taxa seem to maintain their genetic, ecological and morphological distinctiveness outside of the ecotone. Moreover, we show that hybrids are spatially restricted to habitats that are sub-optimal for parental taxa, suggesting that niche separation and adaptation to an ecological gradient confers an important barrier to gene flow among taxa that did not achieve complete reproductive isolation.

5.2.2 Introduction

The observation of natural hybrids with reduced fitness seems to favour the generalized hypothesis that most hybrid zones, particularly in animals, are maintained by selection against hybrids irrespective of the habitat – i.e. post-mating endogenous barriers (Barton & Hewitt, 1989). Yet, in principle, environmental adaptation – i.e. pre- and post-mating exogenous barriers – may also hamper gene flow, by generating equilibrium between migration and natural selection acting along an environmental gradient (Endler, 1977; Doebeli & Dieckmann, 2003). Although it is well recognized that these two selective regimes might co-occur in natural populations, environmental adaptation is difficult to measure, particular in intergrading taxa that have not achieved

[2]This study is presently submitted to an international journal in the science-citation index: Tarroso, P., Pereira, R. J., Martínez-Freiría, F., Godinho, R. & Brito, J.C. Hybridization at an ecotone: ecological and genetic barriers between three Iberian vipers

complete reproductive isolation and occupy an environmental gradient. As a result, it remains unclear whether incipient levels of ecological divergence may result in stable genetic barriers between taxa.

Hybrid zones have been seen as windows on the evolutionary process (Harrison & Others, 1990), and have provided several examples of endogenous forces acting on species formation (Hewitt, 1988). However, hybrid zone studies are usually limited to a single species pair, and so do not directly allow insights into how the nature of species barrier might change over time since speciation. On the contrary, comparative approaches examining more than two hybridizing taxa (e.g. Salzburger *et al.*, 2002; Pinho *et al.*, 2008; Pereira & Wake, 2009) necessarily mix pairs that meet in very different ecological contexts. Study systems such as these are therefore powerful opportunities to examine the nature of species barriers. Moreover, to identify a potential contribution of exogenous barriers to genetic isolation between taxa, one needs to integrate the molecular inference of gene flow with ecological analysis of parental and hybrid individuals in a spatially explicit framework.

The recent increase in availability of environmental variables at higher resolutions and of species distributions data (see review in Kozak *et al.*, 2008) have contributed to an explosion of spatially explicit analyses addressing evolutionary questions that are central for our understanding of the speciation process. Assessing the evolutionary implications of niche divergence remains an important topic of discussion (Nosil, 2012). The observation that allopatric divergence often occurs in ecologically distinct habitats (Mayr, 1963; Wiens & Graham, 2005; Wiens, 2004), and that incipient species often meet in patchy, transitional habitats (Harrison, 1986) suggests that ecological divergence is relevant to speciation both in allopatry and parapatry. Identifying different axes of ecological divergence between parental taxa, and assessing how those axes of exogenous selection may impact hybrid fitness, can provide important insights into this question. Ecological niche-based models are useful to evaluate relationships between species divergence and surrounding environments, but its application under this scope rely mostly on presence-only data (Swenson, 2006; Chatfield *et al.*, 2010; Culumber *et al.*, 2012) or require the classification of continuous data into discreet entities (e.g. population structure to species' presence locations) at a local scale. Thus, current ecological models have limited power to detect niche segregation between closely related taxa that intergrade with each other (e.g. Chatfield *et al.*, 2010). A study of an ecological model for hybridizing taxa would benefit from considering the full gradient from one taxon to the other and thus, methods that accept continuous data will result in more accurate models and will aid in depicting ecological

divergence.

The Iberian Peninsula offers a remarkable opportunity to study barriers between incipient species because of its rich and well-understood biogeographic history. First, paleogeographic events that took place in the Miocene (Duggen *et al.*, 2003) favoured pathways for biodiversity exchange with North-Africa and the posterior isolation led to high levels of species divergence (e.g. Velo-Antón *et al.*, 2012). Second, several lineages evolved in multiple micro-refugia within the Iberian refugium during the glacial and inter-glacial cycles of the Pleistocene (Gómez & Lunt, 2007), resulting in multiple closely related lineages with varying degrees of genetic divergence and leading to a network of secondary contact zones with genetic interactions spanning from full sympatry without gene flow to wide range introgression (e.g. *Podarcis* lizards Pinho *et al.*, 2009). Third, the complex topography of the Iberian Peninsula results in very distinct bioclimatic regions that likely affected both the divergence in the allopatric refugia, where lineages adapted to different environmental settings, and the areas of secondary contact, which often coincide with steep ecotones between the major bioclimatic provinces (Sillero *et al.*, 2009). Such biogeographic dynamics resulted in multiple taxa with varying degrees of genetic and ecological divergence that meet and hybridize, and as such provide powerful opportunities to examine the nature of species barriers.

The three viper species occurring in the Iberian Peninsula - two sibling Mediterranean species (*Vipera aspis* and *V. latastei*) and a third phylogenetically distant Euro-Siberian species (*V. seoanei*) - exhibit varying degrees of morphological, behavioural, ecological and genetic divergence (Saint-Girons, 1980). In general, European vipers diverged allopatrically with significant niche divergence acting as a complete pre-mating barrier (Saint-Girons, 1980; Lenk *et al.*, 2001; Garrigues *et al.*, 2005). The lack of sympatry among European viper species precludes testing whether ecological divergence contributes to maintaining species boundaries, or whether ecological divergence is a by-product of species formation in allopatry. However, for the Iberian species, niche segregation is not complete, since taxa meet upon secondary contact and often overlap (see Duguy *et al.*, 1979; Martínez-Freiría *et al.*, 2008, 2009), generating conditions for the occurrence of gene flow. Along the High course of the Ebro river (hereafter High Ebro), in northern Spain, the two sister species were found in sympatry within a steep ecotone between the Atlantic and Mediterranean climatic provinces, but in allopatry with the third, phylogenetically more divergent, species (Martínez-Freiría *et al.*, 2006, 2008). Subsequent work at this contact zone identified morphologically intermediate individuals relative to parental (extreme) forms, suggest-

ing that the sister species hybridize within the ecotone (Martínez-Freiría *et al.*, 2009). Moreover, significant spatial and temporal niche partitioning and reproductive success differences between parental taxa and hybrids suggested that both exogenous and endogenous selection might contribute to the establishment of species boundaries (Martínez-Freiría *et al.*, 2010). However, an assessment of the genetic structure of viper populations in the contact zone and its possible correlation with environmental variables to investigate the role of the ecological divergence is still lacking.

Here, we focus on the contact zone between the three Iberian vipers to examine how ecological divergence might contribute to the establishment of stable species boundaries. Integrating genetic and geostatistical tools in a spatially explicit framework, we ask: 1) what is the genetic structure of the focal taxa (i.e. genetic differentiation, spatial segregation, gene flow between parental lineages)?, 2) is such genetic structure associated with ecological segregation between taxa?, and 3) is the ecological transition associated with barriers at multiple species traits (mitochondrial, nuclear and morphological)?

5.2.3 Material and Methods

Study area, sampling survey and morphology

The study area of the High Ebro in northern Spain (Fig. 5.8) covers a total of 792 UTM 1 km^2 squares. The landscape is topographically complex, ranging from 590 to 1250 m, with deep valleys excavated by the Ebro River and its tributaries, Rudrón and Sedanillo. The area encompasses the transition between Euro-Siberian and Mediterranean bioregions. The climate is sub-humid Mediterranean with Central European tendency, which is marked by a steep environmental transition (e.g. in precipitation and temperature; Hijmans *et al.*, 2005). The sampling design was based on 1km^2 UTM grids throughout the study area. Between March 2004 and October 2007 we sampled a total of 242 squares (30.5% of the study area), covering the entire range of environmental variability (Fig. 5.8).

A total of 218 vipers found by visual encounter surveys or random road kills were sampled and their geographical location recorded with a Global Positioning System (Fig. 5.8). From these, 210 were genotyped for nuclear and 211 for mitochondrial DNA (203 individuals were common to both datasets). Specimens were preliminarily classified as *V. aspis*, *V. latastei*, *V. seoanei* or as V. sp. (intermediate individuals between *V. aspis* and *V. latastei*) according to the following taxonomically diagnostic traits (n=477; see Martínez-Freiría *et al.*, 2006, 2009): snout elevation, number of apical scales, shape of the dorsal stripe, and number of dorsal marks. Small portions

of tissue (either tail-tips or ventral muscle) were preserved in 96% alcohol for DNA extraction.

DNA extraction, Genetic markers development and genotyping

Total genomic DNA was extracted using the QIAamp DNA Micro kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. A set of 10 unlinked microsatellite loci originally developed for *Vipera berus* (Carlsson *et al.*, 2003; Ursenbacher *et al.*, 2009) were successfully amplified on our three target species in two multiplex reactions (MP1: loci 3, 21, 37, 71 and Vb-B18; MP2: loci 11, 64, Vb-A8, Vb-B10 and Vb-D17). PCR amplifications were performed using the Multiplex PCR Master Mix (Qiagen) on a MyCycler thermocycler (BIO-RAD) based on the Master Mix manufacture instructions and setting the annealing temperature to 54°C (MP1) or 56°C

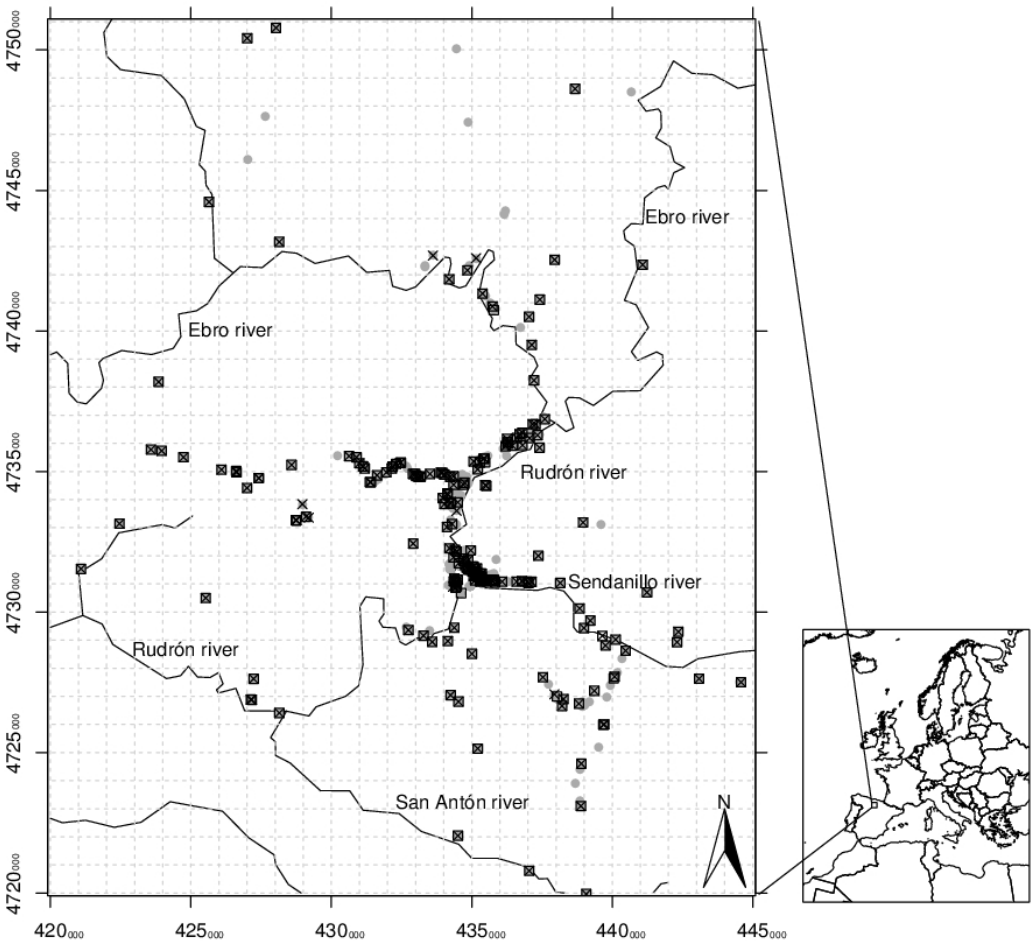


Fig. 5.8 – Study area and sample distribution. The gray dots denote all samples collected in the field (n=477). Black open squares indicate samples used in mtDNA analyses (n=211) and black crosses are samples used in nuclear analyses (n=210). Most samples for mtDNA and nuclear analyses are the same (n=203).

(MP2). Amplifications were performed in a final volume of 10 μl using approximately 10 ng of DNA, 5 μl of Qiagen PCR MasterMix and 1 μl of primer mix. All reactions were accomplished with three primers for each locus, following the M13-tailed primer method (Oetting *et al.*, 1995). Four different universal M13 primers labelled with 6-FAM, VIC, NED and PET were used in the reactions according to the selected dye for each locus at the same concentrations as the reverse primers. Forward primers were 10-fold less concentrated. Fragments were separated together with an internal size standard (Genescan-500 LIZ, ABI) in an ABI3130xl automatic sequencer. Samples were screened using GeneMapper 4.1 software (Applied Biosystems) and alleles were called based on sizing bin windows and re-checked manually. To account for potential null alleles due to the use of non-specific primers, we tested for significant deviations from Hardy-Weinberg equilibrium and heterozygote deficiency across the three taxa using a subset of 12 reference individuals for each species collected further away from the contact zone (appendix Table E). A final group of eight loci (3, 11, 21, 64, 71, Vb-A8, Vb-B10, Vb-B18) were used on the remaining analysis.

mtDNA amplification and the RFLPs

An 846 bp mtDNA sized fragment, including most of the cytochrome b (cyt b) gene, was initially PCR amplified for 18 individuals of the three species (n=5 for *V. aspis* and *V. seoanei* and n=8 for *V. latastei*). Amplifications were performed using primers gluDG and cb3 (Palumbi, 1996) and DNA polymerase EcoTAQ according to manufacturer's instructions. The reaction mixture included 0.1U of Taq, 3mM MgCl2, 12 pmol each primer, 0.2mM of each dNTP and approximately 10 ng of genomic DNA. PCR was carried out in a BioRad MyCycler thermocycler over 38 cycles with annealing temperature set at 46ºC. The resulting single PCR bands were sequenced for both strands with the PCR primers using the BigDye chemistry (Applied Biosystems) and sequencing products were separated in a 3130 XL Genetic Analyzer (Applied Biosystems). Sequences were manually aligned using BioEdit 7.0.9 (Hall, 1999) and the Restriction Map subroutine of the same software was used to locate the site gain or loss for each endonuclease profile. Diagnostic profiles for the three species were obtained with the endonucleases NlaIII and XspI. Although a single enzyme would be enough to distinguish the three species, a second enzyme was used as a control. All samples were PCR amplified for the same cyt b fragment followed by digestion at 37ºC for 3h for both endonucleases. Restriction Fragment Length Polymorphism (RFLP) patterns were visualized under UV light after electrophoretic separation on 2.5% agarose gels for NlaIII or under silver staining after separation on polyacrylamid gel (T9C5) for XspI.

Population structure and gene flow

Genotypes were probabilistically assigned to three clusters with no prior cluster membership using Structure 2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). This Bayesian approach minimizes genetic disequilibria within clusters, identifying groups of randomly mating individuals, i.e. populations or species. We run five pseudo-replicates for 106 iterations after a 105 length burn-in, to assure convergence on the same result. Given the presence of morphologically intermediate individuals (i.e. putative hybrids) in our sampling we chose to run a model allowing for admixture and correlated allele frequencies. The number of K clusters that achieved maximum likelihood was tested. We recorded the cluster membership probabilities (CMP) for each individual and their confidence intervals and imported into R software (R Development Core Team, 2012) for further spatial analysis.

Inverse Distance Weighted (IDW) interpolation implemented in the gstat package (Pebesma, 2004) was used to produce a continuous surface over the study area, from the projected CMP values for each individual. This surface simplifies the visual evaluation of the spatial patterns of population structure and allows testing for correlations with environmental parameters.

Although all our spatial methods take advantage of the full continuum observed between the three taxa and measured by the assignment test, for a simplified visual representation of the nuclear DNA data, a threshold of 0.75 was chosen to distinguish pure from hybrid individuals for each population. In addition, the diagnostic morphological traits were used to assign the genetic clusters to the three known species.

Environmental analysis

Ecological niche-based models were developed for each taxon using the artificial neural networks (ANN) algorithm implemented in Simapse (Tarroso *et al.*, 2012). The machine learning algorithm implemented in this package finds the ecological model supporting a given population structure using a training dataset with population affinity scores and a set of ecogeographical variables (EGV) representing potential environmental features acting in the ecological divergence between parental forms. After the learning process, the trained network is capable of predicting an output to other locations, based on the same set of predictors. In contrast with commonly used presence-only models, the key aspects of the Simapse model are: 1) uses the full continuum of membership to any given number of taxa (or clusters) and thus does not require categorization of individuals into discrete taxa; 2) has optimum performance with non-linear relationships between dependent and independent variables, even in the presence of

Table 5.1 – Environmental factors used in the ecological niche modelling of *Vipera aspis*, *Vipera latastei* and *Vipera seoanei* in the High Ebro contact zone.

Environmental factors	Abbr.	Range and units	Source
Distance to Mixed Forests	DMF	0 to 4000 m	GLC (2003)
Distance to Needle-leaved Evergreen Forests	DNEF	0 to 8544 m	GLC (2003)
Distance to Broadleaved Deciduous Forests	DBDF	0 to 6708 m	GLC (2003)
Distance to Humid Areas	DHA	0 to 5657 m	GLC (2003)
Distance to Herbaceous Vegetation	DHV	0 to 22627 m	GLC (2003)
Distance to Broadleaved Deciduous - Thicket	DBDT	0 to 17889 m	GLC (2003)
Slope	SLP	5 to 45 %	USGS (2004)
Annual Precipitation	AP	698 to 881 mm/year	Hijmans et al., (2005)
Minimum Temperature of the Coldest Month	MTCM	-1.3 to 2.0 ºC	Hijmans et al., (2005)

noise due to the ANN engine (Olden *et al.*, 2008); and 3) uses a very flexible algorithm supporting several degrees of complexity in the model (Olden *et al.*, 2008).

A set of weakly correlated EGVs ($r < 0.45$ in all cases; Table 5.1) were chosen based on their known influence on the viperid snakes distribution (Brito & Crespo, 2002; Santos *et al.*, 2006; Martínez-Freiría *et al.*, 2008), These EGVs fall into three main categories (Table 5.1): 1) topographical – slope derived from the SRTM digital elevation model (`http://srtm.usgs.gov`); 2) climatic – precipitation and temperature grids extracted from Worldclim v1.4 (Hijmans *et al.*, 2005); and 3) habitats – a set of six EGVs representing the distance to a land cover category. The land cover is composed of 14 months (1999-2000) satellite data acquired by the VEGETATION sensor on-board of SPOT 4 satellite (`http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php`). All variables are continuous and have a grid resolution of 1 km^2.

The CMP for the three populations as given by Structure was directly input to Simapse with coordinates. We built twenty five model replicates for each population to yield a consensus model with standard deviation. Each replicate was based on a network with a single hidden layer with 5 nodes; both the learning rate and momentum were set to 0.1, with 250 iterations. The learning rate controls the amount of adjusting of the node connections in the learning process, while the momentum controls the influence of previous history adjustments on the current iteration (Tarroso *et al.*, 2012). These two parameters are important to find a stable solution and to avoid local minima. From the 210 available samples, 25% were selected with random sub-sampling for testing the network and to select the iteration with the best fit, while avoiding overfitting. To test the overall prediction ability of the models, we made linear regressions between CMPs and model predicted values and tested the significance of the predictor with ANOVA. We also tested the predictions against the 477 individuals mor-

phologically attributed to the three parental classes and a fourth group representing intermediate forms (Martínez-Freiría *et al.*, 2009). The mean and standard deviation over replicates were computed for each taxon. The former represents the potential distribution, whereas the latter gives an indication of the uncertainty per pixel between all predictions. The importance of each EGV to the final model was computed based on the squared sum of the partial derivatives of the fully trained network. Response curves were analysed to describe the behaviour of predictions with respect to each EGV (Austin, 1987). Similar profiles for different species would indicate a shared resources usage, i.e. possible sympatry between species, while different responses would be consistent with parapatry or allopatry.

Integration of genetic, morphological and environmental data

The models for each taxon were classified based on the previously used threshold for CMPs, after calibration with the equation derived from the linear regression. This process allowed the threshold to be optimized to predicted values, thus defining an area of presence more accurately. After classification, all models were combined into a single descriptive spatial layer. This was achieved by assigning sequential integers to the area of each pure parental form (*V. latastei*=1, *V. aspis*=2, *V. seoanei*=3). Areas where more than one parental was found were assigned to a fourth category. Areas without classification after this process, representing locations where most non-pure parental forms are found, were accepted as corresponding to the hybrid zone ecotone.

Proportions of individual counts per taxon, plus hybrids were computed for each map cell with species information. This process was repeated for species identification based on each of the three diagnostic traits (nuclear, mitochondrial and morphological). This allowed a visual diagnosis of the spatial distribution resulting from classification based on the different traits and its relation to the models of ecological divergence.

All graphical and text outputs were done with R (R Development Core Team, 2012) using the packages rgdal (Keitt *et al.*, 2012) with the Geospatial Data Abstraction Library (GDAL Development Team, 2011) for spatial analysis, and RSQLite (James, 2011) for interaction with the SQLite (`http://www.sqlite.org`) database holding all data. Additional automations were done using Python (`http://www.python.org`).

Table 5.2 – Assignment of analysed individuals to each cluster after classification of CMPs. Counts are shown based on each trait. Assignments to populations based on nuclear data shown here are average values (and standard deviation) for the individuals in the cluster.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Morphology (total)	**77**	**103**	**11**	**14**
V. latastei	65	5	0	3
V. aspis	2	80	0	4
V. seoanei	0	0	11	0
V. sp	10	18	0	7
mtDNA (total)	**75**	**101**	**13**	**14**
V. latastei	67	18	1	11
V. aspis	7	82	1	3
V. seoanei	1	1	11	0
Nuclear (total)	**78**	**105**	**13**	**14**
Pop 1 mean (sd)	0.968 (0.038)	0.033 (0.050)	0.013 (0.020)	0.500 (0.136)
Pop 2 mean (sd)	0.024 (0.036)	0.958 (0.051)	0.007 (0.003)	0.452 (0.184)
Pop 3 mean (sd)	0.008 (0.007)	0.008 (0.007)	0.980 (0.021)	0.048 (0.081)

5.2.4 Results

Population structure and gene flow

The nuclear DNA data revealed a well defined population structure in three clusters, with higher and lower number of clusters receiving lower likelihood values (appendix E.1, E.2), which match the three morphologically distinct species (Table 5.2, Fig. 5.9). *V. latastei* was attributed to the first population cluster with 78 individuals (84% of the individuals previously identified morphologically), *V. aspis* was attributed to the second population cluster with 105 individuals (78% of the individuals previously found), and *V. seoanei* was attributed to the third cluster with 13 individuals (all individuals identified morphologically).

The individual assignment showed a continuum of membership that is consistent with hybridization between the three species (Fig. 5.9; Table 5.2). When using a threshold of 0.75, 14 intermediate individuals were found, one being intermediate between *V. latastei* and *V. seoanei*, and 13 between the sister species *V. latastei* - *V. aspis*. From the dataset used for mtDNA analysis (n=211), we identified 103 individuals as sharing a mitochondrial lineage corresponding to *V. latastei*, 95 to *V. aspis* and 13 to *V. seoanei*. There was high concordance with identification based on the morphologic traits, however, the cluster assigned as *V. aspis* had 18 individuals carrying *V. latastei* mtDNA (Table 5.2).

Environmental analysis

The ecological models depicted the distribution patterns of the three viper species in the study area and exhibited a limited overlap between areas of suitability (Fig. 5.10). Accordingly, the models showed steep transitions between areas of high and low suitability for the three taxa. The pure parental forms meet in a narrow ecotone that is characterized by environmental features that do not meet the full requirements of any parental.

A good predicting ability was achieved based on the independent dataset, with good discrimination of each morphologically attributed species in the respective model (appendix E.3). Moreover, linear regressions between CMPs and predicted values have demonstrated that the latter is a significant predictor of the former ($p<0.01$ for all models; appendix E.4). Individuals identified as intermediate forms are classified with average prediction values for *V. latastei* and *V. aspis* ($Q1_{Vlat}$ = 0.482, $Q3_{Vlat}$ = 0.536; $Q1_{Vasp}$ = 0.431, $Q3_{Vasp}$ = 0.524), whereas for *V. seoanei* most of these individuals have very low values ($Q1_{Vseo}$ = 0.014, $Q3_{Vseo}$ = 0.017). Standard deviation between model replicates of *V. latastei* and *V. aspis* showed small differences, providing a support of the models (Fig. 5.10). On the other hand, *V. seoanei* models exhibited more spatial discordance, with a fragmented area of low standard deviation and high prediction values.

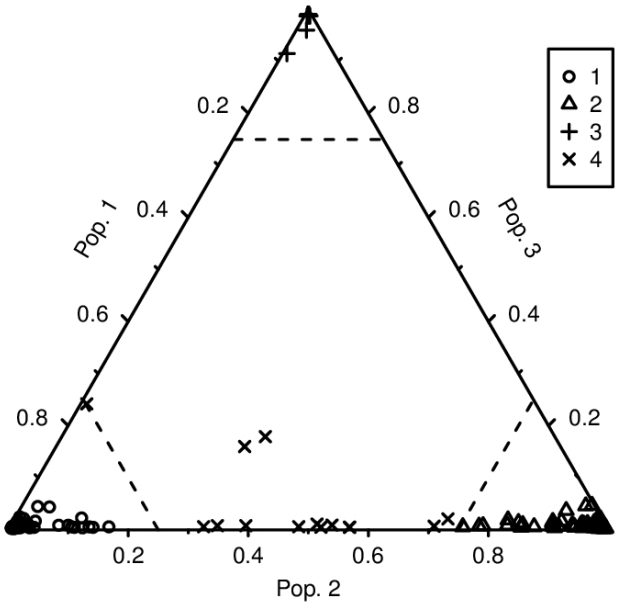


Fig. 5.9 – Ternary plot of structure cluster membership probabilities (CMPs) to each population based on microsatellites from 210 individuals. Dashed lines represent the threshold 0.750, chosen to determine pure individuals. Four sets of data were obtained: one for each parental form (1 to 3 in the figure) and a forth for hybrids (4).

Considering the three multivariate models for each species, we observe that the EGVs have different contributions for each (Fig. 5.11). The annual precipitation (AP) provides high degree of explanation for all models, however, its importance is highlighted in the *V. latastei* model. In this model, AP is the EGV that most contributes, all others having negligible contributions. Distance to humid areas (DHA) and to evergreen forests (DNEF) have high importance to *V. aspis* models, followed by AP. The models suggest that the three species are characterized by strong ecological segregation (Fig. 5.12). The minimum temperature of the coldest month (MTCM) and distance to broadleaved deciduous thicket (DBDT) are mostly related to the distribution of *V. seoanei*. Areas with low precipitation and lower temperatures are preferentially occupied by *V. latastei*, whereas humid and warm areas are more frequently occupied by *V. aspis* and *V. seoanei*. *V. aspis* occupies humid areas near evergreen forests and *V. seoanei* tends to be absent from broadleaved deciduous thicket.

In contrast, the areas where most hybrids occur rarely covered the optimum range of either parental form, and thus are likely ecologically sub-optimal for these (Fig. 5.12). Considering the individual response curves of the EGVs that are most important to predicting habitat suitability for parental taxa provides some insights on the characterization of the ecotone and how it might limit the distribution of the two taxa that hybridize most frequently, *V. aspis* and latastei. Hybrid individuals occurred in an ecotonal region, characterized by some EGVs that either are sub-optimal for both parentals or that affected parental taxa in opposite directions; i.e. while some variables favoured the occurrence of one parental, other variables favour the alternative parental (Fig. 5.12). Concerning DNEF and DBDT, hybrids were found in areas with less suitable values for parental forms, with a tendency of *V. aspis* pure individuals to occupy

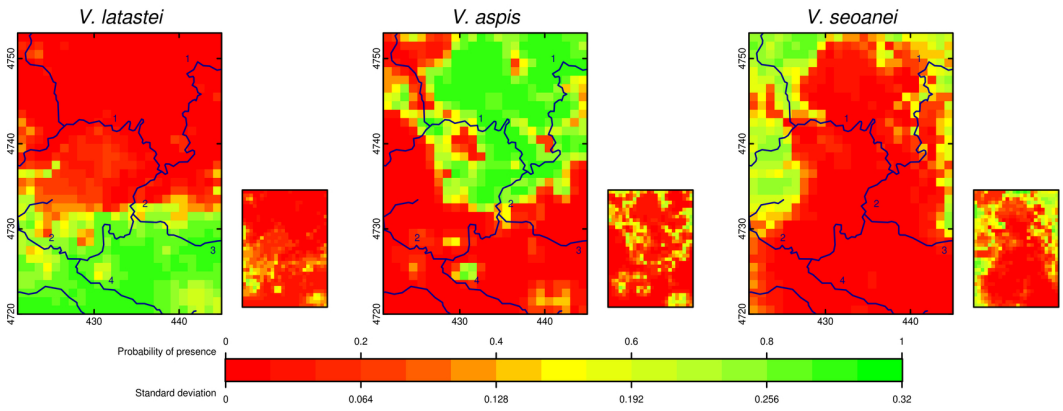


Fig. 5.10 – Ecological niche-based models built with cluster membership probabilities (CMPs). The three models represent the probability of presence for each parental form. Smaller maps at the lower right side of each model represent the standard deviation between replicates.

favourable range of the EGVs. An opposite trend was observed for AP, where hybrids and *V. latastei* pure individuals occurred in areas with low precipitation. For DHA and MTCM, hybrids were found in areas with less optimum values for both parental forms.

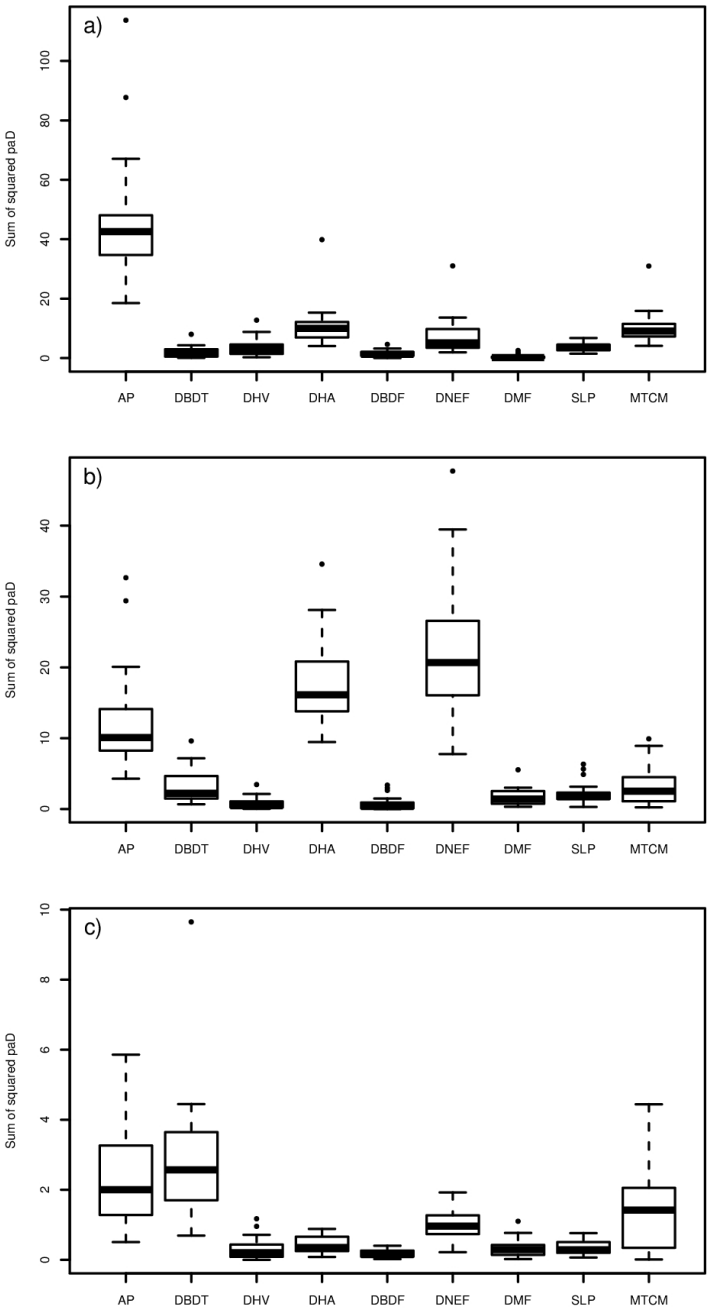


Fig. 5.11 – Importance of environmental variables (EGV) for each ecological niche-based model. Scores of variable importance are based on the squared sum of partial derivatives for each EGV in a trained model. Higher values indicate more influence of the variable to define the final model.

Integration of genetic, morphologic and environmental data

The suitable area for each parental was found after applying a threshold (Fig. 5.13) based on the value chosen for CMPs (0.750, Fig. 5.9), and calibrated using the linear equation for each model (appendix E.2; VL=0.730, VA=0.734, VS=0.674). The geographic variation in nuclear DNA, mitochondrial DNA, and morphological traits revealed a high spatial concordance with the suitable areas predicted by the ecological models (Table 5.3, Fig. 5.13). The area identified by the ecological model after classification for *V. latastei*, *V. aspis* and *V. seoanei* extended over 240, 272 and 86 cells, respectively. The area identified as non suitable for any parental extended over 183 cells, where most hybrids identified by nuclear or morphological data were found (Table 5.3, Fig. 5.13). The transition from *V. latastei* to *V. aspis* habitat occurs within 2km length where the highest densities of individuals were found. Suitable habitat for combinations of two parental species (*V. latastei*-*V. aspis*, *V. latastei*-*V. seoanei*, and *V. aspis*-*V. seoanei*; Fig. 5.13) was found in 11 cells; however, no samples were collected in these locations.

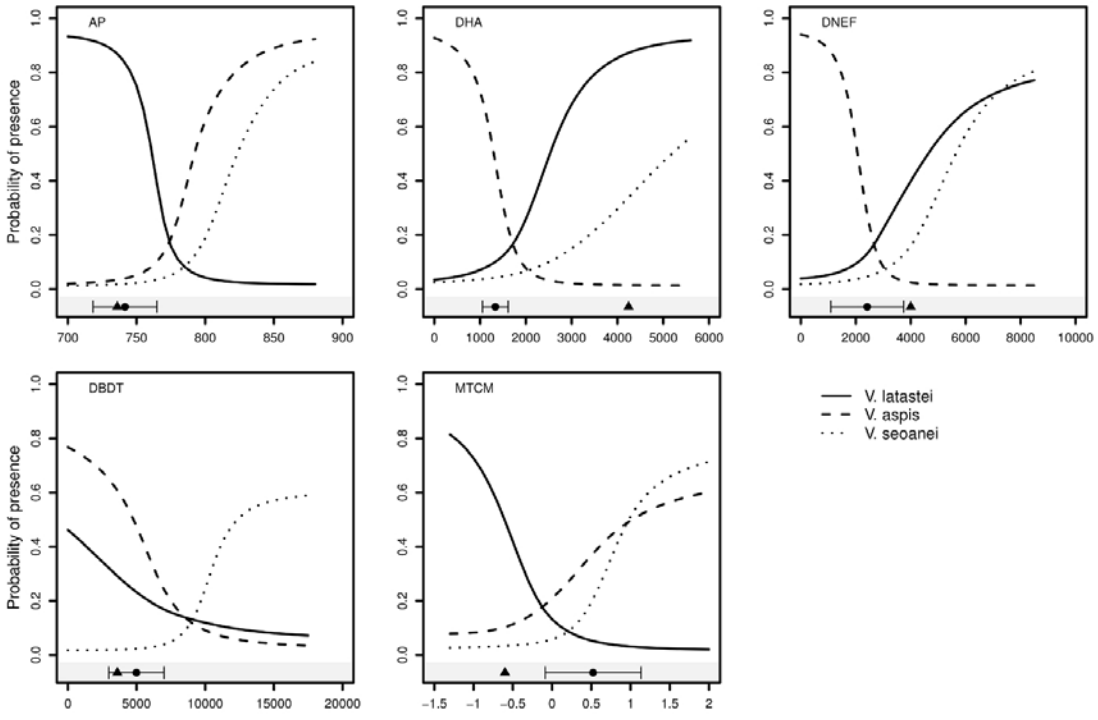


Fig. 5.12 – Response curves for the five most contributing environmental variables (EGV). In the bottom grey rectangle is described the relative position of the hybrids in each EGV. Black circle is the average value and the horizontal lines represent one standard deviation (n=13). The black triangle represents the position in the EGV of the single hybrid with trace of *V. seoanei* genome. See table 5.1 for EGVs details.

