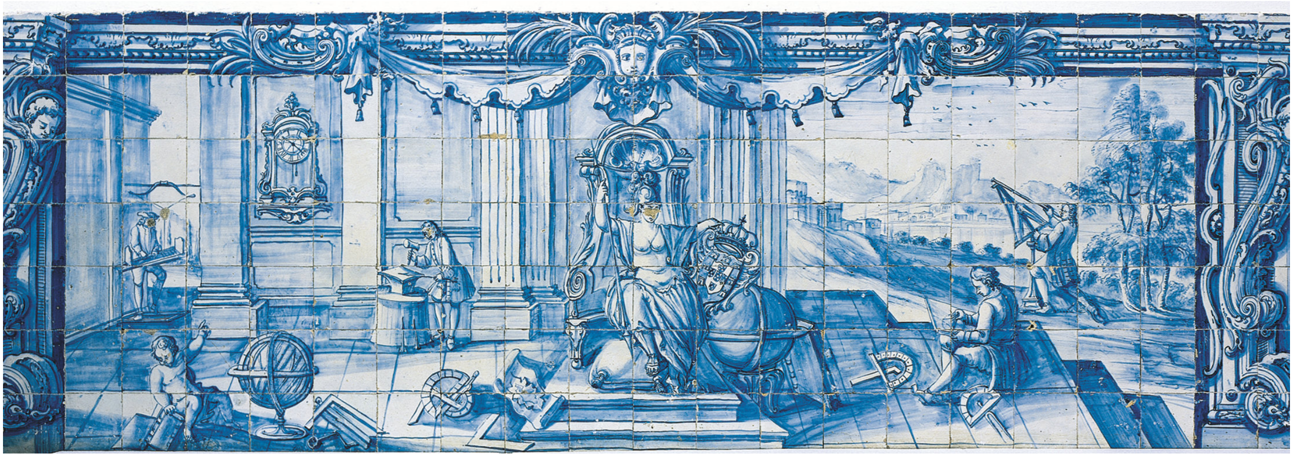
PREDICTING THE DISTRIBUTION OF INVASIVE DECAPODS ACROSS SPACE AND TIME

César Dinis Santos Capinha

A thesis presented to the University of Évora
in partial fulfilment of requirements for the degree of Doctor of Philosophy in
Environmental Sciences

SUPERVISOR: *Dr. Pedro Manuel Anastácio*
CO – SUPERVISOR : *Dr. José António Tenedório*

ÉVORA, DECEMBER, 2012



INSTITUTO DE INVESTIGAÇÃO E FORMAÇÃO AVANÇADA

PH.D THESIS IN ENVIRONMENTAL SCIENCES

Title: Predicting the distribution of invasive decapods across space and time

By: César Dinis Santos Capinha, University of Évora and IMAR – Marine and Environmental Research Centre

Supervisor: Prof. Dr. Pedro Manuel Anastácio, Department of Landscape, Environment and Planning, University of Évora and IMAR – Marine and Environmental Research Centre

Co-supervisor: Prof. Dr. José António Tenedório, Research Centre for Geography and Regional Planning, New University of Lisbon

University of Évora, Évora, 2012

To my grandparents who left and to my son who arrived

‘Para os meus avós que partiram e o meu filho que chegou’

Financial Support

To perform the research presented in this thesis I received financial support from:

- A PhD grant from the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BD/41129/2007)
- The Research project DID (Dispersal of Invasive Decapoda) (PTDC/BIA-BEC/105182/2008) co-financed by the FCT, and the European Union programs COMPETE, QREN and FEDER.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE

GOVERNO DA
REPÚBLICA PORTUGUESA

UNIÃO EUROPEIA - FEDER

I declare no conflict of interest concerning the research conducted within the scope of this thesis.

Aknowledgments

When I started to compile all the work that would eventually shape this thesis I thought this would be the easiest part to write. However, I must have rewritten the opening sentence more than 10 times and also had writers block many minutes during several days in this section. After all this rewriting I now realize that this difficulty comes from the inability to properly acknowledge all the fantastic persons that made these last years so special. Here’s a (very) small and incomplete recognition to them:

- Fist, I have to express my enormous gratitude to Pedro Anastácio, my supervisor. Thank you for reading my e-mail that day; thank you for replying. Thank you for your guidance of excellence and all the freedom that was attached to it.
- A big thank you to Professor Tenedório. Your availability is inspiring.
- Thank you to Lluís, Miguelito, Núria, Vincent, Olatz, Francesc, Elena, David, Jaime, Laura, Evelyn, Iria and all the great people at the Centre Tecnològic. You were absolutely great!
- Brian, I still owe you. Fortran boy and C++ guy, the best of lucks! Erin, David and Ayaz thanks for the warm company in those harsh days.
- To Mandy, Candida, Octavio and so many other special people for turning boring meetings into moments of joy.
- Mónica, this is not the place to say all I should say. You were the Sun in my often shady sky.
- Rodrigo, thank you for making me realize that life is in fact beautiful (and also that we don’t need to sleep that many hours per night).
- Rui and Philip: kayaking; climbing; summer; soccer; red wine; photography; catching crayfish, lets go?
- Mom, dad, sis its almost done. How about a big family dinner?

This page intentionally left blank

Abstract

This dissertation aims at evaluating the reliability of correlative predictions of potential distribution for invasive species and concurrently increase the knowledge about the current and future invasive potential of four problematic decapods: Chinese mitten crab; Red swamp crayfish; Signal crayfish and the Yabby. The reliability of diverse modelling options –including distribution data, explanatory factors and statistical models– was evaluated using traditional and newly developed techniques. Predictions showed to be highly reliant on the modelling options used, however, a methodical selection of calibration data along with a combined analysis of discrimination ability, degree of extrapolation and level of inter-model agreement showed to be a valuable framework for obtaining final predictions. These revealed a strong association between the invader’s current distribution and climate, particularly thermal conditions. Wide uninvaded regions vulnerable to establishment were found across the globe but, concerning the Iberian Peninsula, these are expected to contract as time progresses.

Previsão da distribuição de decápodes invasores no espaço e no tempo

Resumo

Esta tese pretende avaliar a fiabilidade de modelos correlativos de distribuição para espécies invasoras e simultaneamente ampliar o conhecimento acerca do potencial invasor actual e futuro de quatro decápodes problemáticos: Caranguejo-peludo-chinês; Lagostim-vermelho-da-Louisiana; Lagostim-sinal e Lagostim-Australiano. Opções de modelação distintas –incluindo dados de distribuição, factores explicativos e modelos estatísticos– foram avaliadas com recurso a novos métodos e técnicas tradicionais. Verificou-se um ampla dependência entre os resultados e as opções de modelação utilizadas, no entanto, a selecção metódica de dados de calibração conjuntamente com uma análise combinada de concordância entre modelos, capacidade discriminativa e grau de extrapolação revelou-se como um procedimento robusto para a obtenção de modelos finais. Estes indicam uma elevada associação entre a distribuição actual dos invasores e o clima, particularmente condições térmicas. Extensas áreas adequadas ainda por ocupar foram identificadas para múltiplas regiões do globo, na Península Ibérica em particular, estas aparentam contrair à medida que o tempo progride.

Note on the structure of the thesis

The original research developed within the scope of this thesis is here presented in the form of scientific articles submitted or published by peer-reviewed scientific journals (cf. point 1 of the Article 2 of the Service Order Nº 18/2006 of the University of Évora and clarification note of the Scientific Council of the Faculty of Sciences and Technology, 27 May, 2009). Accordingly, the chapters 2 to 5 –each corresponding to one of these articles– maintain the original structure of the information submitted or published by the scientific journals. The only information differing from the original articles refers to the numbering of the appendix information that was sequentially renumbered in accordance to the joint appendix section provided at the end of this thesis.

To date, the following scientific papers were originated from the research activities performed within the scope this thesis:

Capinha C, Anastácio P, Tenedório JA (2012) Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability. *Biological Invasions*. [DOI: 10.1007/s10530-012-0187-z]

Capinha C, Anastácio P (2011) Assessing the environmental requirements of invaders using ensembles of distribution models. *Diversity and Distributions* 17:13-24. [DOI: 10.1111/j.1472-4642.2010.00727.x]

Capinha C, Leung B, Anastácio P (2011) Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. *Ecography* 34:448-459. [DOI: 10.1111/j.1600-0587.2010.06369.x]

Capinha C, Brotons L, Anastácio P (2012) Geographic variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish. Journal of Biogeography. [DOI: 10.1111/jbi.12025]

Ferreira M, **Capinha C,** Banha F, Rabaça J, Anastácio P (in review) Is waterbird-mediated passive dispersal of crayfish (*Procambarus clarkii*) a viable mechanism for invasion?

TABLE OF CONTENTS

Page intentionally left blank

CHAPTER 1

General introduction

1.1 Biological invasions: an overview

Biological invasions are currently considered one of the most important causes of global change, causing strong impacts on biodiversity, economy and human health. In natural environments, invasive species compete, predate or hybridize with native species, alter community structure and ecosystem process, ultimately leading to the homogenization of biota or the extinction of species. A particularly famous example of these impacts is provided by the Nile perch (*Lates niloticus*; Figure 1). In the 1950s this predatory freshwater fish was intentionally introduced in Lake Vitoria, Africa, to boost the lakes' fish stocks which were becoming severely depleted due to overfishing. In the decades after its introduction, its densities grew massively leading to the extinction of nearly 200 endemic species of cichlid fishes (Craig 1992). Other notable examples include the extinction of nearly every native bird species on the Pacific island of Guam caused by the brown tree snake (Wiles et al. 2003) or the extinction of more than one hundred terrestrial gastropods precipitated by the introduction of the predatory Rosy wolf snail (*Euglandina rosea*) in a number of oceanic islands (Régnier et al. 2009).

Figure 1 - The Nile perch (*Lates niloticus*). Photographic credits: Biopix. Reproduced with permission under a Creative Commons License.

The economic costs of invasive species can also be particularly striking. Despite the lack of a consensus on how to quantify the figures (Pimentel et al. 2001; Davis 2009) it is estimated that these costs can reach many billions of Euros in the USA or in Europe (Davis 2009; Hulme et al. 2009). In many cases these costs correspond to direct damages such as those caused on other economically important species (e.g. cinnamon fungus; *Phytophthora cinnamomi*; Robin et al. 2001) or on man-made infrastructures and equipment (e.g. the Zebra mussel; *Dreissena polymorpha*; Aldridge et al. 2005). However, the economic costs of invasive species can also be indirect and thus difficult to quantify, requiring an understanding of the many ways in which ecosystems support economy. Conversely, some invasive species are also known to have originated significant economic incomes. The majority of these cases correspond to the commercial exploration of intentionally introduced species, such as the Nile perch.

Some of the most notable invertebrate invaders worldwide are also vectors of human diseases such as malaria, plague, typhus or yellow fever, causing the incidence of these infections in previously unsuspected areas (Lounibos 2011). Furthermore, there are also invasion events responsible for ecological or landscape changes which have negative implications on human safety. These can assume multiple forms such as the promotions of pathogen outbreaks (e.g. Vanderploeg et al. 2001), or an increase in the vulnerability to natural hazards such as fires (e.g. Berry et al. 2011) or landslides (e.g. Meyer and Medeiros 2011).

Despite the notorious impacts of biological invasions, the emergence of invasion ecology as a scientific discipline is rather recent. Before Charles Elton published the iconic book 'The Ecology of Invasions by Animals and Plants' (Elton 1958) , few scientific works had been devoted to the study of non-native species and even after its publication it was not until the mid 1980s that the field received considerable interest by researchers (Davis 2011). Currently, the study of biological invasions is far-reaching and inter-disciplinary and aims particularly on understanding the causes and consequences of species invasions (Davis 2009; Davis 2011). This recent surge of interest in biological invasions is partly associated with the increasing trends of international travel and trade in recent decades which led to the introduction of many new species (and their impacts) worldwide. In fact, recent studies show that in many areas of the world non-native species are being introduced at accelerating rates (Lockwood et al. 2007). Thus, it is expected that the number of future biological invasions will be largely greater than the one recoded in the present day (Essl et al. 2011). Given these patterns, considerable research effort in invasion ecology is currently devoted to identify the factors which enable a species to establish in new areas and to become invasive.

1.2 Predictive Invasion Ecology

One of the first predictive hypotheses in invasion ecology was the tens rule (Williamson and Fitter 1996). This rule postulated that ≈10% of imported species appear in the wild, from which ≈10% are able to establish viable populations and, from these, ≈10% will become a problem. Currently, it is widely acknowledged that these numbers can be context dependent (e.g. Gherardi 2007) and that multiple and often intricate relationships between distinct factors can dictate invasion success or failure (Gurevitch et al. 2011). These factors can be separated in three categories: (1) propagule pressure, (2) biotic factors and (3) abiotic factors, each potentially influenced by human activity (Catford et al. 2009).

Propagule pressure concerns the number of propagules introduced (i.e. organic structures with the capacity to give rise to new individuals), the number of introduction events carried and the viability of the propagules. Perhaps one of the most consensual relationships so far found on invasion ecology, is a positive relationship between this factor and invasion success (Lockwood et al. 2009). The concept of 'the more you introduce, the more you get' was found on invasive birds (Cassey et al. 2004), fish (Colautti 2005), plants (Von Holle and Simberloff 2005) among others. However, this relationship is contingent on the absence of negative biotic or abiotic factors and thus recent efforts to advance on the predictability of biological invasions have been largely focused on these two last factors.