Table 5.3 – Number of individuals assigned to each species, plus hybrids accordingly to each trait studied and per area attributed in the combined model. Individual counts are shown in the order nuclear / mitochondrial / morphological-based identification. Rows correspond to the individual identifications and columns to the areas of the combined model for each species plus the ecotone where most hybrids are found.

Individuals/Area	*V. latastei*	*V. aspis*	*V. seoanei*	Ecotone
V. latastei	45 / 48 / 86	02/08/04	0 / 0 / 0	31 / 47 / 91
V. aspis	2 / 2 / 6	70 / 65 / 144	0 / 1 / 0	33 / 27 / 46
V. seoanei	0 / 0 / 0	0 / 1 / 2	11 / 10 / 11	2 / 2 / 1
Hybrids	2 / - / 4	3 / - / 11	0 / - / 0	9 / - / 80

5.2.5 Discussion

Integrative methods for the study of hybrid zones

Until very recently we could not address the contribution of ecological divergence in maintaining species boundaries (Kozak *et al.*, 2008). The recent development of new methods and the increasing available data on public databases offer an opportunity to overcome these limitations, so that we can quantify ecological divergence at various stages of species formation. This approach is particularly instructive when applied to the study of hybrid zones as they provide the opportunity to test if exogenous selection by the environment is correlated with barriers to gene flow. Although some efforts have been made in this direction (e.g. Martínez-Freiría *et al.*, 2010; Chatfield *et al.*, 2010; Culumber *et al.*, 2012), current commonly applied methods are designed to operate with discrete units and do not properly characterize the gradient observed at hybrid zones. Here we show that using the full continuous data on population membership resulting from a population structure analysis renders suitable ecological niche-based models (ENM) for species with incomplete reproductive isolation, allowing insights into the contribution of ecological divergence for the establishment of stable species boundaries.

The method we use here presents several advantages over other commonly used procedures. First, by using the population affinity score of each individual rather than the presence of each species, our models for the three viper species were able to include all sampled locations available in the study area. The bias resulting from different sample sizes is, thus, eliminated and the comparability between models is increased. Second, by using a continuous dataset of population assignment rather than a categorization in subjectively defined species, we were able to produce detailed ecological models directly from the population structure, acknowledging the full genetic continuum that characterizes hybrid zones. This procedure also reduces the

uncertainty of the models by eliminating the need to derive (pseudo) absences (Lobo *et al.*, 2010). Third, our method defines the hybrid zone without requiring a model based on the locations of the hybrid individuals. An alternative to our approach is to build separate models for intermediate forms (e.g. Culumber *et al.*, 2012). Yet, this assumes an a priori relation of equilibrium between hybrids and the environments where they are found (Guisan & Zimmermann, 2000), and thus attributes an ecological niche to hybrids even when it might not be biologically meaningful. While this alternative is most appropriate when hybrid genotypes thrive in certain habitats, such as in cases of bounded hybrid superiority (Moore, 1977) or hybrid speciation, we have no support for such phenomenon in the viper system as in the great majority of well-described hybrid zones. Fourth, our approach offers the possibility of using a small-scale area to build ENM to the study of hybrid zones rather than requiring consideration of the entire distribution of parental taxa (e.g. Chatfield *et al.*, 2010). Although approaches considering the complete range of species are appropriate to build the full niche of the

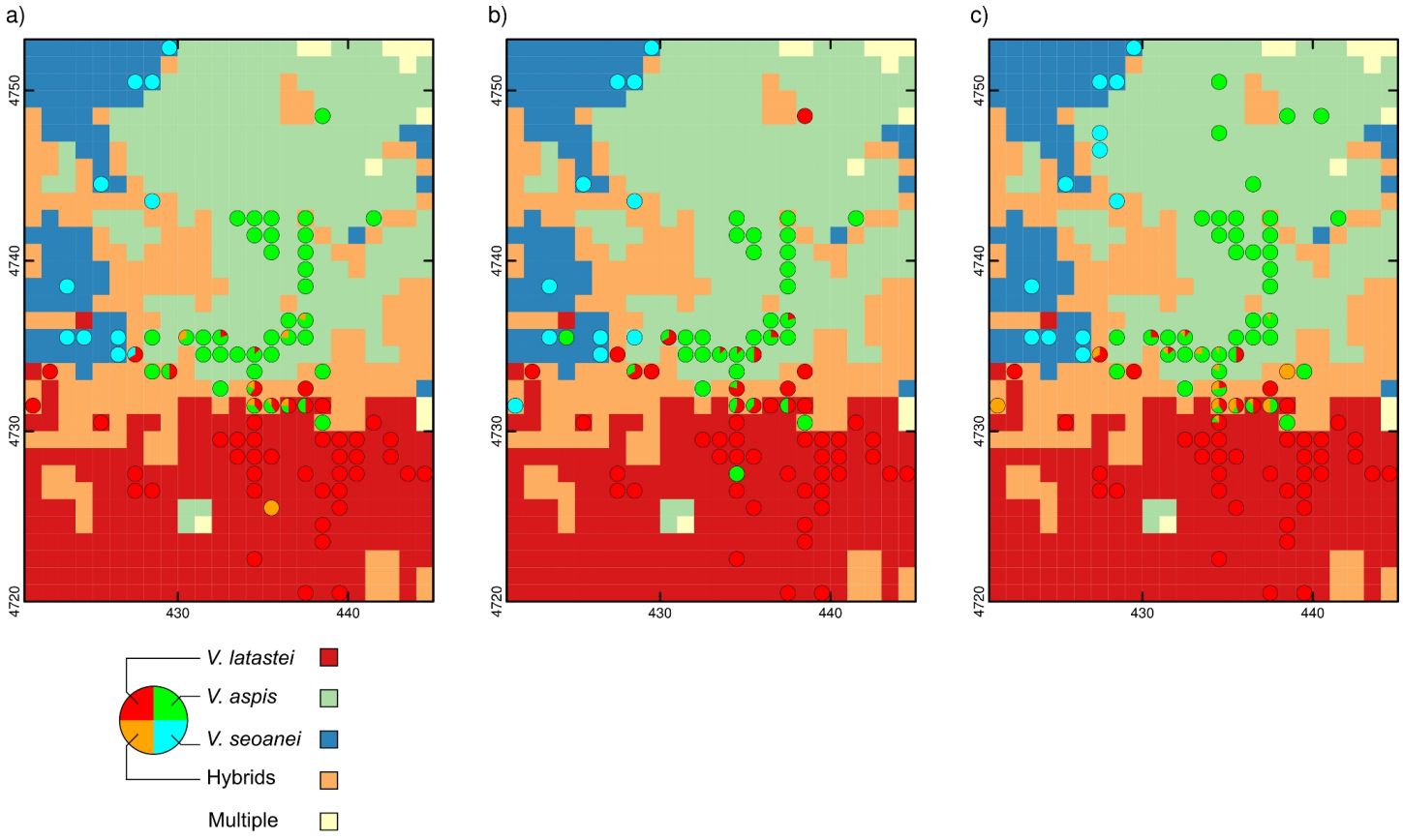


Fig. 5.13 – Combined ecological niche-based models traits. Pie charts represent the proportion of the number of individuals in each cell, as classified by nDNA (a), mtDNA (b), and morphological traits (c). Each parental area is defined after threshold classification of the original model.

species and to guarantee the transferability of the model (Thuiller *et al*., 2004), species at range margins are not necessarily dominated by the same ecological mechanisms like those at the core of the distributions (Bridle & Vines, 2007). The several micro evolutionary processes that may constitute barriers to gene flow in hybrid zones (e.g. selection against hybrids, assortative mating) are relevant at a local population scale rather than the entire species-range scale, and competition between parental forms at the sympatry area may be blurring the expression of the niche (e.g. Saint-Girons, 1975; Martínez-Freiría *et al*., 2010). Fifth, although other algorithms may be used to build models relying on the continuous data (e.g GLMs), the machine learning algorithm we use here is able to detect other relationships between the population structure and environmental variables that are not linear (Olden *et al*., 2008).

A limitation of operating at the spatial scale relevant to the micro-evolutionary processes governing barriers to gene flow is that the spatial resolution of the ecological analysis should match the spatial scale of the organism, as defined by its dispersal ability and home range. In these viper species, we expect that a better definition of ecological segregation at the hybrid zone will result from a finer resolution of the environmental variables, since the average home range of these vipers is small (less than 0.015 km^2 for *V. latastei* and *V. aspis* during the period of maximum activity; Martínez-Freiría *et al*., 2010). The resolution used here (1 km^2) resulted in the grouping of a large number of individuals that were sampled within the same cell, with contrasting population affinity values promoting additional complexity to model fitting. Thus, an ideal model for this organism would require environmental variables at a spatial resolution that was not available. Nevertheless, the resolution used here is approximate to that ideal one, and demonstrates how this method could be applied to other species with higher dispersal, for which environmental data at 1 km^2 resolution is already available worldwide.

Genetic structure of the Iberian vipers

While European viper species have experienced extensive ecological divergence and currently have allopatric or parapatric distributions, the Iberian vipers *Vipera latastei*, *V. aspis*, and *V. seoanei* overlap spatially in a few contact zones (Duguy *et al*., 1979; Martínez-Freiría *et al*., 2008, 2009). One of these occurs along the High Ebro, where morphologically intermediate individuals were found (Martínez-Freiría *et al*., 2006, 2008, 2009). These conditions provide an opportunity to ask whether ecological divergence contributes to stable boundaries between species. However, until now we lack a test of i) the genetic differentiation between parental taxa, ii) the potential hybrid origin

of the morphologically intermediate individuals, and iii) whether hybrids occur at the habitat preferred by parental taxa. The genetic characterization of this triple contact zone allowed us to address these caveats and establish the background to explore the processes maintaining species boundaries.

Our molecular results showed that these three taxa are genetically distinct in both mitochondrial and nuclear markers. Based on the eight microsatellite loci, we found that the pattern of genetic variability of the contact zone is best explained by three genetic clusters (Suppl. Mat. E.1). Moreover, individuals collected away from the contact zone have a high membership coefficient for the respective population cluster (average pp reference individuals > 0.966). By comparing the molecular results to morphological data we show that these three genetic groups are in strong agreement with the morphologically distinct species, indicating that these genetic markers provide adequate power to distinguish the three species, and to distinguish individuals of pure and hybrid origin (Table 5.2).

In agreement with expectations, results show that the morphologically intermediate individuals are often genetically admixed (pp<0.75), and likely result from hybridization between parental species and further backcrossing, providing an opportunity for gene flow between these three taxa. In addition, our results show that hybridization is common between the sister species *V. latastei* and *V. aspis*, and thus, the contact zone between these two taxa is better described as a genetic and morphological gradient between both species. In contrast, hybridization seems to be rare between phylogenetically divergent species (one hybrid between *V. latastei* and *V. seoanei*, Fig. 5.9), confirming that elevated genetic and ecological divergence between the Euro-Siberian *V. seoanei* and the two other Iberian species might preclude genetic interaction and opportunity for gene flow (Bea, 1985; Brito & Crespo, 2002).

Despite the relative high frequency of hybrids in the contact zone between sister species *V. latastei* and *V. aspis*, these admixed individuals are spatially restricted to the area of sympatry between the two taxa, defining a narrow hybrid zone (Fig. 5.10). This genetic continuum and confined distribution of hybrids is consistent with strong barriers to gene flow, and thus provide an opportunity to evaluate the contribution of exogenous and endogenous selection in maintaining species boundaries.

Ecological segregation and species boundaries

The study of hybrid zones, particularly those between several species pairs, can provide important insights into the nature of species barriers and how it might change over the course of speciation (Barton & Gale, 1993). The triple contact between the Iberian

vipers provide an opportunity to ask whether the ecological segregation observed between hybridizing taxa contribute for maintaining the integrity between species.

Our results showed that the three species are ecologically highly divergent, despite not having achieved complete reproductive isolation. Our ecological models show that the suitable habitat for each parental species shows little spatial overlap with each other (Fig. 5.10) and is defined by different sets of explanatory environmental variables (Fig. 5.11). A close analysis of the environmental variables that are most important for each species and the range of values that each taxon is most associated with (Fig. 5.12) offer some insights on the possible niche partitioning between these three species. Annual precipitation (AP) is a major predictor of habitat suitability for the three vipers studied here (Fig. 5.11). Yet, *V. latastei* is mostly associated with drier areas while the other two species are associated with wetter habitats (Fig. 5.12). Other two variables (DHA and DNEF) are important for the definition of *V. aspis* suitable area, with lower importance for the other two species (Fig. 5.11). Likewise AP, the response curves of these two variables show that different viper species are associated with opposite ranges of the same ecological variables (Fig. 5.12). Important variables for *V. seoanei* (DBDT and MTCM) show that this species has a particular habitat that is either unique or related to that of *V. aspis*. Our results suggest that ecological segregation is strong even between the most phylogenetically related taxa *V. latastei* and *V. aspis*, which are frequently associated with opposite values of the same environmental variables. Together, these results suggest that the evolution of parental taxa has already resulted in important ecological trade-offs enhancing the spatial exclusion and constraining the distributions to spatially non-overlapping ranges.

The sister species *V. aspis* and *V. latastei* hybridize frequently in the contact zone, providing an opportunity to test if the ecological segregation observed here might contribute to barriers between species. Thus, it is important to consider the areas where hybrids are located, and how suitable those areas are for the respective parental taxa. Our results show that the hybrid zone between these species is coincident with a steep ecotone, characterized by a sharp transition of suitability between both species (Fig. 5.10) which is evident in the environmental variables that most contributed to define the distribution of parental species (AP, DHA, DNEF; Fig. 5.12). Hybrids are found in areas where the habitat is sub-optimal for parental taxa, either because the relevant environmental variables are unsuitable for either one or both parental species (Fig. 5.12). Although the ecotone extends over 183km^2, secondary contact only occurs where parental species reach high densities, forming a narrow hybrid zone of 4km^2. This contrasts with the slightly wider hybrid zone predicted by ENM

using morphology Martínez-Freiría *et al.*, 2008. This difference, however, probably results from the distinct methods applied in each case; the model using the presences of individuals will generate wider sympatric areas due to the effect of the pure parental forms that are found in the hybrid zone.

The association between neutral genetic structure and different habitats (e.g. mosaic hybrid zones; Harrison, 1986; Nosil *et al.*, 2007) can provide evidence that ecological divergence underlies species formation (Nosil, 2012). Because our ecological models derives from the population structure diagnosed using microsatellite loci, an association between the ecological explanation and the nuclear markers is expected and required for high model performance. Thus, other independent species traits are necessary to validate this hypothesis relating ecological divergence and barriers between species. Our results show that both mitochondrial lineages and morphological traits have steep transitions coincident with the ecotone, with sympatric lineages and intermediate individuals occurring within areas that are less suitable for pure parental taxa (Fig. 5.13). While these observations are largely consistent with an important role of ecological divergence as a significant barrier at several species traits that would otherwise flow across taxa, it is not possible to reject that post-mating endogenous barriers might also contribute for maintaining species integrity. Clines maintained purely by endogenous selection are expected to drift in space and be trapped in population density troughs, such as ecotones (Barton & Gale, 1993), and are indistinguishable from clines maintained by exogenous selection (Kruuk *et al.*, 1999). Noticeably, although the mitochondrial transition is steep and coincident with the ecotone (Fig. 5.13b), it appears to introgress asymmetrically towards the northern parental *V. aspis* (Fig. 5.13; Table 5.3). This pattern is commonly observed in cytonuclear incompatibilities (Turelli & Moyle, 2007) that may quickly arise after short periods of allopatric divergence (Rand *et al.*, 2004; Burton & Barreto, 2012). A more extensive sampling of these genetic transitions and cline analysis would allow testing this hypothesis.

Our observation of a barrier at several diagnostic species traits coincident with habitat preference is consistent either with selection against hybrids (exogenous and endogenous; Barton & Gale, 1993) or with hybrid superiority in habitats that are suboptimal for parental taxa (Endler, 1977; Moore, 1977). Field studies (Martínez-Freiría *et al.*, 2010) showed that hybrid individuals at the ecotone have some demographic advantages (i.e. high abundance and low road mortality) but are reproductively less fit than parental taxa (i.e. lower number of embryos with smaller litters), suggesting that both processes may be acting in this hybrid zone. More field studies and experimental data are needed to accurately address these competing hypotheses.

Together, our results suggest that the ecological segregation of parental taxa is acting as a strong barrier to gene flow, despite the lack of reproductive isolation. Niche segregation is likely acting both as a pre-mating barrier, manifested by the habitat preference of parental individuals, and as a post-mating barrier, during habitat selection on immigrant individuals (pure or of hybrid origin). It is impossible to estimate how much of this barrier is due to endogenous (e.g. genetic incompatibilities) or exogenous (local adaptation) selection, but our results suggest that when genetic divergence is accompanied by strong ecological segregation, both barriers will be strongly expressed in ecotones, and independent species traits might be maintained despite high local rates of hybridization.

5.2.6 Acknowledgments

This study was partially supported by project POCTI/BIA-BDE/55596/2004 from Fundação para a Ciência e Tecnologia (FCT, Portugal). PT, FMF, RG, and JCB are supported by FCT (SFRH/BD/42480/2007, SFRH/BPD/69857/2010, SFRH/BPD/36021/2007, and Programa Ciência 2007, respectively). Acknowledgments extended to “Asociación Sociocultural Hoces del Alto Ebro y Rudrón” (Burgos, Spain) and friends who helped with the field work at the High Ebro, to Susana Lopes for the laboratory support, to Stuart Baird for the insightful discussions about this particular hybrid zone, and to Tessa Pierce for the manuscript revision and the helpful comments.

5.2.7 References

Austin, M.P. (1987) Models for the analysis of species’ response to environmental gradients. *Vegetatio* **69**, 35–45.

Barton, N.H. & Gale, K.S. (1993) Genetic analysis of hybrid zones. *Hybrid zones and the evolutionary process* (ed. R.G. Harrison), pp. 13–45.

Barton, N.N. & Hewitt, G.M.G. (1989) Adaptation, speciation and hybrid zones. *Nature* **341**, 497–503.

Bea, A. (1985) La repartición de las viboras *Vipera aspis* (Linnaeus, 1758) y *Vipera seoanei* (Lataste, 1879), en el Pais Vasco. *Ciencias Naturales* **2**, 7–20.

Bridle, J.R. & Vines, T.H. (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution* **22**, 140–147.

Brito, J. & Crespo, E. (2002) Distributional analysis of two vipers (*Vipera latastei* and *V. seoanei*) in a potential area of sympatry in the Northwestern Iberian Peninsula. *Biology of the Vipers* (eds. G. Schuett, M. Höggren, M. Douglas & H. Greene), pp. 129–138, Eagle Mountain Publishing, Utah.

Burton, R.S. & Barreto, F.S. (2012) A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Molecular Ecology* **21**, 4942–4957.

Carlsson, M., Isaksson, M., Höggren, M. & Tegelström, H. (2003) Characterization of polymorphic microsatellite markers in the adder, *Vipera berus*. *Molecular Ecology Notes* **3**, 73–75.

Chatfield, M.W.H., Kozak, K.H., Fitzpatrick, B.M. & Tucker, P.K. (2010) Patterns of differential introgression in a salamander hybrid zone: inferences from genetic data and ecological niche modelling. *Molecular Ecology* **19**, 4265–4282.

Culumber, Z.W., Shepard, D.B., Coleman, S.W., Rosenthal, G.G. & Tobler, M. (2012) Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: *Xiphophorus*). *Journal of Evolutionary Biology* **25**, 1800–1814.

Doebeli, M. & Dieckmann, U. (2003) Speciation along environmental gradients. *Nature* **421**, 259–264.

Duggen, S., Hoernle, K. & Bogaard, P.V.D. (2003) Deep roots of the Messinian salinity crisis. *Nature* **422**, 602–606.

Duguy, R., Martínez-Rica, J.P. & Saint-Girons, H. (1979) La répartition des vipères dans les Pyrénées et les régions voisines du nord de l'Espagne. *Bulletin de la Société d'Histoire Naturelle de Toulouse* **115**, 359–377.

Endler, J.A. (1977) *Geographic Variation, Speciation and Clines.(MPB-10)*, vol. 10. Princeton University Press.

Falush, D., Stephens, M. & Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–87.

Garrigues, T., Dauga, C., Ferquel, E., Choumet, V. & Failloux, A.B. (2005) Molecular phylogeny of Vipera Laurenti, 1768 and the related genera *Macrovipera* (Reuss, 1927) and *Daboia* (Gray, 1842), with comments about neurotoxic *Vipera aspis aspis* populations. *Molecular Phylogenetics and Evolution* **35**, 35–47.

GDAL Development Team (2011) *GDAL - Geospatial Data Abstraction Library, Version 1.8.0*. Open Source Geospatial Foundation.

Gómez, A. & Lunt, D. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia* (eds. S. Weiss & N. Ferrand), pp. 155–188, Springer Netherlands.

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

Hall, T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98. *Nucleic Acids Symposium Series* pp. 95–98.

Harrison, R.G. & Others (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology* **7**, 69–128.

Harrison, R. (1986) Pattern and process in a narrow hybrid zone. *Heredity* **56**, 337–349.

Hewitt, G. (1988) Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution* **3**, 158–167.

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

James, D.A. (2011) *RSQLite: SQLite interface for R*. URL `http://CRAN.R-project.org/package=RSQLite`.

Keitt, T.H., Bivand, R., Pebesma, E. & Rowlingson, B. (2012) rgdal: Bindings for the Geospatial Data Abstraction Library. URL `http://CRAN.R-project.org/package=rgdal`.

Kozak, K.H., Graham, C.H. & Wiens, J.J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23**, 141–148.

Kruuk, L., Baird, S., Gale, K. & Barton, N. (1999) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**, 1959–1971.

Lenk, P., Kalyabina, S., Wink, M. & Joger, U. (2001) Evolutionary relationships among the true vipers (Reptilia: Viperidae) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **19**, 94–104.

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

Martínez-Freiría, F., Santos, X., Pleguezuelos, J.M., Lizana, M. & Brito, J.C. (2009) Geographical patterns of morphological variation and environmental correlates in contact zones: a multi-scale approach using two Mediterranean vipers (Serpentes). *Journal of Zoological Systematics and Evolutionary Research* **47**, 357–367.

Martínez-Freiría, F., Sillero, N., Lizana, M. & Brito, J.C. (2008) GIS-based niche models identify environmental correlates sustaining a contact zone between three species of European vipers. *Diversity and Distributions* **14**, 452–461.

Martínez-Freiría, F., Brito, J.C. & Avia, M.L. (2006) Intermediate forms and syntopy among vipers (*Vipera aspis* and *V. latastei*) in Northern Iberian Peninsula. *Herpetological Bulletin* pp. 14–18.

Martínez-Freiría, F., Lizana, M., do Amaral, J.P. & Brito, J.C. (2010) Spatial and temporal segregation allows coexistence in a hybrid zone among two Mediterranean vipers (*Vipera aspis* and *V. latastei*). *Amphibia-Reptilia* **31**, 195–212.

Mayr, E. (1963) *Animal species and evolution.* Cambridge,Harvard University Press.

Moore, W. (1977) An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* **52**, 263–277.

Nosil, P. (2012) *Ecological speciation*. Oxford Series in Ecology and Evolution, Oxford University Press.

Nosil, P., Vines, T. & Funk, D. (2007) Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**, 705–719.

Oetting, W., Lee, H., Flanders, D., Wiesner, G., Sellers, T. & King, R. (1995) Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* **30**, 450–458.

Olden, J.D., Lawler, J.J. & Poff, N.L. (2008) Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* **83**, 171–193.

Palumbi, S.R. (1996) Nucleic acids II: the polymerase chain reaction. *Molecular Systematic* (eds. D.M. Hillis, C. Moritz & B.K. Mable), pp. 205–247, Sinauer Associates, Sutherland, Mass.

Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**, 683–691.

Pereira, R.J. & Wake, D.B. (2009) Genetic leakage after adaptive and nonadaptive divergence in the *Ensatina eschscholtzii* ring species. *Evolution* **63**, 2288–2301.

Pinho, C., Kaliontzopoulou, a., Carretero, M.a., Harris, D.J. & Ferrand, N. (2009) Genetic admixture between the Iberian endemic lizards *Podarcis bocagei* and *Podarcis carbonelli* : evidence for limited natural hybridization and a bimodal hybrid zone. *Journal of Zoological Systematics and Evolutionary Research* **47**, 368–377.

Pinho, C., Harris, D.J. & Ferrand, N. (2008) Non-equilibrium estimates of gene flow inferred from nuclear genealogies suggest that Iberian and North African wall lizards (*Podarcis* spp.) are an assemblage of incipient species. *BMC Evolutionary Biology* **8**, 63.

Pritchard, J., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rand, D.M., Haney, R.a. & Fry, A.J. (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology & Evolution* **19**, 645–653.

Saint-Girons, H. (1975) Coexistence de *Vipera aspis* et de *Vipera berus* en Loire-Atlantique: un probleme de competition interspecifique. *La terre et la Vie* **29**, 590–613.

Saint-Girons, H. (1980) Biogéographie et évolution des vipéres européennes. *Comptes Rendus de la Société de Biogéographie* **496**, 146–172.

Salzburger, W., Baric, S. & Sturmbauer, C. (2002) Speciation via introgressive hybridization in East African cichlids? *Molecular Ecology* **11**, 619–625.

Santos, X., Brito, J.C., Sillero, N., Pleguezuelos, J.M., Llorente, G.A., Fahd, S. & Parellada, X. (2006) Inferring habitat-suitability areas with ecological modelling techniques and GIS: A contribution to assess the conservation status of *Vipera latastei*. *Biological Conservation* **130**, 416–425.

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

Swenson, N.G. (2006) Gis-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *Journal of Evolutionary Biology* **19**, 717–725.

Tarroso, P., Carvalho, S.B.S. & Brito, J.C. (2012) Simapse - Simulation Maps for Ecological Niche Modelling. *Methods in Ecology and Evolution* **3**, 787–791.

Thuiller, W., Brotons, L., Araújo, M.B.M. & Lavorel, S. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* **27**, 165–172.

Turelli, M. & Moyle, L.C. (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics* **176**, 1059–1088.

Ursenbacher, S., Monney, J.C. & Fumagalli, L. (2009) Limited genetic diversity and high differentiation among the remnant adder (*Vipera berus*) populations in the Swiss and French Jura Mountains. *Conservation Genetics* **10**, 303–315.

Velo-Antón, G., Godinho, R., Harris, D., Santos, X., Martínez-Freiría, F., Fahd, S., Larbes, S., Pleguezuelos, J. & Brito, J. (2012) Deep evolutionary lineages in a Western Mediterranean snake (*Vipera latastei*/*monticola* group) and high genetic structuring in Southern Iberian populations. *Molecular Phylogenetics and Evolution* **65**, 965–973.

Wiens, J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution* **58**, 193–197.

Wiens, J.J. & Graham, C.H. (2005) Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics* **36**, 519–539.

Chapter 6

General Discussion

> But actually it is the whole (four-dimensional) pattern of the 'phenotype', the visible and manifest nature of the individual, which is reproduced without appreciable change for generations, permanent within centuries - though not within tens of thousands of years - and borne at each transmission by the material structure of the nuclei of the two cells which unite to form the fertilized egg cell. That is a marvel - than which only one is greater; one that, if intimately connected with it, yet lies on a different plane. I mean the fact that we, whose total being is entirely based on a marvellous interplay of this very kind, yet possess the power of acquiring considerable knowledge about it.
>
> — ERWIN SCHRÖDINGER, *What is Life?*

6.1 Applications for ecological niche modelling

The general interest on ecological niche modelling (ENM) tools has been increasing in the last decade as a result of its application to a wide range of questions than those originally suited to this tool. From the estimation of potential species' distributions and quantification of environmental response curves, the application has spread to diverse areas of research, including the study of potential impacts of climate change (e.g. Araújo *et al.*, 2006; Carvalho *et al.*, 2010), species invasiveness (e.g. Peterson & Vieglais, 2001; Rödder & Lötters, 2009), and niche dynamics in an evolutionary context (e.g Peterson *et al.*, 1999; Kozak & Wiens, 2006). We have shown here that this vast range of applications still offers a large opportunity for innovation, providing new perspectives on the study of macro- and micro-scale biological processes.

This general interest is accompanied by an extraordinary increase of software

packages availability for ENMs. The classical approaches based mainly on linear and logistic regressions have expanded to other algorithms, such as the ecological niche factor analysis (Hirzel *et al.*, 2002) or machine learning algorithms (e.g. GARP, Stockwell & Peters, 1999 and Maxent, Phillips & Dudík, 2008). In this scope, we provide a new application for ENM based on artificial neural networks (ANN; section 3.2). Although this method does not present an innovation per se (the beginnings of ANN date as far as 1943, with the first description of the biologically inspired artificial neuron, McCulloch & Pitts, 1943), our package offers a comprehensive solution for using this algorithm with species data and provides state of the art sensitivity analysis of the ANN model in a spatially explicit framework. Other available ENM packages including ANN require some degree of computer programming skills (e.g. BIOMOD, Thuiller, 2003) hampering its application to conservation solutions outside an academic background. Nevertheless, in recent years there has been a proliferation of studies within applied conservation using more intuitive packages (e.g. Peterson *et al.*, 2004; Kozak & Wiens, 2006; Swenson, 2006; Brito *et al.*, 2009; Rödder & Lötters, 2009; Chatfield *et al.*, 2010; Rebelo *et al.*, 2010; Culumber *et al.*, 2012). As a tentative to reduce the uncertainty associated to each model, the usage of several algorithms to build a single consensus model has recently increased in the scientific literature (Thuiller, 2003; Araújo & New, 2007). The combination of several algorithms in one package poses, however, an additional problem: each algorithm has a set of sensitivity analyses that are unique to it (e.g. partial derivatives for ANN) and the generalization of these packages is often done at expense of this procedures. Single algorithm packages (e.g Simapse, MaxEnt or ENFA) usually provide specific techniques for the analysis of the output. Therefore, we expect a general adoption of Simapse, especially under an ensemble forecasting perspective, for conservation purposes and evolutionary studies.

The accuracy of Simapse models was exhaustively tested here. Elith *et al.* (2006) did an extensive review of ecological modelling methods concluding that the newest algorithms, most related to machine learning, outperform the others. However, they did not consider ANNs, despite being previously included in multi-algorithm packages like BIOMOD (Thuiller, 2003). Moreover, the method performance was based on the area under the receiver operating characteristic curve (AUC) that was later criticized due to its dependency on the method and due to inheritance of the uncertainty related to absence data (Peterson *et al.*, 2007; Lobo *et al.*, 2007). Simapse models were tested thoroughly with virtual and real species representing different ecological requirements, and within a framework containing positive and negative controls, allowing a complete quantification of model accuracy under different degrees of network

complexity. We show that AUC value was dependent on ecological requirements of the species as suggested before (Lobo *et al.*, 2007), but lower AUC values do not necessarily stand for *bad* models. Addictionaly, for presence-only models, AUC was shown to be dependent on network complexity and data availability, especially for generalist species. The inclusion of a random species, without any relation with the predictors used, constituted a negative control to model building, demonstrating that Simapse does find a relation between species data and predictors used, and its high predictive power does not fit random noise.

On the side of predictor variables, the technological advances are towards an increase in the diversification of variables available, higher spatial resolution and improvements on models predicting future climates. These immense data are usually stored under specific formats that hampers the direct use in conservation studies addressing climate change impacts. The current trend in the ecological modelling community is to provide huge databases with data stored in common formats (e.g. WorldClim, Hijmans *et al.*, 2005 and CliMond, Kriticos *et al.*, 2011). Our objective in the development of E-Clic tool was to provide an easy access to climate predictions for the current century, avoiding duplication of databases where only the format changes. The dataset used as E-Clic input includes observed and predicted climate variables (Mitchell *et al.*, 2004) that were proven useful to assess impacts of climate change by means of ENM (Araújo *et al.*, 2006; Rebelo *et al.*, 2010; Xavier *et al.*, 2010). We expect that the availability of these data in ready-to-use formats will make the application more common.

The two software packages we describe here are open-source and freely available. Although both packages offer a graphical interface to easy the interaction with the software, the availability of the source code is an important feature in the scientific process allowing testing and, more importantly, reproducibility of the results (Ince *et al.*, 2012).

6.2 Macro-scale: Integrating space and time in biodiversity studies

The main achievement of the current work is the integration of space and time in the study of the Iberian biodiversity patterns (Fig. 6.1). This resulted from the combination of all data and methods presented in the chapters 3 and 4. At a regional scale, our study benefit from the immense potential resulting from the combination of palaeoecological data and ecological modelling (Anderson *et al.*, 2006; Nogués-Bravo, 2009).

This integration led to a complete description of the past dynamics in the Iberian herpetofauna and predicted future scenarios of change, by calculating two velocity metrics that reflect the changes in species richness.

In order to analyse past distributions, we have created a set of spatial layers representing past climate variables reconstructed from a biological proxy. The combination of improved methods and spatial interpolation to build spatial climate layers for every 1,000 years are, to our knowledge, integrated for the first time in a comprehensive study on the dynamic patterns of biodiversity. Sandel *et al.* (2011) studied the past dynamics using the LGM and present climates, without including the climate variation within this period. We have shown that the variation had strong impacts over wide areas, driving contractions and expansions of species ranges, ultimately leading to different velocities of species richness change. Moreover, we show that, as suggested by Loarie *et al.* (2009), the velocities of climate change do not directly relate to impacts on species due to the climate niche breath: high velocities of climate change will force migration or led to extinction only if the change of climate is larger than the tolerance breath of the species.

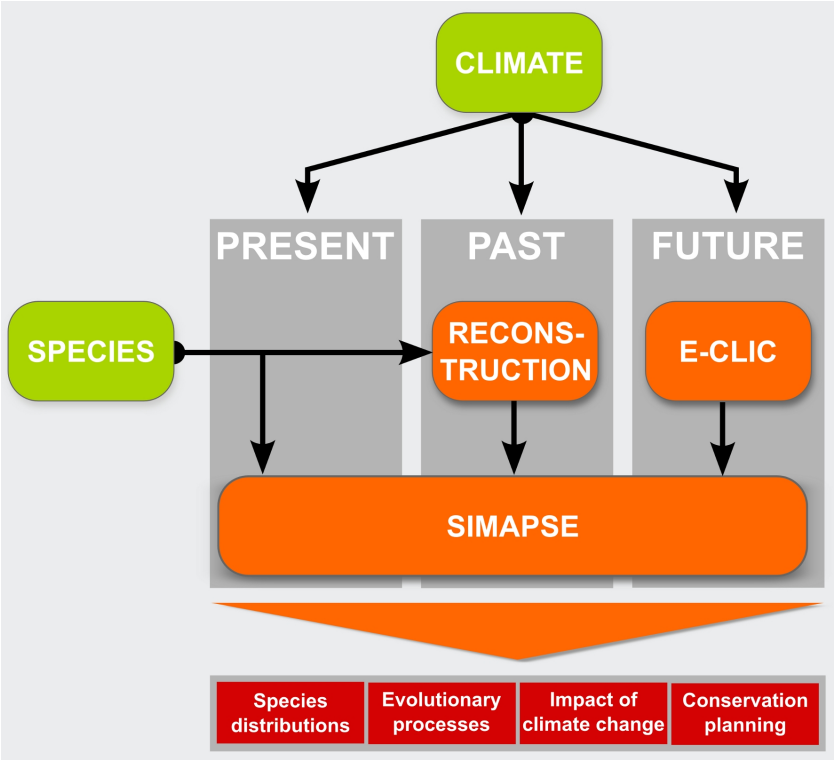


Fig. 6.1 – Integration of multidisciplinary data in different scales of space and time. Data from species' distributions and climate data (green boxes) were used in an integrative framework as input to Simapse to assess biodiversity patterns at macro- and micro-scales in different periods (dark grey boxes). Orange boxes represent the position of data and methods developed in the present work. Possible applications of this framework are highlighted in red.

Our work introduces two new metrics of velocity that can be applied to the study of species compositional changes. The change in species richness is measured as the nearest distance that richness values migrates each temporal layer to maintain the value for each location (Vm); and reflecting the shortest path that the richness value traveled during the time frame (Vs). Low velocities depicted areas where species composition had fewer changes over the period between 15,000 and 3,000 years BP, and high velocities showed areas with frequent changes in species richness suggesting common migration routes for herpetofauna. Higher velocities due to species migratory responses to a changing climate may led to secondary contacts between reptiles and amphibians lineages that previously diverged in the glacial refugia (Alexandrino *et al.*, 2000; Sequeira *et al.*, 2005; Godinho *et al.*, 2006; Gómez & Lunt, 2007). These glacial refugia have been identified with molecular studies and correspond to areas of high genetic diversity mostly located in the surroundings of low velocities. These findings support the hypothesis of a network of micro-refugia within the broader European refugia, such as the case of the Iberian Peninsula proposed by Gómez & Lunt, 2007. Unfortunately, our climate data covers only the last 15,000 years BP due to the restricted availability of fossil pollen data covering longer time ranges limiting, thus, the discussion about glacial refugia. However, the temperature rising characteristic of the late-Quaternary occurs after 15,000 years BP in Iberia (Naughton *et al.*, 2007) and the abrupt changes observed since then in Europe, and particularly in Iberia as shown in chapter 4, might have forced extreme range contractions (Hewitt, 1996; Naughton *et al.*, 2007).

The relation of past changes and predicted future changes is difficult to establish. There is evidence of high resilience to extreme temperature changes in the past (Willis *et al.*, 2010), and also that adaptation to current climate change is already happening (Parmesan, 2006). However, predictions indicate dramatic changes in a few decades that will pose major challenges to species (Loarie *et al.*, 2009). We show that changes to come affecting herpetofauna species richness will pose many conservation issues, even in the best case scenario, corroborating other ecological modelling studies (Araújo *et al.*, 2006; Carvalho *et al.*, 2010).

6.3 Micro-scale: environmental divergence in an hybrid zone

Our study on three viperid snakes in northern Iberia suggests the occurrence of ecological divergence in the contact zone. The use of ecological niche modelling to address evolutionary questions in a hybrid zone provides insightful results (e.g. Chatfield *et al.*, 2010; Culumber *et al.*, 2012; Swenson, 2006), especially in a environmental

perspective using a multi-trait analysis such as we present here. Swenson (2008) emphasized the importance of spatial analysis in evolutionary studies suggesting, however, few algorithm alternatives for ENM. We have shown the advantages of using Simapse under this context (see section 5.2 for details). The use of a continuous dataset revealed to be one major leap forward in the analysis of the ecological divergence between three hybridizing vipers species.

The contact zone between *Vipera latastei*, *V. aspis* and *V. seoanei* reveals a major hybrid zone between the first species pair. The area where they met represents a transitional habitat where none of the parental forms seems to have advantage over the other. Our method addresses the existence of an ecotone from a different perspective of previous studies (see section 5.2). The major differences are: 1) we do not model the hybrids because there was no a priori evidence that they constitute a viable population and, by applying a ENM, it is assumed a relation of equilibrium with the environment (Nogués-Bravo, 2009); 2) as input data to ENM we used only individuals with microsatellite data, thus avoiding the use of the full species distribution. Although such procedure may increase the detectability of differences in ecological requirements within the study area, these are likely to be generated by specific phenomena occurring at species range margins (Bridle & Vines, 2007).

6.4 Future prospects

Our analysis provided comprehensive results on both macro- and micro-scale factors affecting biodiversity patterns in the Iberian Peninsula. Future research may benefit from multiple aspects that will increase the generalization of results to wider geographical areas and other species groups. In this line, the creation of variables representing past climates should be extended to greater spatial scales. Although a global synthesis of past climate is theoretically possible, it will be extremely challenging because of the required database on detailed current plant distributions needed for climate reconstruction. At a European scale, there is extensive data from fossil pollen sites in published works and public databases (European Pollen Database, `http://www.europeanpollendatabase.net/`) and also from distributional atlases and databases of current plant distributions. Building spatial explicit layers of climate variables for Europe, covering a period since the LGM, will fuel studies of biodiversity patterns, provide new data to use under evolutionary hypotheses-testing of the European biota, and will help predicting impacts of climate change for the current century with less uncertainty, contributing to global conservation efforts.

We have shown the potential of calibrated and downscaled scenarios of future

climate. Nevertheless, studies will have significant improvements using newest climate models with the same time frequency as those we used here. Climate models predicting future climate change are periodically improved, with increased resolution and complexity by adding new parameters, such as the physical interactions affecting climate (IPCC, 2007). The uncertainties of these models, although impossible to fully eliminate due to the unpredictable effects of anthropogenic impacts on global climate, are being reduced with these parameters and benefiting from the fine tuning of measured climate change in the last decades. These data are often obscured in complex databases and unmanageable formats, hampering biodiversity related studies due to a lack of computer knowledge on how to deal with such intricate information. New tools to gather these data and to convert it to common formats will be of much value to help researchers better forecasting the likely impacts of climate change on biodiversity patterns.

Our software application, Simapse, introduced many advantages to the research on biodiversity patterns and to the study of ecological divergence on hybridizing species. Future improvements of this application should increase the velocity of model building and automating of tasks for multi-species models. Other improvements to be addressed in future versions should include; 1) new sub-sampling strategies to avoid modelling pitfalls, especially those related to auto-correlation of the presence signal in the multidimensional space of the predictors; 2) an adaptive learning rate to increase model performance; and 3) optimization and increased diversity of algorithms to find the best network that would improve speed. Simapse was used at micro- and macro-scale analysis of biodiversity patterns, generating an immense quantity of data. An interesting future research line is the application of Simapse to the study of niche conservatism. This has been under active development with the recent arising of new metrics to test the degree of niche similarity and equivalence (Warren *et al.*, 2008). Such developments should also be implemented in Simapse. An important limitation of the studies addressing ecological roles in evolutionary analyses is the spatial resolution of variables. The current trend is to increase the resolution at which these variables are available, as result of improved sensors in remote sensing and also the increase of processing power to derive high-resolution spatial data from complex interpolation algorithms using immense quantities of information.

The knowledge about dynamic processes shaping current biodiversity patterns are increasing at a fast pace. Insights come from the most diverse areas as landscape ecology, molecular evolution, climate modelling and paleoecology. We provide an integrative approach of data generated in all these fields of research, but much is

still to be achieved. Our broad expectation is that this increasing knowledge will help humanity to understand and deal with the potential impacts of climate change in all living organisms.

6.5 References

Alexandrino, J., Froufe, E., Arntzen, J.W. & Ferrand, N. (2000) Genetic subdivision, glacial refugia and postglacial recolonization in the golden-striped salamander, *Chioglossa lusitanica* (Amphibia: urodela). *Molecular Ecology* **9**, 771–781.

Anderson, N.J., Bugmann, H., Dearing, J.A. & Gaillard, M.J. (2006) Linking palaeoenvironmental data and models to understand the past and to predict the future. *Trends in Ecology & Evolution* **21**, 696–704.

Araújo, M.B., Thuiller, W. & Pearson, R.G. (2006) Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**, 1712–1728.

Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution* **22**, 42–47.

Bridle, J.R. & Vines, T.H. (2007) Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution* **22**, 140–147.

Brito, J.C., Acosta, A.L., Álvares, F. & Cuzin, F. (2009) Biogeography and conservation of taxa from remote regions: An application of ecological-niche based models and GIS to North-African canids. *Biological Conservation* **142**, 3020–3029.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2010) From climate change predictions to actions - conserving vulnerable animal groups in hotspots at a regional scale. *Global Change Biology* **16**, 3257–3270.

Chatfield, M.W.H., Kozak, K.H., Fitzpatrick, B.M. & Tucker, P.K. (2010) Patterns of differential introgression in a salamander hybrid zone: inferences from genetic data and ecological niche modelling. *Molecular Ecology* **19**, 4265–4282.

Culumber, Z.W., Shepard, D.B., Coleman, S.W., Rosenthal, G.G. & Tobler, M. (2012) Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: *Xiphophorus*). *Journal of Evolutionary Biology* **25**, 1800–1814.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle,

B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., 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.

Godinho, R., Mendonça, B., Crespo, E.G. & Ferrand, N. (2006) Genealogy of the nuclear beta-fibrinogen locus in a highly structured lizard species: comparison with mtDNA and evidence for intragenic recombination in the hybrid zone. *Heredity* **96**, 454–463.

Gómez, A. & Lunt, D. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of Southern European Refugia* (eds. S. Weiss & N. Ferrand), pp. 155–188, Springer Netherlands.

Hewitt, G. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247–276.

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

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

Ince, D.C., Hatton, L. & Graham-Cumming, J. (2012) The case for open computer programs. *Nature* **482**, 485–488.

IPCC (2007) *Climate Change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

Kozak, K.H. & Wiens, J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution* **60**, 2604–2621.

Kriticos, D.J., Webber, B.L., Leriche, A., Ota, N., Macadam, I., Bathols, J. & Scott, J.K. (2011) CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* **3**, 53–64.

Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature* **462**, 1052–1055.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2007) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* **17**, 145–151.

McCulloch, W. & Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology* **5**, 115–133.

Mitchell, T.D., Carter, T.R., Jones, P.D., Hulme, M. & New, M. (2004) A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901-2000) and 16 scenarios (2001-2100).

Naughton, F., Sanchez Goñi, M., Desprat, S., Turon, J.L., Duprat, J., Malaizé, B., Joli, C., Cortijo, E., Drago, T. & Freitas, M. (2007) Present-day and past (last 25000 years) marine pollen signal off western Iberia. *Marine Micropaleontology* **62**, 91–114.

Nogués-Bravo, D. (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography* **18**, 521–531.

Parmesan, C. (2006) Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics* **37**, 637–669.

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

Peterson, A.T., Martinez-Meyer, E. & Gonzalez-Salazar, C. (2004) Reconstructing the Pleistocene geography of the *Aphelocoma* jays (Corvidae). *Diversity and Distributions* **10**, 237–246.

Peterson, A.T., Papes, M. & Soberón, J. (2007) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecology* **3**, 63–72.

Peterson, A. & Vieglais, D. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience* **51**, 363–372.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175.

Rebelo, H., Tarroso, P. & Jones, G. (2010) Predicted impact of climate change on European bats in relation to their biogeographic patterns. *Global Change Biology* **16**, 561–576.

Rödder, D. & Lötters, S. (2009) Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (*Hemidactylus turcicus*). *Global Ecology and Biogeography* **18**, 674–687.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R.G., Gaston, K.J., Sutherland, W.J. & Svenning, J.C. (2011) The influence of Late Quaternary climate-change velocity on species endemism. *Science* **334**, 660–664.

Sequeira, F., Alexandrino, J., Rocha, S., Arntzen, J..W. & Ferrand, N. (2005) Genetic exchange across a hybrid zone within the Iberian endemic golden-striped salamander, *Chioglossa lusitanica*. *Molecular Ecology* **14**, 245–254.

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

Swenson, N.G. (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology* **21**, 421–434.

Swenson, N.G. (2006) Gis-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *Journal of Evolutionary Biology* **19**, 717–725.

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

Warren, D.L., Glor, R.E. & Turelli, M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**, 2868–2883.

Willis, K.J., Bennett, K.D., Bhagwat, S.A. & Birks, H.J.B. (2010) 4°C and beyond: what did this mean for biodiversity in the past? *Systematics and Biodiversity* **8**, 3–9.

Xavier, R., Lima, F.P. & Santos, A.M. (2010) Forecasting the poleward range expansion of an intertidal species driven by climate alterations. *Scientia Marina* **74**, 669–676.

Chapter 7

Conclusions

Ah, que ninguém me dê piedosas
intenções,
Ninguém me peça definições!
Ninguém me diga: 'vem por aqui'!
A minha vida é um vendaval que se soltou,
É uma onda que se alevantou,
É um átomo a mais que se animou...
Não sei por onde vou,
Não sei para onde vou
Sei que não vou por aí!

— JOSÉ RÉGIO, *Cântico negro*

In this work, we provided an integrative framework to analyse macro- and micro-scale biological processes in the Iberian Peninsula. The main conclusions are:

1. The integration of past climate reconstruction from palaeoecological data, predicted climate change for the current century, ecological niche modelling and molecular analyses provided a deeper understanding of macro- and micro-scale biological processes occurring in the Iberian Peninsula. This region is part of a major biodiversity hotspot and, thus, results are likely to have a wide impact in conservation issues, particularly under scenarios of climate change and persistence of threatened species.

2. The climate data converter was proven to be useful to convert large climate datasets to ready to use formats for ecological niche modelling. The availability

of E-Clic is expected to facilitate the access to these data in studies related to climate change.

3. The application Simapse is a user friendly application, virtually available in all platforms, to assess the distribution of a species based on presence data. It provides state of the art analysis of the model built with artificial neural networks.

4. Simapse provides high quality models from virtual and real species data. The framework developed for the test provided exhaustive analyses with proper controls and can be extended to other algorithms. Additionally, potential misinterpretations of commonly used metrics for model performance were identified under specific conditions.

5. The reconstruction of past climate variables in a spatially explicit framework poses many challenges. We contributed with method improvements for climate reconstruction using fossil pollen as biological climate proxy, resulting in a time-series of spatial layers of three climate variables that are potentially useful in a wide variety of studies.

6. The Iberian Peninsula climate was described with a general warming trend between 15,000 and 3,000 years BP with abrupt events that had a wide spatial impact. Areas that shared similar climate evolution were summarized using statistical methods and showed high concordance with a network of potential glacial refugia within the Iberian Peninsula.

7. Comparative analysis of past and future dynamic processes alerted for the conservation challenge resulting from the changes predicted to occur in the current century. Past changes in herpetofauna species richness occurred at a slower pace than those predicted for the following decades.

8. Two metrics of velocity of change of species composition were created and used to derive areas where extreme change occurred in the past and to define areas that will suffer major changes in the future. In the past, areas with low velocities were broadly coincident with glacial refugia described from molecular studies of herpetofauna species.

9. New methods based on Simapse were applied to the study of ecological processes in a hybrid zone between three Iberian vipers. Main methodological achievements were the use of the genetic population structure to describe local patterns and the avoidance of a priori assumptions about an ecological niche of

the hybrid individuals. These methods represented a step forward in integrating genetic and ecological analysis in the description of contact zones.

10. The sister species *V. latastei* and *V. aspis* met and hybridized in a narrow contact zone. Only one hybrid was detected between *V. latastei* and the phylogenetically more distant *V. seoanei*. The hybrid zone constitutes a strong barrier to gene flow in the studied area.

11. We show that each parental species shows little spatial overlap with each other, and the niches are defined by different sets of environmental variables. The ecotone where most intermediate individuals were found is located in a ecological gradient, where none of the parental has advantage over the others, thus, sugesting an ecological barrier.

12. Future predictions of climate change are as threatning to biodiversity as they are chalenging to analyse and mitigate. The results presented here, with a primary look into the past changes and with two different scales of analyses, are expected to have a wide application on the study of the impacts of climate.

Chapter 8

Appendices

– I was reading about how countless species are being pushed toward extinction by man's destruction of forests.
– Sometimes I think the surest sign that intelligent life exists elsewhere in the universe is that none of it has tried to contact us.

— CALVIN [BILL WATTERSON], *Calvin & Hobbes*

Sid: You know? This whole ice age thing is getting old. You know what I could go for? A global warming.
Diego: Keep dreaming.
Sid: No really...

—Ice Age, 2002

Appendix A

Supplementary material for section 3.2

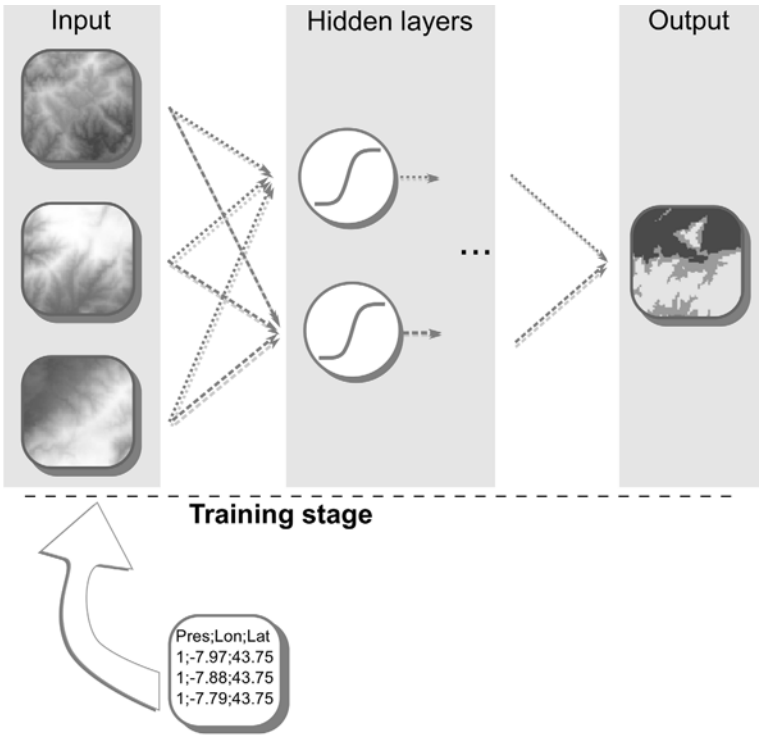


Fig. A.1 – The Back-Propagation Artificial Neural Network is a structure of layers with neurons connected by weights (dashed arrows). The basic structure has three main parts: the input layer where values from different variables are fed to the network; the hidden layer(s) where neurons with some activation function (usually sigmoid) receive information from the previous layer, process it and send it to the neurons on the next layer; and the output layer where the results of the network are returned to the user. During the training phase (learning stage), a set of known locations with presence, presence/absence or abundance of a species is fed to the network with the respective values of the chosen variables. This allows to process an output error at each iteration that is back-propagated through the network for the weights adjustment based on the learning rules (back-propagation algorithm). Once trained, the network is used to predict presence of the species to locations without presence data.

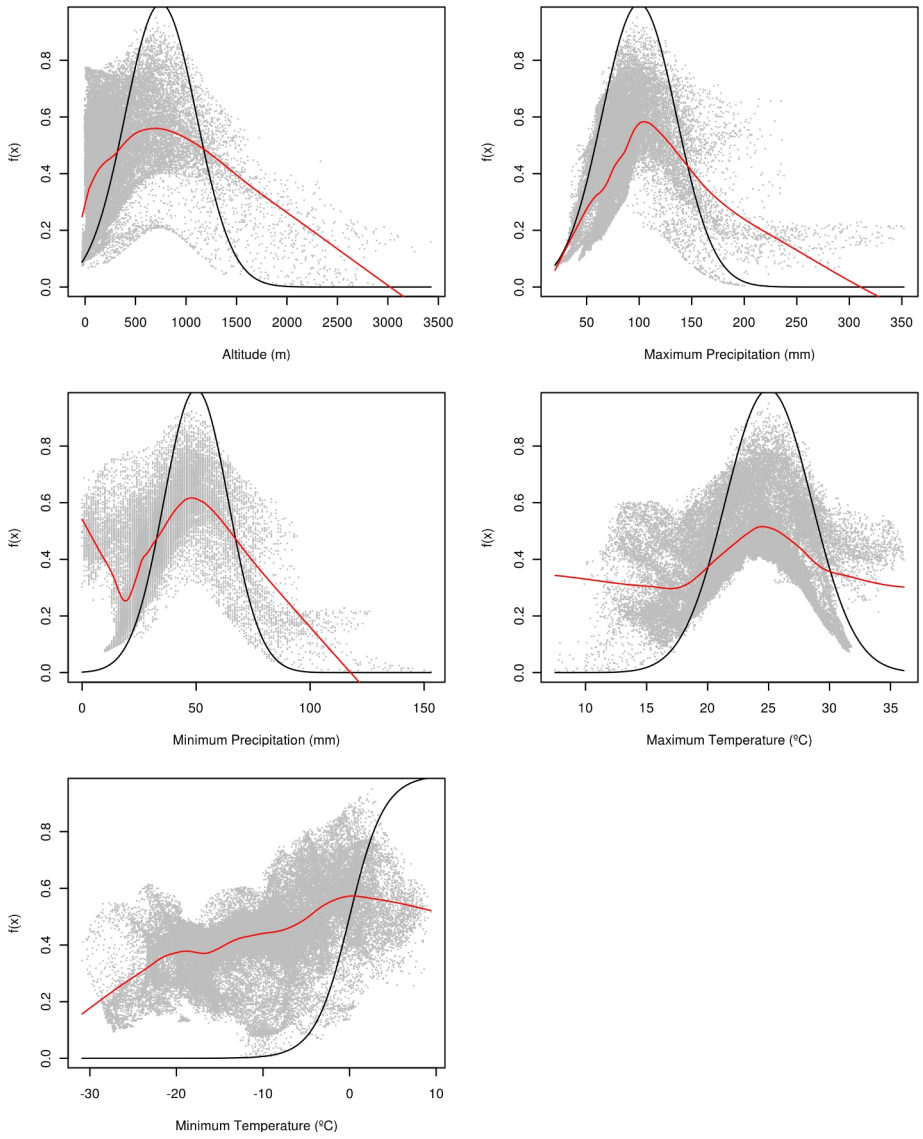


Fig. A.2 – Variables used to create the virtual species with a Gaussian (altitude, maximum and minimum precipitation and maximum temperature) or logistic (minimum temperature) function applied. The black line represents the function applied to the variables and the red line is the locally-weighted polynomial regression of the grid cells of the variable and the presence of the virtual species (grey dots).

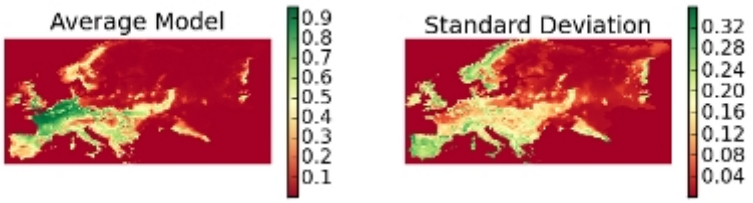


Fig. A.3 – Consensus averaged prediction and standard deviation, as shown by Simapse as a preview of the built model. These results are saved as text raster files in the output directory.

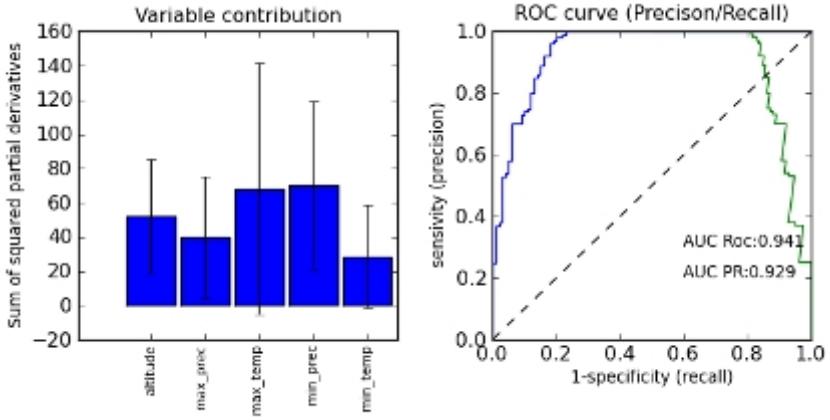


Fig. A.4 – Variable importance, ROC and precision-recall plots as provided by Simapse.

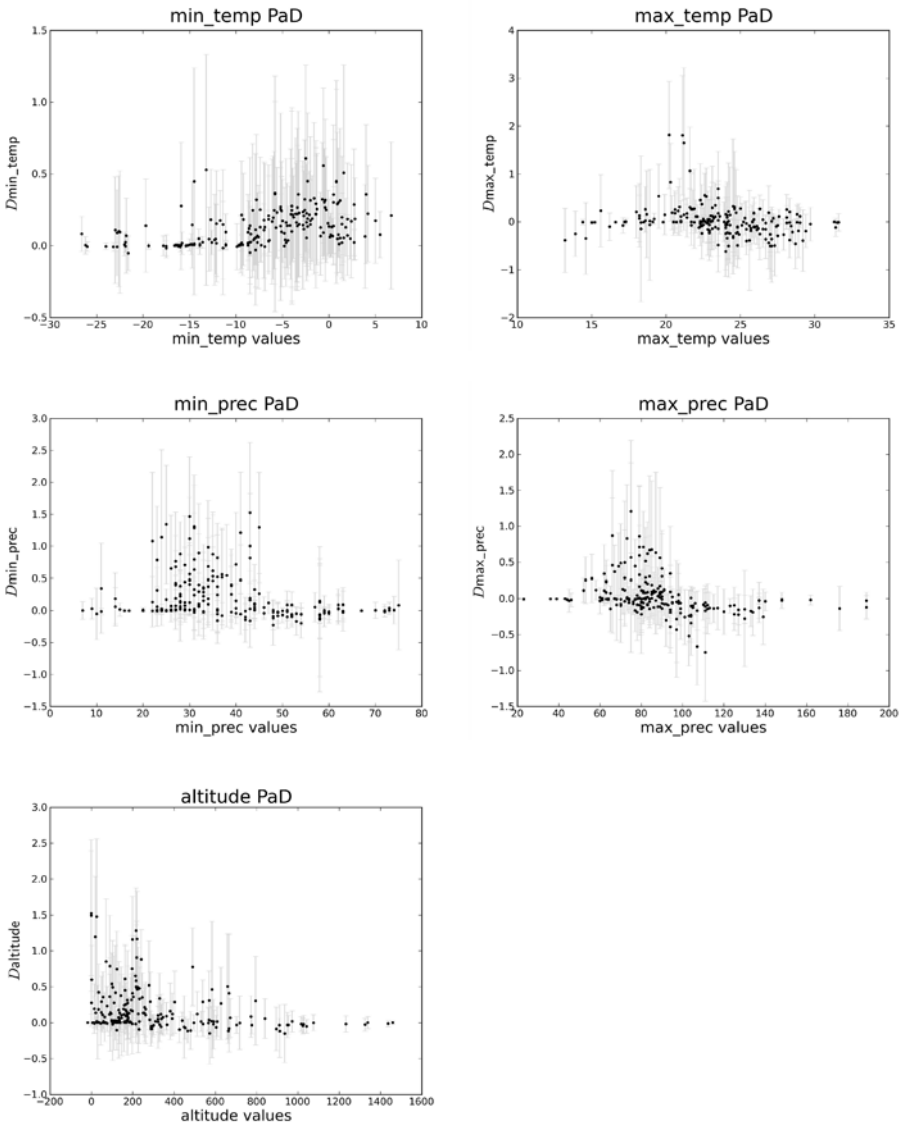


Fig. A.5 – Partial derivatives plots as provided by Simapse.

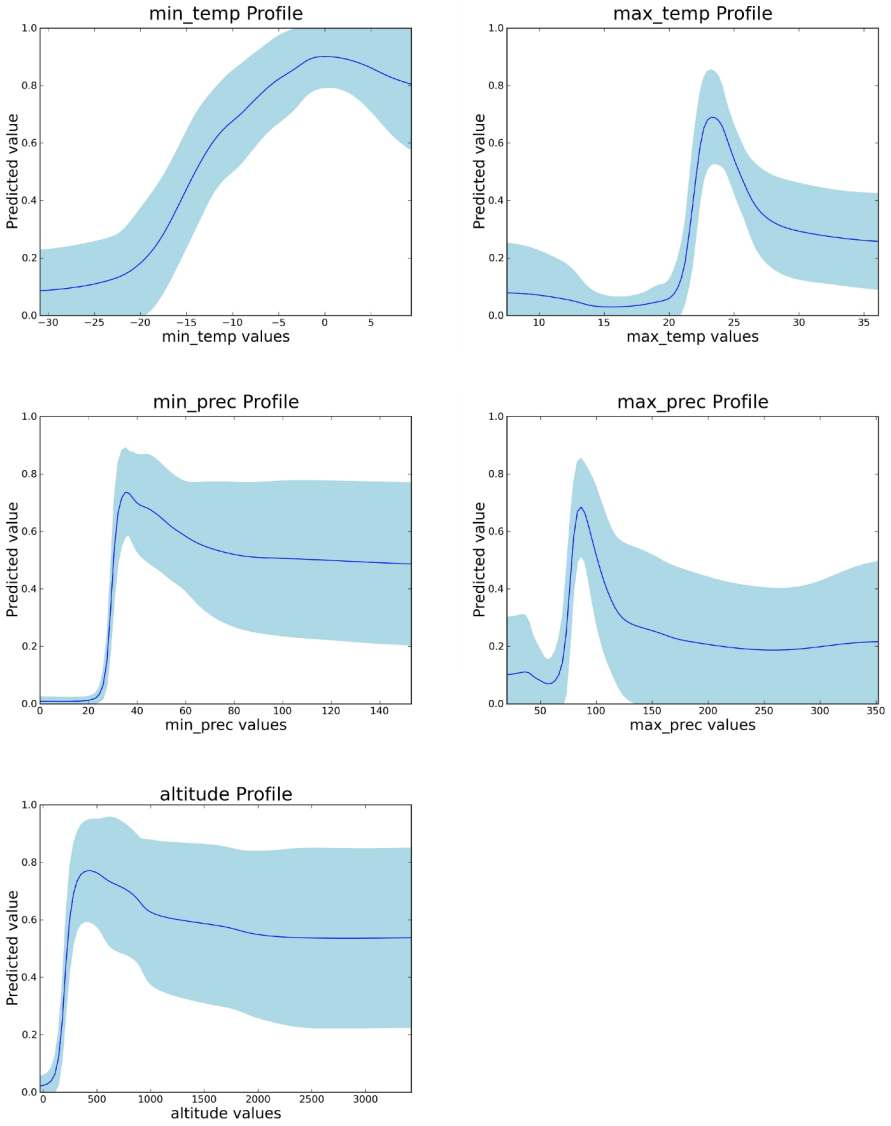


Fig. A.6 – Profile plots as provided by Simapse.

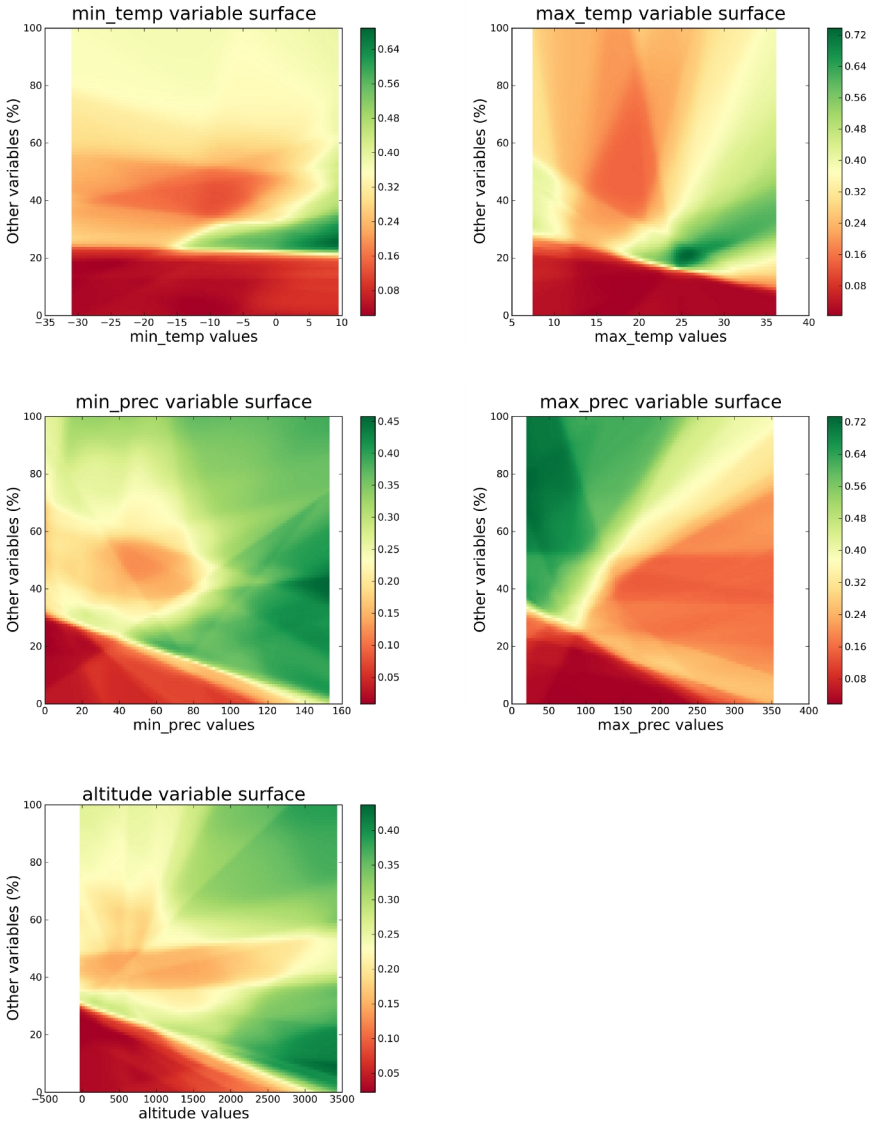


Fig. A.7 – Variable surfaces as provided by Simapse.

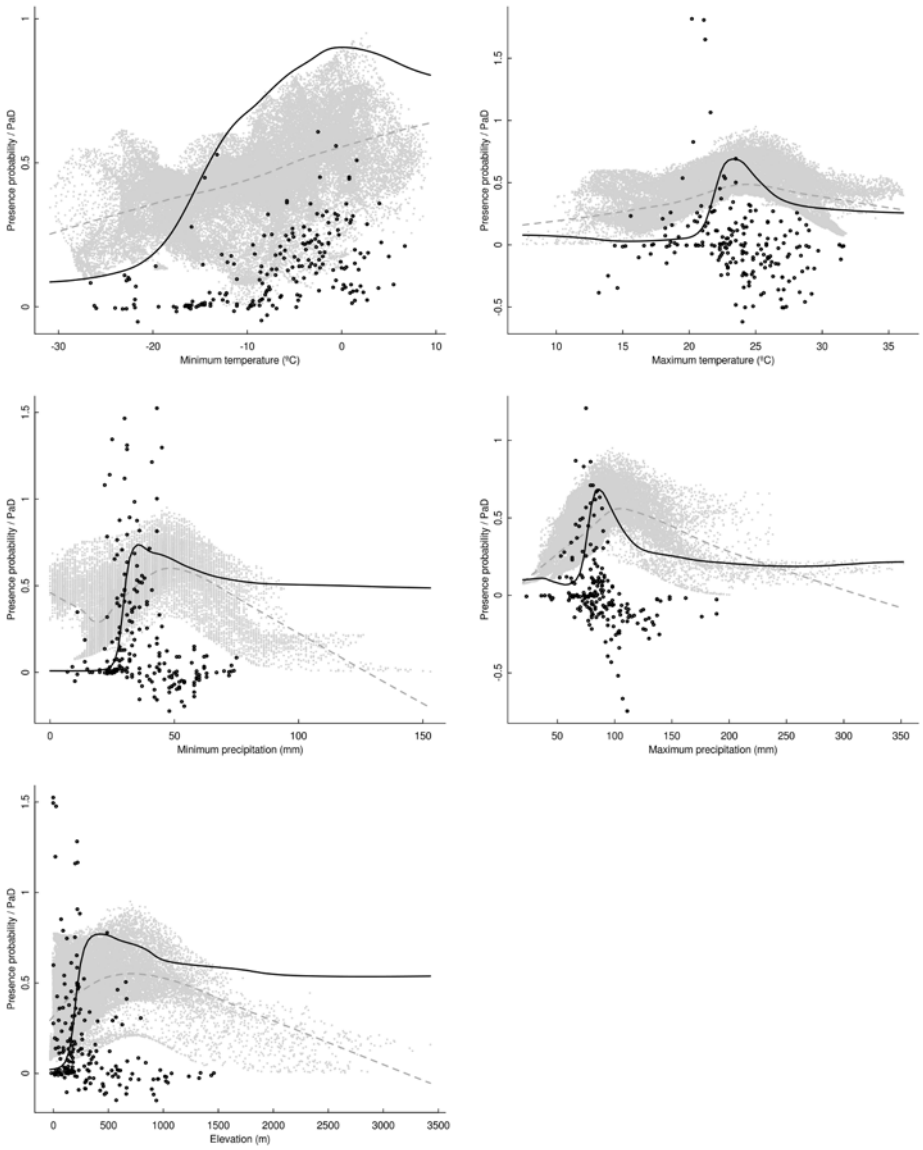


Fig. A.8 – The grey dots represent each variable values with the virtual presence and the dashed grey line represents the locally-weighted polynomial regression. The black line is the average profile and the black dots are the average partial derivatives, as given by Simapse in the text results. It is possible to see that the model, with presence-only data can fit the original variation of the virtual species.