One of the main questions asked by current invasion ecologists is if some species are more invasive than others, and if so, which of their biological attributes are associated with invasive behaviour. This possibility has received much attention in recent years and the current consensus is that very few species-level characteristics are significant predictors of invasiveness and that these are often taxa-specific (Hayes and Barry 2008 and references therein). The reason for this idiosyncrasy has mostly been attributed to the distinct characteristics of the receiving areas, among which the receiving community. In fact, the variety of ways by which biotic interaction can interfere with the invasion process as led to the development of an array of invasion hypotheses (see Catford et al. 2009). For instance, the ‘enemy release hypothesis’ (Keane and Crawley 2002) predicts that the invasibility of a community by a certain species will be higher if the natural enemies of that species (e.g. predators, competitors or parasites) are absent from the new area. One supporting example was provided by DeWalt et al. (2004) for the neotropical shrub *Clidemia hirta.* This plant does not occur in forests in its native range but is a vigorous invader of tropical forests in its introduced range. DeWalt et al. (2004) found that after spraying against fungus and insects the survival of populations of this species in native understory areas increased dramatically. However, similarly to what happens for many other hypotheses, a number of contradictory examples can be found. For example, Pearson et al. (2000) found that the larvae of two introduced gall flies (*Urophora affinis* and *U*. *quadrifasciata*), became a highly favoured meal of a native deer mice (*Peromyscus maniculatus*) in Montana (Figure 2), United States, a process Colautti et al. (2004) named as the ‘enemy inversion hypothesis’. Similarly, many other contradicting hypotheses have also shown to be valid, depending on the biological group or area targeted (Davis 2009).

Equally important for predicting invasions is the ability to accurately identify the suitability of the abiotic conditions. Regardless of its biological attributes, a species will only be able to maintain populations in the long term in the areas that meet its physiological requirements (i.e. its fundamental niche; Hutchinson 1957). Temperature, humidity, salinity, or oxygen levels are all examples of abiotic factors that either alone or in interaction with other factors (either biotic or abiotic) can dictate the success of a new population. Given that these factors vary across the geographical space, knowing a species fundamental niche would allow knowing where it can invade. It is thus unsurprising that most research aiming to predict the influence of abiotic factors on invasions is devoted to quantifying the invaders fundamental niches. Overall, two approaches have been applied for this purpose: mechanistic models (also known as process-based models) and niche based models (also known as species distribution models, ecological niche models or habitat suitability models).

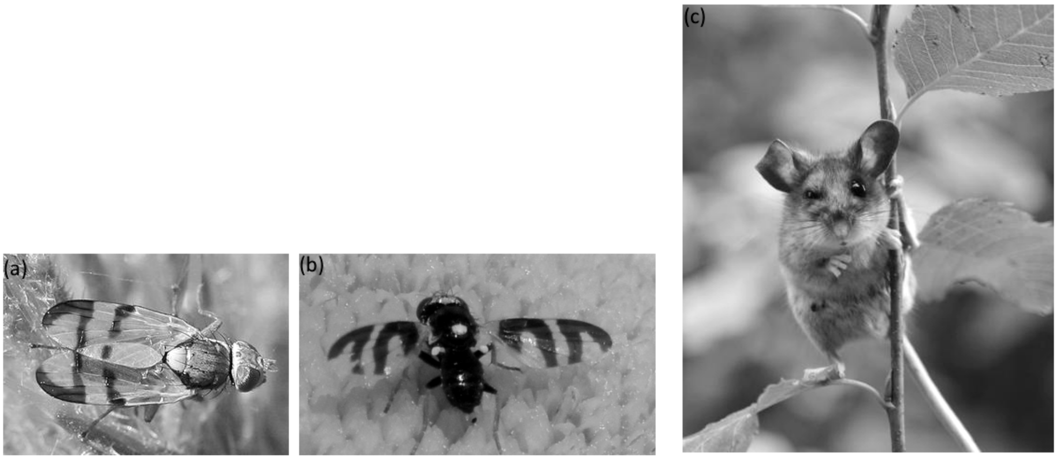


Figure 2 – The gall flies (a) *Urophora affinis* and (b) *Urophora quadrifasciata* and the (c) deer mice (*Peromyscus maniculatus*). Photographic credits of left image: entomart, middle image: Didier-35 and right image: Phil Myers. All images reproduced with permission under a Creative Commons License

Mechanistic models are based on direct assessments of the species physiological tolerances. For example, Kearney et al. (2008) used these models for predicting the potential distribution (i.e. the geographic expression of suitable abiotic conditions) of the invasive Cane toad (*Bufo marinus*) in Australia. To perform this task they initially evaluated the fitness of the specie's movements, survival and reproduction under distinct abiotic conditions using laboratory experiments, field measurements and previously published parameters. Next, they used these parameters to calibrate biophysical models simulating the species fundamental niche. Finally, these models were projected into the geographical space under current and future climatic conditions for predicting the specie's current and future invasive potential. Despite the scientific soundness of the approach, mechanistic models have a number of known limitations and uncertainties. A major difficulty is the need of capturing all the abiotic factors constraining a species' range which can be a complex task given that fundamental niches are multidimensional and that limiting conditions often result from an interaction of a variety of factors (Kearney and Porter 2009). As a results, and since this evaluation requires a daunting effort and a large number of resources, the application of these models remains restricted to a small number of species.

Distinctly from mechanistic models, niche based models (NBM) estimate environmental suitability indirectly. This is done by correlating the species known distribution with a set of spatially explicit environmental variables that can reasonably be expected to affect its physiology (Guisan and Zimmermann 2000). Several factors have merged recently to make this the current reference tool for predicting environmental suitability for species (not only invaders). First, nowadays there is a large availability of the data needed to build these models (species occurrences and spatially explicit environmental data) which, in many cases, can be obtained online and free of cost. Second, the statistical tools necessary to build these models also became increasingly available to the public, particularly in the form of free software with user-friendly interfaces (e.g. Phillips and Dudík 2008). Finally, and despite its simpler framework in comparison to the mechanistic approach, NBM showed to be able to provide reliable predictions (e.g. Elith et al. 2006) and examples of their application to invasive species are now common (e.g. Ficetola et al. 2007). However, these models also have a known number of limitations and uncertainties particularly when species are not in distributional equilibrium in a region, as is often the case of biological invasions (Elith and Leathwick 2009; Jiménez-Valverde et al. 2011). Most of these are particularly related to the fact that species distributions are shaped not only by limiting physiological conditions but also from biotic and historical factors such as competitive exclusion or dispersal limitations (i.e. the realized niche; Hutchinson 1957). Thus, NBM which based their inference on observed species distributions may provide an underestimated description of the abiotic conditions suitable for the species (see also the next subchapter).

Despite all the challenges mentioned above, many researchers agree that the current ability to predict new and ongoing invasions has significantly improved since the beginning of this field of research (Davis 2009; Lockwood et al. 2009; Essl et al. 2011) and a general expectation is that this ability will continue to increase in the future (Davis 2009). However, and unlike in the beginning, it is now widely acknowledged that no single factor can provide 'universal' estimates of invasion potential (Gurevitch et al. 2011). Thus, future research efforts should continue untangling the many and intricate complexities that intervene in the invasion process.

1.3 The use of niche based models for invasive species: challenges and uncertainties

As discussed in the previous chapter, knowing the environmental conditions under which a species can persist can be extremely useful to predict invasions. Furthermore, this information can be projected into the geographical space allowing to know where they can take place. In recent years, niche based models (NBM) has become the reference tool for performing this type of assessments. However, a key problem arises when these models are used for invaders. The most fundamental assumption behind NBM is that the species being modelled are in distributional equilibrium with the environment (Guisan and Thuiller 2005). This means that the information used for model calibration (species distribution data and environmental predictors) has to be representative of the conditions for which predictions are made. Thus, the equilibrium assumption can only be met when NBM are used to predict the specie’s current distribution and within their native ranges. Consequently, the risks arising from predictions for non-equilibrium settings (i.e. biological invasions or under climate change) can be various including possible changes on the abiotic or biotic factors responsible for driving distributions, predicting for novel environmental conditions (i.e. environmental extrapolation), or changes on the species itself, through evolutionary adaptation (Elith and Leathwick 2009; Hoffmann and Sgrò 2011). Any of these situations can severely harm the reliability of the predictions, and thus, possible approaches for improving NBM under these conditions have been the object of investigation in recent years (reviewed in Elith and Leathwick 2009).

One of the topics receiving largest focus has been the type of occurrence data needed to predict invasions. Using the currently invaded range allows to take into account possible changes on the drivers of distribution in the new area, however, it may also underestimate the species niche, because the full range of suitable conditions may not yet be occupied. On the other hand, using just data of the native range may not yield the full niche either, as the original range may not exhibit all possible climate and site combinations that could possibly be inhabited by a species. Recent studies recommend using both ranges simultaneously (Broennimann and Guisan 2008; Beaumont et al. 2009a), however, this may not be possible due to issues of data availability and also because if we want to predict invasions that have not occurred, only native data is available. In addition, the use of absence data can also be an issue, since this type of records is frequently unavailable and because a species may be absent from an area due to factors other than detrimental abiotic conditions (e.g. negative biotic interactions or dispersal barriers; Colwell and Rangel 2009). A common practice to circumvent this problem has been to use pseudo-absences instead of 'true absences'. However, these records can be generated in multiple ways and the best option for invasive species remains unknown (Chefaoui and Lobo 2008; Phillips et al. 2009; VanDerWal et al. 2009).

Environmental extrapolation is also problematic because the native (or the current) distribution of a species alone provides no information on the suitability of the new (i.e. unsampled) conditions (Thuiller et al. 2004; Fitzpatrick and Hargrove 2009). This is of particular relevance when working with models for invasive species since it is very difficult not to encompass new environmental conditions when models explicitly aim at predicting for new areas. In addition, if the predictions are projected across time (e.g. under scenarios of climate change) the dissimilarities between calibration and prediction conditions can be even greater. Despite its manifest importance, so far few advances have been made regarding on how to deal with environmental extrapolation or how it affects the reliability of the predictions (Elith and Leathwick 2009).

Besides the challenges originating from violating the equilibrium assumption, NBM for invaders also have to deal with the more general conceptual and methodological problems of NBM. Conceptual issues have been less addressed in the literature and are mostly centred on the discussion of what NBM effectively model (Soberón and Nakamura 2009) or on the exact definitions of popular terms in the field such as the 'niche' itself (Araújo and Guisan 2006). Methodological issues, on the other and, have received a larger attention and include: selection of spatial scale (i.e. extent of the study area and spatial resolution of the data) (Austin and Van Niel 2011), issues of sample size and prevalence (Wisz et al. 2008; Santika 2011), spatial autocorrelation (Dormann et al. 2007), model selection and evaluation (Liu et al. 2011b) or variable selection methods (Austin and Van Niel 2011). Zimmermann et al. (2011), refer that in the last decade large advances were made regarding these issues. Nonetheless, and despite the advances, the preferable methodological framework for a given application (biological invasions included) is still far from consensus (e.g. Jiménez-Valverde et al. 2008; Elith et al. 2010).

To account for some of this methodological uncertainty, the use of ensembles of multiple, but equally plausible, models was advocated (Araújo and New 2007). This suggestion is based on the principle that the errors of distinct models will tend to cancel each other out and thus more reliable predictions can be obtained (Araújo and New 2007). In addition, this approach also allows assessing the uncertainty of predictions through the examinations of the level of agreement between the multiple models. Given these merits, ensembles of predictions are recognized to be also of great promise for dealing with the inflated methodological uncertainties of NBM predictions for invaders (Davis 2009). Despite this recognition, so far, ensembles in invasion ecology have only been applied to reduce the predictive variability arising from to the use of distinct correlative algorithms (e.g. Roura-Pascual et al. 2009) and its potential for dealing with the uncertainty originated by other sources of variability such as calibration data remains untested.

Besides its use for predictions, NBM also hold a great potential for providing insights on the drivers of invasive distributions. For example, Muñoz and Real (2006) using a NBM framework, found that the distribution of the invasive Monk parakeet (*Myiopsitta monachus*) in Spain was mostly driven by human-related factors whereas climate was shown to be of reduced influence. This type of studies can be highly helpful in further understanding the complexities of the invasion process. However, and similarly to what happens with its predictions, this type of results can be dependent on the modelling procedures used. In this context, ensembles of NBM can be also promising in exploring arrays of explanatory results coming from distinct models. Nonetheless, the potential improvements of this approach remain unexplored.

2.1 Studied species

Decapods (Crustacea) are amongst the world's most successful and harmful invaders (Ricciardi 2011). In Europe alone there are currently about 15 non-native species established in inland waters (Holdich and Pöckl 2007), including prawns (Caridea), crabs (Brachyura) and crayfish (Astacidea). Based on a criterion of geographic proximity, this thesis is focused on four of the most notorious invasive decapods known for the inland waters of the Iberian Peninsula (Cabral and Costa 1999; Souty-Grosset et al. 2006): the Chinese mitten crab (*Eriocheir sinensis* H. Milne-Edwards, 1853); the Yabby (*Cherax destructor* Clark, 1936); the Signal crayfish (*Pacifastacus leniusculus* Dana, 1852) and the Red swamp crayfish (*Procambarus clarkii* Girard, 1852).

2.2 The Chinese mitten crab

The Chinese mitten crab is native from the temperate and subtropical regions of the Yellow Sea region bordering China and Korea, in Eastern Asia. It is a small-sized crab (up to ≈ 8 cm of carapace width), usually of brown colour which can be distinctively recognized by the furry patches present on its claws (that look like mittens, hence the common name) (Figure 3). It has highly opportunistic feeding habits which can include algae, plants, detritus, fishes, and a wide variety of invertebrates (Veilleux and De Lafontaine 2007; Dittel and Epifanio 2009). It is a catadromous species, spending most of its life in freshwater habitats (both lentic and lotic) but it needs to migrate to brackish waters of coastal areas to reproduce.