Appendix B

Supplementary material for section 3.3

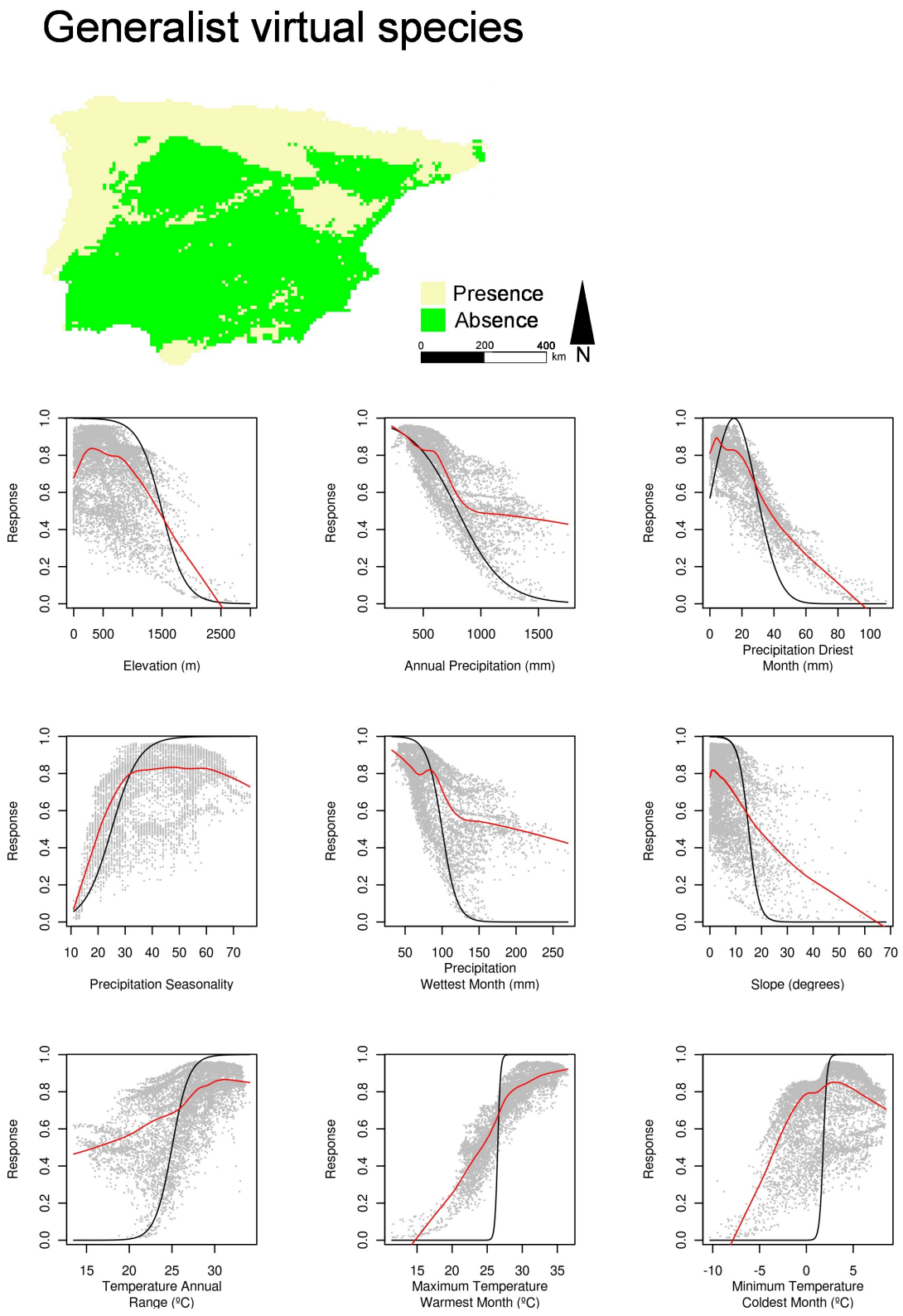
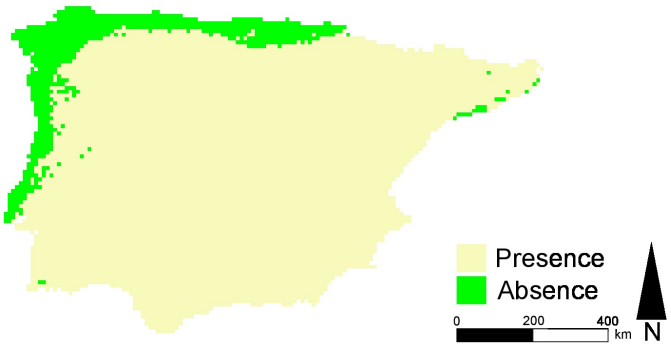
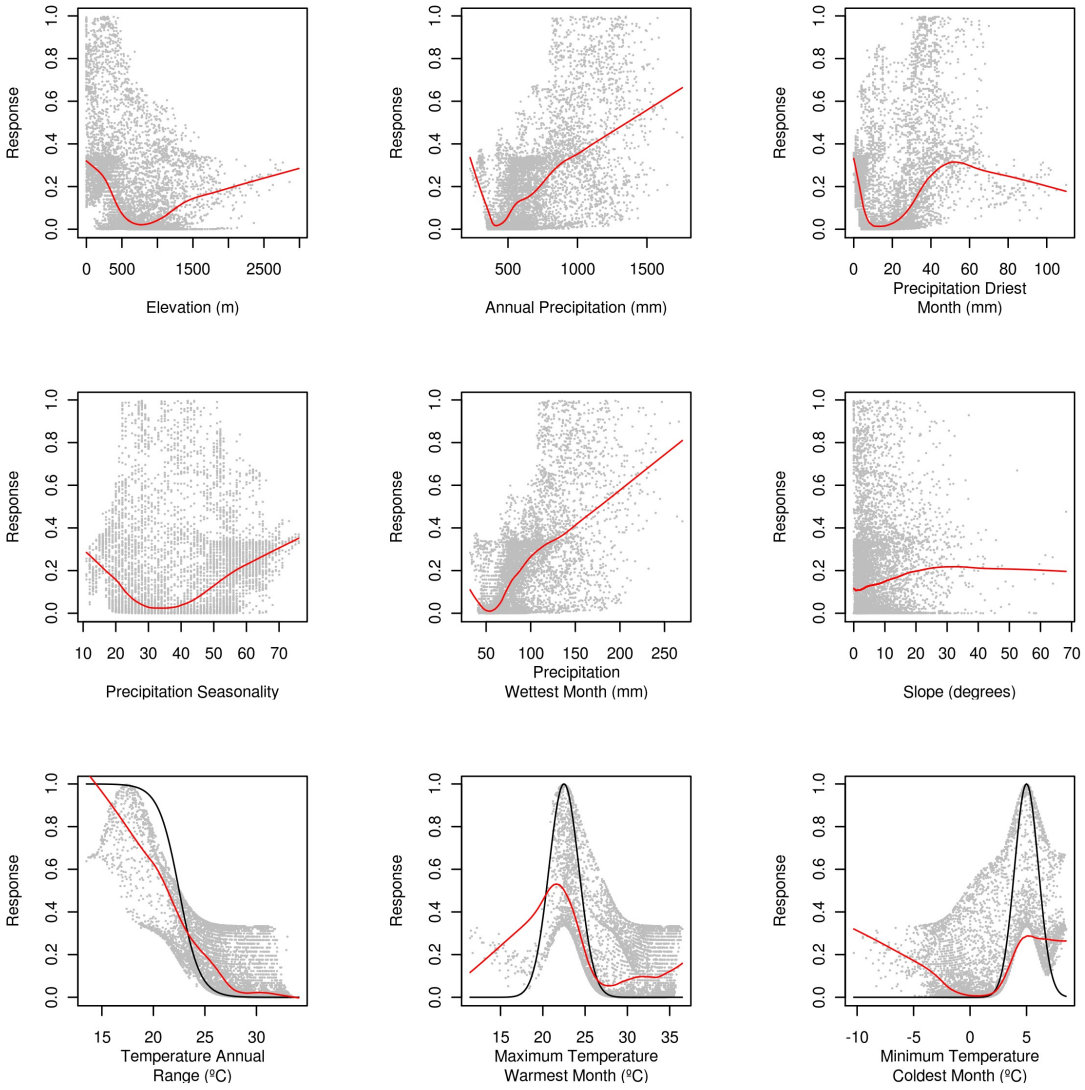


Fig. B.1 – Variables and responses curves of the virtual species. The black line represents the function used to simulate a response of the virtual species to the variable (with SVS only temperatures had a response); the dots represent the available cells in the study area, relating the final variable response (after averaging all available responses) to the value, and the red line represents a locally-weighted polynomial regression of the dots.

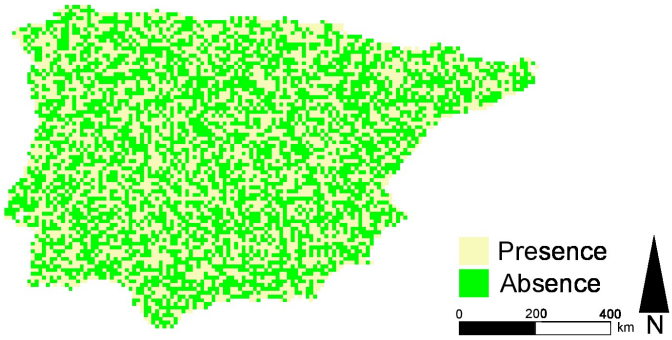
Specialist virtual species

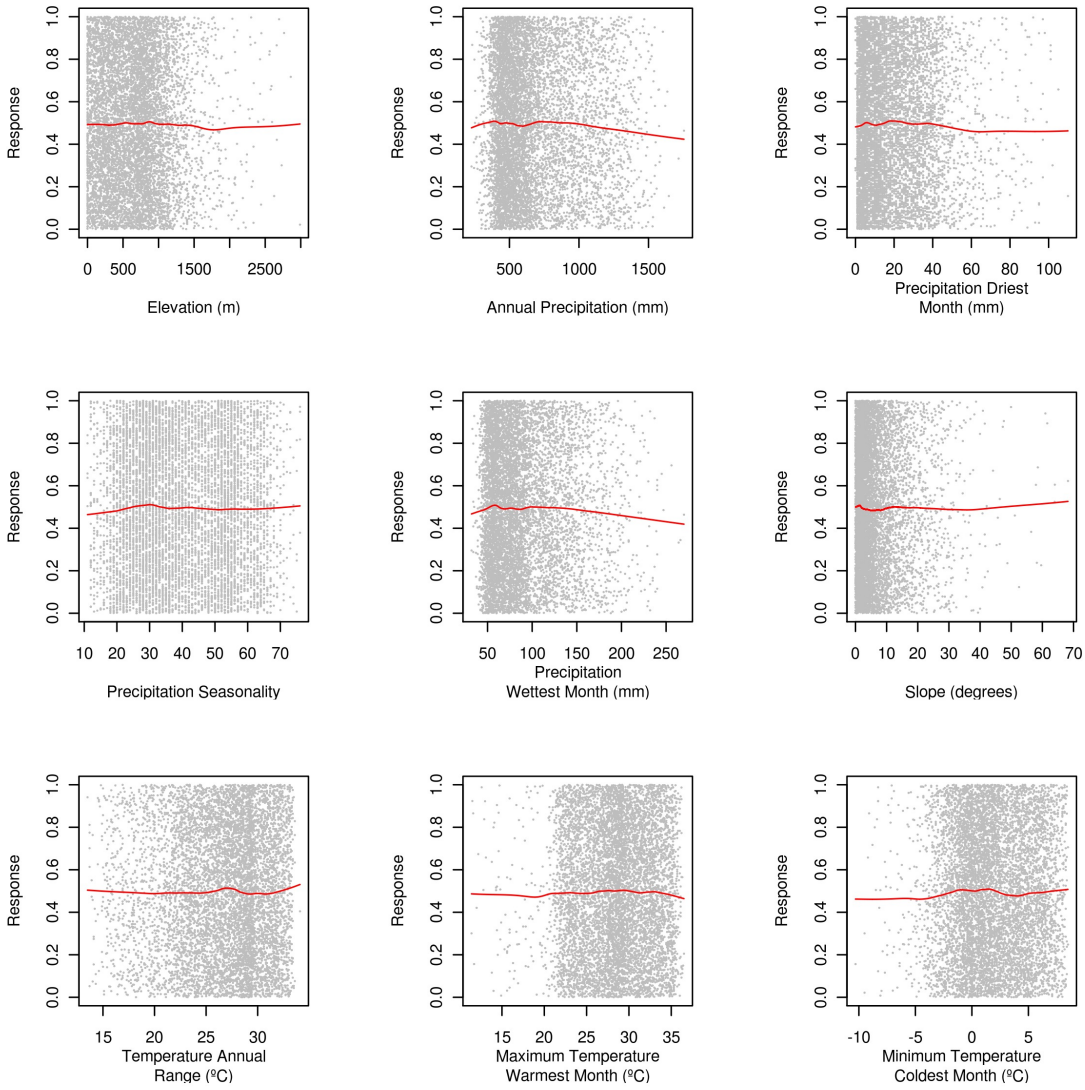




Continued from previous page.

Random virtual species





Continued from previous page.

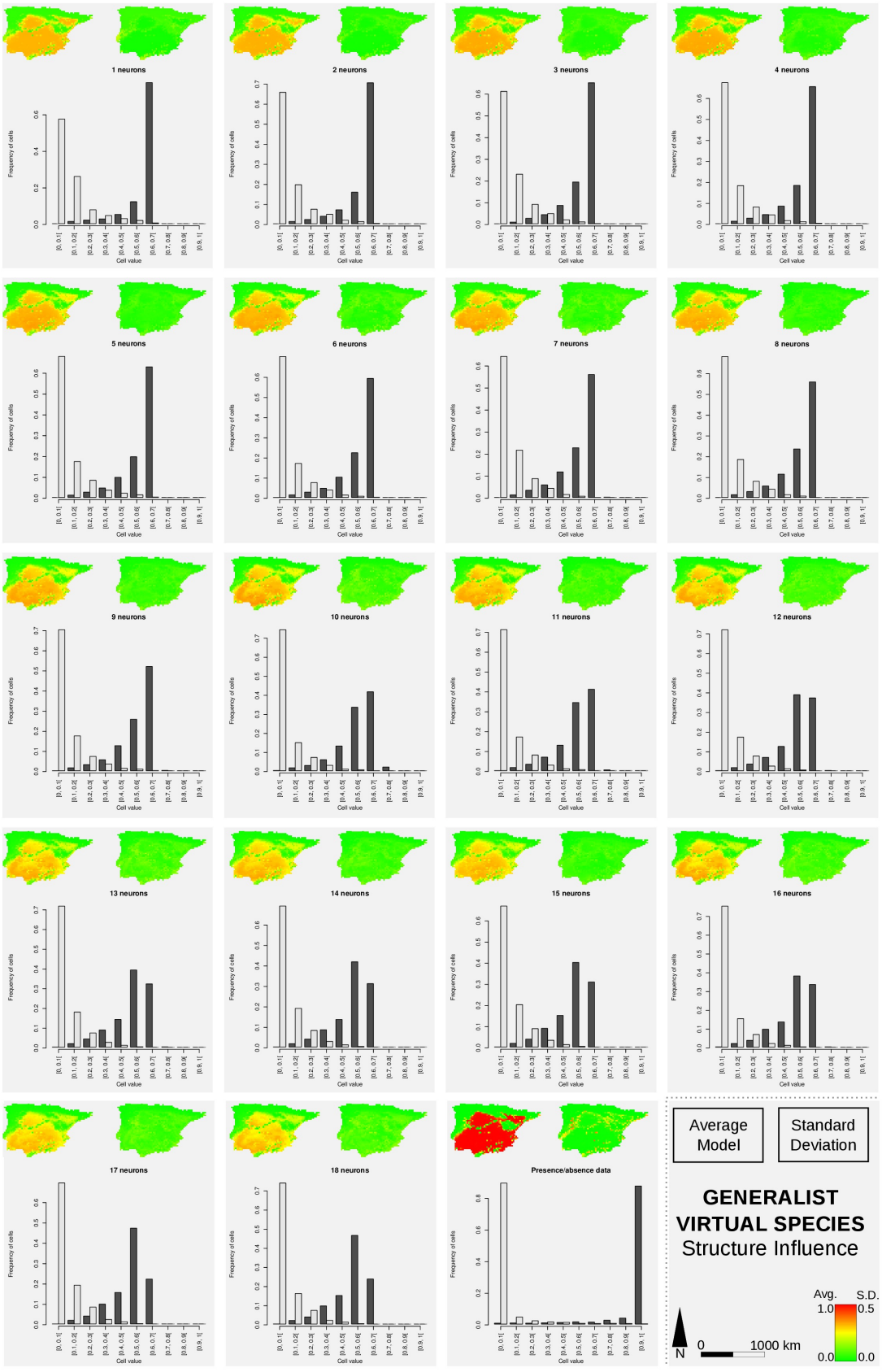
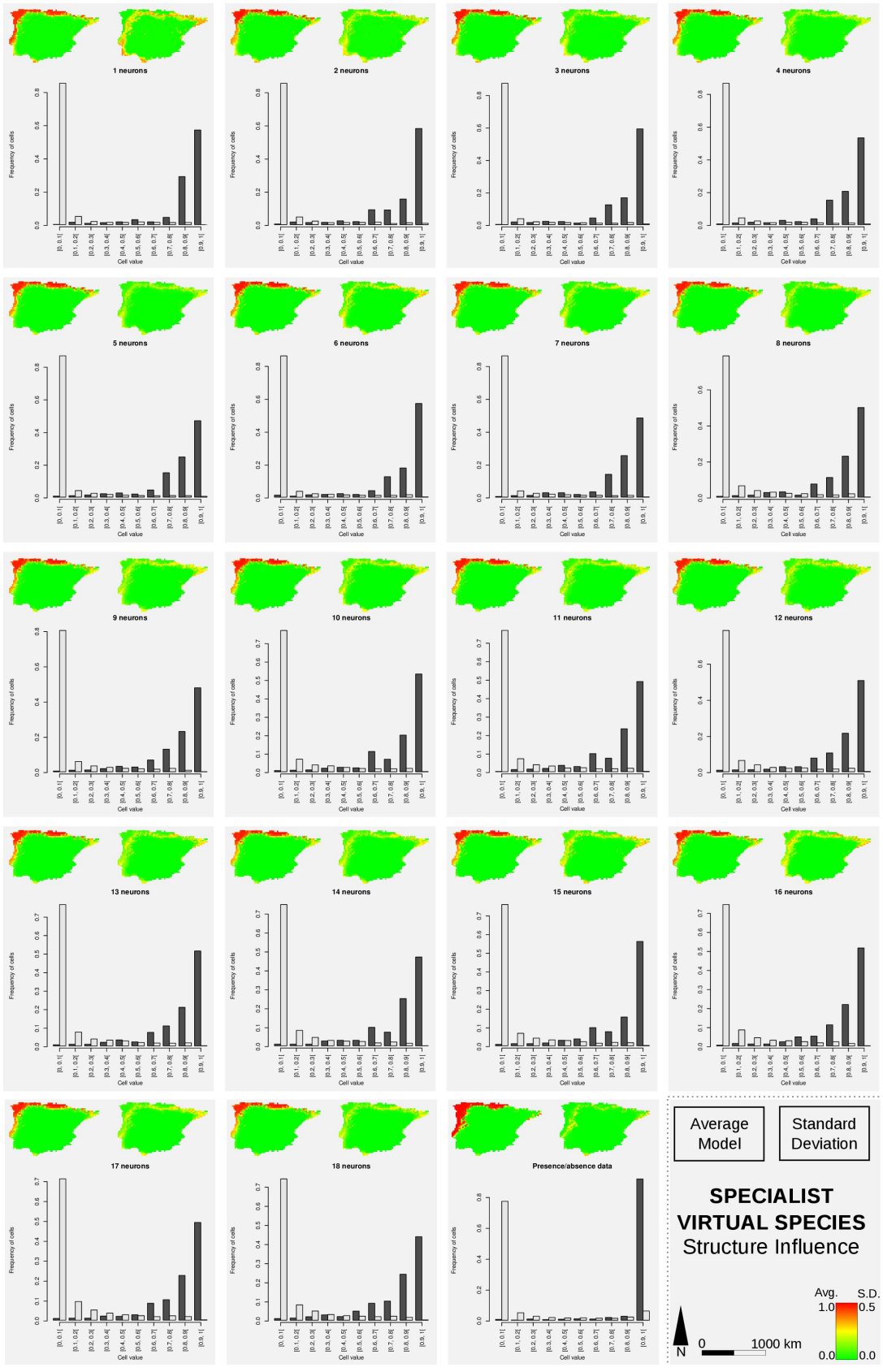
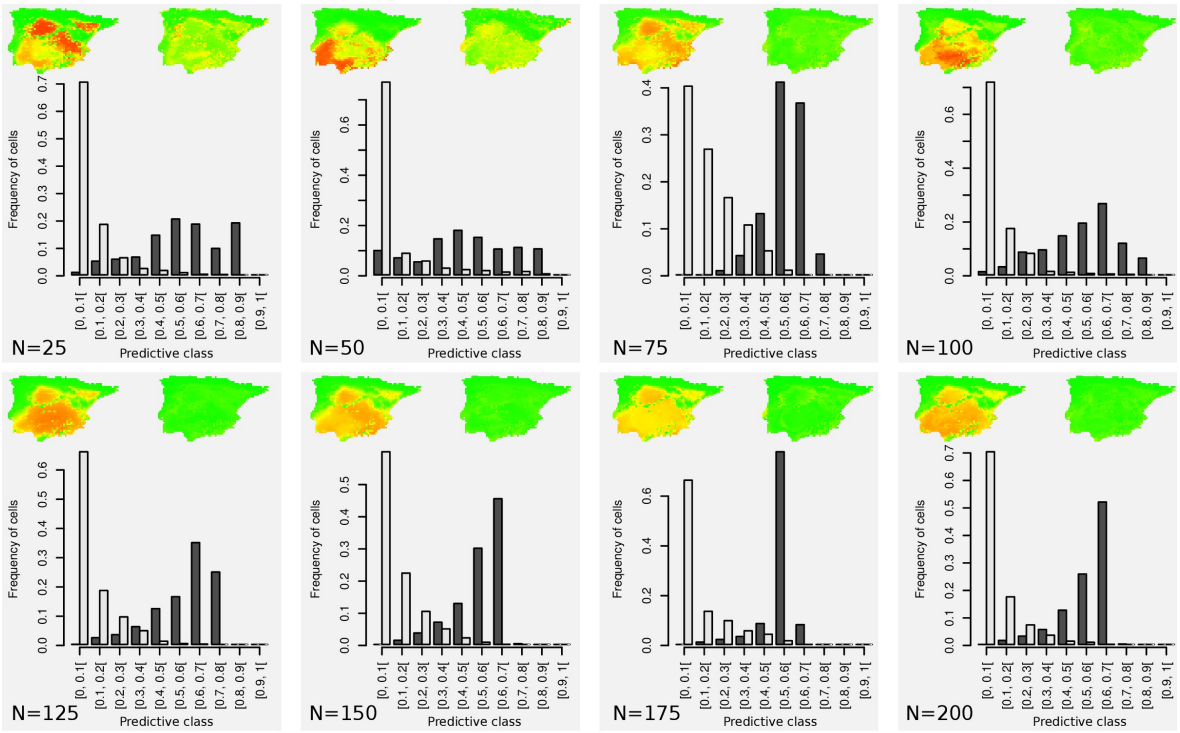


Fig. B.2 – Potential distribution models, standard deviation of the final consensus model, and plot of the proportion of presences and absences per predictive class for the virtual species. The effects of network structure and sample size are shown.

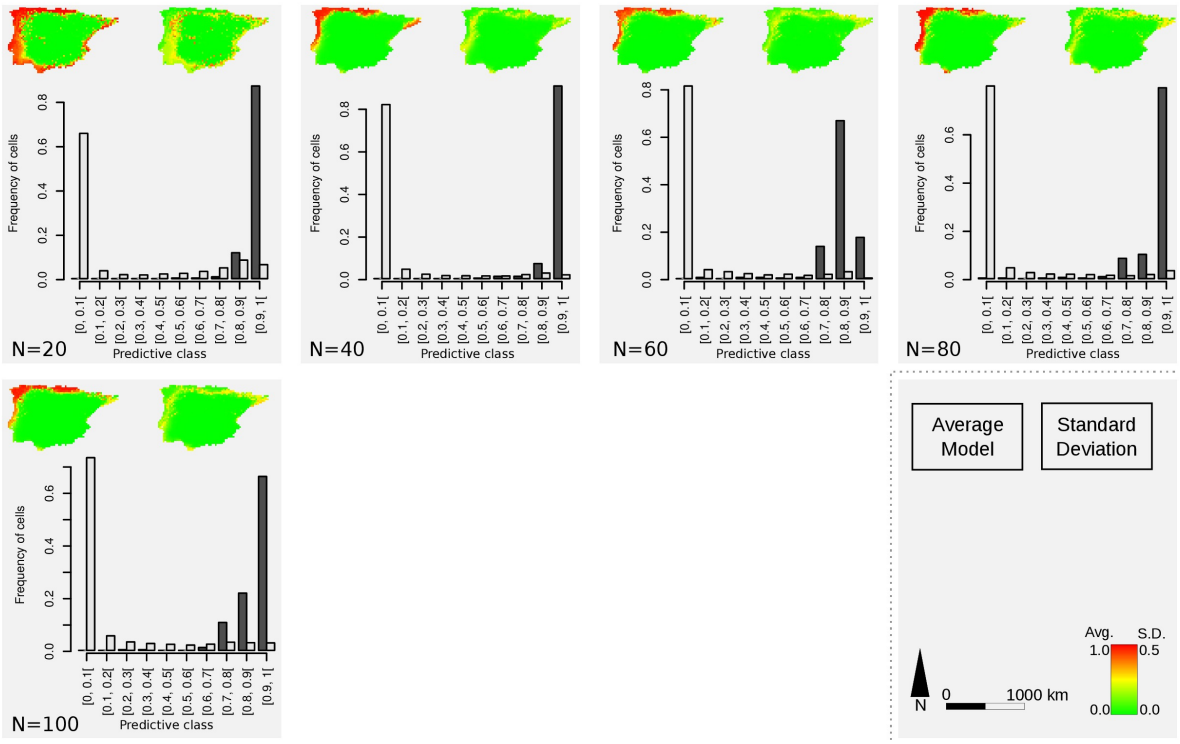


Continued from previous page.

GENERALIST VIRTUAL SPECIES
Number of samples



SPECIALIST VIRTUAL SPECIES
Number of samples



Continued from previous page.

Appendix C

Supplementary material for section 4.1

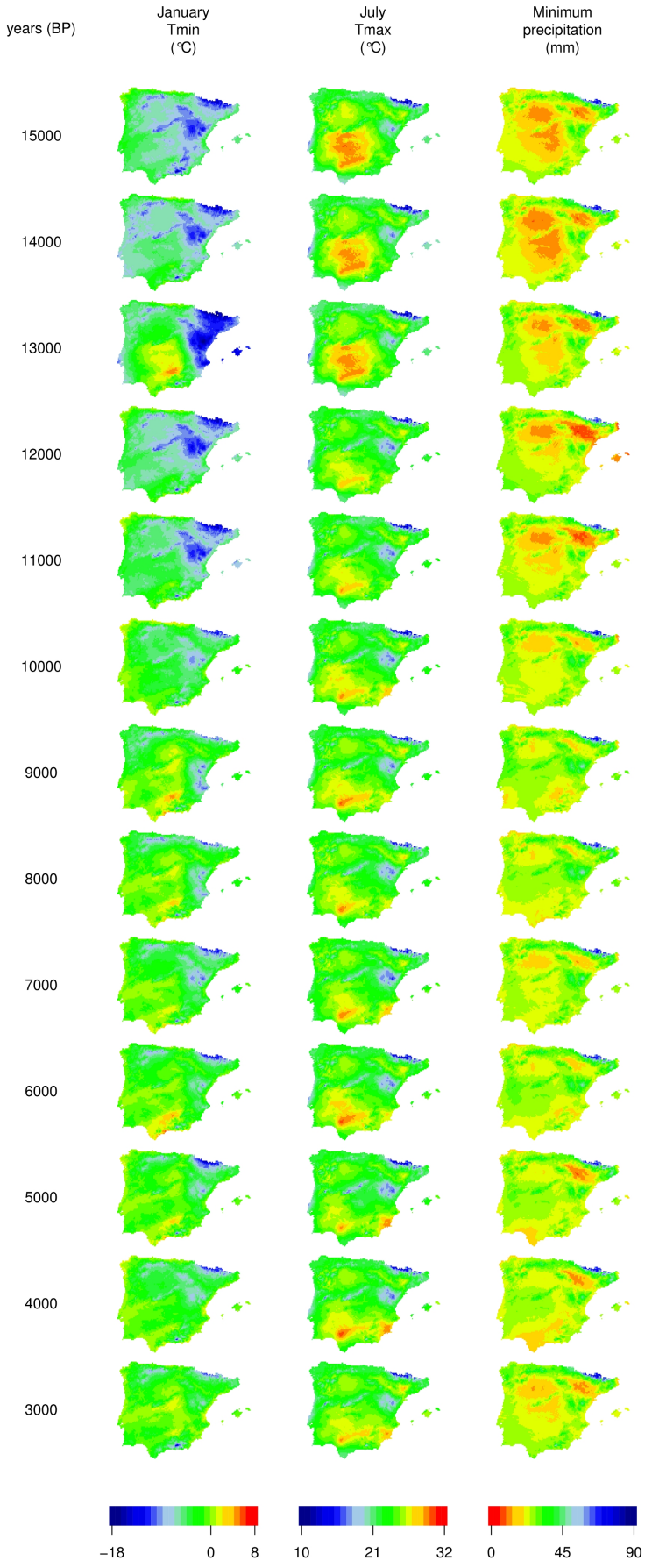


Fig. C.1 – Maps of Tjan, Tjul and Pmin in the Iberian Peninsula and the Balearic Islands in the past 15 ka.

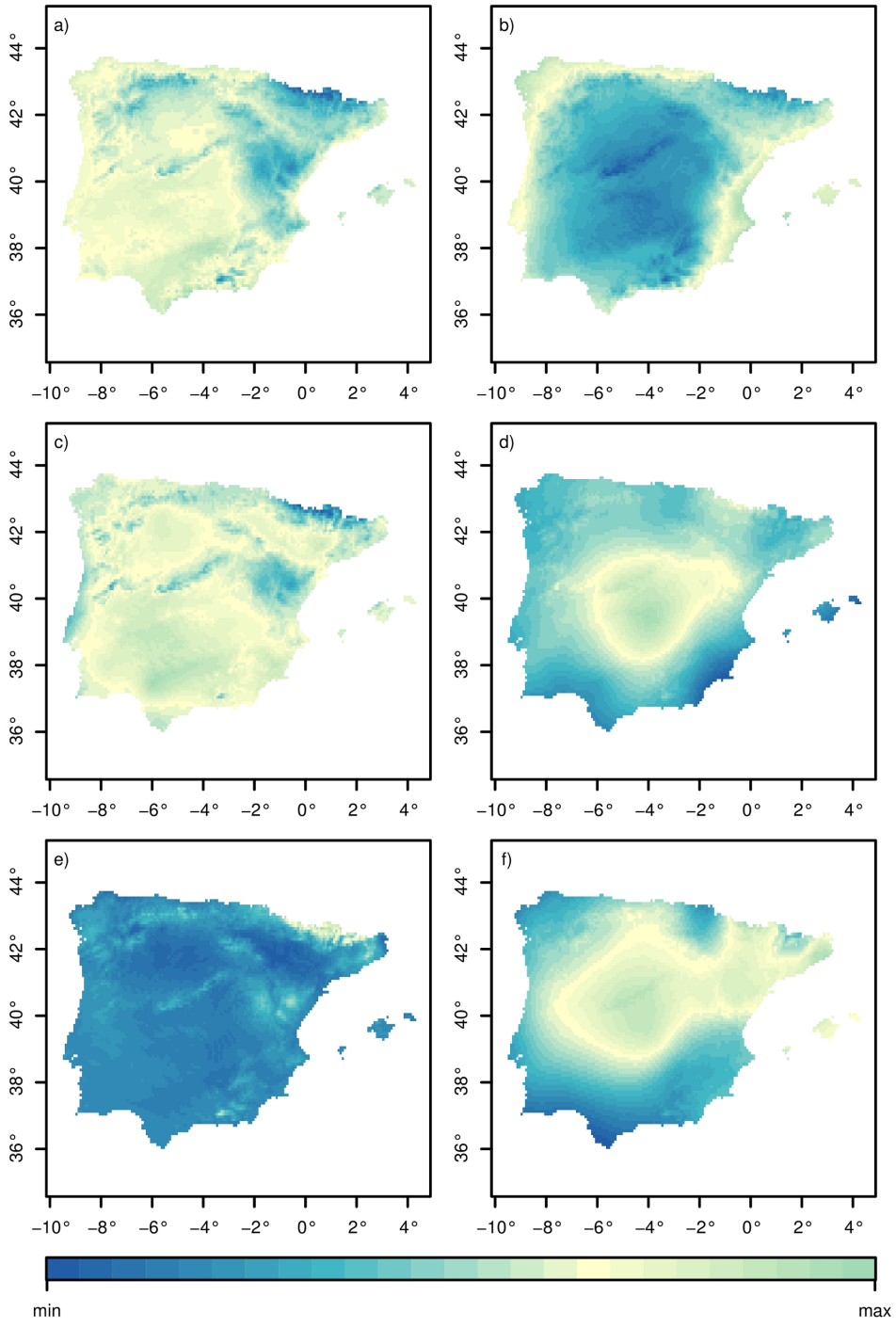


Fig. C.2 – Spatial representations of the fPCA score for the first and second components of Tjan (a and b, respectively); Tjul (c and d) and Pmin (e and f) in the Iberian Peninsula and the Balearic Islands in the past 15 ka.

Appendix D

Supplementary material for section 5.1

Table D.1 – Detailed model results for the current species distributions. Learning rates (LR), ranging from 9x10^{-4} to 6x10^{-3}, were calculated for each species based on the complexity of the model species as measured by the average nearest neighbour (NN) index, which measures the degree of clustering in the presence data. The threshold shown here is the 10th percentile value of the predictions for presences, used to transform probability results to binary. The true positives, true negatives, false positives and false negatives (TP, TN, FP and FN, respectively) are given in number of pixel counts. The area under the receiver operating characteristic curve (AUC) is also shown. The absences pixels are divided in clustered absences (CAb) and total absences (Ab). Clustered absences were defined if, at least, 6 neighboring pixels in a 3x3 window around a absence were also absences. It is also given the proportion between clustered and total absences (CAb/Ab).

Species	NN	LR	Thres.	TP	TN	FP	FN	AUC	CAb	Ab	CAb/Ab
Algyroides marchi	1.2	0.0031	0.1	53	8011	835	6	0.971	8823	8846	0.997
Alytes cisternasii	1.1	0.0028	0.588	1685	5080	1954	186	0.878	6408	7034	0.911
Alytes dickhilleni	0.9	0.0021	0.249	254	7526	1096	29	0.955	8518	8622	0.988
Blanus cinereus	1	0.0025	0.683	2688	4111	1810	296	0.856	4719	5921	0.797
Chalcides bedriagai	0.8	0.0018	0.421	1454	2593	4699	159	0.731	6377	7292	0.875
Chalcides striatus	0.8	0.0018	0.454	2450	1670	4514	271	0.682	4797	6184	0.776
Chioglossa lusitanica	0.8	0.0018	0.689	577	7782	483	63	0.978	8061	8265	0.975
Discoglossus galganoi	1	0.0025	0.345	1945	3235	3510	215	0.78	5675	6745	0.841
Discoglossus jeanneae	0.7	0.0015	0.367	923	3714	4165	103	0.775	7373	7879	0.936
Iberolacerta cyreni	0.9	0.0021	0.059	61	7992	845	7	0.975	8801	8837	0.996
Iberolacerta galani	1.5	0.0041	0.09	18	8536	349	2	0.981	8872	8885	0.999
Iberolacerta martinezricai	2.1	0.0060	0.036	7	8406	491	1	0.971	8892	8897	0.999
Iberolacerta monticola	0.6	0.0012	0.333	316	7479	1075	35	0.945	8398	8554	0.982
Lacerta schreiberi	0.8	0.0018	0.428	1408	5696	1644	157	0.922	6866	7340	0.935
Lissotriton boscai	1.1	0.0028	0.216	2212	3613	2834	246	0.844	5701	6447	0.884
Mauremys leprosa	0.9	0.0021	0.607	3094	3556	1911	344	0.854	4284	5467	0.784
Pelobates cultripes	1	0.0025	0.493	3192	1998	3360	355	0.724	3953	5358	0.738
Pelophylax perezi	1.3	0.0034	0.956	7249	327	559	770	0.756	186	886	0.21
Pleurodeles waltl	1	0.0025	0.636	2650	3455	2506	294	0.84	4874	5961	0.818
Podarcis bocagei	0.7	0.0015	0.404	675	7160	995	75	0.965	7938	8155	0.973
Podarcis carbonelli	0.5	0.0009	0.219	144	7835	911	15	0.957	8678	8746	0.992
Podarcis hispanica	1.2	0.0031	0.836	5834	683	1763	625	0.679	865	2446	0.354
Psammodromus algirus	1.2	0.0031	0.843	5599	1275	1412	619	0.784	1483	2687	0.552
Psammodromus hispanicus	0.9	0.0021	0.497	2650	2615	3345	295	0.726	4433	5960	0.744
Rana iberica	1	0.0025	0.493	1232	6175	1361	137	0.941	7218	7536	0.958
Rana pyrenaica	1	0.0025	0.428	38	8737	125	5	0.992	8845	8862	0.998
Rhinechis scalaris	1.1	0.0028	0.676	4712	1553	2116	524	0.711	1847	3669	0.503
Timon lepidus	1.2	0.0031	0.918	6259	573	1378	695	0.64	516	1951	0.264
Vipera latastei	0.7	0.0015	0.404	1754	2239	4718	194	0.724	5987	6957	0.861
Vipera seoanei	0.9	0.0021	0.467	653	7378	801	73	0.964	7874	8179	0.963

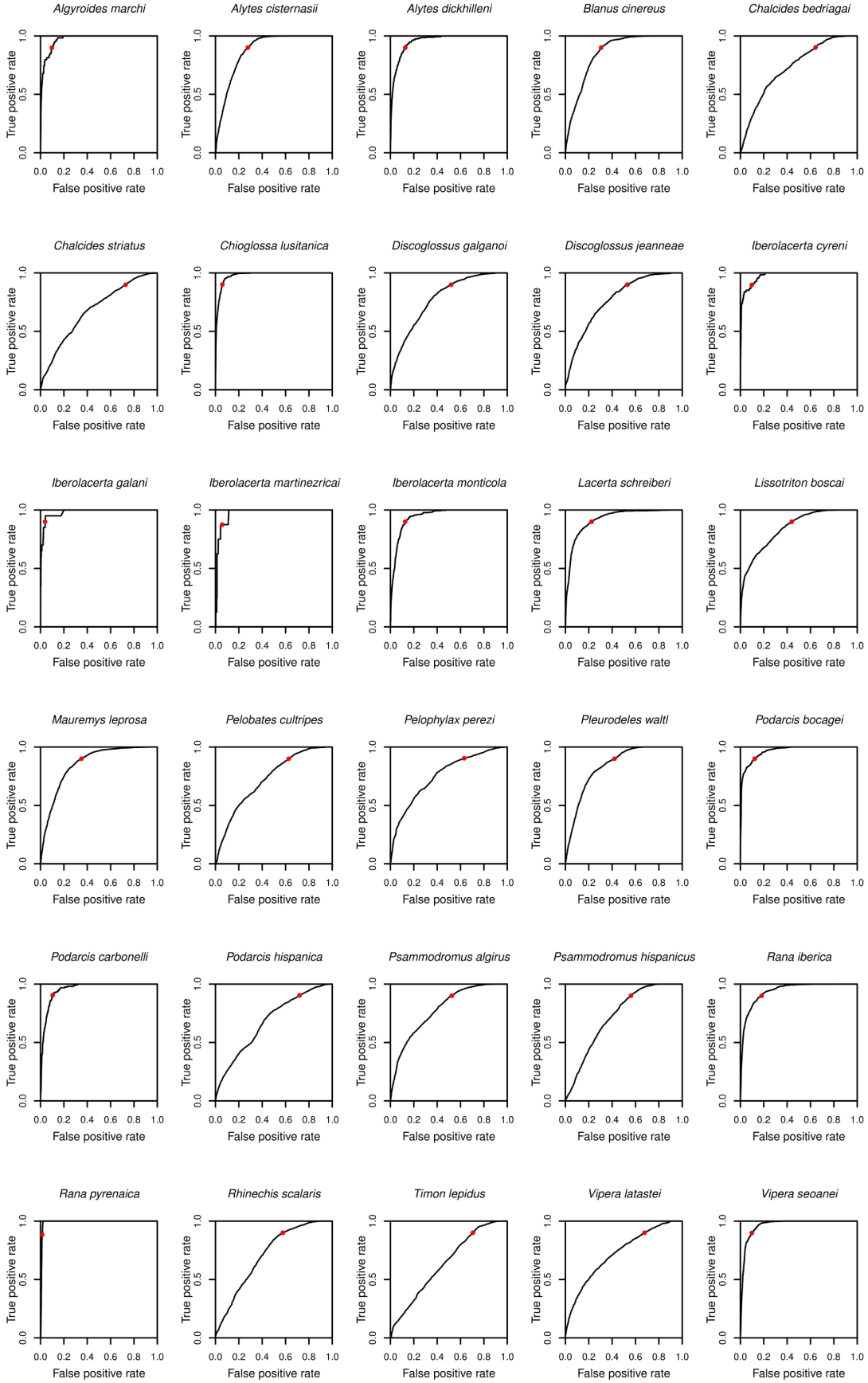


Fig. D.1 – ROC plots for each species model. Red dot indicates the position of the chosen threshold (10^{th} percentile) for binary classification of the continuous model. AUC values are detailed in table D.1.

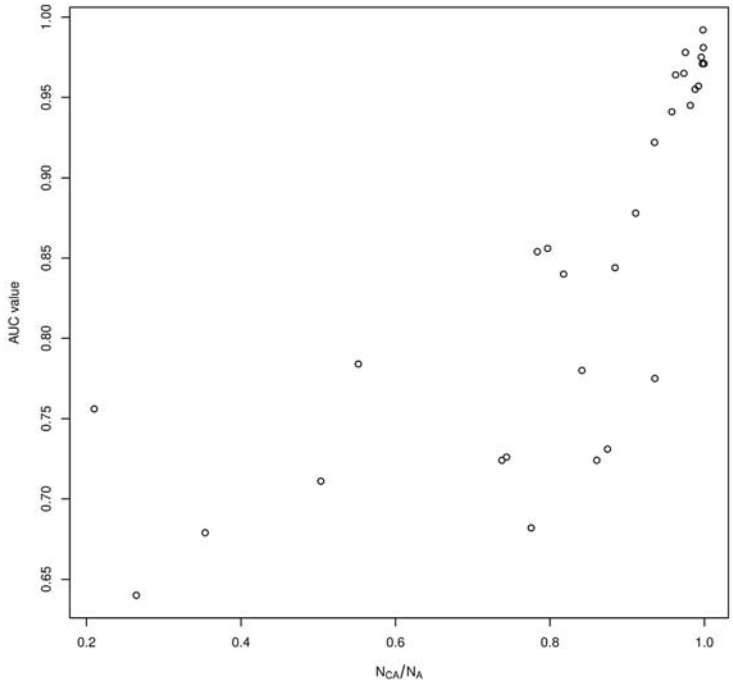


Fig. D.2 – Relation between AUC values and proportion of clustered absences. A clustered absence is found if, at least, 6 neighbours are absences. Proportion is found by dividing the number of clustered absences (NCA) by the total absences (NA).

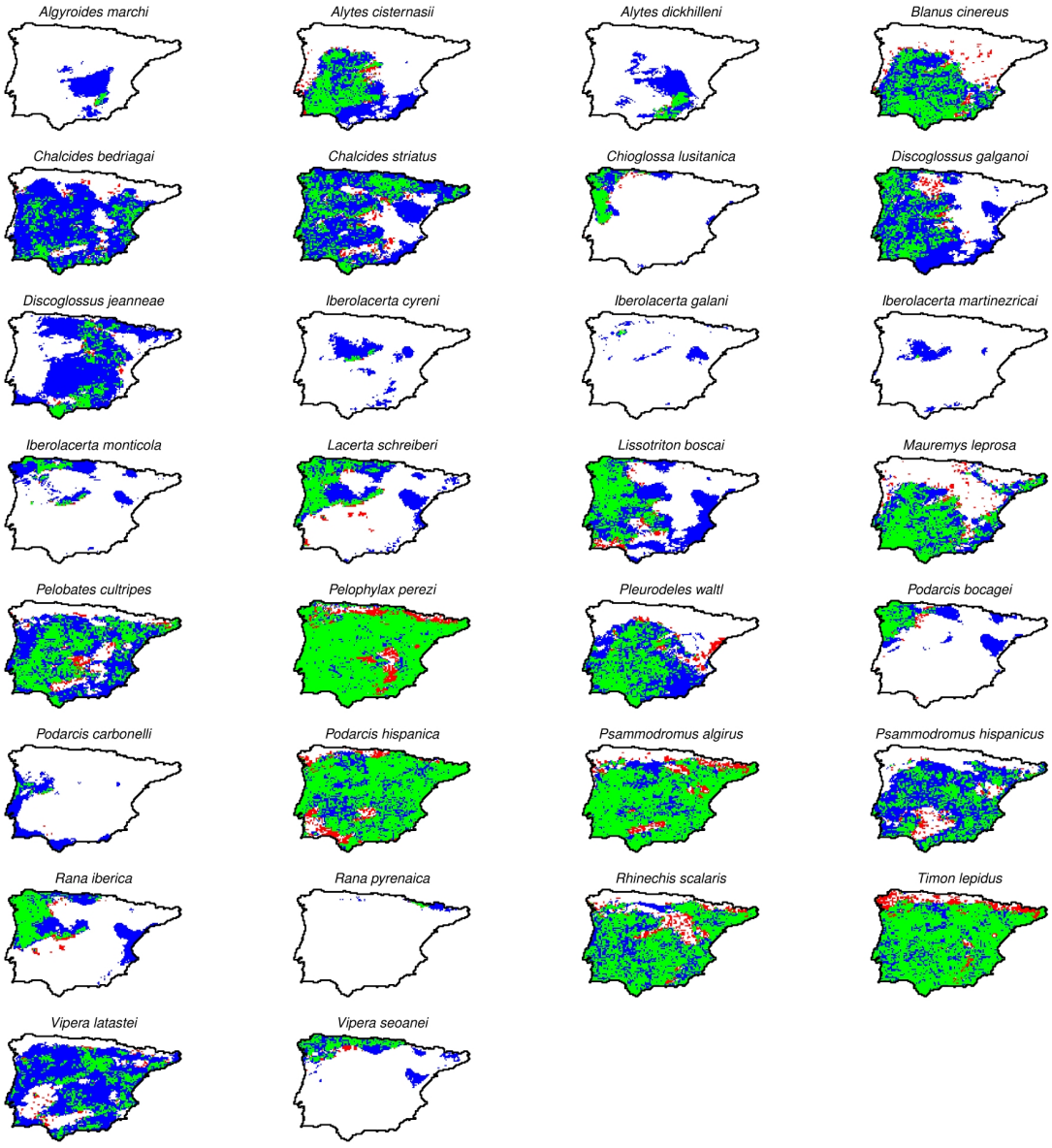


Fig. D.3 – Results of individual species models. Green cells corresponds to the area predicted as presence by the model and where presences of species were found (true positives). Blue area is the predicted area by the model corresponding to a generalization based on the presences of the species but without direct observation of the species (false positives). Red cells represent locations where species was found but left out by the binary model (false negatives). White area corresponds to locations where neither the model predicted presences nor the species was found (true negatives).

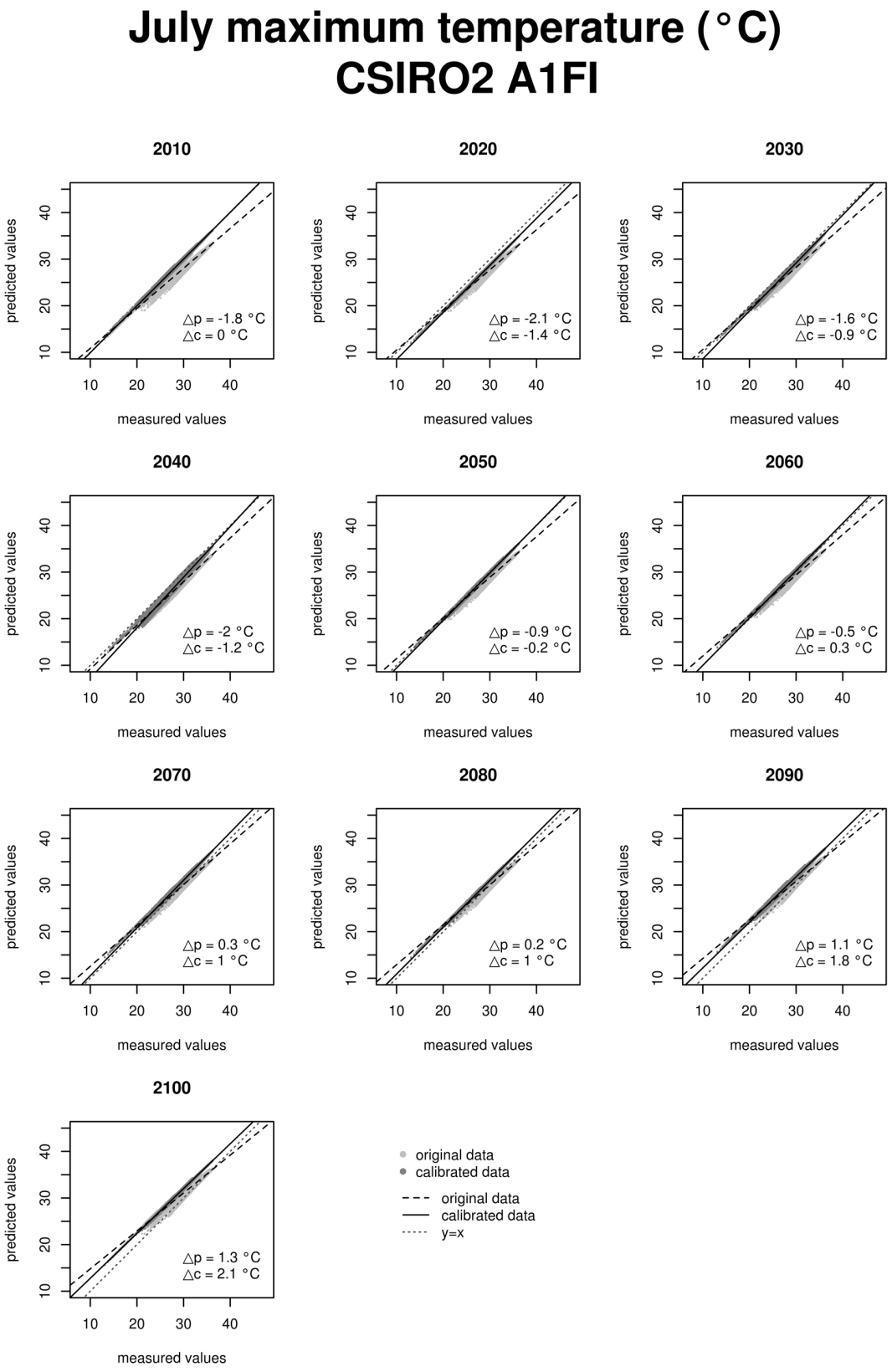
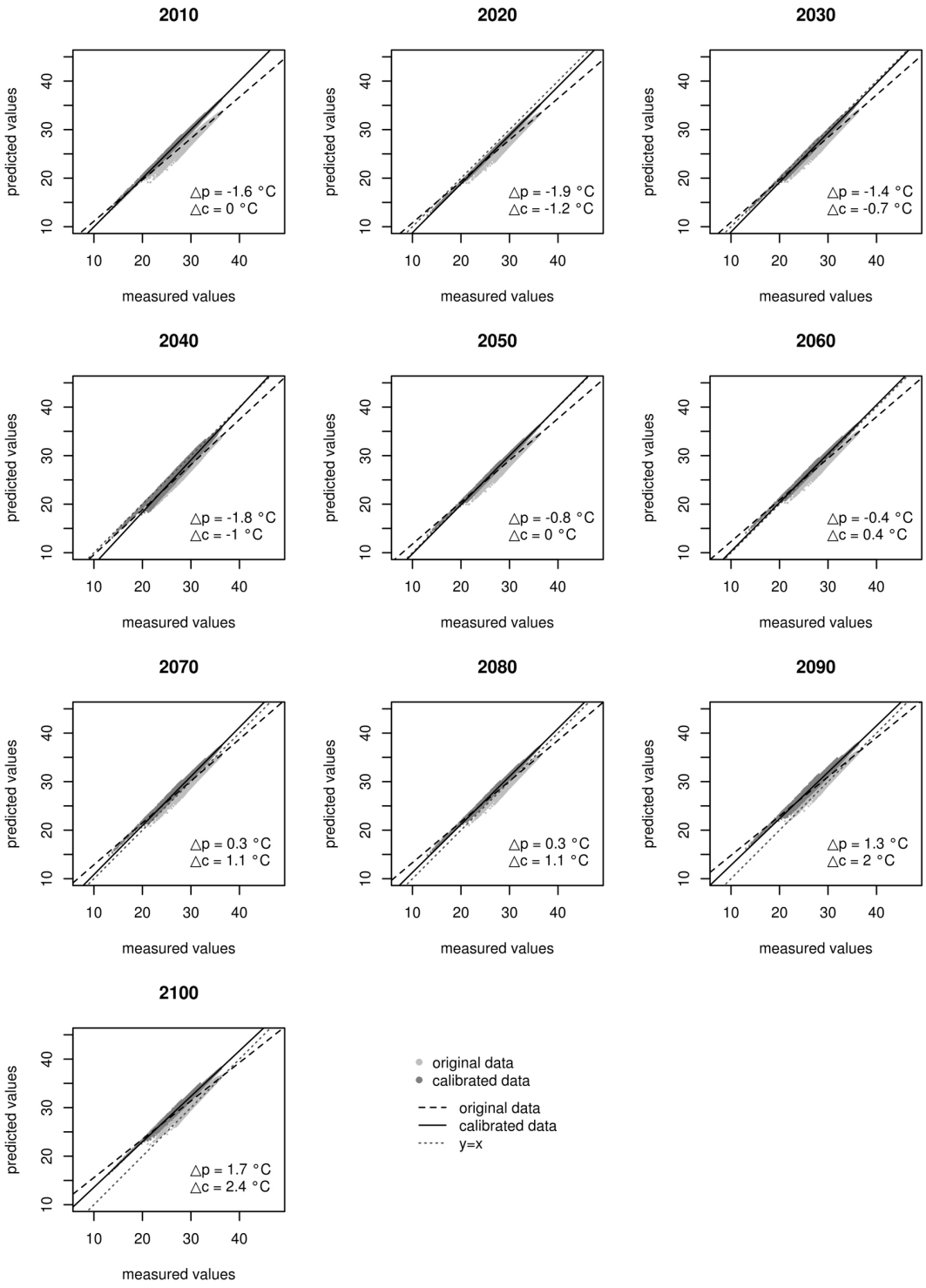


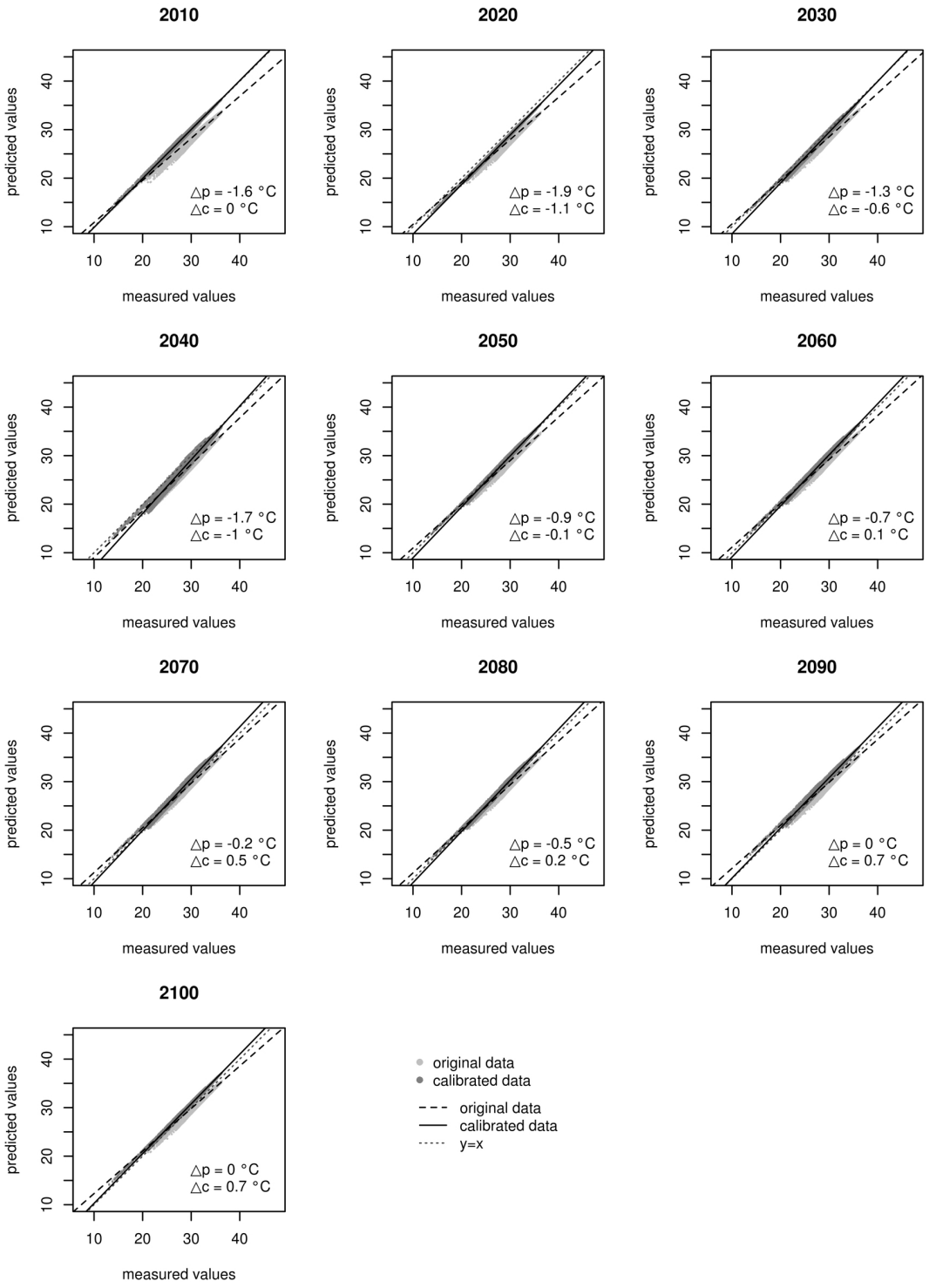
Fig. D.4 – Calibration results of predicted future climate data for each combination of model and emission scenario. Calibration values of the slope and intercept for temperature and only slope for precipitation (intercept=0) were obtained with 2010 data and used to adjust next decades data. Measured values are the current climate values based on WorldClim data and predicted values are given by the combination model/emission scenario for each decade. The difference between predicted and measured climate averages for non-calibrated (Δp) and calibrated data (Δc) are given in the plot. Solid and dashed black lines correspond to calibrated and non-calibrated linear regression. Dotted line is the y=x line.

July maximum temperature (°C)
CSIRO2 A2



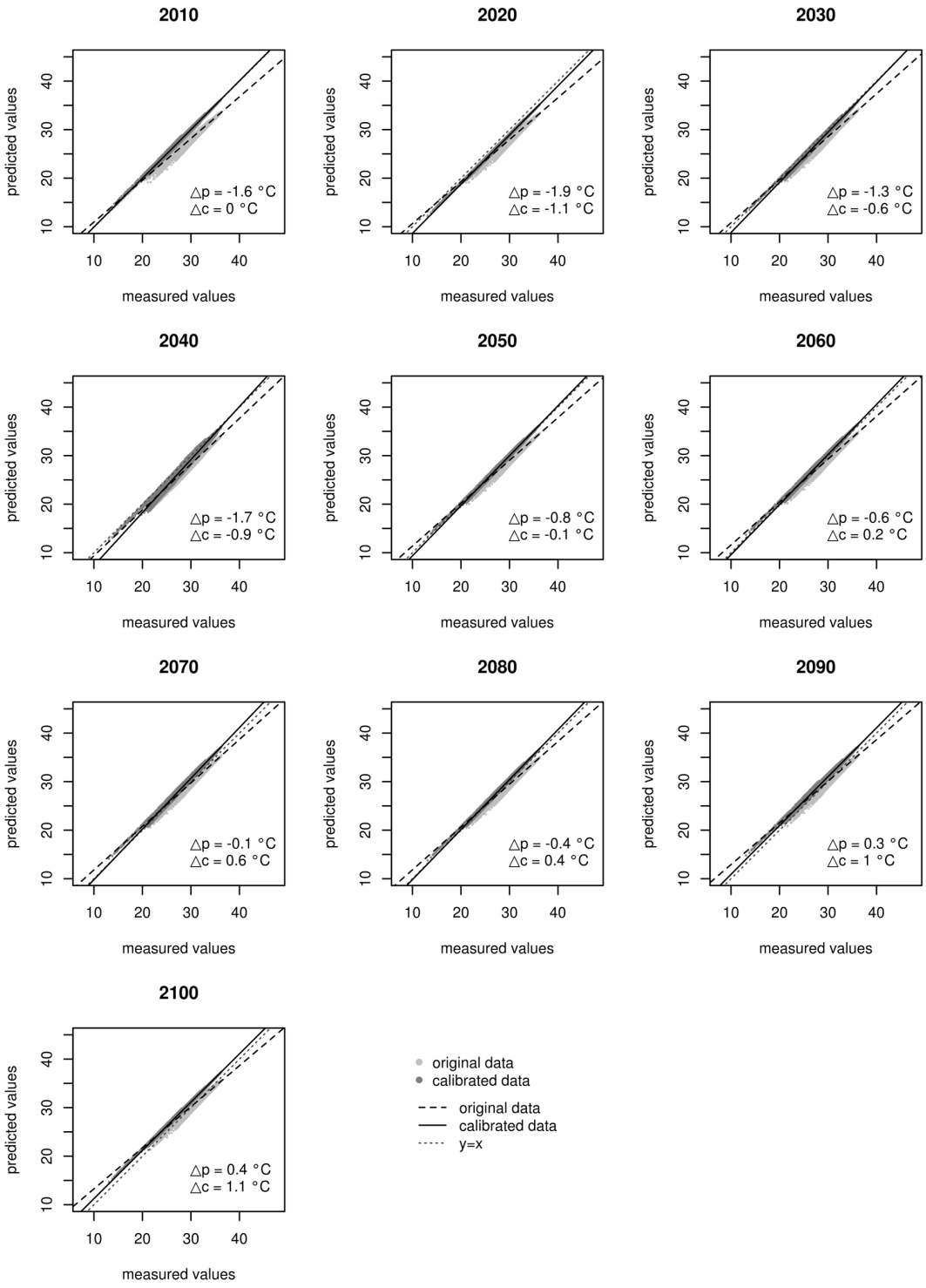
Continued from previous page.

July maximum temperature (°C)
CSIRO2 B1



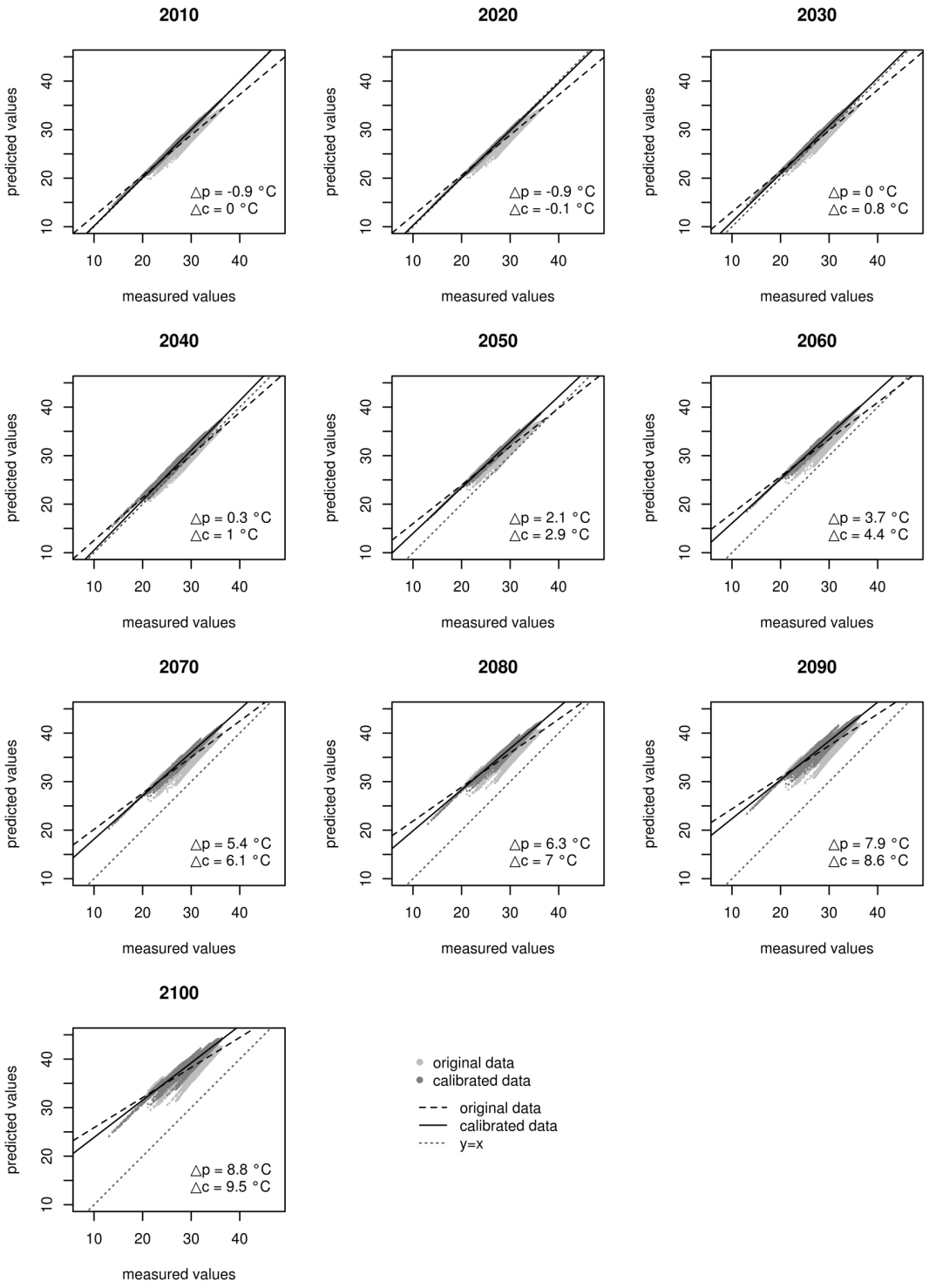
Continued from previous page.

July maximum temperature (°C)
CSIRO2 B2



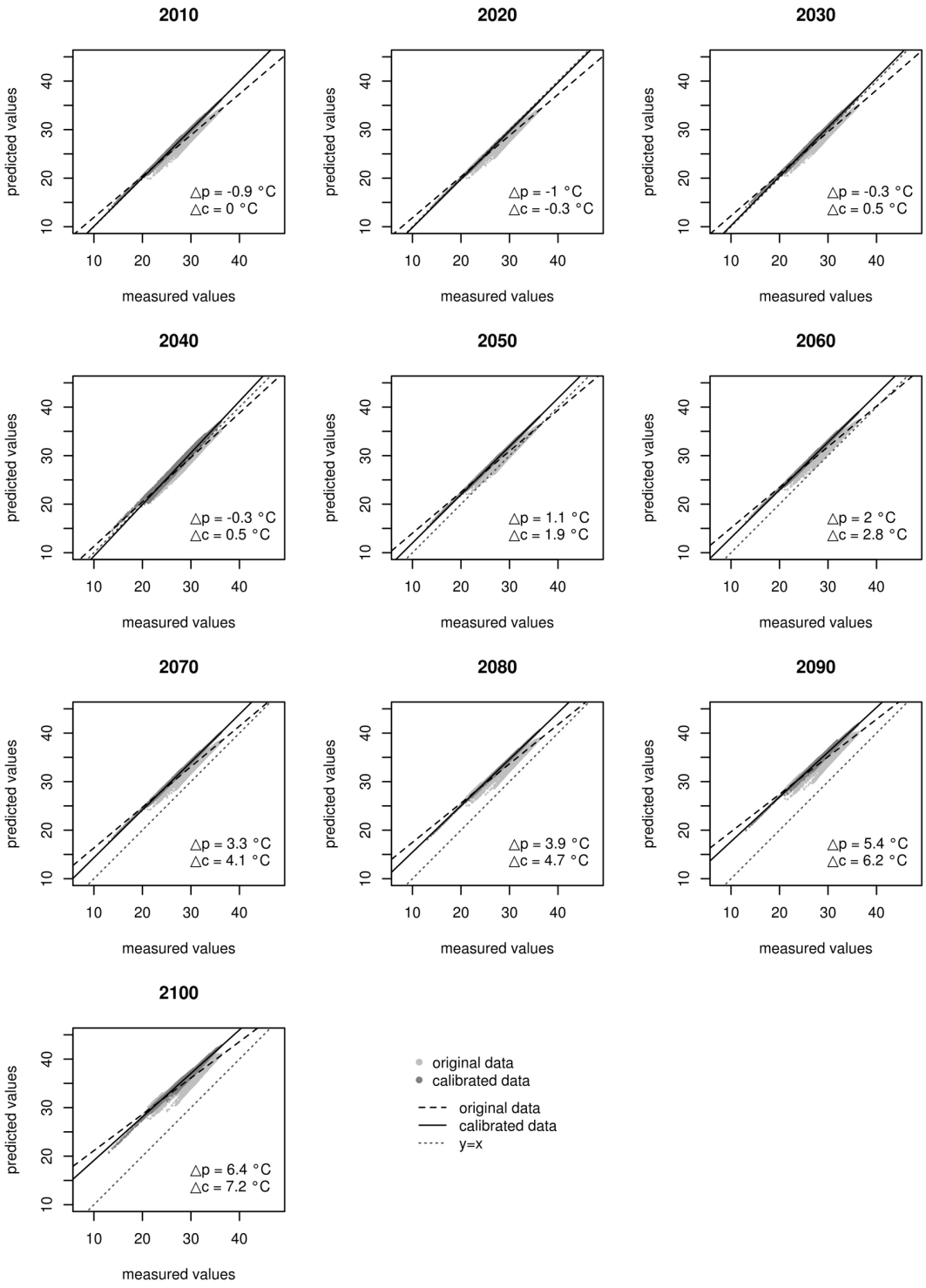
Continued from previous page.