This species was first found outside its native range in the Aller River, a tributary of the Weser River, Germany, in 1912 (Panning 1939). It is thought that from there it reached the Baltic Sea where it was first found in 1926. Since then it has colonized all adjacent countries and spread westwards as far as the Atlantic coast of Southern Spain (Veilleux and De Lafontaine 2007). Outside Europe this species was also recently found on the Pacific and Atlantic coasts of North America (Veilleux and De Lafontaine 2007) and in Iraq (Hashim 2010). Two main vectors are considered responsible the introduction of this species in the new areas: the active transport and introduction into the wild in order to obtain a new food source, and the unintentional release of propagules via ship ballast water discharges (Cohen and Carlton 1997).

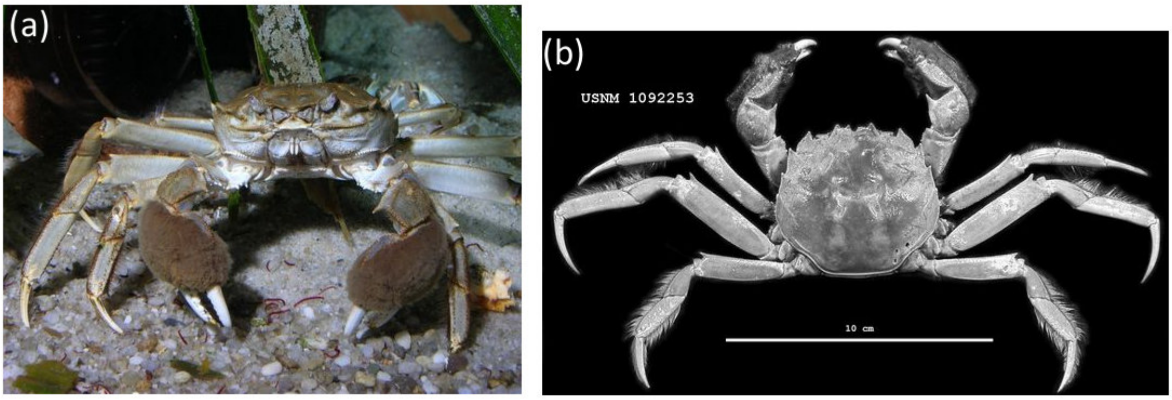


Figure 3 – The Chinese mitten crab. Photographic credits of left picture: Biopix and right picture: The Smithsonian Institution. All images reproduced with permission under a Creative Commons License.

This species has several known negative impacts in invaded areas and also poses a risk to human health. Negative impacts on invaded areas refer mostly to an acceleration of river bank erosion and instability due to burrowing activities and also to direct damages caused on human equipment such as fishing gear and plumbing systems (Panning 1939). This species is also an intermediate host of the oriental lung fluke, *Paragonimus westermani* (Veilleux and De Lafontaine 2007) which can be transmitted to humans by the ingestion of poorly cooked crabs. Nonetheless, this parasite has not yet been found in invasive populations (Veilleux and De Lafontaine 2007). Impacts on native biota are still poorly documented, however given its behaviour and feeding habits it is likely to cause negative impacts on local population through predation and/or competition (e.g. Gilbey et al. 2008). Known positive effects of this species refer almost exclusively to its harvest by small scale commercial fisherman (e.g. in the Tagus Estuary; author personal observation).

2.3 The Red swamp crayfish

The red swamp crayfish is native to subtropical regions of north-eastern Mexico and south-central USA. It can reach up to 15 cm of body length having a usually dark red to brownish body colour and red claws covered with spines and tubercles (Figure 4). This species feeds on plants and organic sediments but also on animal food such as other crayfish, adult insects and their larvae, fish, tadpoles and snails (Correia 2003; Souty-Grosset et al. 2006). This species has a short lifespan (12 -18 months in nature), a high fecundity (as many as 600 eggs at a time) and a highly plastic reproductive cycle (Gherardi 2006). They also have a wide ecological plasticity, being tolerant to fluctuating water levels and to high temperatures mainly due to their ability to excavate burrows (Gherardi 2006; Souty-Grosset et al. 2006).

Figure 4 – The Red swamp crayfish. Photographic credits and copyright: César Capinha

Currently, this is the most widely introduced crayfish in the world being found on every continent except Australia and Antarctica (Gherardi 2006). Its widest invasive range is found in Europe. It was first introduced in 1973 in southern Spain for harvesting in the wild and from there it was soon introduced to new areas of Spain, Portugal, France and Italy. Currently it is known present in European countries, including some islands such as São Miguel in the Azores or Tenerife in the Canary Islands (Holdich et al. 2009). Several pathways of introduction are known to have contributed for the spread of this species. Perhaps the most important has been its intentional introduction for harvesting in the wild or in aquaculture facilities which was motivated by the success of commercial aquaculture in its native range (Gutiérrez-Yurrita et al. 1999; Henttonen and Hunner 1999). The Red swamp crayfish is also a popular as an aquarium pet, for ornamenting artificial ponds or as a study model, uses that are also considered to be promoting its spread to new areas (Larson and Olden 2008).

This species is responsible for multiple negative impacts on native species, ecosystems and economic activities in the new ranges. Invasive populations have led to declines in native crayfish species in Europe through competition and because they can carry the fungus *Aphanomyces astaci* (Holdich et al. 2009). Its polytrophic and opportunistic feeding habits can also profoundly modify the trophic structure of freshwater communities leading to a decline of several native species (Gherardi 2006). The red swamp crayfish is also a well known agricultural pest, responsible for large damages on rice plantations by feeding on the young rice plants (Anastácio et al. 2005). Its burrowing activity can also contribute to the destabilization of river banks and dams (Correia and Ferreira 1995). On the other hand, some positive effects have been recognized to this invasive species. In some areas this crayfish became an important food source for some vertebrate predators, potentially leading increases in their populations (Clavero et al. 2010). Additionally, its commercial value as a food item is also explored by a number of fisherman and aquaculture industries worldwide (Souty-Grosset et al. 2006).

2.4 The Signal crayfish

The Signal crayfish is native from the cool temperate regions of north west America, west of the Rocky Mountains. This crayfish can reach up to 16 cm and have a usually bluish/brown to reddish/brown carapace colour (Figure 5). It has large and robust claws each having a white patch near the claw hinge. It consumes a wide variety of food sources including other crayfish, insects, small fish, snails, aquatic plants, tree leaves and detritus (Bondar et al. 2005; Souty-Grosset et al. 2006). This species has a long lifespan (up to 20 years), reaches maturity at an age of 2-3 years and, each year, females can generate from 200 to 400 eggs (Souty-Grosset et al. 2006). The Signal crayfish is considered to be relatively tolerant to several environmental stressors such as brackish water and temperature variations; however it has fairly high oxygen demands (Souty-Grosset et al. 2006).

Figure 5 – The Signal crayfish. Photographic credits and copyright: César Capinha

Currently this is the most widespread invasive crayfish in Europe, present in 27 countries (Holdich et al. 2009). The large majority of this success is due to intentional (legal and illegal) introductions in the wild for harvesting, since this species is a valued food item. Introductions began in Sweden in the 1960s and soon expanded to several other European countries. High levels of country-level introduction effort are reported for Sweden, Finland, United Kingdom, Spain, Austria and Luxemburg. Nonetheless, currently this invader is also widely dispersed in France, Germany and Belgium.

The most severe negative effect of the invasive Signal crayfish is the extermination of several populations of native European crayfish species, caused by the transmission of the deadly crayfish fungus plague *Aphanomyces astaci* (Holdich et al. 2009). It is also reported to compete with some of these species for food and shelter (Lindqvist and Huner 1999; Westman et al. 2002) and to cause reproductive interference with the native Noble crayfish (*Astacus astacus*) (Westman et al. 2002). Unsurprisingly, given the strong human interest in its introduction, this species has a high commercial value in some countries, particularly Sweden and Finland (Henttonen and Hunner 1999).

2.5 The Yabby

The Yabby is a crayfish native to the warm to semi-desertic regions of south east Australia. This crayfish can reach up to 15 cm and its carapace colours can vary from olive-green to brown, but can also be blue, yellow, red or black depending on the habitat, location and individual (Figure 6). Generally, the Yabby's diet is dominated by plant material and detritus (Souty-Grosset et al. 2006). However, this species can also be highly predatory (Beatty 2006). This species has an average lifespan of 3 up to 6 years and females can generate from 100 to 1000 eggs up to five times a year (Withnall 2000; Souty-Grosset et al. 2006). This species is tolerant to a wide range of temperatures but water temperatures above 35°C may be lethal and it falls into a state of partial hibernation when temperatures drop below 15/16°C (Withnall 2000; Souty-Grosset et al. 2006). It also tolerates a wide range of dissolved oxygen levels and high salinities.

Currently, this species is invasive in the south west of Australia and in some regions of Spain and Italy (Holdich et al. 2009). Similarly to the Signal crayfish, this species has been translocated to new locations mainly for harvesting. Many of these enterprises have been conducted in isolation from natural water bodies, but some escaped individuals have been able to establish wild populations (Souty-Grosset et al. 2006). This species is also used in the aquarium trade, which may promote its wider spread in the future (Holdich et al. 2009).

Figure 6 – The Yabby. Photographic credits of left picture: Neil Amstrong and right picture: Rudie Kuiter. All images reproduced with permission under the terms of use of the ARKive project.

The ecological impacts of this species in invaded areas remain less studied than of the three previously mentioned species. Nonetheless, studies conducted in Western Australia support that this species can compete with and predate native species (Beatty 2006; Lynas et al. 2007). Economy is again on the opposite side of the balance. Yabbies are harvested commercially in a large number of areas outside its native range and despite an usually low farmer's profit, this activity also has low operating costs (Morrissy 1994).

3.1 Objectives of the thesis

Given the challenges previously described, it is a general objective of this thesis to contribute for the advancement of the field of predictive invasion ecology. Towards that end, the thesis is particularly focused on evaluating and improving the reliability of predictions of niche based models for invasive species. As a result, several of the most currently debated topics in this field of research were investigated. These included the data requirements for building accurate models (Capinha et al. 2011), the potential improvements arising from the use of ensembles of distribution models (Capinha and Anastácio 2011) or the reliability of predictions coupling both invasive species and scenarios of climate change (Capinha et al. 2012). The obtained knowledge was mainly discussed within the context of (re)framing guidelines of model building and results interpretation for invasive species.

At the same time, this thesis also intended to expand the current knowledge about the invasive potential of the four problematic decapods described above (Chinese mitten crab, Red swamp crayfish, Signal crayfish and the Yabby). This included assessing their current worldwide potential distribution (Capinha et al. 2011), disentangling the main drivers of their macro-scale invasion patterns (Capinha et al. unpublished), quantifying their fine-scale environmental requirements (Capinha and Anastácio 2011) and evaluating how climate change may alter their potential distributions (Capinha et al. 2012). On the basis of the obtained results, the potential implications for managers and stakeholders are also analysed and discussed.

The research conducted is described in detail in the four following chapters (chapters 2, 3, 4, and 5). Each of these corresponds to a research paper that was published or is being considered for publication in an international peer-reviewed scientific journal in the field of ecology.

Chapter 2 investigates the implications of using distinct distribution data in predicting the worldwide environmental suitability for invasive species. It evaluates the predictive consequences of using only occurrence data from native or from invasive ranges or from both ranges simultaneously along with distinct sampling strategies for extracting pseudo-absences. It also explores the potential usefulness of consensus predictions in comparison to predictions from single-models. This research was based on data from the four invasive decapods and estimates of worldwide environmental suitability for each were also obtained and examined.

Chapter 3 investigates the relative importance of human, biological and climatic factors in determining the current macro-scale pattern of distribution of the Signal crayfish and the Red swamp crayfish in Europe. Here, the Chinese mitten crab and the Yabby were not considered due to the inexistence of adequate distribution data. Results were analysed in light of four main invasion hypotheses (propagule pressure, climate matching, biotic resistance and human disturbance) and used for anticipating future range expansions of the invaders in Europe.

Chapter 4 investigates the use of ensembles of distribution models to predict the fine-scale environmental suitability of the Iberian Peninsula to the four invasive decapods and simultaneously examine the uncertainty arising from the use of distinct statistical techniques. Subsequently, it also investigates how consensus predictions can be used to provide insight on the species-environment relationships of the invaders. Acquired knowledge was further discussed in terms of species ecology and management implications.

Chapter 5 investigates how climate change may alter the environmental suitability of the Iberian Peninsula to the four invasive decapods in the mid to long-term. In particular, it explores how the reliability of the predictions differs among the four species in view of the model’s ability to predict current ranges, variability among predictions and degree and extent of environmental extrapolation encompassed. The indicators of predictive reliability and the avenues for possible improvements are also discussed.

After these chapters, a general discussion and conclusion follows (chapter 6), where the main results and implications of this research are discussed and future perspectives are outlined.

Page intentionally left blank

CHAPTER 2

Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets

This chapter was published in:

Capinha C, Leung B, Anastácio P (2011) *Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. Ecography 34:448-459.*

PREDICTING WORLDWIDE INVASIVENESS FOR FOUR MAJOR PROBLEMATIC DECAPODS: AN EVALUATION OF USING DIFFERENT CALIBRATION SETS

César Capinha[1]; Brian Leung[2] and Pedro Anastácio[1]

[1]IMAR, Centro de Mar e Ambiente c/o Departamento de Paisagem, Ambiente e Ordenamento, Universidade de Évora, Rua Romão Ramalho, n.º 59, 7000-671 Évora, Portugal.

[2]Department of Biology, McGill University, Montreal, QC, H3A 1B1, Canada.