July maximum temperature (°C)
HadCM3 A1FI



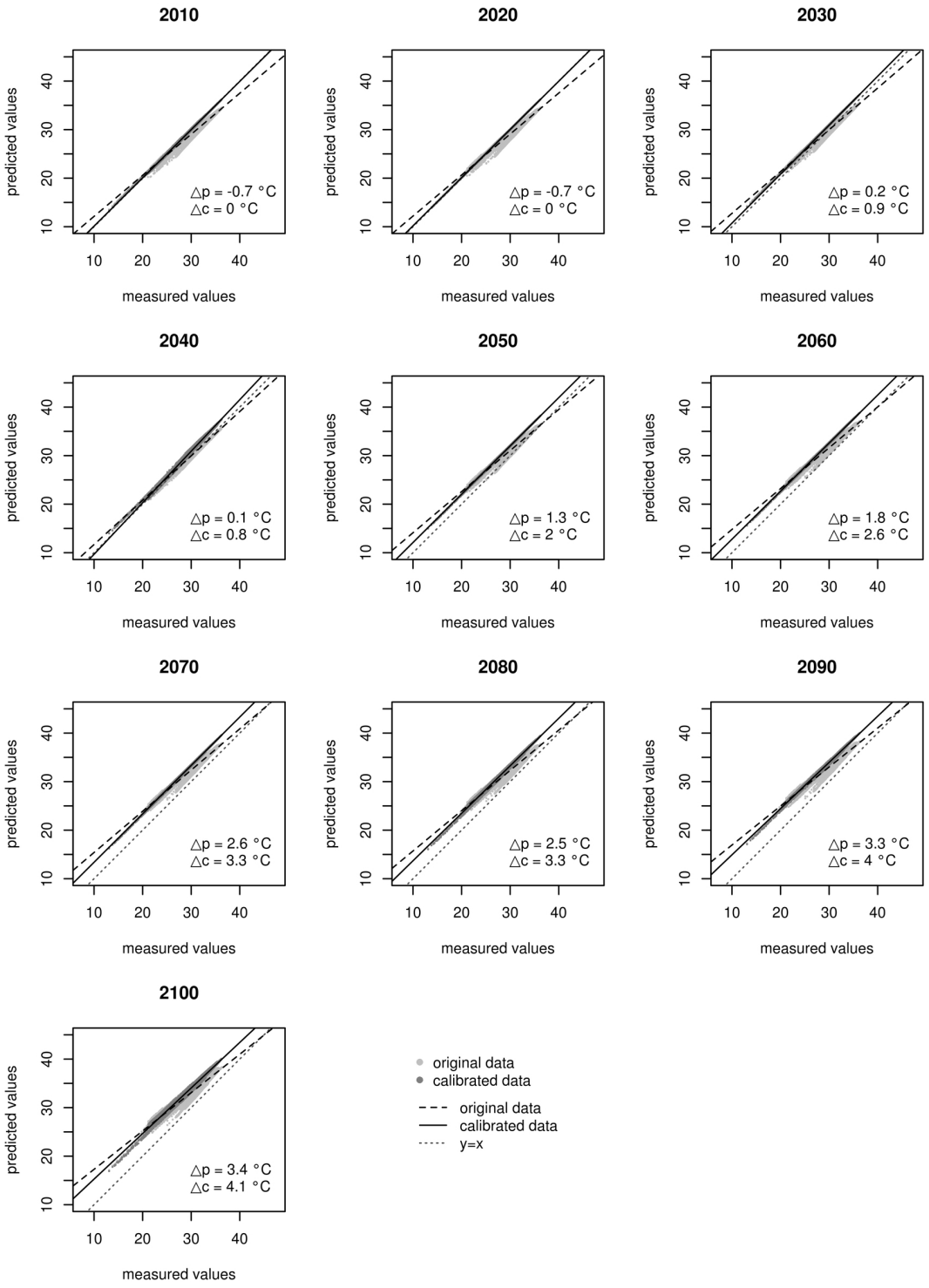
Continued from previous page.

July maximum temperature (°C)
HadCM3 A2



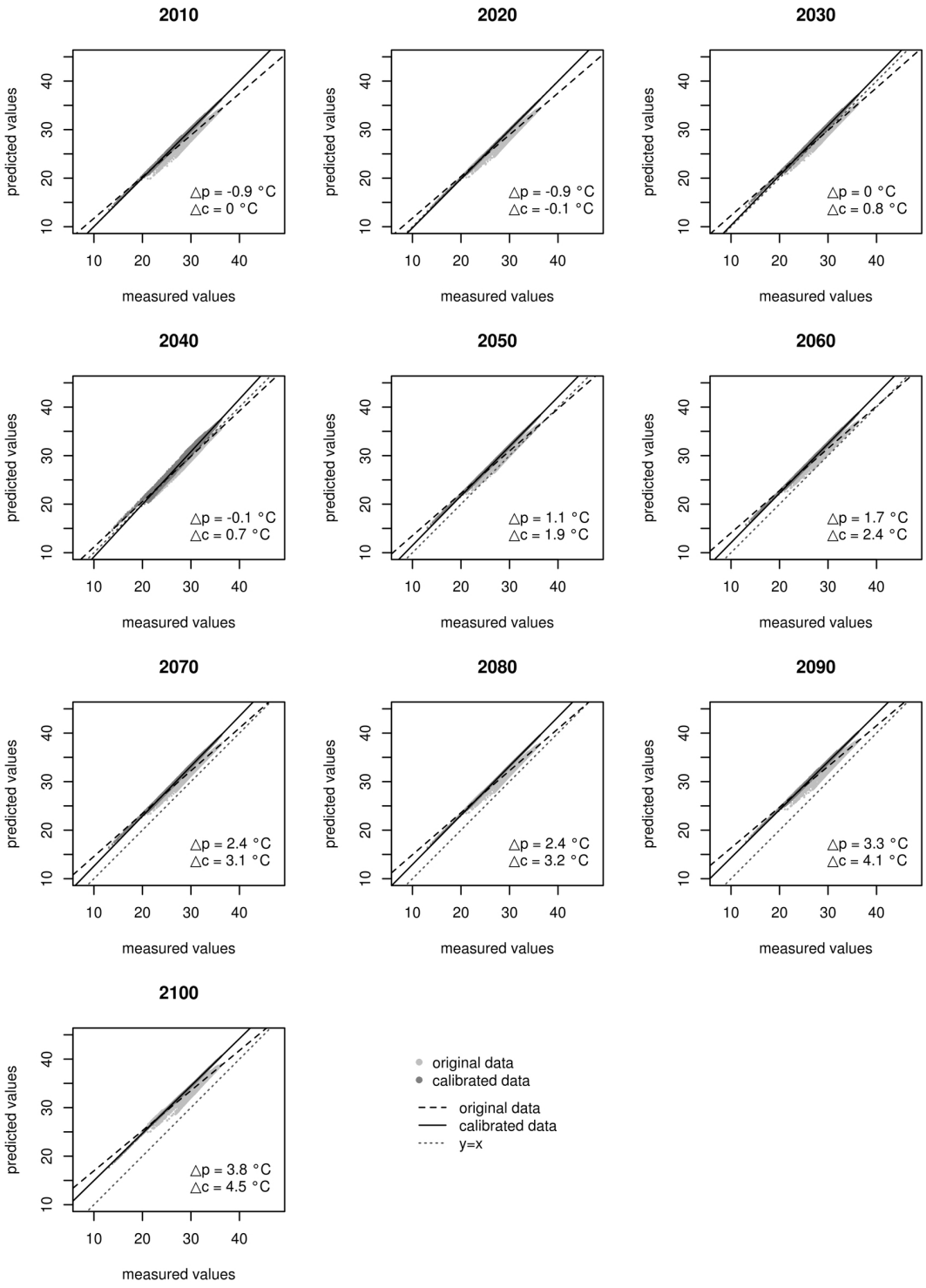
Continued from previous page.

July maximum temperature (°C)
HadCM3 B1



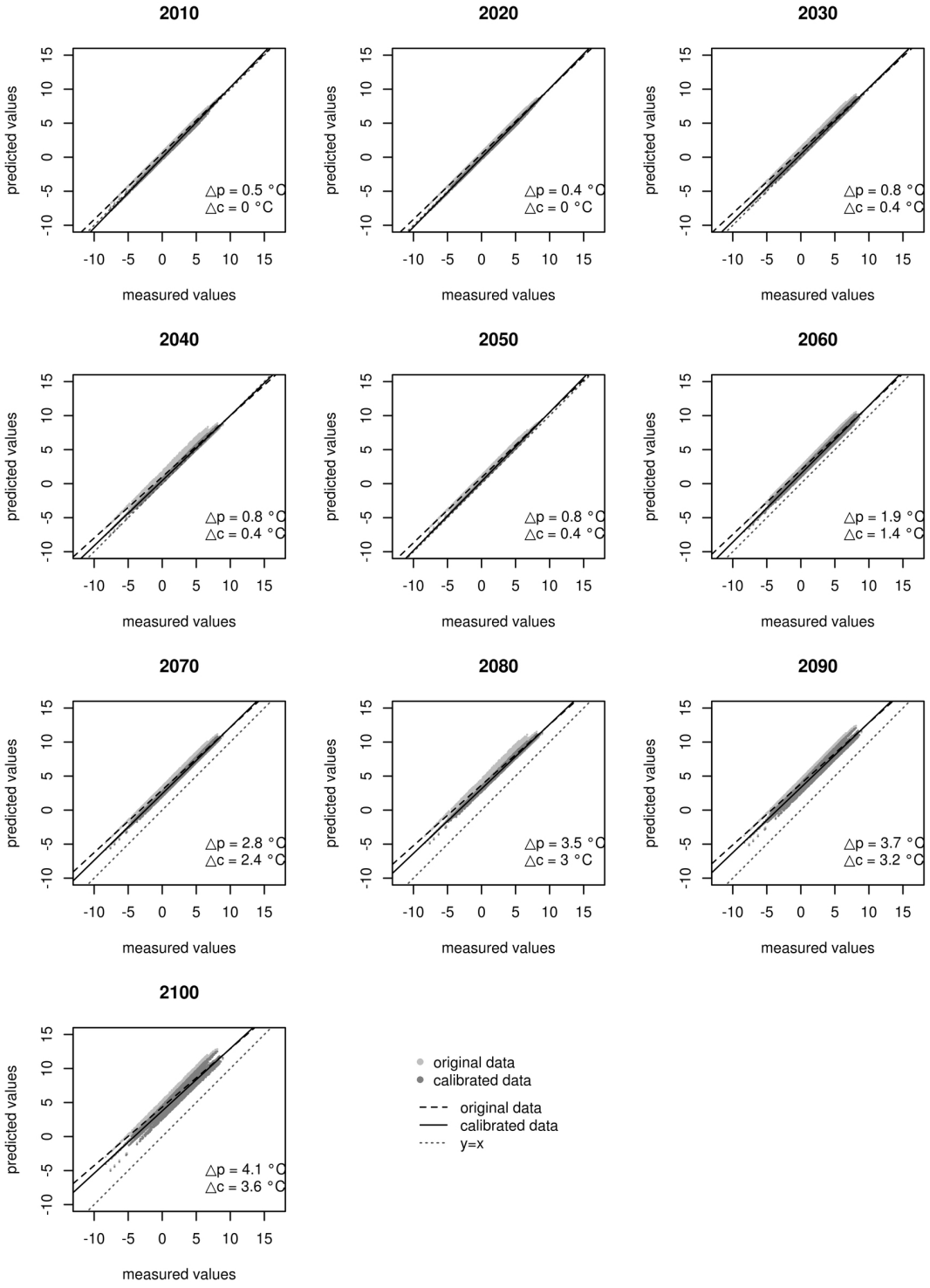
Continued from previous page.

July maximum temperature (°C)
HadCM3 B2



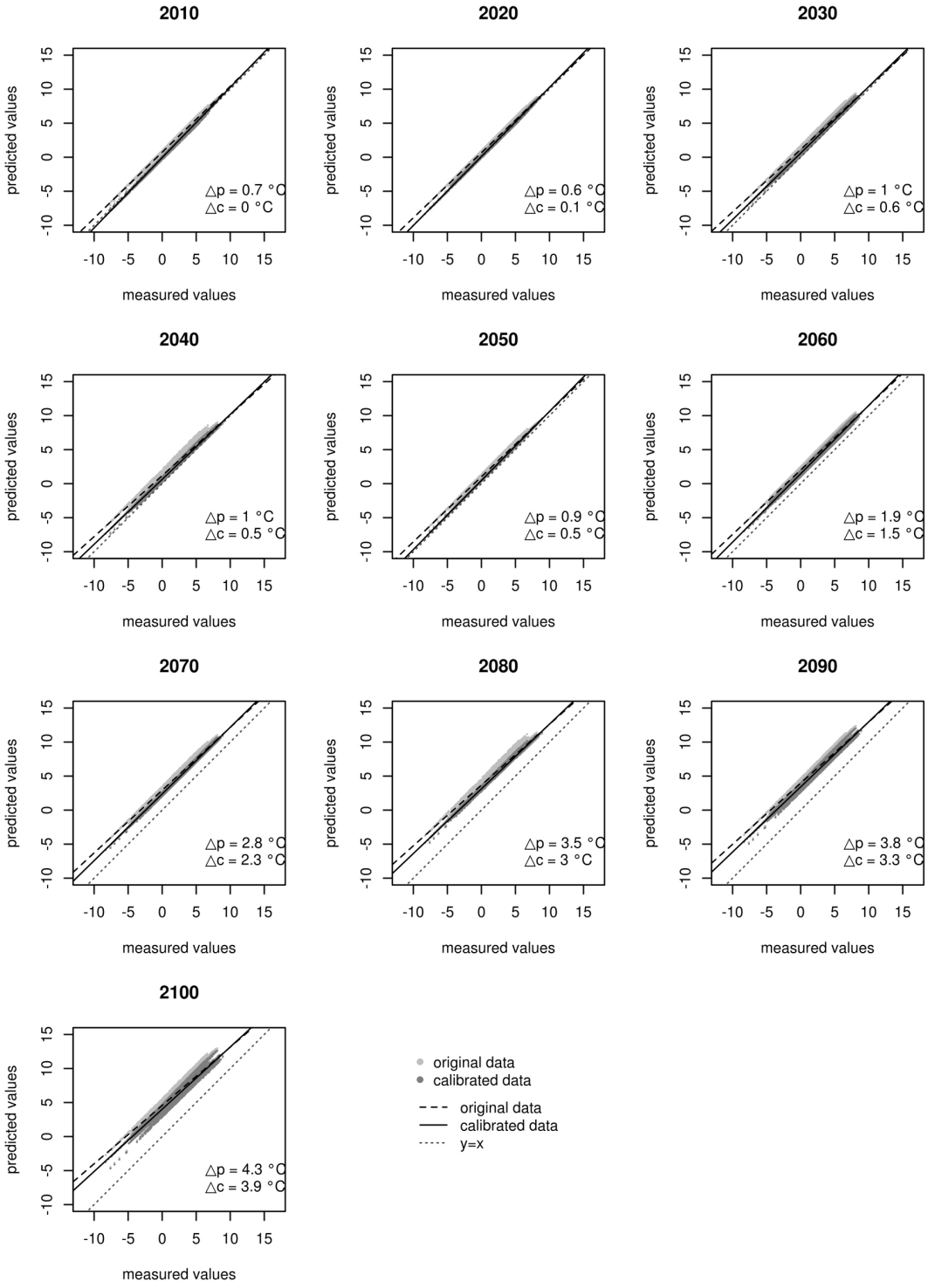
Continued from previous page.

January minimum temperature (°C)
CSIRO2 A1FI



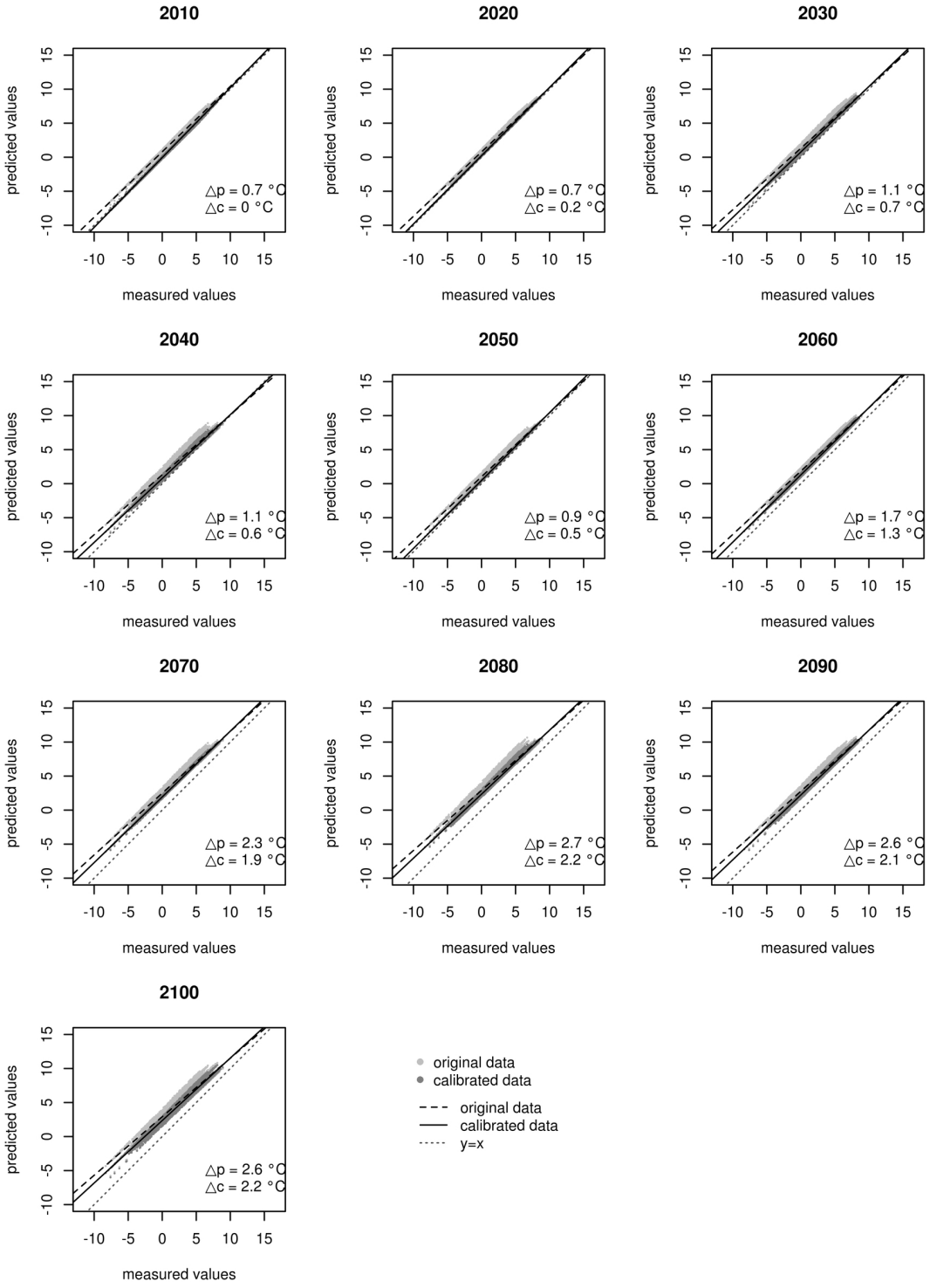
Continued from previous page.

January minimum temperature (°C)
CSIRO2 A2



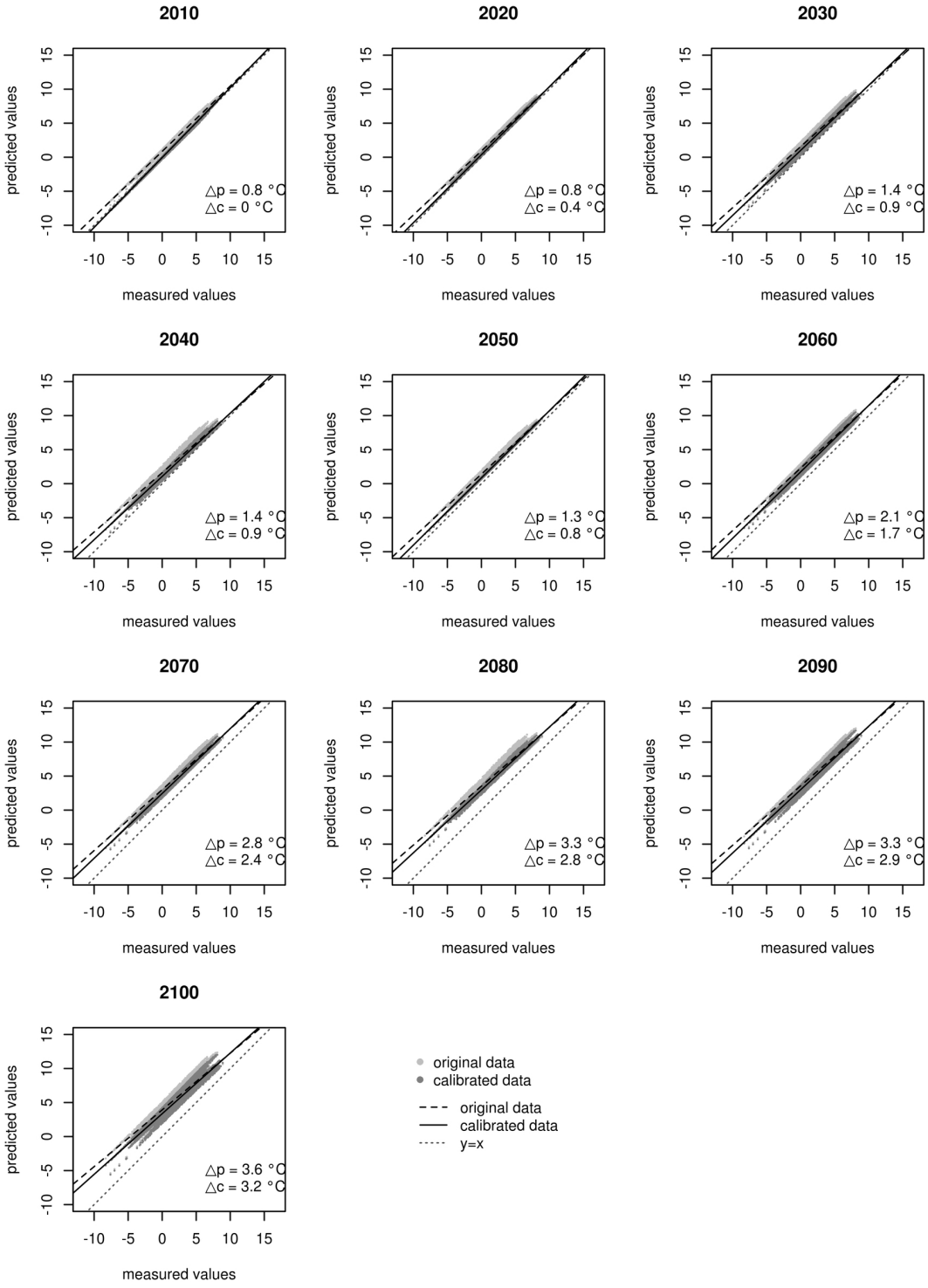
Continued from previous page.

January minimum temperature (°C)
CSIRO2 B1



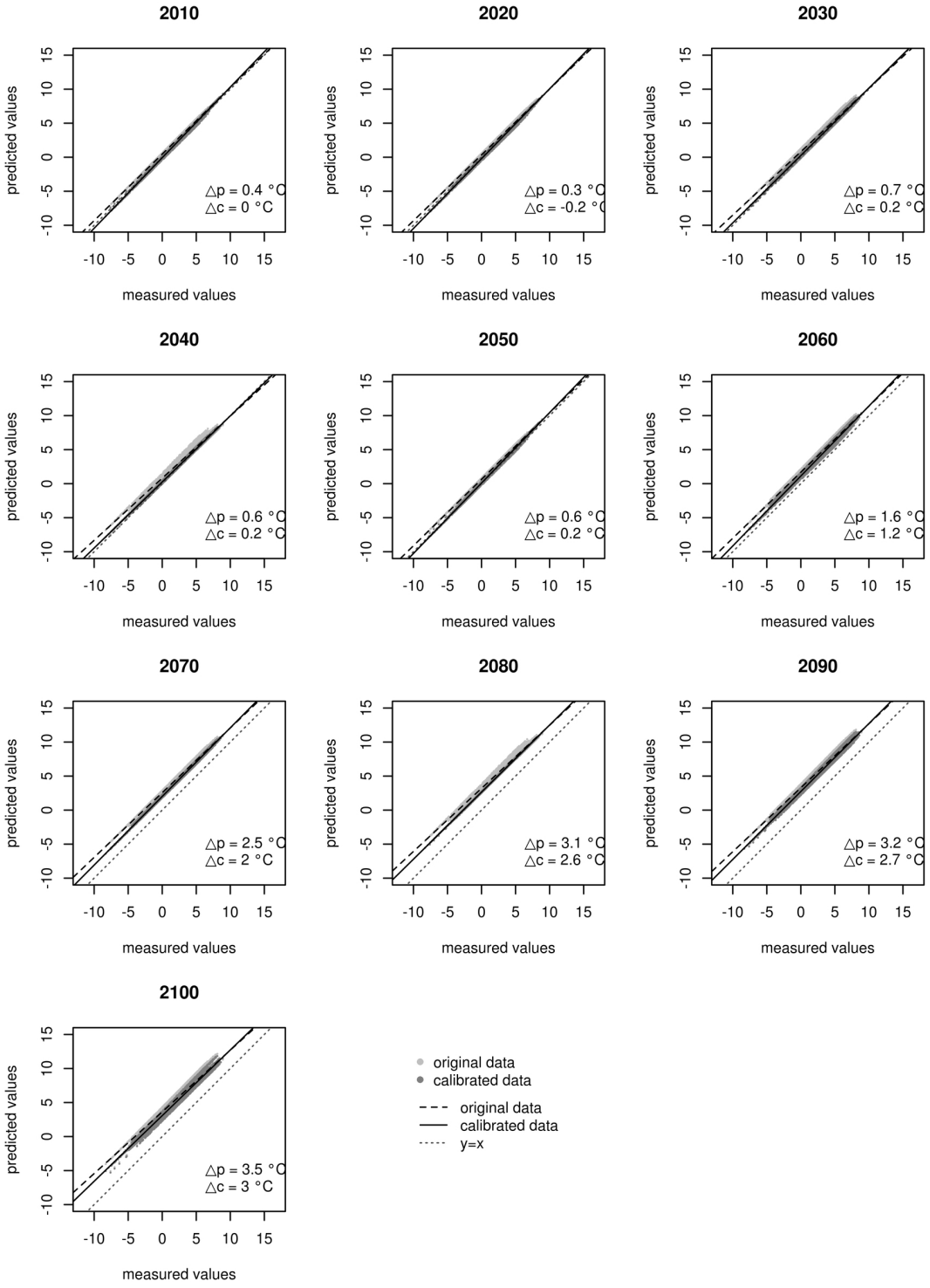
Continued from previous page.

January minimum temperature (°C)
CSIRO2 B2



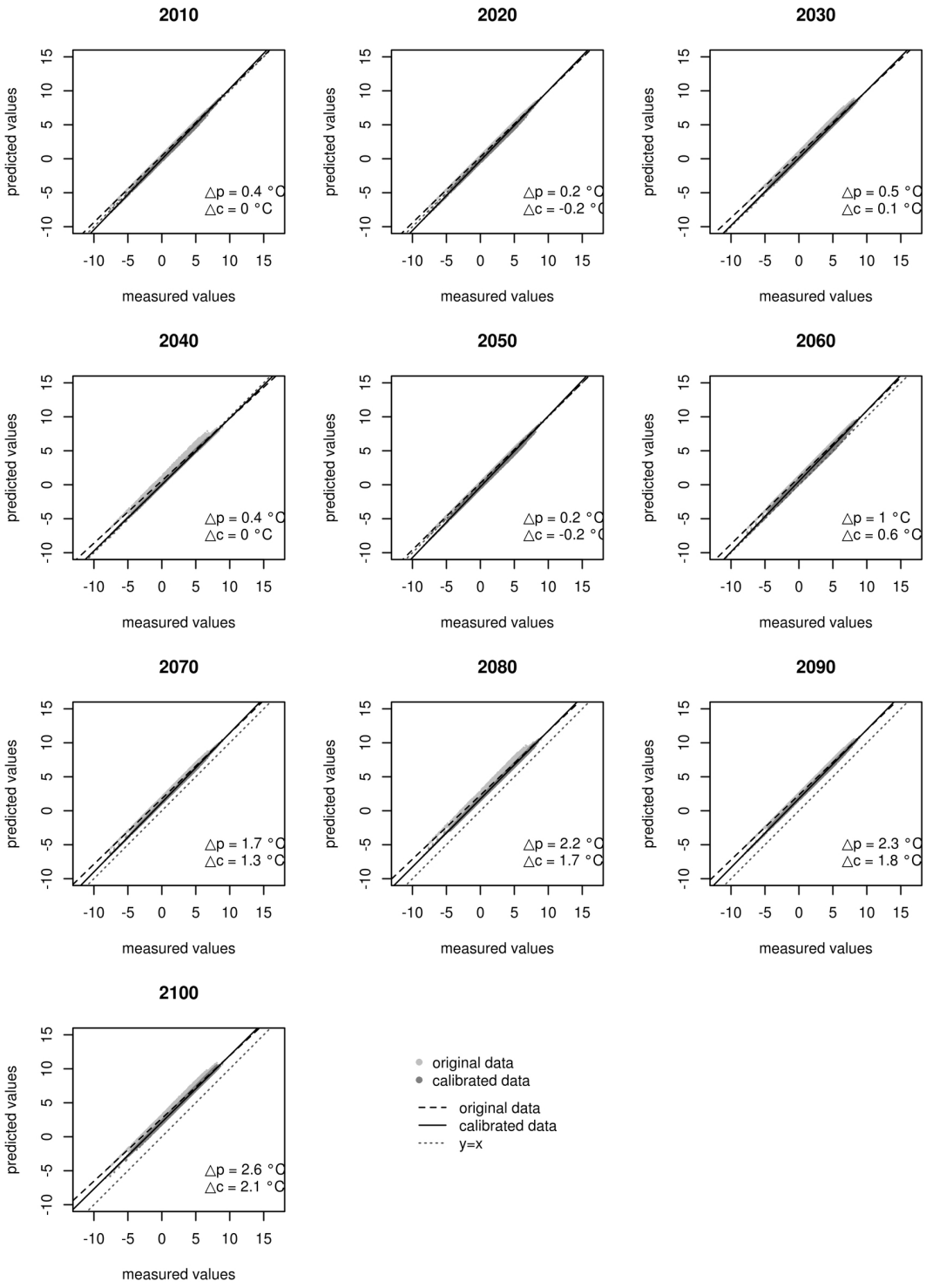
Continued from previous page.

January minimum temperature (°C)
HadCM3 A1FI



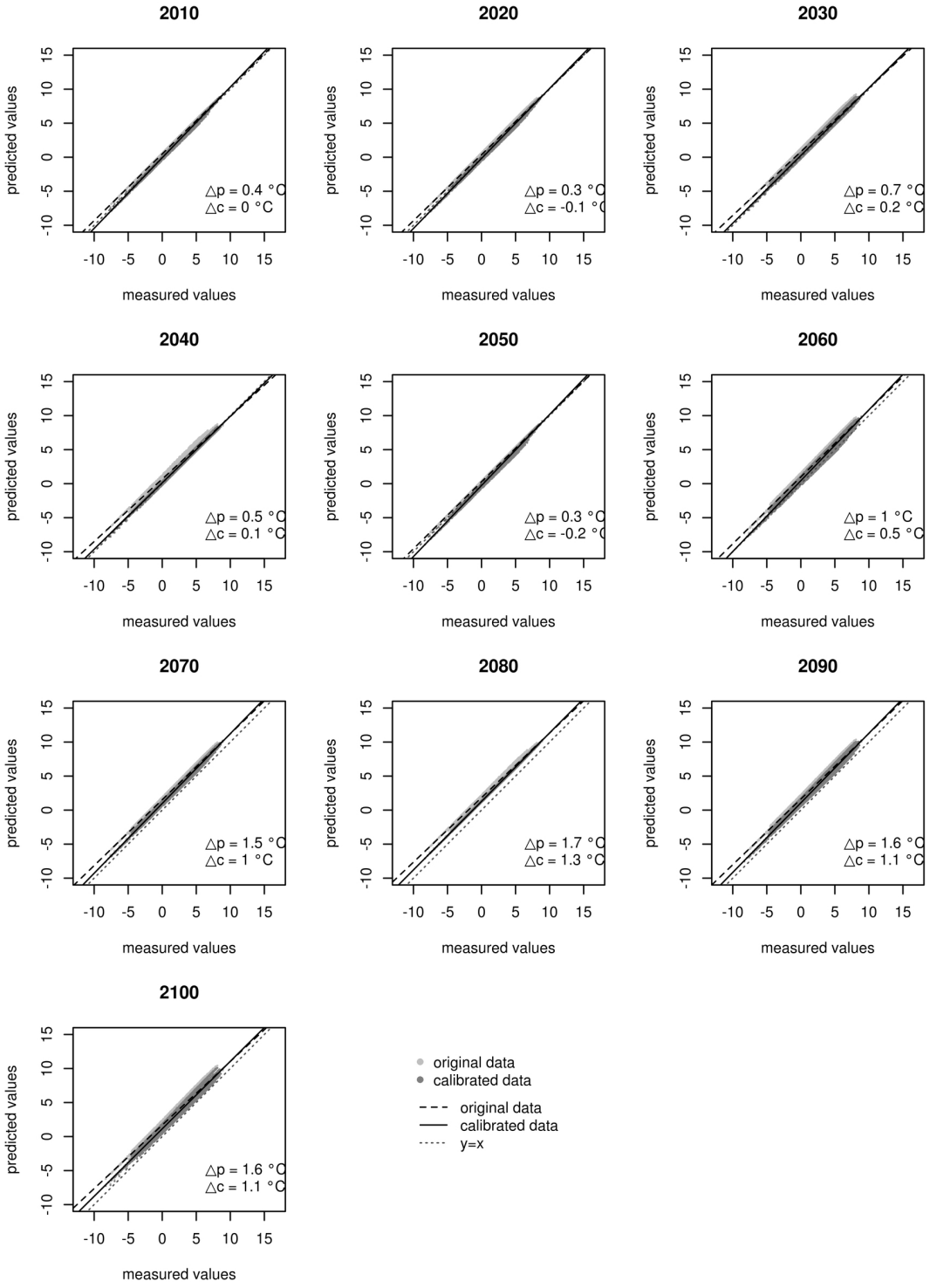
Continued from previous page.

January minimum temperature (°C)
HadCM3 A2



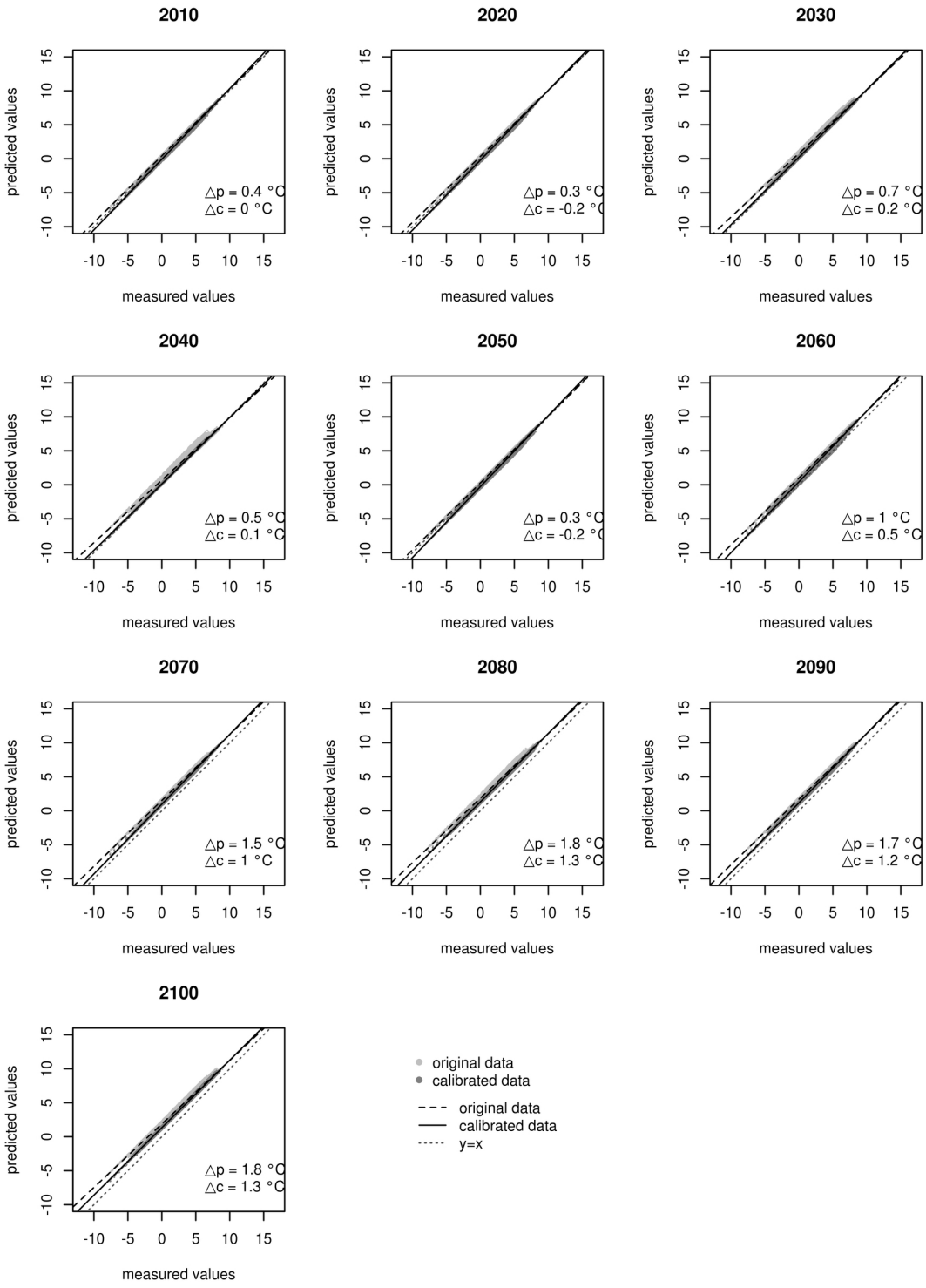
Continued from previous page.

January minimum temperature (°C)
HadCM3 B1



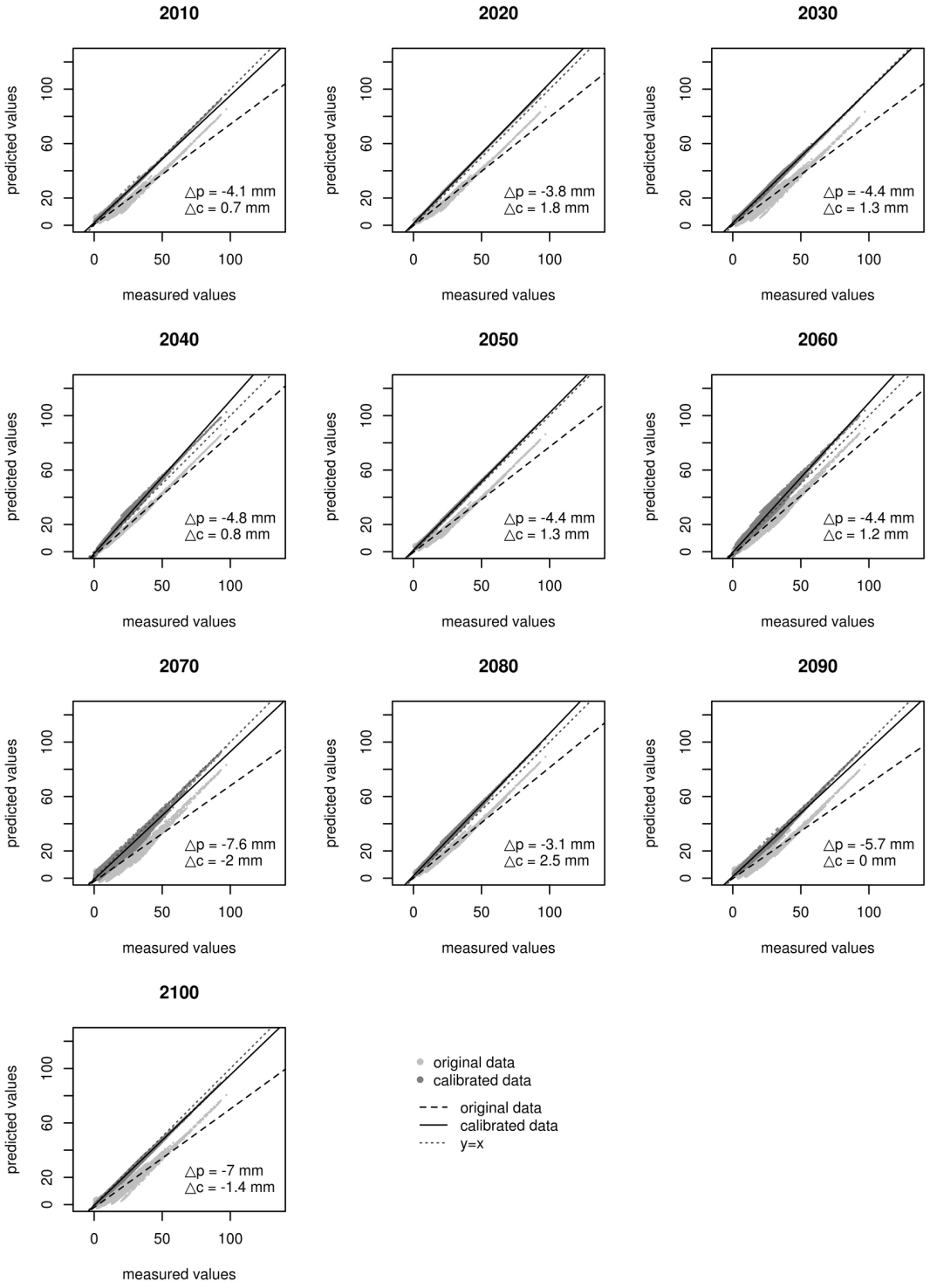
Continued from previous page.

January minimum temperature (°C)
HadCM3 B2



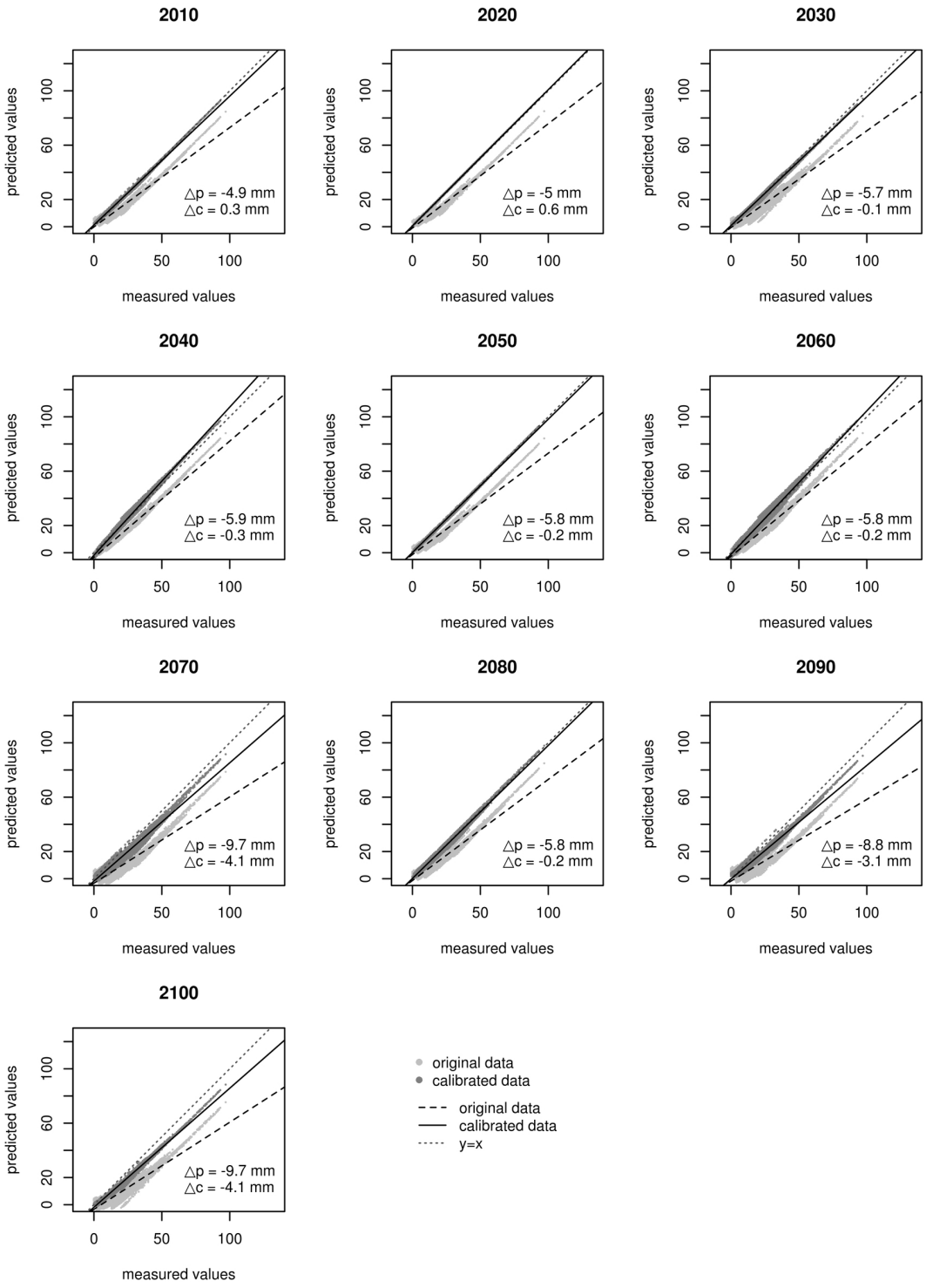
Continued from previous page.

Minimum annual precipitation (mm)
CSIRO2 A1FI



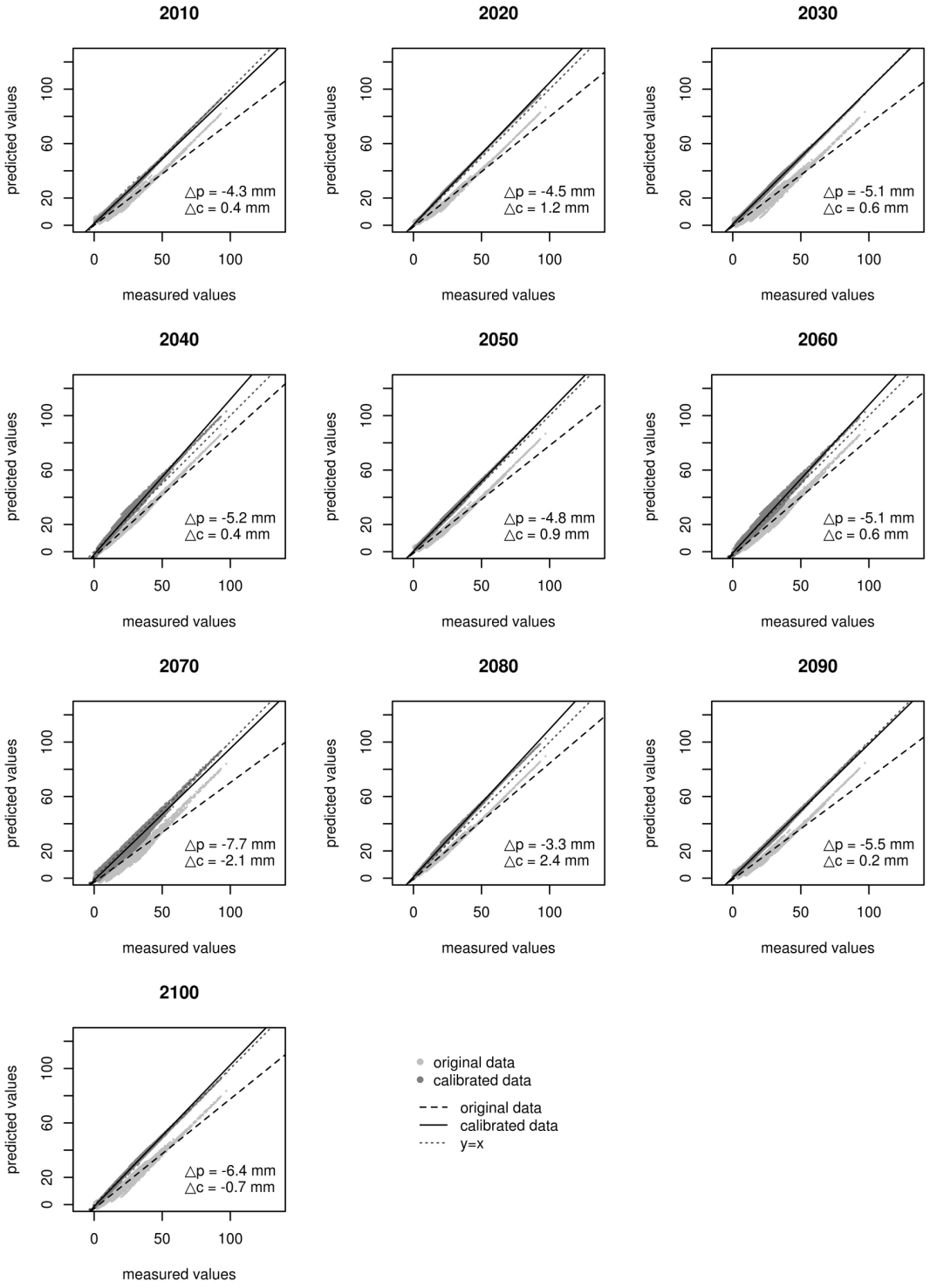
Continued from previous page.

Minimum annual precipitation (mm)
CSIRO2 A2



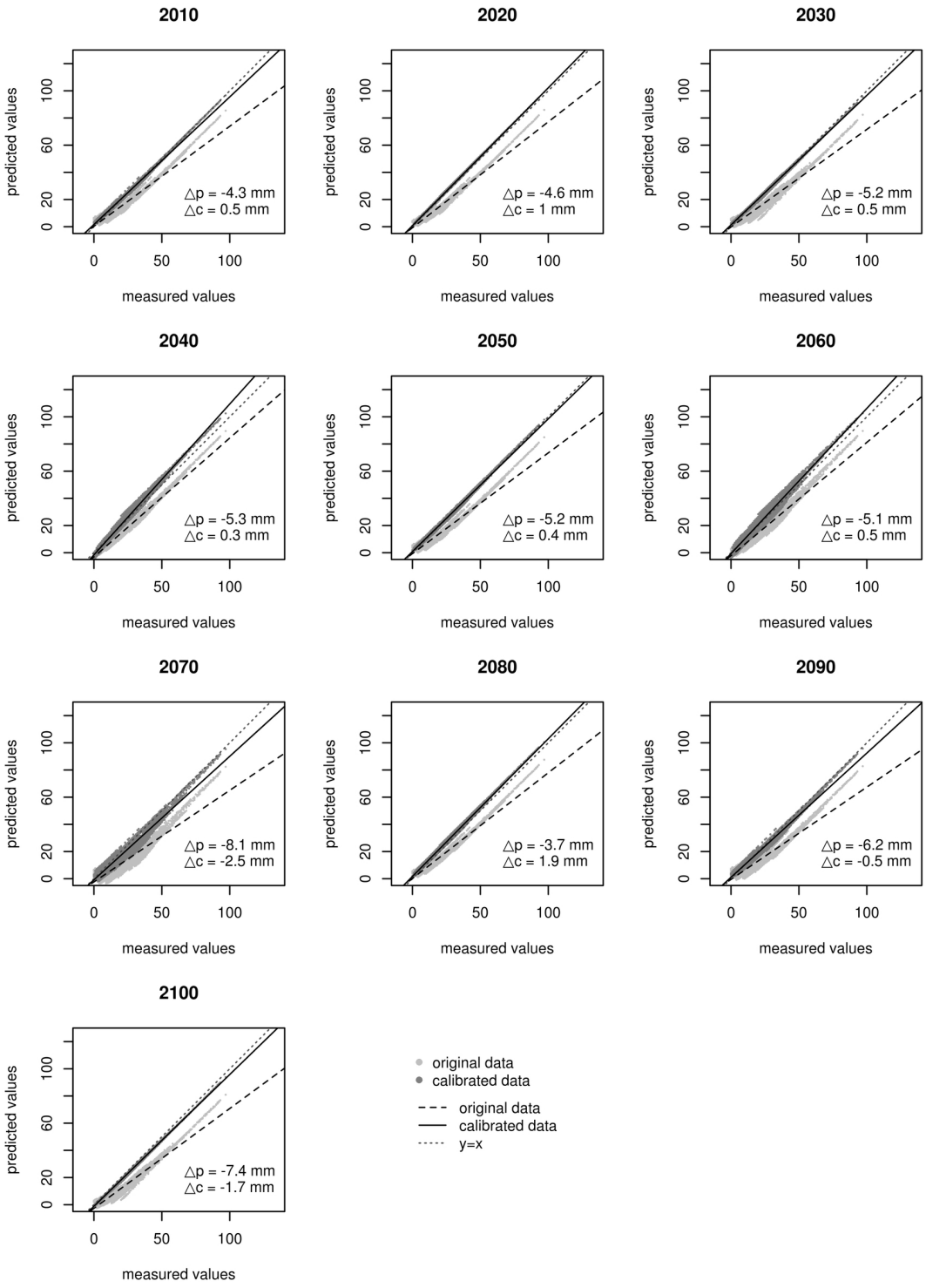
Continued from previous page.

Minimum annual precipitation (mm)
CSIRO2 B1



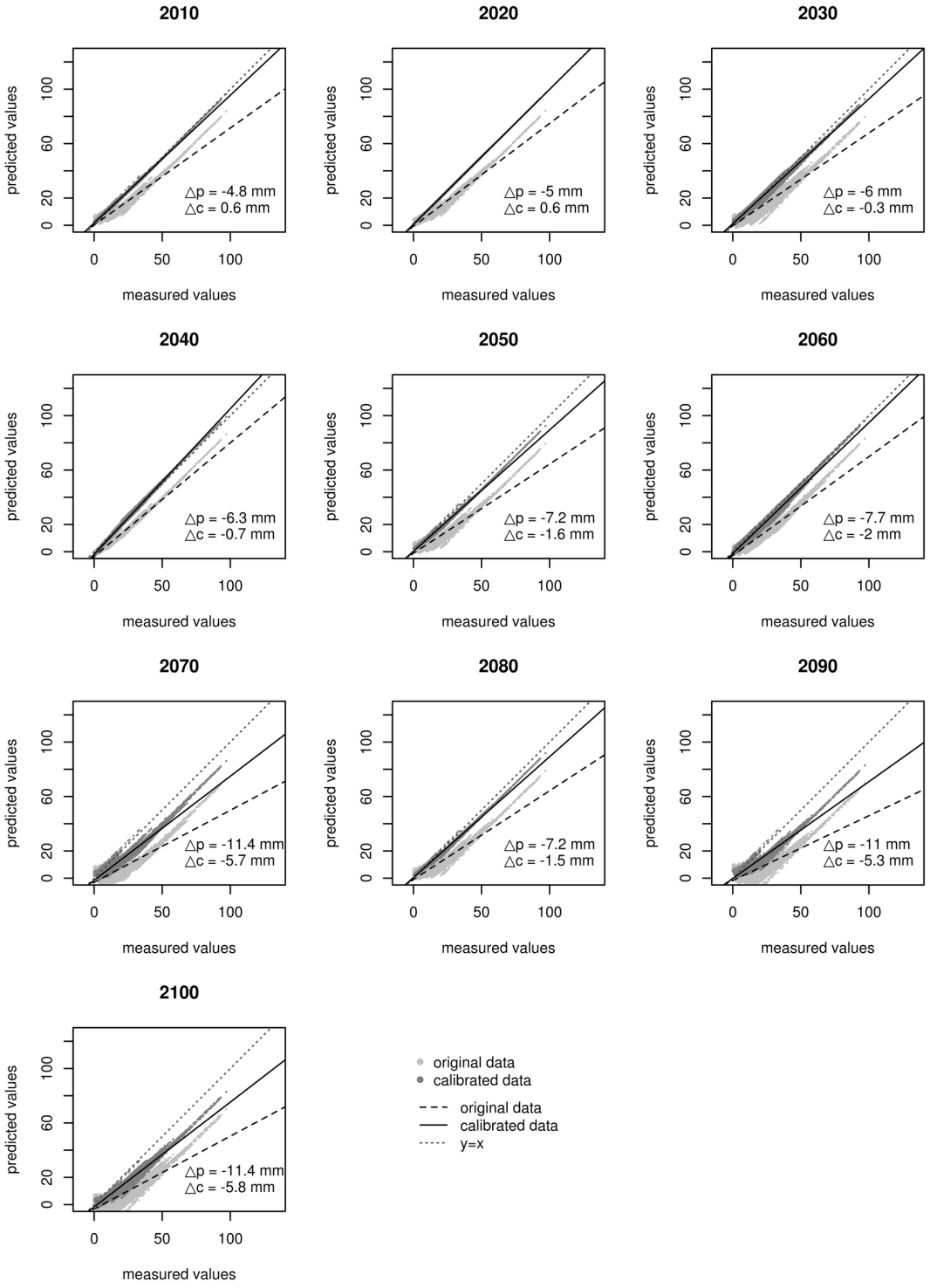
Continued from previous page.

Minimum annual precipitation (mm)
CSIRO2 B2



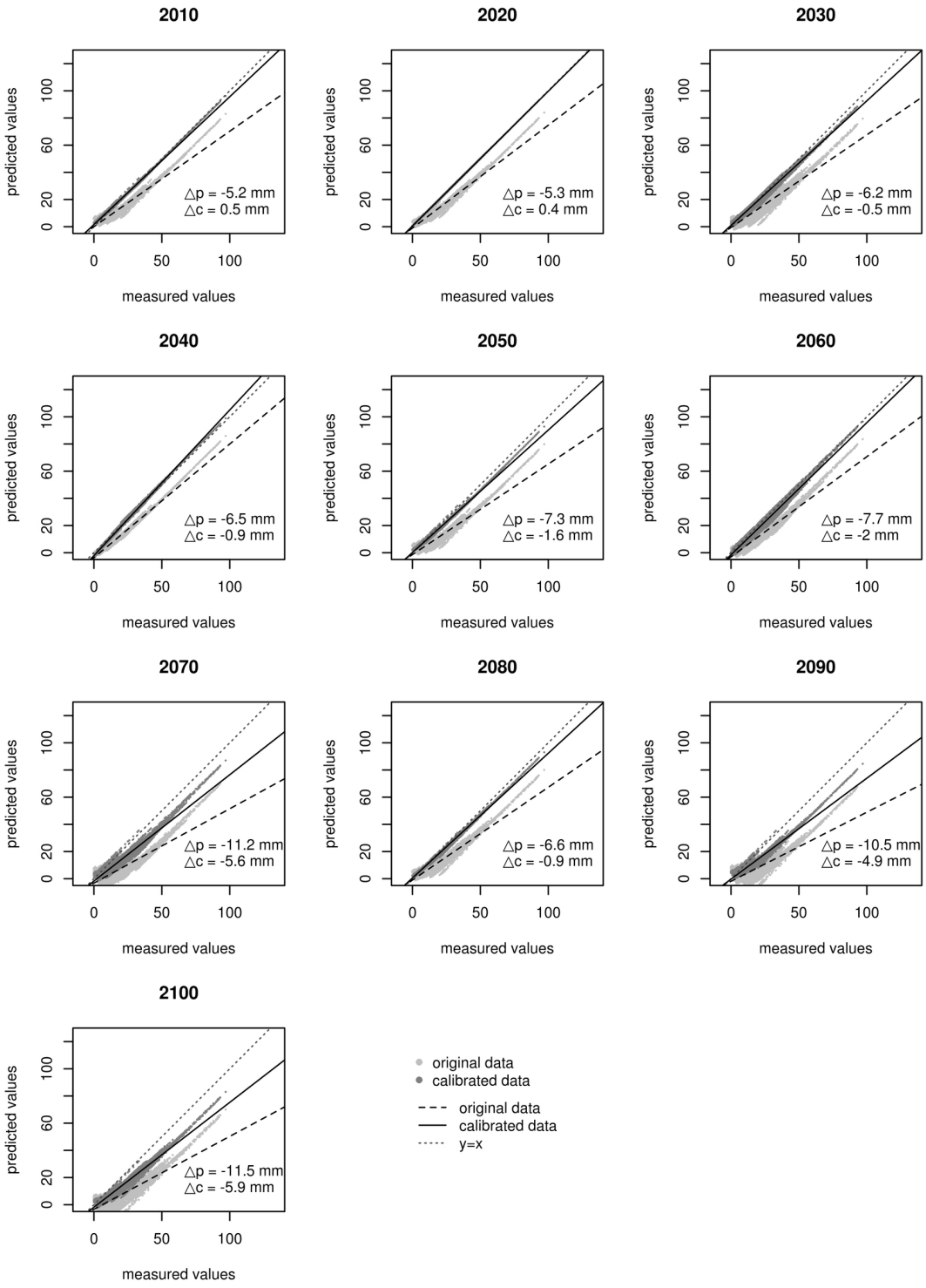
Continued from previous page.

Minimum annual precipitation (mm)
HadCM3 A1FI



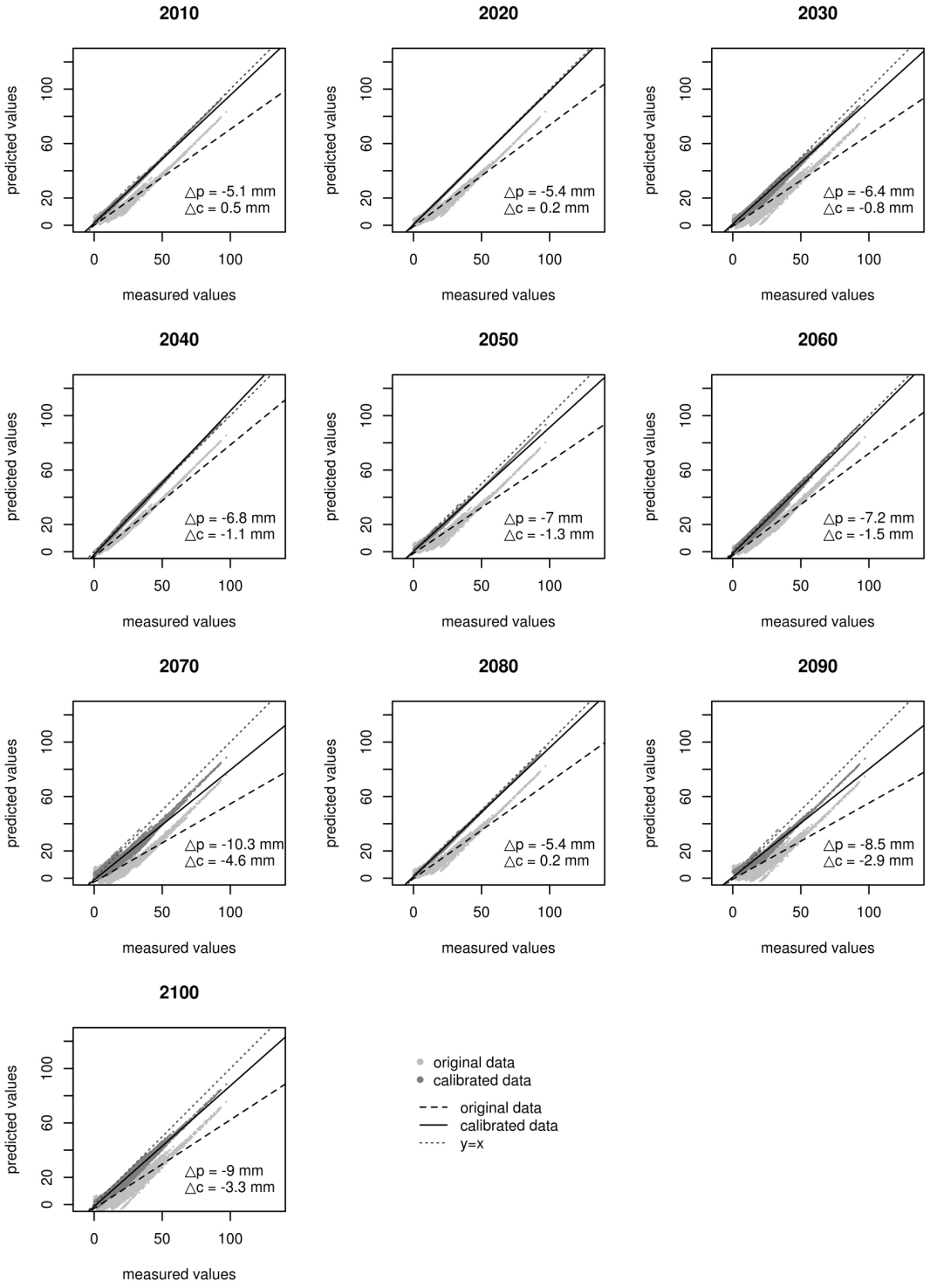
Continued from previous page.

Minimum annual precipitation (mm)
HadCM3 A2



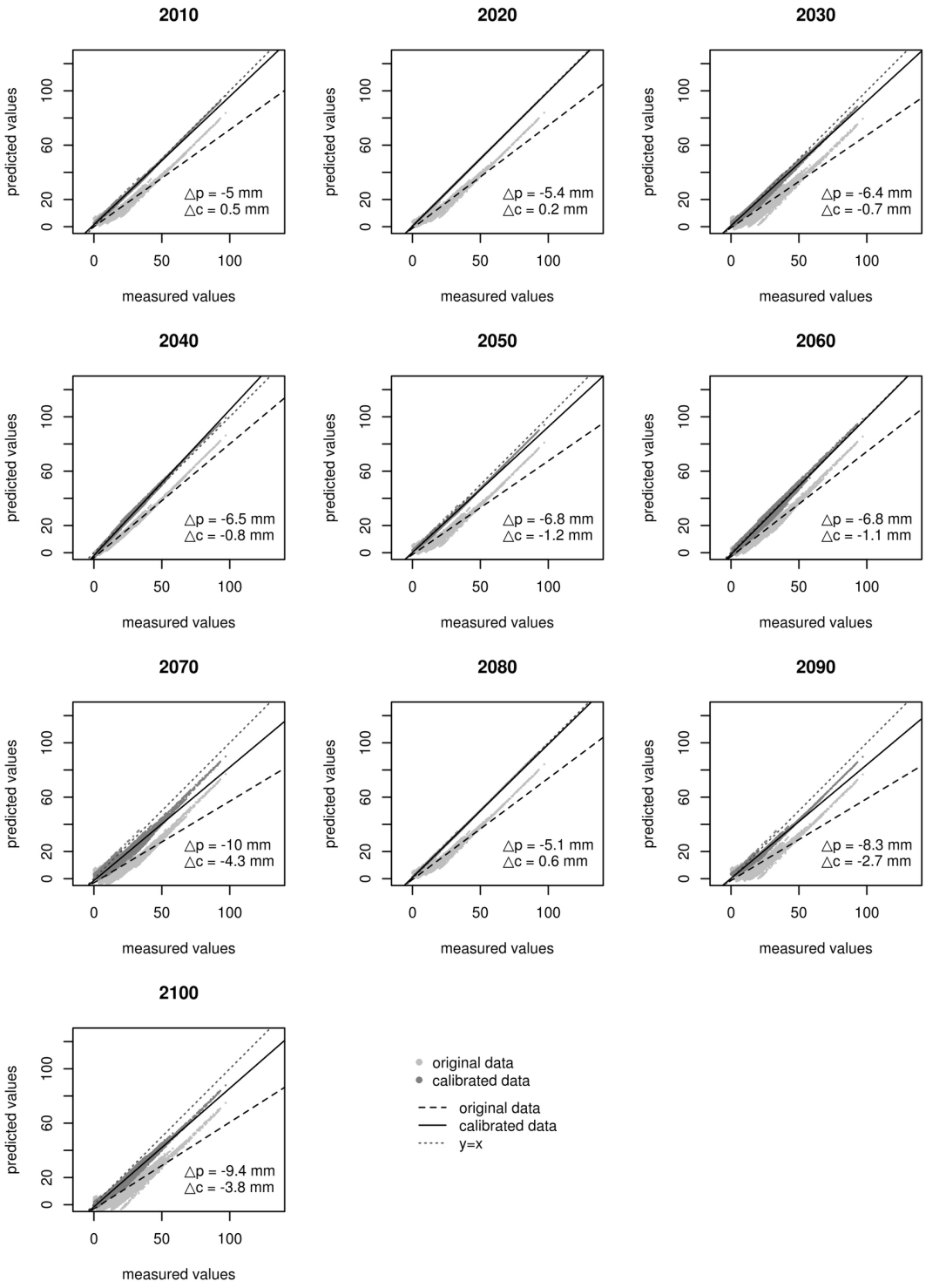
Continued from previous page.

Minimum annual precipitation (mm)
HadCM3 B1



Continued from previous page.

Minimum annual precipitation (mm)
HadCM3 B2



Continued from previous page.

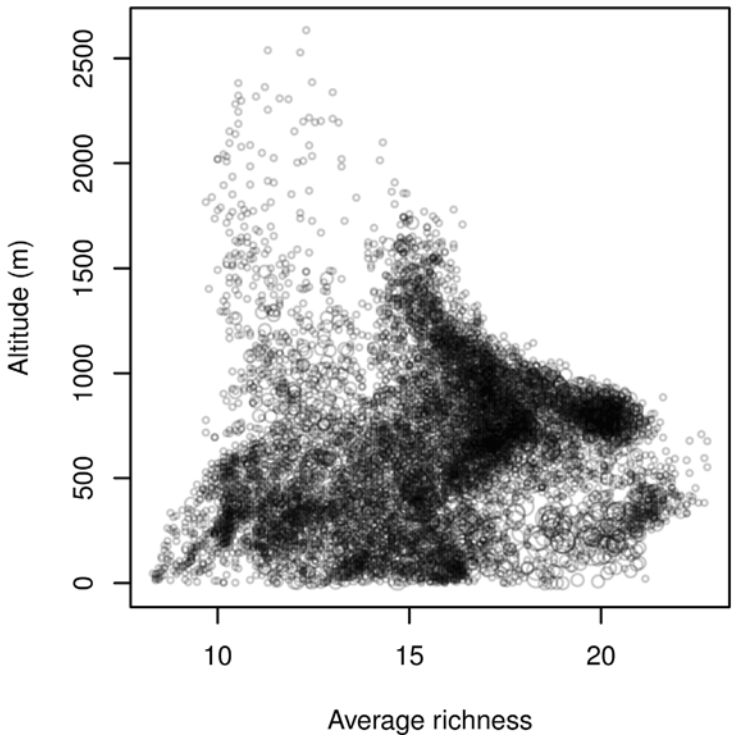


Fig. D.5 – Relation between average species richness and altitude for past model results. Circle size is proportional to the variance of species richness throughout time (see fig. 5.3).