Abstract

Recently, there as been much debate whether niche based models (NBM) can predict biological invasions into new areas. These studies have chiefly focused on the type of occurrence data to use for model calibration. Additionally, pseudo-absences are also known to cause uncertainty in NBM, but are rarely tested for predicting invasiveness. Here we test the implications of using different calibration sets for building worldwide invasiveness models for four major problematic decapods: *Cherax destructor*; *Eriocheir sinensis*, *Pacifastacus leniusculus* and *Procambarus clarkii*. Using Artificial Neural Networks models we compared predictions containing either native range occurrences (NRO), native and invasive occurrences (NIO) and invasive only (IRO) coupled with three types of pseudo-absences - based on sampling only 1) the native range (NRA), 2) native and invasive ranges (NIA), and 3) worldwide random (WRA). We further analysed the potential gains in accuracy obtained through averaging across multiple models. Our results showed that NRO and IRO provided the best predictions for native and invaded ranges, respectively. Still, NIO provided the best balance in predicting both ranges. Pseudo-absences had a large influence on the predictive performance of the models, and were more important for predictiveness than types of occurrences. Specifically, WRA performed the best and NRA and NIA performed poorly. We also found little benefit in combining predictions since best performing single-models showed consistently higher accuracies. We conclude that NBM can provide useful information in forecasting invasiveness but are largely dependent on the type of initial information used and more efforts should be placed on recognizing its implications. Our results also show extensive areas highly suitable for the studied species worldwide. In total these areas reach from three to nine times the species current ranges and large portions of them are contiguous with currently invasive populations.

Keywords: *Invasiveness, calibration data, pseudo-absences, niche based models, decapods*

Introduction

Accurate information concerning the risk of a species becoming established outside its native range can provide a solid foundation for justifying preventive measures and has been a subject of increasing focus. As such, invasion biologists have sought predictive methods to forecast invasions (Côté and Reynolds 2002; for a recent review see Hayes and Barry 2008). Specifically, niche based models (NBM), which estimate the degree of environmental compatibility for the species in new areas, have become increasingly popular in the last few years. For example, Thuiller et al. (2005) found a close similarity between worldwide NBM invasiveness predictions and South African plant invaders distribution. Additionally, others have used NBM predictions coupled with propagule pressure estimates in order to provide final predictions of risk of establishment of a species (e.g. Leung and Mandrak 2007).

Recently, there has been much debate concerning the use of NBM for invasive species predictions (Mau-Crimmins et al. 2006; Loo et al. 2007, Broennimann and Guisan 2008; Pearman et al. 2008; Steiner et al. 2008; Beaumont et al. 2009a). This has been chiefly focused on the type of occurrence data to use for model calibration. Some studies using native-based models were able to provide accurate invasiveness predictions (e.g. Welk et al. 2002; Thuiller et al. 2005), but others found reduced predictability of the entire invaded ranges (Mau-Crimmins et al. 2006; Fitzpatrick et al. 2007; Broennimann and Guisan 2008; Beaumont et al. 2009a). Most recent studies argue that the use of occurrence data from the invaded ranges improves predictions (e.g. Mau-Crimmins et al. 2006; Loo et al. 2007; Broennimann and Guisan 2008; Beaumont et al. 2009a), since invaders may not conserve their niches across space (Broennimann et al. 2007; Pearman et al. 2008). Still, invasive occurrences would be constructed only after the invasion had already taken place and could have issues with data availability. Moreover, if the invasion process is not complete the equilibrium assumption underlying NBM may be violated and cause underestimation of the full invasiveness potential (Wilson et al. 2007). Further, native range data for invasive species is also often unavailable or difficult to collect (Mau-Crimmins et al. 2006). Thus, while coupling native and invasive occurrences may allow the best overall characterization of species niches, there could be issues with data availability. As such, a direct comparison of the quantitative differences and marginal gains of using different occurrence data would be useful.

The inclusion of absence data is arguably also an issue of NBM for invaders. These data are usually hard to obtain from common sources of species distribution data such as museums or biodiversity databases (Chefaoui and Lobo 2008). To deal with these difficulties, the use of pseudo-absences has been common since it allows the use of "group-discrimination techniques" considered to provide more accurate predictions than presence only "profile techniques" (Hirzel et al. 2001; Brotons et al. 2004; Segurado and Araújo 2004). Still, the use of pseudo-absences is a known cause of uncertainty in NBM (Lobo 2008; Phillips 2008) and the best way to obtain them is still far from consensus (Chefaoui and Lobo 2008; VanDerWal et al. 2009; Wisz and Guisan, 2009). Moreover, extraction techniques have mostly been tested with species for which equilibrium with the environment is assumed (i.e. species within their native ranges) (for a review see Pearce and Boyce 2006). However, invasion biology is interested in estimating suitability in new areas and, thus, it is unclear how best to obtain pseudo-absences. Options include: 1) use sites located within the native range. The assumption is that there has likely been sufficient time for propagules to reach these sites. 2) Use sites within native and invasive ranges simultaneously since new ranges may provide additional information. However, this would not necessarily take into account the range of environments possible. 3) Use random points across the world. The consequences of these different forms of pseudo-absences for predictive ability of invasions have not been examined and are currently unknown.

For dealing with uncertainty in the predictions, researchers have been increasingly adopting the use of consensus methods - "ensembles" of single-model NBMs, with different architectures or different assumptions (i.e. averaging across the results of single models). Model ensembles have been applied for predicting distributions of threatened species (e.g. Marmion et al. 2009), impacts of climate change on species distributions (e.g. Araújo et al. 2005) or potential distributions of invasive species (e.g. Roura-Pascual et al. 2009; Stohlgren et al. 2010). Ensembles have been mostly used to deal with the uncertainty caused by the use of distinct correlative models. Despite their potential to reduce uncertainties coming from the use of different calibration data, to our knowledge, ensembles have not been applied for dealing with this source of uncertainty in invasiveness predictions.

In this study, we build habitat suitability models and test the consequences of occurrence type and pseudo-absence type. Specifically we examine the use of occurrence data from their native range (NRO), invasive range (IRO) and both native and invasive occurrences (NIO) in the algorithm's calibration sets. Simultaneously we also evaluate the effect of three different pseudo-absence methods based on sampling: 1) the native range (NRA), 2) native and invasive range (NIA), and 3) woldwide random (WRA). Finally, we also explore the use of consensus methods as a possibility for dealing with the uncertainty coming from the use of different calibration information.

We focus on four important invasive species: *Cherax destructor*; *Eriocheir sinensis*; *Procambarus clarkii* and *Pacifastacus leniusculus.* These are wide-ranging invasive decapods, for which either populations or individuals are being systematically found in new areas. Their impacts in the invaded ecosystems are numerous (e.g. predation and competition with native species, habitat alteration and agricultural damage) and the major mechanisms of introduction have been identified. As such, it will be most effective for managers to target habitat suitability models to these species. Still, the uncertainty coming from the use of inadequate calibration information can undermine this objective.

Methods

Invaders and distribution data

Cherax destructor Clark, 1936 (Yabby) is a crayfish indigenous to Eastern Australia that currently invades several areas in Western Australia and Iberian Peninsula. As for other invaders, occurrence records for *C. destructor* are scarce and most have a low spatial accuracy (e.g. Souty-Grosset et al. 2006). For reducing the uncertainty of their spatial location of collected data we opted for using a cell resolution of 50km. We used only one record per grid cell and gathered a total of 154 occurrence records for this species, 103 referring to its native range and 51 from invaded areas. This information was mostly collected from the Museum Victoria collections and several published works.

Eriocheir sinensis Milne-Edwards, 1853 (Chinese mitten crab) is an invasive crab included in the 100 "World's Worst" invaders by the World Conservation Union (Lowe et al. 2000). The native range of this catadromous crab encompasses eastern China, Japan and eastern Russia, being presently invasive in several coastal areas of North America and with particular expression in Europe (Gollasch 2006). For this species we obtained a total of 295 occurrence records, 101 from its native range and 194 from invaded areas. This information was collected using the global biodiversity information facility (GBIF) (http://www.gbif.org/) and a vast number of published works referring to this species.

Procambarus clarkii Girard, 1852 (Red swamp crayfish) is a commercially harvested crayfish, native from northeast Mexico to south-central USA. This species currently has invasive populations across 5 continents (Africa, Asia, Europe, North America and South America) and was recently quoted as one of the 100 "Most invasive alien species in Europe" (Drake 2009). For this crayfish we collected a total of 598 occurrence records, 173 from its native range and 425 from invaded areas. Its distributional data were collected from the Smithsonian Institution National Museum of Natural History, the Illinois Natural History Survey, the Atlas of Crayfish in Europe (Souty-Grosset et al. 2006), GBIF and several published works.

Pacifastacus leniusculus Dana, 1852 (Signal crayfish) is native from the north-western USA and south-western Canada and currently invades large portions of the European continent, south-western USA and some Japanese regions (Souty-Grosset et al. 2006). A total of 565 occurrence records were collected for this species, 125 from its native range and 440 from invaded areas. The data sources used where the same as for *P. clarkii*.

Environmental factors

To summarize the world environmental characteristics (Antarctica excluded) we considered 10 spatial coverages. All environmental predictors used were not collinear (Pearson's $|r|<0.8$). Eight climatic variables concerning the period 1961-2000 were included: near surface annual mean temperature (*amtemp*); near surface mean maximum temperature of the warmest month (*maxtwm*); near surface mean minimum temperature of the coldest month (*mintcm*); near-surface mean diurnal temperature range (*trange*); mean number of frost days (*frost*); mean total annual precipitation (*anpre*); mean total precipitation of the wettest month (*prewm*) and mean total precipitation of the driest month (*predr*). These predictors were built using information from the CRU TS2.1 climate dataset (Mitchell and Jones 2005). Two physiographic variables were also included: altitude (*alt*) and in-stream distance to ocean (*disto*). Altitude was included since it can act as a surrogate of several environmental factors important for our species such as stream velocity and size – usually faster and smaller at higher elevations. In-stream distance to ocean was employed only for *E. sinensis* due to its catadromous nature. The digital elevation model was acquired on the United States Geological Survey (USGS) HYDRO1k geographical dataset (Verdin and Jenson 1996), from which in-stream distance to ocean was calculated using ILWIS 3.5 Open (http://52north.org/). All variables were resampled to a 50km cell resolution using a bicubic method and projected to a Mollweide equal area world projection.

Pseudo-absences extraction

Pseudo-absences can be seen as a sample of the available conditions (e.g. Philips et al. 2009) or as indicator of unsuitable conditions (e.g. Chefaoui and Lobo 2008). For pseudo-absences used in this study, we excluded all cells having occurrence of the species in order to potentially maximize the representativeness of unsuitable conditions. Following this principle, one approach for generating pseudo-absences is a simple spatially-random generation of records along the entire study area, except from where presences are known (WRA).

Our second approach was entirely based on pseudo-absences in the native range distribution of the species (NRA). Given our relatively intense search for occurrence data and the fact that all four species have well documented native distributions, we assumed the areas without the species presence within native range boundaries were reliable representatives of unsuitable conditions. For NRA, we limited our sampling area to the inner boundary of the convex-hull defined by the occurrence records. Our third approach consisted of sampling both native and invasive ranges (NIA). By this we assume that the unoccupied areas within the invasive range can provide additional information regarding unsuitable conditions. Sampling was made within the inner boundary of the convex-hull defined by the occurrence records of each range. To avoid wide and unrealistic sampling areas distinct invasive populations of each species were delimited by independent convex-hulls.

Dataset assembly

Before calibration datasets were built we retained 20% of each species' occurrence to validate our predictions (i.e. they were not used to build the model). The remaining occurrence records were used to build nine different types of calibration datasets: 1) native range occurrences versus worldwide random pseudo-absences; 2) native range occurrences versus native range pseudo-absences; 3) native range occurrences versus native and invasive ranges pseudo-absences; 4) native and invasive ranges occurrences versus worldwide random pseudo-absences; 5) native and invasive ranges occurrences versus native range pseudo-absences; 6) native and invasive ranges occurrences versus native and invasive ranges pseudo-absences; 7) invasive range occurrences versus worldwide random pseudo-absences; 8) invasive range occurrences versus native range pseudo-absences and 9) invasive range occurrences versus native and invasive ranges pseudo-absences. For increasing the representation of the environment captured by the pseudo-absences we created 20 calibration datasets for each combination. Each of these had an independently drawn set of pseudo-absences. To avoid biasing predictions towards a more prevalent response each calibration dataset had a number of pseudo-absences equal to the number of occurrences (See Table 1 in the Appendix A).

Model selection and predictions

NBM has been built using many distinct correlative models with several new approaches receiving great promise (Elith et al. 2006). For this study we have chosen to use Artificial Neural Networks (ANN) for predicting the probability of environmental suitability in each cell. ANN is a method used regularly in NBM and has also been recognized as one of the best performing techniques (Segurado and Araújo 2004). Moreover ANN are particularly appropriate when the relations between variables are not well known, which is often the case with ecological data (Lek and Guégan 1999). We used feedforward multilayer perceptron with back-propagation ANN models (MLP-ANN). MLP-ANN is one of the most common types of supervised ANN's used in ecology, being normally structured in one input layer representing the predictors (environmental variables), one or more hidden layers, each with a variable number of nodes, and one output layer representing the dependent variable (presence/absence).

Due to its large flexibility, ANN models are prone to overfitting, making the model less generalizable and decreasing their predictive power. To avoid overfitting in our models we used both a cross-validation procedure during the training episodes and tested different network configurations in order to optimize their degree of complexity and number of training cycles (Özesmia et al. 2006). While higher complexity increases the risk of overfitting, oversimplification can also result in poor fits. In the same way, excessive training of the network is prone to overfit the data while the inverse may result in failure to capture its regularities.

Before building final predictions we tested for the more appropriate network configuration for each of the dataset types. To do this we compared the performances of single hidden layer MLP networks using three different levels of complexity. According to Burnham and Anderson (2002), the available data sample should be at least ten times larger than the number of parameters in a model. We adopted this principle for establishing the maximum complexity allowed in each of the tested models. Medium complexity networks were also considered, each containing half the hidden nodes of the previous models. Finally, for the least complex models we tested the performance of MLP-ANN containing no hidden nodes, which are equivalent to Generalized Linear Models.

Models were built using Weka 3.6 (Witten and Frank 2005). To comply with the binary response of the dependent variable, hidden nodes were automatically set to sigmoid functions. All training sessions included a weight decay function of the learning rate by dividing the starting value by the cycle number, forcing a low learning rate and by so reducing the risk of data overfit. A stopping rule was also included in order to avoid overtraining. Models were allowed to train for a total of 4000 cycles as long as the predictions did not exceed more than 500 consecutive cycles without performance improvement.

To analyse the predictive power of each of the three network configurations we used a 10 fold cross-validation procedure. That is, all models were calibrated using 90% of cases for model calibration while the remaining 10% were left-out for comparison with predicted values. This procedure was then repeated 10 times until the entire dataset had been compared against the predictions. These comparisons were evaluated using the mean values of the root mean squared error (RMSE) automatically supplied by Weka. The network configurations achieving lower mean RMSE for each dataset type were then selected and applied for predicting along the entire range of worldwide environmental conditions. Due to the use of 20 independent sets of pseudo-absences, the final prediction for each dataset type corresponded to the mean value obtained by these 20 calibration datasets.

Ensemble predictions

For dealing with the variability of single predictions, the combination of ensemble models has been adopted in studies with invasive species (e.g. Roura-Pascual et al. 2009; Stohlgren et al. 2010). While the majority of efforts have been focused on the variability caused by the use of distinct modeling methods, this logic could apply to reduce the uncertainty coming from the use of different sets of calibration data. Here we explored the possibility of improving invasiveness predictions using ensemble models. We examined three types of ensembles – predictions based on a weighted average of all single models (*WA(all)*), averaged within each occurrence type (*WA(NRO)*; *WA(NIO)*; *WA(IRO)*), and averaged within each pseudo-absence type (*WA(NRA), WA(NIA), WA(WRA)*). For each, all ensembles were obtained through averaging single-models by their relative accuracy value (Marmion et al. 2009). In order to attain a fair comparison against the single-models predictions, relative accuracy value was based on the RMSE obtained from the 10-fold cross-validation process used in the network configuration selection process as supplied by Weka (See Table 2 in the Appendix A). The weighted averages of the single-models were performed as given by Eqn 1

$$WA_i = \frac{\Sigma((1 - RMSEpj_i) \times pj_i)}{\Sigma_j(1 - RMSEpj_i)}$$

(Eqn 1)

where pj_i was the probability of environmental suitability for the *i*th decapod species in each of the *j*-selected single-models.

Predictions validation

After predictions were made for both single and combined-models their evaluation was performed. We used the 20% of each type of occurrence records (i.e native and invasive) initially excluded from the calibration datasets. These were complemented with an equal number of worldwide random sample of areas without native or invaded occurrences. For increasing the representativeness of these areas in the evaluation datasets we made 10 datasets for each type of occurrence records. Each of these had an independently drawn sample of areas without native or invaded occurrences (See Table 3 in the Appendix A). Validation records were compared with the predicted values using the area-under-the-curve of the receiver-operating characteristic (ROC-AUC) (Hanley and McNeil 1982) and Cohen's Kappa (*k*) (Cohen 1960). Kappa was calculated across a range of thresholds along the 0 to 1 interval using a 0.05 amplitude increment and its maximum value selected (Elith et al. 2006). Both native and invasive ranges were evaluated. Final evaluation values were obtained by averaging the scores of the 10 replicate evaluation datasets. For assessing variability in predictive performance we also calculated the standard deviation of the obtained evaluation scores. For qualitatively describing the predictions values of *k*, we established the following classes: $k < 0.2$ poor; $0.2 < k < 0.4$ fair; $0.4 < k < 0.6$ moderate; $0.6 < k < 0.8$ good and $k > 0.80$ as very good (modified from Landis and Koch 1977). For ROC-AUC we considered ROC-AUC < 0.8 as poor accuracy; 0.8 < ROC-AUC < 0.9 moderate; 0.9 < ROC-AUC < 0.95 good and ROC-AUC > 0.95 as very good (adapted from Thuiller et al. 2005).

Species environmental space

We also examined environmental similarities between native and invaded ranges. If environmental conditions differ, this may indicate that a niche shift has occurred. Following Mau-Crimins et al. (2006), we used a Principal Components Analysis (PCA) to simplify the species niche dimensionality and compared the position occupied by each occurrence record. By distinguishing native and invasive occurrences this procedure allowed us to verify the degree of environmental overlap between the two ranges occupied by each species. This was made through a score plot of the two Principal Components. Further we used the same method to compare the position of the species occurrences with the overall best performing pseudo-absences extraction method. For improving visual interpretation we only included 300 randomly selected pseudo-absence records for plotting.

Results

Predictive performances

For single-models, native range occurrences and invasive range occurrences provided the best predictions for native and invaded ranges, respectively (Tables 1 and 2). Still, combined native and invasive occurrences allowed a good balance between the two, with best models attaining good to very good accuracy values in both ranges (Tables 1 and 2). Interestingly, native range occurrences allowed extrapolation to the invaded ranges of three of the four species examined with moderate to good accuracy values (*C. destructor* NRO best model: k = 0.63 and ROC-AUC = 0.89; *E. sinenis* NRO best model: k = 0.9 and ROC-AUC = 0.95; *P. leniusculus* NRO best model k = 0.89 and ROC-AUC = 0.93, Table 2). In contrast, best predictions using native range occurrences for the invasiveness potential of *P. clarkii* were worse (k = 0.57; ROC-AUC = 0.71). Models using invasive range occurrences were relatively modest in predicting the native ranges for all species (Table 1). Variability in accuracy was similar for the three types of occurrences in predicting native (mean SD of NRO ROC-AUC: 0.029; mean SD of NIO ROC-AUC: 0.033; mean SD of IRO ROC-AUC: 0.031) or invasive distributions (mean SD of NRO ROC-AUC: 0.033; mean SD of NIO ROC-AUC: 0.030; mean SD of IRO ROC-AUC: 0.029) (See Tables 4 and 5 in the Appendix A).

Table 1. Predicting distribution in native range: single-models validation results of kappa statistic (k) and area under the curve of receiver-operating characteristic (ROC-AUC) for native ranges using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA).

		NRO		**NIO**		**IRO**	
Species	Pseudo-absences	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC
Cherax destructor	NRA	0.51	0.68	0.28	0.49	0.21	0.38
	NIA	0.51	0.67	0.52	0.63	0.17	0.29
	WRA	0.91	0.99	0.85	0.98	0.64	0.90
Eriocheir sinensis	NRA	0.81	0.89	0.52	0.63	0.32	0.15
	NIA	0.78	0.84	0.55	0.72	0.12	0.32
	WRA	0.82	0.90	0.81	0.86	0.74	0.86
Pacifastacus leniusculus	NRA	0.68	0.75	0.50	0.60	0.49	0.35
	NIA	0.51	0.63	0.45	0.55	0.24	0.44
	WRA	0.95	0.99	0.88	0.96	0.82	0.91
Procambarus clarkii	NRA	0.79	0.85	0.39	0.38	0.02	0.07
	NIA	0.79	0.86	0.52	0.74	0.05	0.21
	WRA	0.93	0.99	0.90	0.97	0.78	0.86

Table 2. Predicting distribution in introduced range: single-models validation results of kappa statistic (*k*) and area under the curve of receiver-operating characteristic (ROC-AUC) for invaded ranges using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA).

		NRO		**NIO**		**IRO**	
Species	Pseudo-absences	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC
Cherax destructor	NRA	0.28	0.45	0.42	0.67	0.32	0.59
	NIA	0.23	0.39	0.28	0.42	0.64	0.72
	WRA	0.63	0.89	0.74	0.93	0.74	0.97
Eriocheir sinensis	NRA	0.57	0.67	0.80	0.90	0.68	0.81
	NIA	0.45	0.61	0.78	0.89	0.76	0.88
	WRA	0.90	0.95	0.90	0.96	0.91	0.97
Pacifastacus leniusculus	NRA	0.58	0.70	0.66	0.78	0.83	0.85
	NIA	0.46	0.63	0.76	0.88	0.85	0.93
	WRA	0.89	0.93	0.90	0.96	0.91	0.96
Procambarus clarkii	NRA	0.56	0.70	0.76	0.86	0.72	0.84
	NIA	0.33	0.56	0.80	0.91	0.81	0.95
	WRA	0.57	0.71	0.87	0.94	0.88	0.97

Type of pseudo-absence had a pronounced effect on predictive performance. Specifically, models using native range pseudo-absences and native and invasive ranges pseudo-absences had the lowest predictive performances (Tables 1 and 2). Best performances were achieved unanimously with worldwide random pseudo-absences. Worldwide random pseudo-absences also provided the lowest mean variability of accuracy in predicting both native (mean SD of ROC-AUC: 0.019) and invasive distributions (mean SD of ROC-AUC: 0.024) (Tables 4 and 5 in the Appendix A). We found that the weighted average procedure provided poor to moderate predictive accuracy for both native or invaded ranges for *WA(all)* and *WA(NIO) (*See Table 6 in the Appendix A for a measure of variability among models). The *WA(NRO)* method provided moderate accuracy for predicting native distributions but was clearly unable to capture the species invasive ranges. Inversely, the *WA(IRO)* method provided good predictions concerning the species invasive distributions except for *C. destructor* (k = 0.64 and ROC-AUC = 0.74), but was unable to predict native ranges. The *WA(NRA)* and *WA(NIA)* methods provided the lowest accuracies for predicting the species native ranges but provided moderate to good accuracy in predicting the invasive range of three species (*E. sinensis*; *P. leniusculus* and *P. clarkii*). Finally, the *WA(WRA)* model provided the best accuracies in both ranges. However, in general we found little benefit of combined models when compared to the best performing single models, specifically native and invasive occurrences coupled with worldwide random pseudo-absences.

Species environmental space

The PCA analysis allowed comparison of environmental space of presences versus pseudo-absences (Figure 1), and between regions occupied in the native and exotic ranges (Figure 2). Not surprisingly, the comparison between occupied ranges and best performing pseudo-absence extraction method (worldwide random) illustrated a much larger range of environmental conditions for pseudo-absences compared to occurrence records (Figure 2). Interestingly, however, focusing on the occurrence data demonstrated that all species were occupying different environmental space in the naturalized areas compared to the exotic ranges (Figure 2). For *C. destructor* the difference occurred along Component 1, which was primarily a temperature composed gradient. The multivariate space occupied by *E. sinensis* in the invaded range was mainly differentiated along Component 2, primarily associated with both near surface mean minimum temperature of the coldest month and mean total precipitation of the driest month. For *P. clarkii*, despite the existence of some overlapping environmental conditions, native and invasive populations were differentiated along a gradient dominantly composed by the altitude and mean number of frost days variables. The PCA clouds for *P. leniusculus* indicated a less clear differentiation between the two ranges than for the previous species, but still had large non-overlapping portions along both Components.

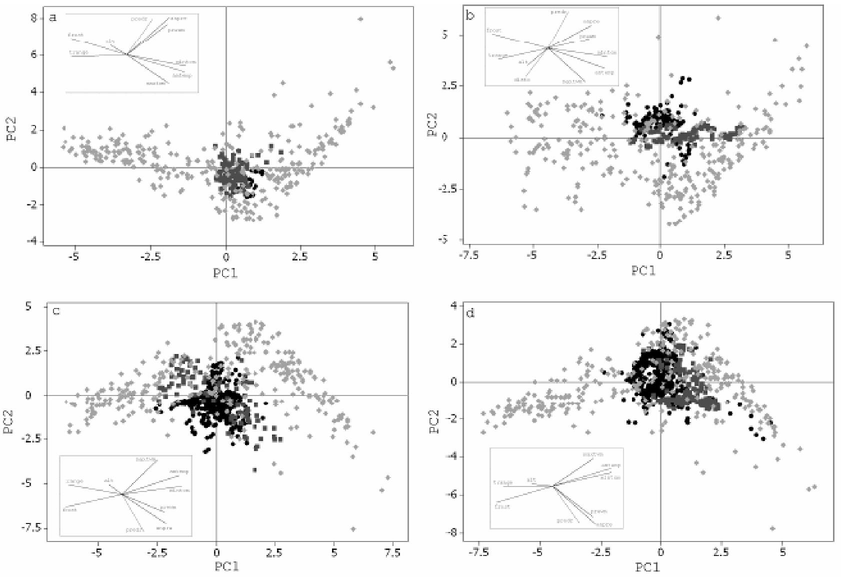


Figure 1 - Position in the environmental space of native (grey squares) and invasive (black dots) populations and worldwide random pseudo-absences (light grey diamonds). A Score plot of the two Components was made from a PCA containing all environmental variables. PC1 and PC2 for (a) *Cherax destructor* with worldwide random pseudo-absences explained 76% of total variance, for (b) *Eriocheir sinensis* with worldwide random pseudo-absences explained 70% of total variance, for (c) *Pacifastacus leniusculus* with worldwide random pseudo-absences explained 75% of total variance, for (d) *Procambarus clarkii* with worldwide random pseudo-absences explained 73% of total variance. Components loadings are represented in the interior boxes.