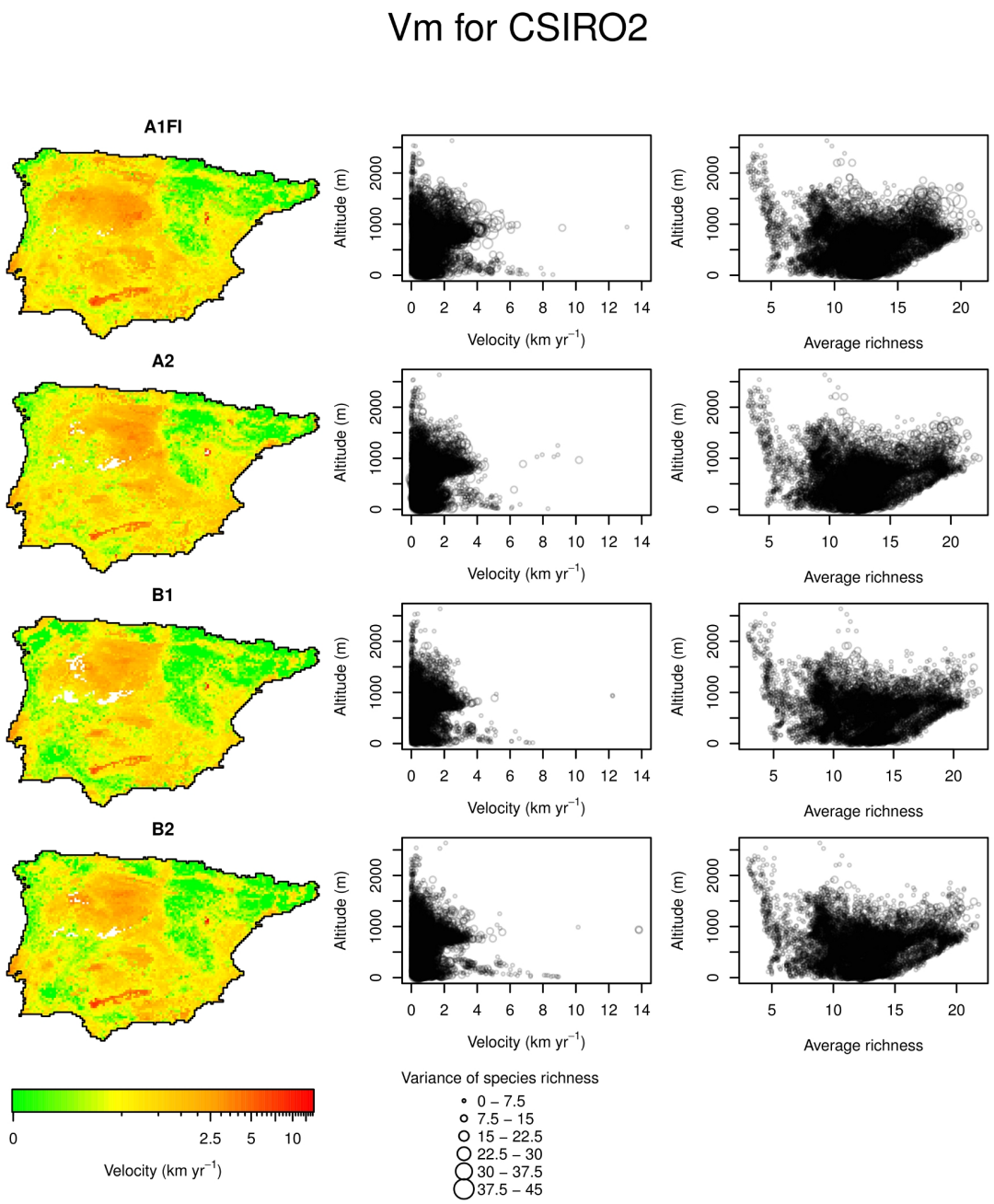
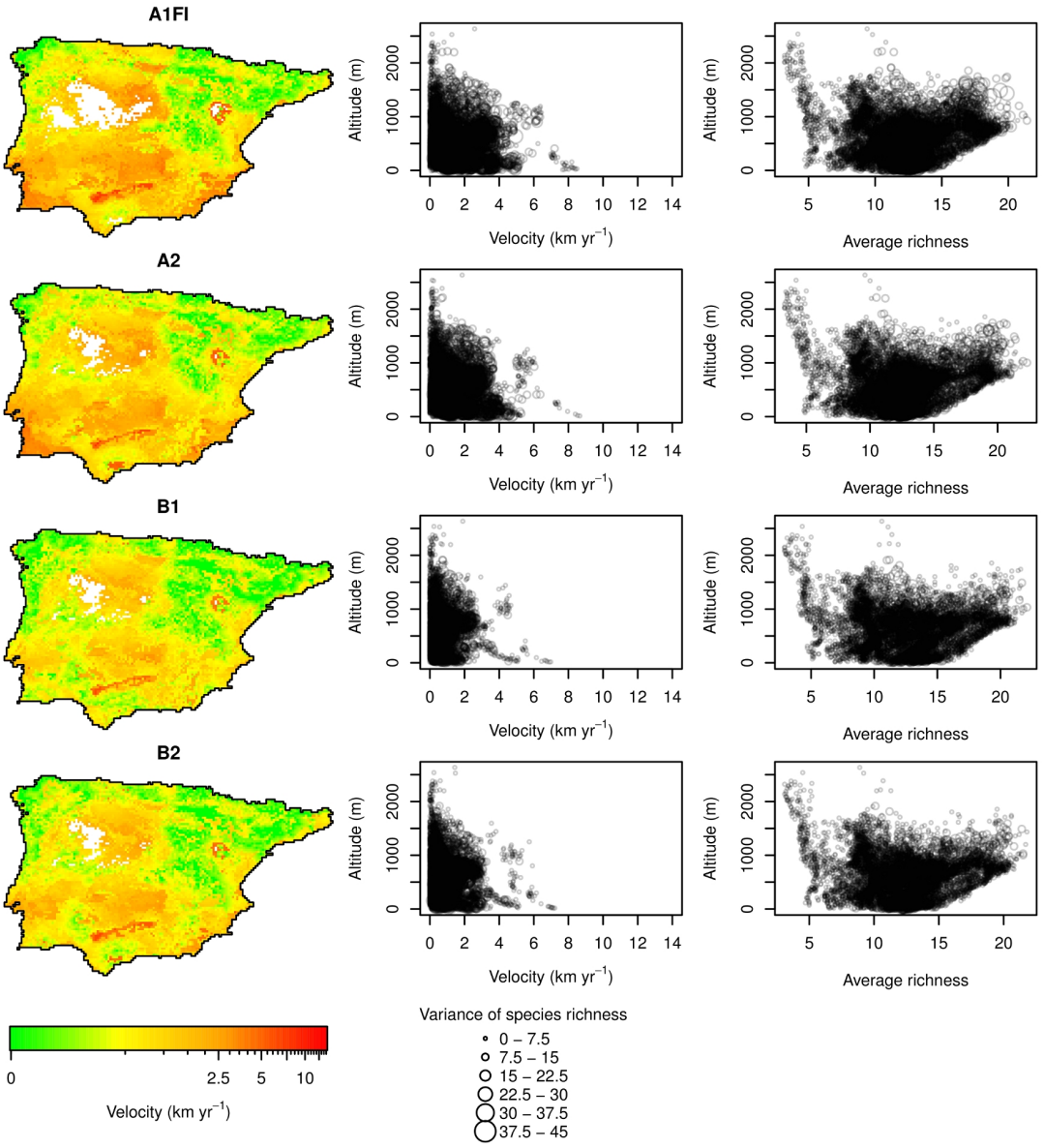


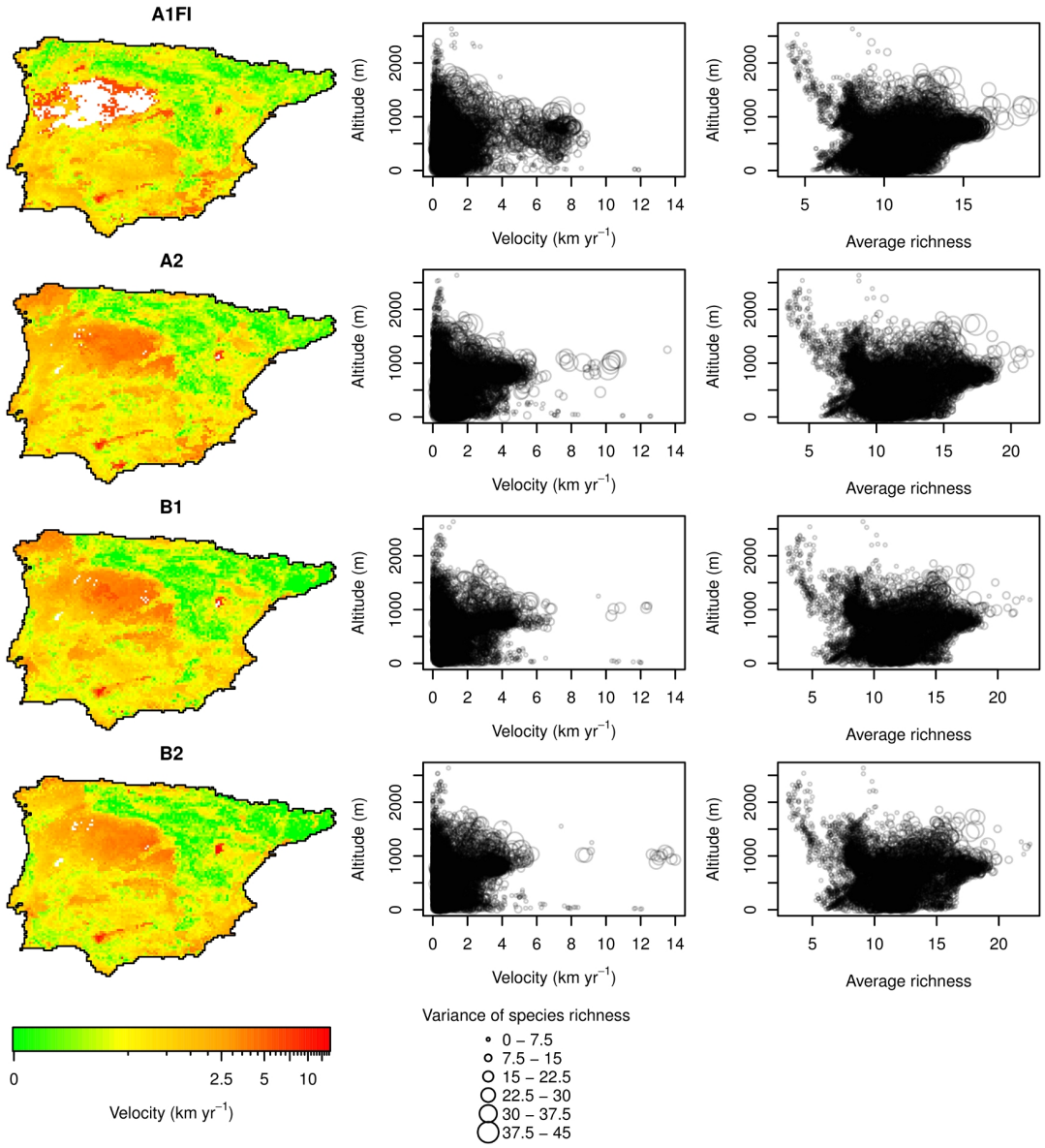
Fig. D.6 – Velocities of change in species composition for all combinations models and emission scenarios. Velocity of pixel maintenance (Vm) and velocity of richness shift (Vs) are shown. White areas in the map correspond to infinite velocity values. Relations between altitude and velocity and also between average species richness and altitude are shown for each model/emission scenario. Circle size is proportional to the variance by location of species richness through time.

Vs for CSIRO2



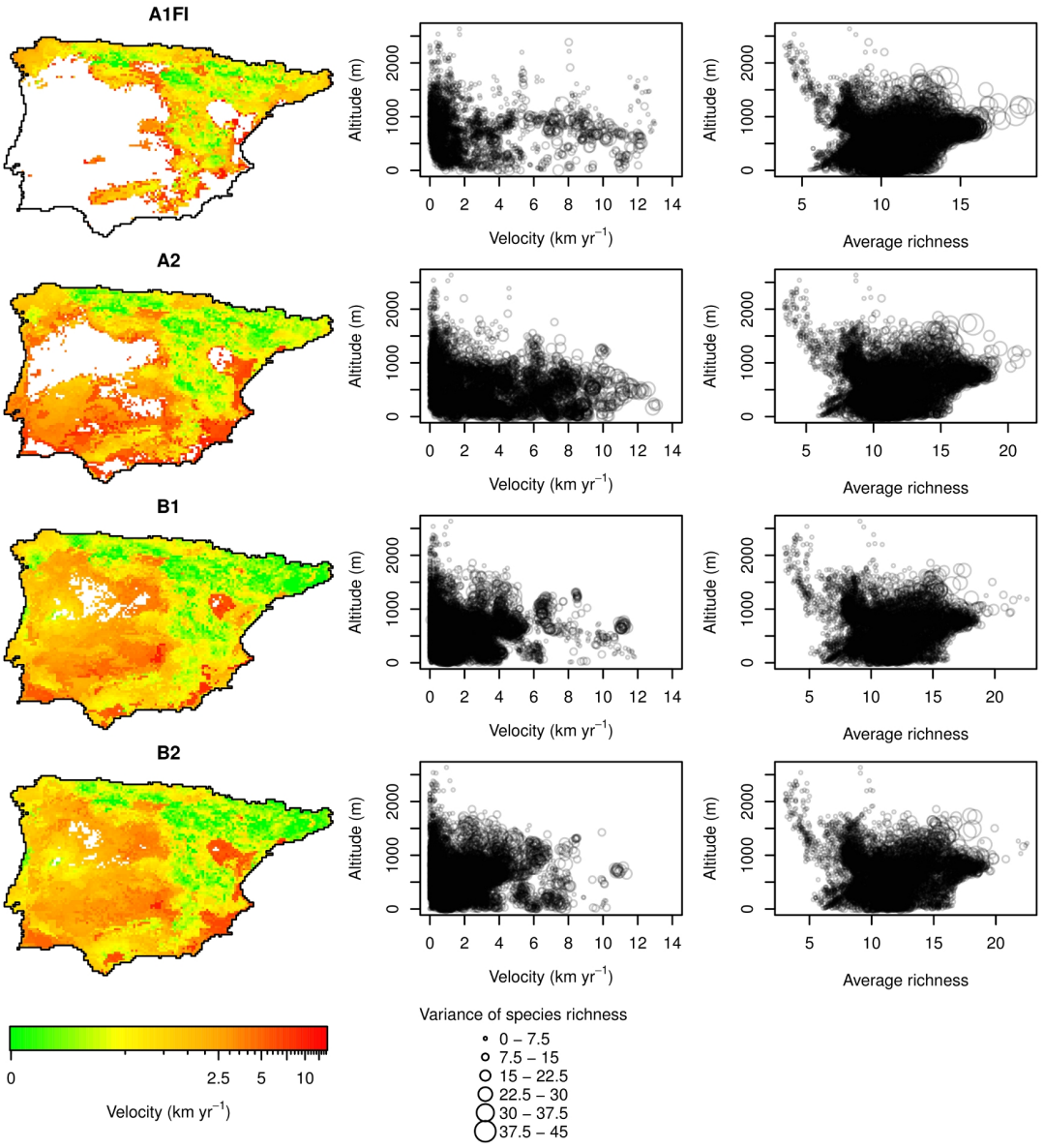
Continued from previous page.

Vm for HadCM3



Continued from previous page.

Vs for HadCM3



Continued from previous page.

Appendix E

Supplementary material for section 5.2

Table E.1 – Sampled individuals with morfological identifications, cluster membership probabilities and mtDNA results. Rows marked with an asterisk are reference individuals used in Hardy-Weinberg equilibrium test.

Number	morphology	pop1	pop2	pop3	NlaIII	XspI	HW
1	*Vipera aspis*	0.03	0.962	0.008	VA	VA	
2	*Vipera aspis*	0.004	0.991	0.006	VA	VA	
3	*Vipera aspis*	0.005	0.986	0.009	VA	VA	
4	*Vipera aspis*	0.005	0.991	0.005	VA	VA	
5	*Vipera aspis*	0.007	0.983	0.01	VL	VL	
6	*Vipera aspis*	0.007	0.986	0.008	VA	VA	
7	*Vipera aspis*	0.015	0.979	0.006	VA	VA	
8	*Vipera aspis*	0.007	0.989	0.005	VA	VA	
9	*Vipera aspis*	0.212	0.78	0.007	VL	VL	
10	*Vipera aspis*	0.004	0.992	0.004	VA	VA	
11	*Vipera aspis*	0.004	0.993	0.003	VL	VL	
12	*Vipera aspis*	0.085	0.901	0.014	VL	VL	
13	*Vipera aspis*	0.015	0.939	0.046	VA	VA	
14	*Vipera aspis*	0.004	0.978	0.017	VA	VA	
15	*Vipera aspis*	0.006	0.982	0.012	VA	VA	
16	*Vipera aspis*	0.031	0.961	0.008	VA	VA	
17	*Vipera aspis*	0.058	0.936	0.007	VA	VA	
18	*Vipera aspis*	0.006	0.985	0.008	VA	VA	
19		0.009	0.985	0.006	VA	VA	
20		0.009	0.983	0.008	VA	VA	
21	*Vipera latastei*	0.984	0.008	0.008	VL	VL	
22	*Vipera latastei*	0.986	0.009	0.005	VL	VL	
23	*Vipera latastei*	0.482	0.339	0.179	VL	VL	
24	*Vipera latastei*	0.085	0.899	0.016	VL	VL	
25	*Vipera latastei*	0.991	0.004	0.005	VL	VL	
26	*Vipera latastei*	0.988	0.005	0.007	VL	VL	
27	*Vipera latastei*	0.992	0.004	0.004	VL	VL	
28	*Vipera latastei*	0.991	0.005	0.004	VL	VL	
29	*Vipera latastei*	0.969	0.02	0.011	VL	VL	
30	*Vipera latastei*	0.977	0.007	0.017	VL	VL	
31	*Vipera latastei*	0.982	0.013	0.005	VL	VL	
32	*Vipera latastei*	0.976	0.016	0.008	VL	VL	

Continued on next page

Table E.1 – *Continued from previous page*

Number	morphology	pop1	pop2	pop3	NlaIII	XspI	HW
33	*Vipera latastei*	0.99	0.005	0.005	VL	VL	*
34	*Vipera latastei*	0.979	0.015	0.006	VL	VL	
35	*Vipera latastei*	0.992	0.004	0.004	VL	VL	
36	*Vipera latastei*	0.987	0.006	0.007	VL	VL	
37	*Vipera latastei*	0.973	0.021	0.006	VL	VL	
38	*Vipera latastei*	0.991	0.005	0.004	VL	VL	*
39	*Vipera latastei*				VL	VL	
40	*Vipera latastei*	0.989	0.007	0.004			
41	*Vipera latastei*	0.989	0.005	0.005	VL	VL	
42	*Vipera latastei*	0.988	0.006	0.006	VL	VL	
43	*Vipera latastei*	0.989	0.004	0.007	VL	VL	
44	*Vipera latastei*	0.985	0.008	0.007	VL	VL	
45	*Vipera latastei*	0.978	0.018	0.004	VL	VL	
46	*Vipera latastei*	0.987	0.01	0.003	VA	VA	
47	*Vipera latastei*	0.964	0.027	0.009	VL	VL	
48		0.991	0.004	0.006	VL	VL	
49	*Vipera latastei*	0.955	0.041	0.005	VL	VL	*
50	*Vipera latastei*	0.991	0.005	0.004	VL	VL	
51	*Vipera latastei*	0.991	0.004	0.005	VL	VL	*
52	*Vipera latastei*	0.989	0.005	0.006	VL	VL	*
53	*Vipera latastei*	0.968	0.007	0.025	VL	VL	
54	*Vipera latastei*	0.98	0.015	0.005	VL	VL	
55	*Vipera latastei*	0.975	0.008	0.018	VL	VL	
56	*Vipera latastei*	0.988	0.005	0.007	VL	VL	*
57	*Vipera latastei*	0.748	0.01	0.242	VL	VL	*
58		0.008	0.005	0.987	VS	VS	*
59	*Vipera* sp.	0.988	0.006	0.006	VL	VL	
60	*Vipera* sp.	0.088	0.903	0.009	VL	VL	
61	*Vipera* sp.	0.865	0.111	0.023	VL	VL	
62	*Vipera* sp.	0.429	0.566	0.005	VL	VL	
63	*Vipera* sp.	0.928	0.027	0.045	VS	VS	
64	*Vipera* sp.	0.017	0.978	0.005	VA	VA	
65		0.078	0.007	0.915	VL	VL	
66	*Vipera* sp.	0.033	0.962	0.005	VL	VL	
67	*Vipera aspis*	0.456	0.535	0.009	VL	VL	
68	*Vipera aspis*	0.829	0.165	0.006	VL	VL	
69	*Vipera aspis*	0.6	0.392	0.008	VL	VL	
70	*Vipera aspis*	0.005	0.992	0.003	VA	VA	
71	*Vipera aspis*	0.287	0.706	0.007	VA	VA	
72	*Vipera aspis*	0.007	0.976	0.017	VA	VA	*
73	*Vipera aspis*	0.008	0.97	0.021	VA	VA	
74	*Vipera aspis*	0.052	0.911	0.037	VS	VS	
75	*Vipera aspis*	0.013	0.983	0.004	VA	VA	
76	*Vipera aspis*	0.006	0.99	0.004	VA	VA	*
77	*Vipera aspis*	0.014	0.98	0.006	VA	VA	
78	*Vipera aspis*	0.006	0.988	0.006	VA	VA	
79	*Vipera aspis*	0.008	0.988	0.005	VA	VA	
80	*Vipera aspis*	0.01	0.985	0.005	VA	VA	
81	*Vipera aspis*	0.006	0.987	0.006	VA	VA	

Continued on next page

Table E.1 – *Continued from previous page*

Number	morphology	pop1	pop2	pop3	NlaIII	XspI	HW
82	*Vipera aspis*	0.009	0.983	0.008	VA	VA	
83	*Vipera aspis*	0.041	0.95	0.01	VL	VL	
84	*Vipera aspis*	0.017	0.969	0.014	VA	VA	
85	*Vipera aspis*	0.012	0.981	0.007	VA	VA	
86	*Vipera aspis*	0.006	0.972	0.022	VA	VA	
87	*Vipera aspis*	0.036	0.953	0.011	VA	VA	
88	*Vipera aspis*	0.004	0.991	0.004	VA	VA	
89	*Vipera aspis*	0.005	0.99	0.005	VA	VA	
90	*Vipera aspis*	0.035	0.96	0.005	VA	VA	
91	*Vipera aspis*	0.032	0.959	0.008	VA	VA	
92	*Vipera aspis*	0.007	0.987	0.005	VA	VA	
93	*Vipera aspis*	0.014	0.981	0.005	VA	VA	
94	*Vipera aspis*	0.012	0.981	0.007	VA	VA	
95	*Vipera aspis*	0.006	0.99	0.004	VA	VA	
96	*Vipera aspis*	0.013	0.98	0.007			
97	*Vipera aspis*				VA	VA	
98	*Vipera aspis*	0.056	0.935	0.009	VA	VA	*
99	*Vipera aspis*	0.007	0.988	0.005	VA	VA	*
100	*Vipera aspis*	0.03	0.959	0.011			*
101	*Vipera aspis*	0.005	0.99	0.005			*
102	*Vipera aspis*	0.003	0.993	0.003	VA	VA	
103	*Vipera aspis*	0.121	0.873	0.006	VA	VA	
104	*Vipera aspis*	0.005	0.989	0.006	VA	VA	
105	*Vipera aspis*	0.021	0.975	0.004	VA	VA	
106	*Vipera aspis*	0.004	0.949	0.047	VA	VA	
107	*Vipera aspis*	0.006	0.99	0.005	VA	VA	
108	*Vipera aspis*	0.006	0.988	0.006	VA	VA	
109	*Vipera aspis*	0.004	0.992	0.004	VA	VA	
110	*Vipera aspis*	0.874	0.122	0.004	VA	VA	
111	*Vipera aspis*	0.004	0.992	0.004	VA	VA	
112	*Vipera aspis*	0.007	0.99	0.003	VA	VA	
113	*Vipera aspis*	0.004	0.989	0.007	VA	VA	*
114	*Vipera aspis*	0.088	0.908	0.004	VA	VA	
115	*Vipera latastei*					VL	
116	*Vipera latastei*	0.981	0.01	0.009	VL	VL	
117	*Vipera latastei*	0.992	0.004	0.004	VL	VL	
118	*Vipera latastei*	0.982	0.007	0.011	VL	VL	
119	*Vipera latastei*				VL	VL	
120	*Vipera latastei*	0.988	0.008	0.004	VL	VL	
121	*Vipera latastei*	0.983	0.009	0.008	VL	VL	
122	*Vipera latastei*	0.982	0.014	0.004			
123	*Vipera latastei*	0.138	0.853	0.01	VL	VL	
124	*Vipera latastei*	0.648	0.345	0.008	VL	VL	
125	*Vipera latastei*	0.99	0.006	0.005	VL	VL	
126	*Vipera latastei*	0.963	0.015	0.022	VL	VL	
127	*Vipera latastei*	0.987	0.008	0.006	VL	VL	
128	*Vipera latastei*	0.991	0.005	0.004	VL	VL	
129	*Vipera latastei*	0.946	0.037	0.017	VA	VA	
130	*Vipera latastei*	0.989	0.005	0.006	VL	VL	

Continued on next page

Table E.1 – *Continued from previous page*

Number	morphology	pop1	pop2	pop3	NlaIII	XspI	HW
131	*Vipera latastei*	0.964	0.024	0.012	VL	VL	
132	*Vipera latastei*	0.987	0.008	0.005	VL	VL	*
133	*Vipera latastei*	0.986	0.01	0.004	VL	VL	
134	*Vipera latastei*	0.908	0.047	0.044	VL	VL	
135	*Vipera latastei*	0.054	0.939	0.007	VL	VL	
136	*Vipera latastei*	0.989	0.005	0.006	VL	VL	*
137	*Vipera latastei*	0.961	0.034	0.004	VL	VL	
138	*Vipera latastei*	0.992	0.004	0.003	VL	VL	
139	*Vipera latastei*	0.99	0.006	0.003	VL	VL	
140	*Vipera latastei*	0.992	0.004	0.004	VL	VL	*
141	*Vipera latastei*	0.891	0.103	0.006	VL	VL	
142	*Vipera latastei*	0.961	0.028	0.011	VA	VA	
143	*Vipera latastei*	0.961	0.031	0.008	VL	VL	
144	*Vipera latastei*	0.855	0.139	0.005	VL	VL	
145	*Vipera latastei*	0.991	0.005	0.004	VL	VL	*
146	*Vipera latastei*				VL	VL	
147	*Vipera seoanei*	0.005	0.005	0.99	VS	VS	*
148	*Vipera seoanei*	0.004	0.006	0.99	VS	VS	*
149	*Vipera seoanei*	0.005	0.005	0.99	VS	VS	*
150	*Vipera seoanei*	0.007	0.005	0.988	VS	VS	*
151	*Vipera seoanei*	0.006	0.004	0.99	VS	VS	*
152	*Vipera seoanei*	0.008	0.005	0.987	VS	VS	*
153	*Vipera seoanei*	0.004	0.009	0.986	VS	VS	*
154	*Vipera seoanei*	0.006	0.007	0.987	VS	VS	*
155	*Vipera seoanei*	0.023	0.017	0.959	VA	VA	*
156	*Vipera seoanei*	0.006	0.005	0.989	VS	VS	*
157	*Vipera seoanei*	0.008	0.005	0.987	VS	VS	*
158	*Vipera* sp.	0.205	0.787	0.008	VA	VA	
159	*Vipera* sp.	0.48	0.509	0.011	VA	VA	
160	*Vipera* sp.	0.887	0.108	0.005	VA	VA	
161	*Vipera* sp.	0.911	0.08	0.009	VL	VL	
162	*Vipera* sp.	0.526	0.314	0.16	VL	VL	
163	*Vipera* sp.	0.08	0.913	0.007	VA	VA	
164	*Vipera* sp.	0.158	0.823	0.019	VA	VA	
165	*Vipera* sp.	0.004	0.989	0.007	VA	VA	
166	*Vipera* sp.	0.672	0.323	0.006	VL	VL	
167	*Vipera* sp.	0.968	0.025	0.007	VL	VL	
168	*Vipera* sp.	0.003	0.993	0.004	VA	VA	
169	*Vipera* sp.	0.476	0.518	0.005	VA	VA	
170	*Vipera* sp.	0.029	0.959	0.012	VL	VL	
171	*Vipera* sp.	0.164	0.827	0.009	VL	VL	
172	*Vipera* sp.	0.038	0.95	0.012	VA	VA	
173	*Vipera* sp.	0.979	0.014	0.007	VL	VL	
174	*Vipera* sp.	0.976	0.017	0.007	VL	VL	
175	*Vipera* sp.	0.09	0.902	0.007	VL	VL	
176	*Vipera* sp.	0.257	0.722	0.021	VL	VL	
177	*Vipera* sp.	0.143	0.843	0.014	VA	VA	
178	*Vipera* sp.	0.24	0.755	0.006	VA	VA	
179	*Vipera aspis*	0.006	0.981	0.013	VA	VA	

Continued on next page

Table E.1 – *Continued from previous page*

Number	morphology	pop1	pop2	pop3	NlaIII	XspI	HW
180	*Vipera aspis*	0.054	0.941	0.005	VA	VA	
181	*Vipera aspis*	0.013	0.983	0.004	VA	VA	
182	*Vipera aspis*	0.011	0.983	0.006	VL	VL	
183	*Vipera aspis*	0.007	0.99	0.003	VA	VA	
184	*Vipera aspis*	0.003	0.975	0.021	VA	VA	
185	*Vipera aspis*	0.07	0.913	0.017	VL	VL	*
186	*Vipera aspis*	0.006	0.982	0.012	VA	VA	*
187	*Vipera aspis*	0.015	0.982	0.004	VA	VA	
188	*Vipera aspis*	0.006	0.99	0.004	VA	VA	
189	*Vipera aspis*	0.513	0.481	0.006	VL	VL	
190	*Vipera aspis*	0.016	0.98	0.003	VA	VA	
191	*Vipera aspis*	0.022	0.97	0.008	VA	VA	
192	*Vipera aspis*	0.005	0.986	0.009	VA	VA	*
193	*Vipera aspis*	0.007	0.988	0.005	VA	VA	*
194	*Vipera aspis*	0.005	0.99	0.005	VA	VA	
195	*Vipera aspis*	0.007	0.987	0.005	VA	VA	
196	*Vipera aspis*	0.004	0.991	0.005			
197	*Vipera latastei*	0.976	0.012	0.011	VL	VL	*
198	*Vipera latastei*	0.143	0.851	0.005	VL	VL	
199	*Vipera latastei*	0.986	0.009	0.005	VL	VL	
200	*Vipera latastei*	0.023	0.972	0.004	VA	VA	
201	*Vipera latastei*				VL	VL	
202	*Vipera latastei*	0.984	0.01	0.006	VL	VL	
203	*Vipera latastei*	0.987	0.009	0.003	VL	VL	
204	*Vipera latastei*	0.862	0.132	0.006	VL	VL	
205	*Vipera latastei*	0.985	0.007	0.007	VL	VL	
206	*Vipera latastei*	0.99	0.005	0.005			
207	*Vipera* sp.	0.429	0.567	0.005	VL	VL	
208	*Vipera* sp.	0.041	0.955	0.004	VA	VA	
209	*Vipera* sp.	0.896	0.096	0.009	VA	VA	
210	*Vipera* sp.	0.979	0.014	0.007	VA	VA	
211	*Vipera* sp.	0.143	0.851	0.006	VL	VL	
212	*Vipera* sp.	0.015	0.981	0.003	VL	VL	
213	*Vipera* sp.	0.01	0.987	0.003	VA	VA	
214	*Vipera aspis*	0.007	0.989	0.003	VA	VA	*
215	*Vipera aspis*	0.007	0.983	0.01	VA	VA	
216	*Vipera aspis*				VA	VA	
217	*Vipera aspis*	0.005	0.991	0.003	VA	VA	
218	*Vipera* sp.				VL	VL	

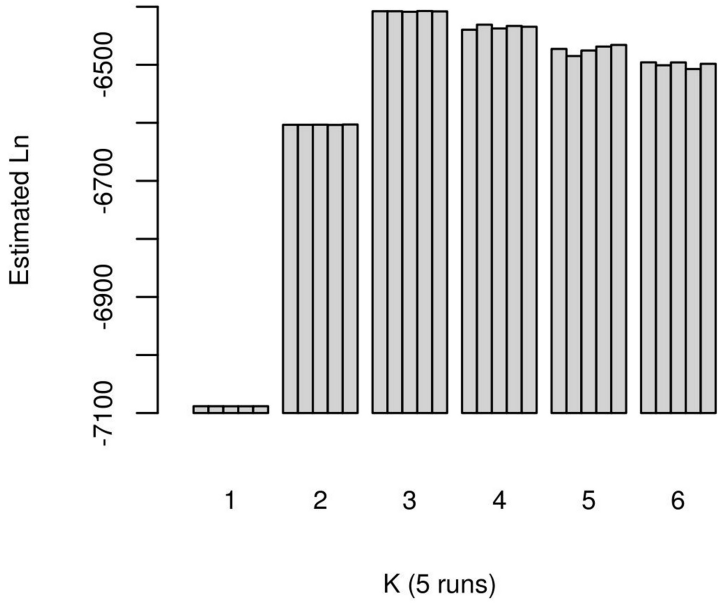


Fig. E.1 – Estimated likelihood values for five runs with K set to values between one and five.

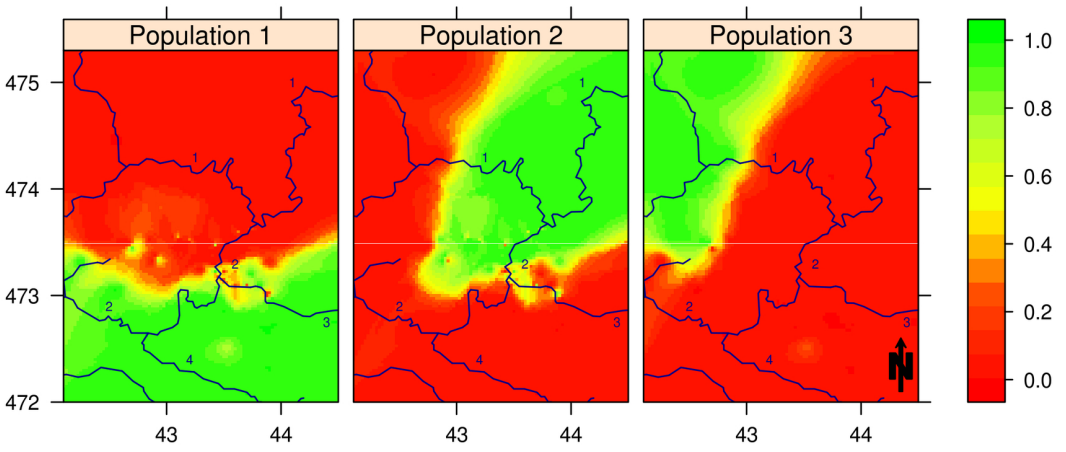


Fig. E.2 – Interpolation of cluster membership probabilities (CMPs) with inverse distance weighting algorithm (IDW). The populations 1, 2 and 3 coincide with samples of *V. latastei*, *V. aspis* and *V. seoanei*. More details in table 2 in the text.

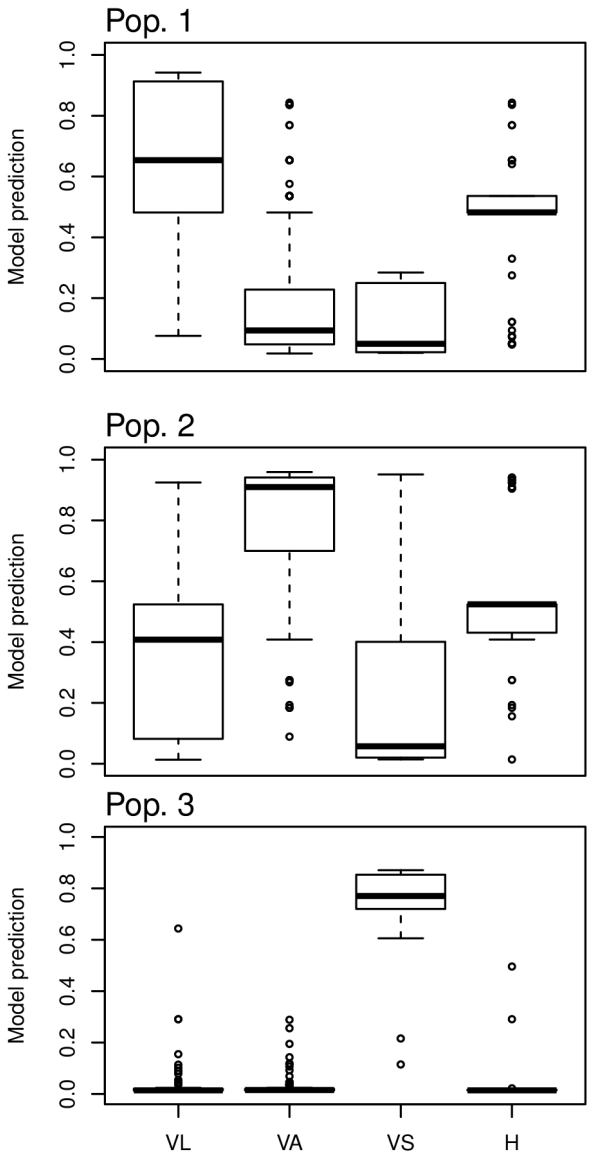


Fig. E.3 – Ecological niche-based model accuracy when predicting a set of morphological classified individuals (n=477). Most individuals of population 1 were classified as *V. latastei*. Population 2 predicted high values for most individuals classified as *V. aspis*, and population 3 for *V. seoanei*. Those individuals classified as hybrids have average values for both population 1 and 2. This fact, along with the high degree of overlap of *V. latastei* and *V. aspis* both in populations 1 and 2, suggests a shared space between these species and frequent hybridization.

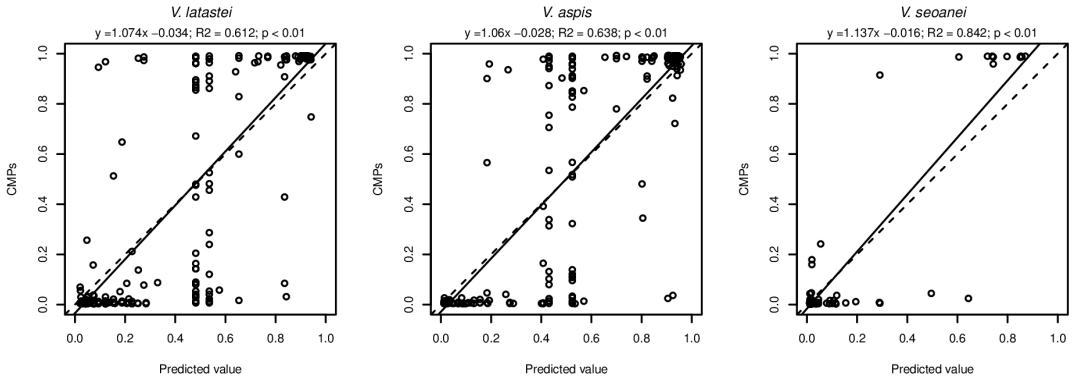


Fig. E.4 – Linear regressions between cluster membership probabilities (CMPs) and values predicted by the ecological niche-based models. Analyses of variance (ANOVA) indicate that the model is a significant predictor of the CMPs. Extreme values are well defined in each model. The middle values have a less predicted accuracy due to the space share between parental in the hybrid zone, i. e. the same pixel in the hybrid area holds many different individuals.