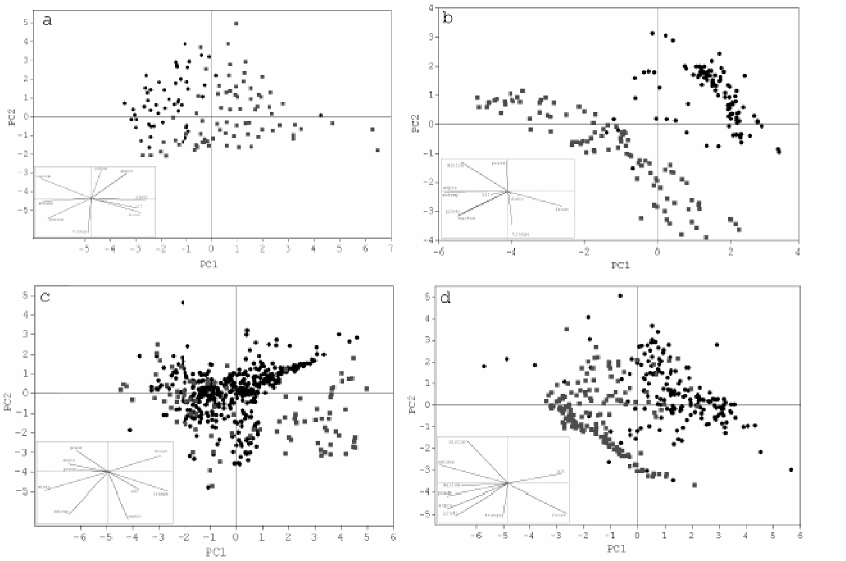


Figure 2 - Position in the environmental space of native (grey squares) and invasive (black dots) populations. A Score plot of the two Components was made from a PCA containing all environmental variables. PC1 and PC2 for (a) *Cherax destructor* explained 69% of total variance, for (b) *Eriocheir sinensis* explained 66%, for (c) *Pacifastacus leniusculus* 77% and (d) *Procambarus clarkii,* 67%. Components loadings are represented in the lower left corner.

Table 3. Validation results for consensus methods using kappa statistic (*k*) and area under the curve of receiver-operating characteristic (ROC-AUC) metrics. It compares for, each species, a weighted average of all single-models *WA(all)*, a weighted average of all single-models using native range occurrences *WA(NIO)*, a weighted average of all single-models using native and invasive occurrences *WA(NIO)*, a weighted average of all single-models with invasive range occurrences *WA(IRO)*, a weighted average of all single-models using native range pseudo-absences *WA(NRA)*, a weighted average of all single-models using native and invasive ranges pseudo-absences random extraction *WA(NIA)* and a weighted average of all single-models using worldwide random pseudo-absences *WA(WRA)*.

		WA(all)		**WA(NRO)**		**WA(NIO)**		**WA(IRO)**		**WA(NRA)**		**WA(NIA)**		**WA(WRA)**	
Species	**Range**	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC
Cherax destructor	Native	0.58	0.72	0.71	0.81	0.57	0.69	0.47	0.57	0	0.15	0.17	0.32	0.9	0.98
	Invaded	0.61	0.73	0.53	0.64	0.66	0.8	0.64	0.74	0.23	0.52	0.27	0.4	0.79	0.96
Eriocheir sinensis	Native	0.64	0.73	0.75	0.87	0.65	0.79	0.34	0.38	0.12	0.49	0.38	0.55	0.62	0.88
	Invaded	0.73	0.85	0.64	0.74	0.79	0.89	0.79	0.9	0.7	0.89	0.71	0.82	0.86	0.97
Pacifastacus leniusculus	Native	0.61	0.73	0.75	0.85	0.59	0.73	0.27	0.56	0.11	0.55	0.53	0.61	0.74	0.94
	Invaded	0.73	0.84	0.58	0.7	0.74	0.88	0.74	0.92	0.77	0.95	0.63	0.76	0.87	0.97
Procambarus clarkii	Native	0.62	0.77	0.8	0.91	0.61	0.7	0.27	0.32	0	0.34	0.12	0.24	0.84	0.98
	Invaded	0.69	0.84	0.55	0.67	0.73	0.88	0.76	0.92	0.67	0.85	0.67	0.79	0.79	0.94

Discussion

In this study we aimed to test different yet plausible calibration data for predicting invasions for four major problematic decapods. Researchers have generally used either data from the species native or invaded ranges (e.g. Welk et al. 2002; Thuiller et al. 2005; Roura-Pascual et al. 2009). More recent studies have argued that due to possible changes in the species niches, coupling both native and invasive occurrence data in the calibration datasets might be the preferable option (Broeninman and Guisan 2008; Beaumont et al. 2009a). In the same way pseudo-absences are also a known source of uncertainty in NBM (Lobo 2008; Phillips 2008), but to which little attention has been given in invasiveness predictions. Further, in invasion biology, NBM are typically applied to wide extents (e.g. continental or global scales), comprising a broad variety of environmental conditions that occur with different spatial frequencies and for which propagule pressure is usually unknown. Our results demonstrate that a large variability in predictive power can arise from the choice of both occurrence and pseudo-absence data.

Invasiveness predictability and the role of calibration data

Niche shifts occur when a species is occupying different environmental conditions in new areas or time periods than the ones found in initial populations. These shifts may be due to changes in the species realized niche (e.g. when a natural competitor is absent in new areas or the species moved to new environmental combinations in the invaded regions) or in its fundamental niche caused by changes in the species physiology (e.g. due to evolutionary change); either may undermine the ability of NBM to predict new suitable areas for invaders (see Pearman et al. 2008 for a review). Our results are consistent with recent arguments that incorporating information from both native and introduced ranges yields the best estimate of the invasion potential (e.g. Broennimann and Guisan 2008; Beaumont et al. 2009a), in that NBM built with both sets of occurrence data was able to simultaneously predict invasions in both ranges well. Thus, if such data are available, it should be used to make future predictions. However, while we should certainly be aware of the effects of possible niche shifts, our results also show that predictions based on native distributions could accurately forecast the invaded areas for three out of four species examined: *C. destructor*; *E. sinensis* and *P. leniusculus*, but not for *P. clarkii*, using the best performing pseudo-absence type (WRA, discussed below). These results suggest that, while caution is warranted, in the absence of information in the invaded range, NBM based on native distributions can still be useful in invasiveness forecasting, especially for forecasting the possible range of very new invasions or potential invasions that have not yet occurred. This is particularly relevant for invasive species modeling, where much of the interest in NBM has been its promise for forecasting invasions into new areas, before they actually occur. Further, despite environmental conditions differing between native and introduced ranges for all species (Figure 2), NBM retained its predictive abilities. We argue that this occurred because NBM was able to identify which environmental variables were important for species establishment, and down weighted those that were not important (Table 7 in the Appendix A).

While the majority of effort has been focused on the type of occurrence data (e.g. Mau-Crimmins et al. 2006; Loo et al. 2007; Broennimann and Guisan 2008; Beaumont et al. 2009a), our results suggest that NBM models are even more sensitive to the type of pseudo-absences used. Native and native and invasive extraction methods attained the lowest performances for all species and types of occurrence data (Tables 1 and 2). Although the ranges sampled by these pseudo-absences likely had the greatest chance of having received propagule pressure over time, this benefit appears to have been outweighed by having restricted environmental conditions which could not be extrapolated to new areas.

Worldwide random pseudo-absences had the stronger predictiveness achieving the highest evaluation scores for all predictions in both ranges (Tables 1 and 2). This result seems supportive of this method. Still, the implications of sampling areas with unequal or unknown propagule pressure require further research. Consensus methods showed poor to high performances and could be a good alternative to deal with the wide array of possible outcomes. Still, their best performances in terms of predictiveness was no better than the best fitting single model – native and invasive occurrences with worldwide random pseudo-absences. While applying consensus methods for models obtained by distinct algorithms (e.g. regressions, classification models or machine learning) has been shown to improve predictive accuracy (Marmion et al. 2009; Stohlgren et al. 2010), we did not find obvious improvements for combining distinct calibration data. Still, the use of these ensembles may prove useful for cases where it is not possible to identify a single best performing model. Further, as we visually assessed in our predictions, single-models with similar accuracies may provide different spatial patterns of predictions. In such cases ensembles may also be of use by providing a consensual spatial pattern or model uncertainty.

Species invasiveness potential and conservation concerns

We used the best performing models in both native and introduced ranges to analyze their invasiveness potential. We found that each of the studied species still has large extents of suitable areas unoccupied (Figure 3). This result is of high conservation concern, since the direct impact of these species in biological diversity is known to be high. Conservation problems caused by these decapods have been reported including disease transmission, competition, and active predation of native species and changes in the trophic webs of the invaded ecosystems (Gutierrez-Yurrita et al. 1999; Lynas et al. 2004; Cruz et al. 2008; Correia and Anastácio 2008; Dittel and Epifanio 2009). For example, Cruz et al. (2008) related the strong decline of both abundance and diversity of amphibian populations in a Portuguese protected wetland with the establishment of *P. clarkii*. These authors found that since the initial establishment of this invader, the number of amphibian species in the area was reduced from 13 to 6. It is thus worrying that the four species show suitable unoccupied areas that surpass the extents of the currently known invasive ranges (Figure 4). Using the suitability threshold achieving higher kappa value to discretize predictions we found that higher suitability areas outside the species current ranges (both native and invaded) are nearly three times larger for *P. leniusculus*, four times for *P. clarkii*, seven times for *E. sinensis* and about nine times for *C. destructor*. Although also dependent on propagule pressure, these values indicate a large invasive potential for these species and detailed predictions should be conducted in areas of interest. Suitability for *C. destructor* and *P. clarkii* are noticeably similar (Figure 3). This is a clear reflection of the similarity between the ecological preferences of these two species (Nyström 2002). Several biodiversity hotspots

(Myers et al. 2000) fall under their environmental requirements. The Mediterranean basin and southwest Australia are of special concern for *P. clarkii* and *C. destructor* respectively. The Mediterranean basin while being highly suitable for *P. clarkii* (Figure 3) also encompasses the majority of *P. clarkii* invasive range in Europe (mostly found in the Iberian Peninsula). It is thus worrying that new nonadjacent invasions are also taking place here, such as in the Nile River (Cumberlidge 2009). Under this context the Mediterranean region may be particularly important because it contains a largely endemic biota and there is a great potential for invasion due to the high propagule pressure and environmental suitability. Likewise the currently largest invaded area for *C. destructor* occurs mostly within Western Australia – an area of high biological diversity – and its impact on the endemic crayfish species here is a concern (Lynas et al. 2004).

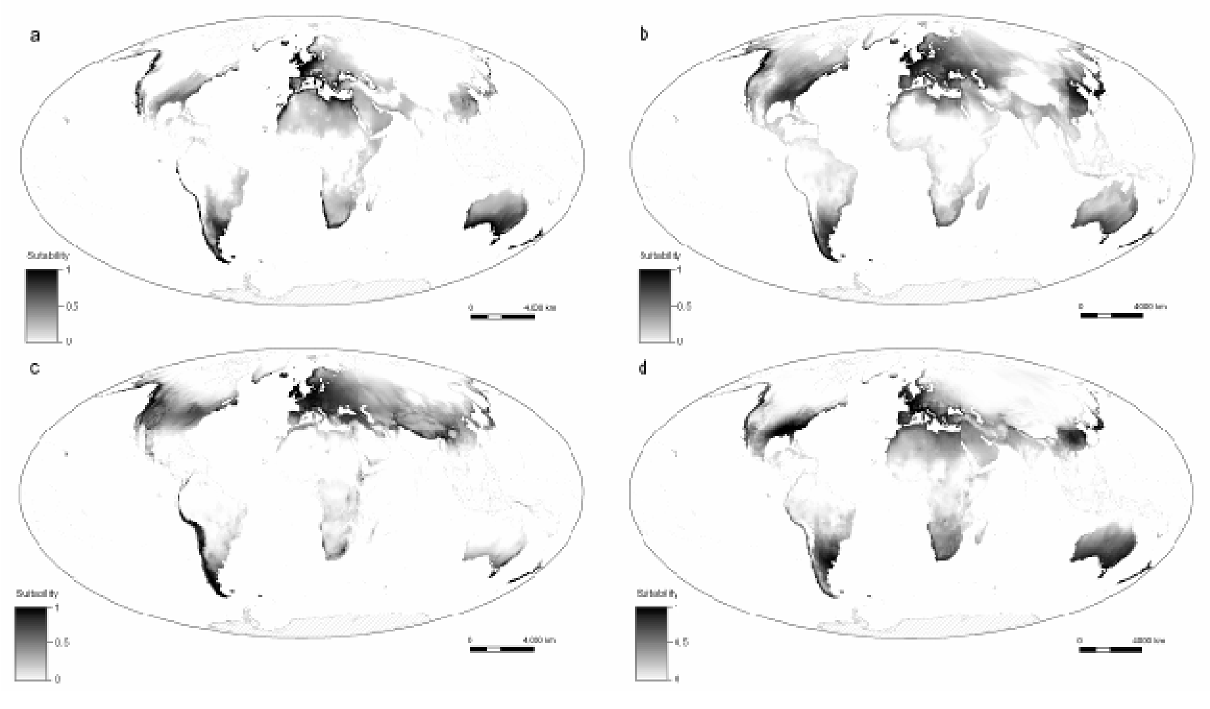


Figure 3 - Best performing suitability models in both native and invaded areas for (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii*.

It is also worth noting that for these two species some high latitude areas such as several southern regions of Iceland and Greenland, Aleutian Islands and the southern tip of South America appear as suitable for this warm-water species. These areas, despite presenting colder temperatures than the ones verified in the majority of their distribution ranges, are under influence of the oceans moderating effect and their mean minimum temperatures reach fairly higher values than many other areas within the same latitudinal ranges. Although low temperatures have a known influence on some biological traits of *P. clarkii* (for a compilation see Anastácio et al. 1999) and *C. destructor* (Semple et al. 1995), these relatively unsuspicious areas are possibly on the edge of the thermal regimes required for these species to persist. Another possibility is that these values result from extrapolation errors. Whilst new methods for assessing that possibility exist (Elith et al. 2010) this evaluation would require a different modeling framework, beyond the scope of this work.

Also *P. leniusculus* show a large potential for future invasions. Despite already being the most widespread invasive crayfish in Europe, its invasion here seems to still be far from finished. The invasiveness projected for this continent shows a wide extent of suitable, yet uninvaded, areas mostly in Eastern Europe, the Balkans and Turkey (Figure 3). This potential expansion is of particular concern for native crayfish populations. Mostly due to competition and its role on transmission of the crayfish plague (*Aphanomyces astaci*), *P. leniusculus* is considered to have a major role in the extirpation of native crayfish populations (Souty-Grosset et al. 2006). Thus, crayfish conservationists should be aware that this species still has a large extent of environmentally suitable areas in contiguity to the existing invasive populations.

For *E. sinensis* five large suitable areas emerge as particularly vulnerable to new large-scale invasions: surroundings of Black and Caspian Seas, the Mediterranean basin (especially on the European side), and both eastern and western North American coasts (Figure 3). In fact, besides presenting high suitability, small populations or isolated individuals have already been reported here (Dittel and Epifanio 2009). In contrast, it is not clear why *E. sinensis* has not been found in the Southern Hemisphere. Our results also show that there are considerable extents of suitable environments in the southern areas of Australia, Africa and South America (Figure 3). Ballast waters are the main vector of introduction of this species and some of these areas also have extensive international shipping connections. We therefore recommend that attention should be given to the possible existence of unknown populations or new introductions of *E. sinensis* in these areas since its establishment potential is high.

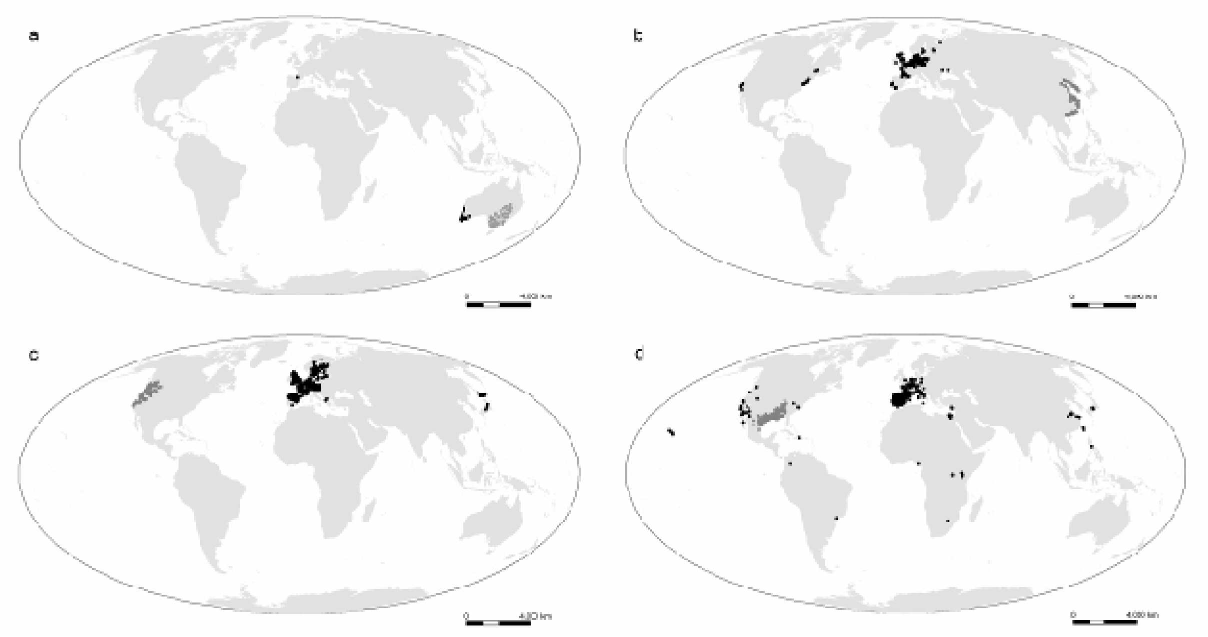


Figure 4 - Native (grey circles) and invasive occurrences (black circles) of (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* used for model calibration.

Caveats and future directions

While pseudo-absences are of current use in NBM, a large number of its implications remain poorly understood. We compared three distinct types of extraction methods and found substantial differences in accuracy. Still, more detailed comparisons should be addressed in the future. More specifically, different evaluation dataset types should also be tested. Our predictive evaluation relied on the use of worldwide pseudo-absences and, while to our mind this allows to test predictions for the entire study area, it may also provide overoptimistic scores for models calibrated with the same type of information. Further, increasing the number of invaders tested and more robust distribution datasets would also potentially increase the confidence of the obtained results. A comparison with other extraction approaches (and their paradigms) addressing issues such as sampling bias on the occurrence data (e.g. Philips et al. 2009) or the use of presence-only NBM for driving extraction (e.g. Chefaoui and Lobo 2008) would also be highly valuable future directions, although they can require additional information that may not be available (e.g., some metric of sampling intensity or bias, Philips et al. 2009).

Future work should also consider the effects that known issues of NBM such as spatial autocorrelation have in its predictive performance. While previous studies have found that simple spatial autocorrelation models can perform as well as NBM (Bahn and McGill 2007), this only indicates that we cannot distinguish between habitat suitability and spatial autocorrelation. However, because we were able to extrapolate predictions from the native range to the exotic range, we were able to show predictive power in new spatially uncorrelated areas. Still, additional analyses on this topic are warranted. Other issues known to affect NBM such as unequal propagule pressure worldwide or potential lacks in our occurrence data (e.g. due to incomplete knowledge about the species invasive range) are of importance to our study and should also be explored in the future. One possibility to address the issues of unequal propagule pressure and spatial autocorrelation is to include a propagule pressure model (which provides a mechanistic model of spatial autocorrelation) together with our NBM (e.g. Leung and Mandrak 2007). Finally, obtained potential distributions for our invaders would also benefit from the availability of distribution data with higher spatial accuracy. Such data would allow the use of more detailed environmental data and thus increasing the detail of predictions which would also increase their value for managing purposes.

Despite the previous caveats, we present the first description of the worldwide invasiveness potential for four important invasive decapods. We also determined that the obtained projections are highly dependent on the type of data used for model calibration. While the type of occurrence data used could be important, accurate predictions were still obtained based solely on the occurrences in the native range for three of the four species examined. Moreover, our results suggest that pseudo-absences extraction methods were even more influential than the type of occurrence data used. Finally, despite its good results in other situations, the use of consensus models had limited benefit, with the best single modeling approach achieving consistently higher accuracies. Our results are supportive that worldwide predictions of invasiveness should be based on both native and invasive data, when available, and that worldwide random pseudo-absences seems the more favourable option if real absence information is lacking.

Acknowledgements

We thank the many researchers and institutions that provided species distribution data and three anonymous reviewers and the Subject Editor for their helpful comments.This research was performed within the scope of the DID (Dispersal of Invasive Decapoda) (PTDC/BIA-BEC/105182/2008) research project funded by the Portuguese Fundação para a Ciência e Tecnologia (FCT) and co-financed by FEDER, trough the Eixo I of the Programa Operacional Factores de Competitividade (POFC) from QREN and a research grant provided to C. Capinha (SFRH / BD / 41129 / 2007) (FCT).

Page intentionally left blank

CHAPTER 3

Geographic variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish

This chapter is submitted to an international peer-reviewed journal:

Capinha C, Brotons L, Anastácio P (in review) Geographic variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish

GEOGRAPHIC VARIABILITY IN PROPAGULE PRESSURE AND CLIMATIC SUITABILITY EXPLAIN THE EUROPEAN DISTRIBUTION OF TWO HIGHLY INVASIVE CRAYFISH

César Capinha[1]; Lluís Brotons[2] and Pedro Anastácio[1]

[1]IMAR, Centro de Mar e Ambiente c/o Departamento de Paisagem, Ambiente e Ordenamento, Universidade de Évora, Rua Romão Ramalho, n.º 59, 7000-671 Évora, Portugal.

[2]Centre de Recerca Ecològica i Aplicacions Forestals (CREAF) and Àrea de Biodiversitat, Centre Tecnològic Forestal de Catalunya (CTFC), Ctra. Sant Llorenç de Morunys km 2, E-25280 Solsona, Spain.

Abstract

Aim: In this paper, we assess the role of human, biological and climatic factors in determining the colonisation success of two highly invasive freshwater decapods, the Signal crayfish and the Red swamp crayfish.

Location: Europe

Methods: We used boosted regression trees to evaluate the relative influence and relationship between the invader's current pattern of distribution and a set of spatially explicit variables considered influential to their colonisation success. These variables are related to four well-known invasion hypotheses: propagule pressure, climate matching, biotic resistance from known competitors and human disturbance.

Results: Model predictions attained a high accuracy for the two invaders (mean ROC-AUC ≥ 0.91). Propagule pressure and climatic suitability were identified as the primary drivers of colonisation but the former had a much higher relative influence for the Red swamp crayfish. Climate matching was shown to have limited predictive success since climatic suitability models based on occurrences from other invaded areas attained a consistently higher relative explanatory power. Biotic resistance and human disturbance were also shown to be of reduced importance in shaping the distribution pattern of the two invaders.

Main conclusions: The results are important for advancing the general knowledge of the factors that enable certain species to become notable invaders. Being primarily driven by propagule pressure and climatic suitability, we expect that, given their continued dispersal, the future distribution of these problematic decapods in Europe will increasingly become a representation of their fundamental climatic niche.

Keywords: *Biotic resistance; Climate matching; Europe; Freshwater invasions; Human disturbance; Propagule pressure, Red swamp crayfish; Signal crayfish*

Introduction

As a result of human activity, species are now being introduced to new areas at unprecedented rates (Ricciardi 2007). These events have led to the establishment of many highly invasive species with often dramatic effects in the recipient regions (Clavero et al. 2009; McGeoch et al. 2010). Nonetheless, it is also well-acknowledged that the large majority of introductions fail to produce viable populations and, from those which succeed, only a small portion become invasive (Williamson 1996). Given this apparent idiosyncrasy, unravelling the factors that determine invasion success and the prediction of future impacts has long been a central issue of invasion ecology and conservation biology (Kolar and Lodge 2001; Van Kleunen et al. 2010).

Understanding the causes behind the establishment success of invaders is a complex task and efforts towards that end have led to the development of many hypotheses. Among these, the propagule pressure hypothesis (Lockwood et al. 2005) is one of the most widely acknowledged. Higher propagule pressure increases the population’s ability to overcome Allee effects (Drake and Lodge 2006), endure environmental and demographic stochasticity (Simberloff 2009), and adapt to local conditions (due to a wider gene pool) (Ahlroth et al. 2003). Thus, a positive relationship between propagule pressure and the likelihood of invader establishment is expected. Another common expectation is that invaders will more likely establish in areas with climates similar to that of their native range. This assumption underlies the ‘climate matching’ hypothesis (Williamson 1996) which has received strong support in recent years (Wiens and Graham 2005; Peterson 2011). Biotic interactions are also expected to play an important role in the invasion process. The biotic resistance hypothesis is a prime example of this expectation, predicting that areas with higher competitor diversity will have a lower probability of invasion (Stachowicz and Tilman 2005). This will probably be especially relevant in the case of taxonomically similar species (e.g. DeRivera et al. 2005). Ultimately, anthropogenic disturbances can change the abiotic and/or biotic factors of an area, leading to an increase in vulnerability to invasion (see Catford et al. 2009 and references therein). This ‘human disturbance’ hypothesis is based on the potential role of facilitative mechanisms, such as the release of resources, the alteration of biotic interactions and the creation of vacant space (Lockwood et al. 2007). Besides these four main hypotheses, several others – not necessarily mutually exclusive and encompassing a relatively vast array of biotic, abiotic or human factors – have also been proposed to explain the establishment success or failure of non-native species (see Catford et al. 2009).

Although the value of individual invasion hypotheses is usually supported by a number of studies, increasing evidence is accumulating that their relative importance is highly context-dependent (Catford et al. 2009; Gurevitch et al. 2011). For example, recently, Nuñez et al. (2011) analysed an 80-year dataset of tree introductions in an Argentinean national park and found that detailed estimates of propagule pressure provided little predictability on the current patterns of invasion of the area. They suggest that other factors, such as biotic resistance, may be playing a more important role and thus reducing the invasibility of the area. This influence of 'external' factors has led to suggestions that future research should focus more on untangling the relative role of the distinct factors involved in the process instead of simply testing the validity of isolated hypotheses (e.g. Thuiller et al. 2006; Pyšek et al. 2010). However, despite this recognition, few studies have thus far provided insight into the interaction of multiple factors during an invasion process. Additionally, most existing studies have focused on finding generalities among a large number of invasions (e.g. Thuiller et al. 2006; Pyšek et al. 2010) and therefore an acute knowledge of the mechanisms driving many of the most notable invaders is still lacking.

In this study, we focus on two of the most widespread and harmful invaders so far reported in European freshwaters, the Signal crayfish (*Pacifastacus leniusculus* Dana) and the Red swamp crayfish (*Procambarus clarkii* Girard). Previous studies referring to the establishment of these invaders have identified a number of mechanisms explicitly related to the four abovementioned hypotheses (propagule pressure, climate matching, biotic resistance and human disturbance). For instance, they refer to the existence of distinct efforts of introduction among countries (Souty-Grosset et al. 2006; Holdich et al. 2009), high tolerance to environmental heterogeneity (Geiger et al. 2005), competitive interactions with other co-occurring crayfish species (Gherardi and Cioni 2004; Dunn et al. 2009) and the positive impact of human activities on the environment (Lindqvist and Huner 1999). However, and although commonly acknowledged, the way these mechanisms interact to determine their colonisation success or failure remains unknown, particularly on a continental scale.

We examine how propagule pressure, climatic suitability, biotic resistance and human disturbance interact to shape the current distribution of the Signal and the Red swamp crayfish in Europe. Specifically, we relate the spatial variability of these factors to the pattern of distribution of the invaders using an ensemble statistical model that assesses the relative contribution of each factor and its relationship with the species' probability of occurrence. The results of this analysis are intended: (1) to provide the first quantitative evaluation of the relative role of distinct human, climatic and biological factors in explaining the pattern of distribution of these two notorious invaders; (2) to compare the relative role and type of relationships between the tested factors and the species' occurrence to the postulations of the four hypotheses described above; and (3) to provide insight into how the distribution of these invaders will progress on the European continent.

Methods

Species and distribution data

The Signal crayfish and the Red swamp crayfish are both native to North America and have been imported to Europe since the 1960s and the 1970s respectively. Currently, the Signal crayfish is the most widespread invasive crayfish on this continent, having been found in 27 countries, while the Red swamp crayfish occurs in 10 (Holdich et al. 2009). These species are also invasive in other parts of the world. Non-native populations of Signal crayfish are found in Japan and the southwest United States, while the Red swamp crayfish, which has a much broader invasive distribution, is established in several areas of Africa, Asia, North America and South America (Souty-Grosset et al. 2006).

For the purpose of this study, we defined our study area as the countries represented in the Atlas of Crayfish in Europe (Souty-Grosset et al. 2006), which provided the bulk of the distribution data for Europe. From these, we excluded those referred to as strongly under-sampled (Figure 1). We then collected current native and invasive occurrences of the Signal crayfish and the Red swamp crayfish worldwide from the Atlas, natural history museum collections, the Global Biodiversity Information Facility (GBIF; http://data.gbif.org) and recent literature. From these sources, we also obtained current European occurrences of six other crayfish species known to establish competitive interactions with either of the two invaders (see list of species in Table 1 in the Appendix B). All distribution data was re-sampled to a 50 × 50 km UTM grid, and allowing for only one record per grid square, we accumulated 597 records for the Signal crayfish (454 in Europe, 52 in invaded areas outside Europe, and 91 in the native range; Figure 1a; Figure 1 in Appendix B), 631 for the Red swamp crayfish (377 in Europe, 80 in invaded areas outside Europe, and 174 in the native range; Figure 1b; Figure 2 in Appendix B) and 2 587 for the crayfish species identified as competitors.

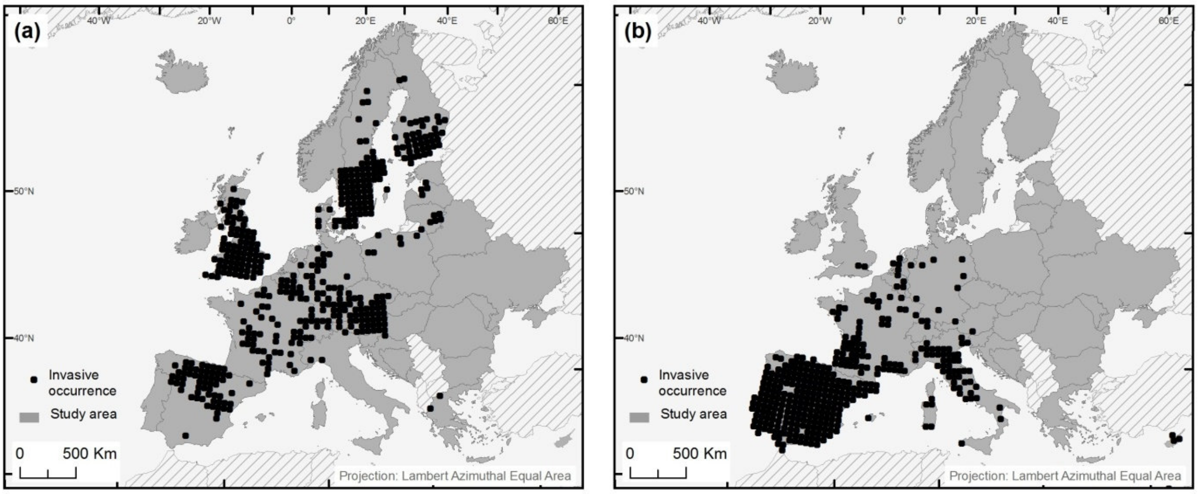


Figure 1 - Current distribution of (a) the Signal crayfish and (b) the Red swamp crayfish in the study area at a 50 × 50 km cell resolution.

Explanatory variables

Both the Signal crayfish and the Red swamp crayfish were initially introduced into Europe for the purpose of being harvested in the wild or in aquaculture facilities, since they were a valued food item. The success achieved by most soon motivated a large number of introductions into the wild in many new regions and countries. Currently, this is considered to be their major dispersal pathway through Europe (Hogger 1986; Gutiérrez-Yurrita et al. 1999; Gherardi 2006; Souty-Grosset et al. 2006). To obtain estimates of propagule pressure, we consulted a comprehensive collection of literature reporting on the history of introduction and human-mediated dispersal of the Signal crayfish and the Red swamp crayfish in Europe. Based on this information, using GIS, we categorised each country into one of three categories of introduction effort: (1) countries where the introduction of the species into the wild was promoted by governmental entities and/or commercial associations, and introductions were reported for more than one region (‘strong’ category); (2) countries where the introduction of the species into the wild was not supported by organised actions or was reported for only one region (‘moderate’ category) and (3) countries for which no attempt at introducing the species for harvesting into the wild is reported (‘null’ category).

Climatic suitability models (Guisan and Thuiller 2005) were used as a measure of suitability of climatic conditions. To build these models, we collected five non-redundant (pairwise Pearson correlations ≤ ± 0.8) climatic variables from the WorldClim project (Hijmans et al. 2005): mean diurnal temperature range; mean temperature of warmest quarter; mean temperature of coldest quarter; precipitation of the wettest quarter; and precipitation of the driest quarter. To be compatible with the spatial accuracy of our distribution data, these variables were re-sampled from their original resolution (10 × 10 km) to a 50 × 50 km cell resolution.

To quantify the potential effect of biotic resistance, we used the distribution data of competing crayfish species to calculate, for each invader, the richness of co-occurring competitors. This was done in GIS and consisted simply of the sum of all competitors occurring in each UTM square comprising the study area.

To represent human disturbance, we used the human footprint index (Sanderson et al. 2002). This index merges spatial data concerning population density, land transformation, accessibility and electrical power infrastructure to infer the degree of human influence on the land surface, and is available at the global scale using a 1 × 1 km resolution. To include this variable in our analysis, we calculated the mean value within each of the UTM squares comprising the study area.

Three distinct climatic suitability models were built for each of the two invaders based on: (1) native range occurrences; (2) invasive occurrences outside the study area; and (3) these two ranges combined. To calibrate these models, we coupled the respective occurrences with pseudo-absences randomly generated worldwide, excluding areas with the species occurrence (cf. Capinha et al. 2011). To maximise the pseudo-absences' representation of the available climatic conditions, we created 15 replicate datasets for each model. Each of these had the species occurrence records coupled with an equal number of independently drawn pseudo-absences.

Statistical analyses

Climatic suitability models can be built using a large variety of methods; however, there is still little agreement on which one delivers the 'best' results (e.g. Elith et al. 2006). As an alternative, an ensemble of distinct but equally plausible models can be produced and their array of results analysed for agreement (Araújo and New 2007; Capinha and Anastácio 2011). Accordingly, we used three distinct statistical models to predict climatic suitability to the invaders: boosted regression trees (BRT), generalised additive models (GAM) and random forests (RF). These statistical models are amongst the most frequently used in niche modelling exercises where they have consistently revealed high predictive ability (e.g. Elith et al. 2006; Capinha and Anastácio 2011). Models were implemented in R-CRAN (R Development Core Team 2011) using the BIOMOD package (Thuiller et al. 2009). Model parameters were kept at default values, with the exception of the number of trees in BRT which was set to 2 000, and the degrees of freedom of the GAM smooth functions which was set to four. The discrimination ability of each replicate model was evaluated in terms of the area under the receiver operating curve (AUC). The AUC was calculated by comparing predictions with 20% of the distribution data that was withheld from model development. Final models of climatic suitability were obtained by averaging the 45 replicate predictions (15 replicates × 3 modelling algorithms) into a single prediction.

In order to prevent data redundancy in the explanatory variables, we measured the degree of correlation between all pairs of independent continuous variables collected, which revealed the inexistence of high correlation levels ($|r| < 0.8$). Next, we created three distinct sets of explanatory variables for each of the two invaders. Each of these had one of the three climatic suitability models made for each species, together with the remaining explanatory variables: country-level introduction effort; richness of co-occurring competitors; and human footprint (Figures 3 and 4 in Appendix B). To analyse the influence of these explanatory variables on the presence or absence of the invaders, we used boosted regression trees (BRT) (Friedman et al. 2000; Friedman 2001; Elith et al. 2008). BRT consists of a machine-learning technique that builds an ensemble prediction by iteratively fitting a chain of multiple decision trees. This method has shown a consistently high predictive ability in several studies (e.g. Elith et al. 2006; Guisan et al. 2007), and is also particularly appropriate in exploring the combined influence of a set of predictors due to its ability to fit non-linear responses and automatic detection of interaction effects between independent variables (Elith et al. 2008). We used the gbm package version 1.6-3.1 (Ridgway 2010) for R-CRAN and the R functions provided by Elith et al. (2008) to run all BRT models. Before analysing the datasets, we tested for the optimal parameters of the BRT models by running multiple models of varying complexity and comparing their predictive accuracy with data set aside from model calibration. For that purpose, five different learning rates (0.05, 0.025, 0.01, 0.005 and 0.0001) and six levels of tree complexity (1, 3, 6, 9, 12 and 15) were tested. Additionally, the optimal number of trees for each combination was automatically assessed using the R code provided by Elith et al. (2008). For each combination of parameters, we ran 100 replicate models as a way of dealing with the variability of BRT. The performance of each of these was tested using a 10-fold cross validation and measured using the AUC. The mean predictive performance across the 100 replicates was calculated and, from the replicates of the best performing combination, we extracted the mean relative influence of each explanatory variable. This relative influence is automatically calculated by the gbm package by relating the number of times a variable is used with the improvement it generates in the model (Friedman and Meulman 2003).

We were also interested in assessing how the probability of occurrence of each invader changes along the gradient of each predictor. For that purpose, we extracted partial dependence plots, which show the marginal effect of each predictor in the species' probability of occurrence (Freidman 2001; Friedman and Meulman 2003). Partial dependencies were evaluated by calculating the mean response of a model to a single value of the predictor of interest while keeping the remaining predictors at their original values. This procedure was repeated with new values until a representation of how the mean response varies along the gradient of the predictor of interest was obtained. By taking the mean response, the effect of the other predictors entering the model was removed to some extent, which allowed for a representation of the effect of that single predictor. Partial dependence plots are sensitive to strong correlations and interactions among predictors. While the former were assessed prior to modelling (see above), the strength of interactions among pairs of predictors was evaluated through exploratory analyses in R and the nature of the relationships was visually assessed using joint partial dependence plots (cf. Elith et al. 2008).

Results

The mean AUC of all climatic suitability models achieved fair to good accuracy scores. For the Signal crayfish, this accuracy ranged from 0.82 (± 0.04) for models based on invasive occurrences, to 0.84 for the models based on native occurrences (± 0.02) and on both occurrences (± 0.03). For the Red swamp crayfish, the mean AUC of models based solely on invasive occurrences was 0.90 (± 0.02), for models using only invasive occurrence, 0.93 (± 0.03) and, for models using both occurrences, 0.91 (± 0.03).

All boosted regression tree models attained a good to very good predictive ability (mean AUC ≥ 0.91) (Table 1) which is shown to be obtained through a combined effect of multiple factors (Figure 2). All models showed strong (> 5%) pairwise interactions between introduction effort and climatic suitability, while the models for the Signal crayfish also showed strong interactions between propagule pressure and richness of co-occurring competitors. However, visual analyses of joint partial dependence plots between these pairs of predictors allowed for the verification of the inexistence of contradictory effects of these interactions on the general responses shown by the partial dependence plots for single predictors.

Table 1 - Accuracy of boosted regression trees models in predicting the current pattern of presence/absence of the Red swamp crayfish and the Signal crayfish in Europe, measured by the area under the curve of the receiver operating characteristic (AUC). Three groups of models were evaluated using predictions of climatic suitability based on: (1) the species' native range (native); (2) the species' invasive range outside Europe (invasive); and (3) both ranges simultaneously (both). Mean values were calculated from a set of 100 replicate models. Standard deviation of mean AUC values was < 0.01 for all models.

Range	**Red swamp crayfish**	**Signal crayfish**
native	0.95	0.92
invasive	0.97	0.91
both	0.97	0.93

The models for the Signal crayfish identified climatic suitability as the most important predictor (≥ 35.6 %), followed by introduction effort and richness of co-occurring competitors (≥ 32.4 and ≥ 17.9% respectively) (Figure 2a). Human footprint receives the lowest relative contribution (≤ 11.9%). Also noticeable was a slightly higher relative influence of climatic suitability models using information from invasive ranges. A combined analysis of these results and the relationships shown by the partial dependent plots (Figure 3a; Figures 5 and 6 in Appendix B) indicates that the invasive populations of the Signal crayfish are found mainly in areas with suitable climate, predominantly within countries that had strong levels of introduction effort and, interestingly, where a large number of competing species occurs. A (slight) tendency to avoid highly disturbed areas is also suggested.

The current pattern of distribution of the Red swamp crayfish is shown to be largely explained by country-level introduction effort (≥ 63.5%), followed by climatic suitability (≥ 13.6%). Concerning this last predictor, a significant increase in its explanatory power was found when the invaded areas outside Europe were used to represent the species' climatic niche. Models based on the species' native range achieved a mean relative influence of 13.5% (± 0.5) whereas those based on the invasive range outside the study area achieved 19.6% (± 0.6) and, those based on both ranges, 20.5% (± 0.6). The combined analysis of the relative influence of the explanatory variables and their marginal effect on the species' probability of occurrence (Figure 3b; Figures 6 and 7 in Appendix B) indicates the Red swamp crayfish is found mainly in the climatically suitable areas of countries with strong to moderate introduction effort. In addition, the occurrence of this invader also seems to be slightly favoured by high levels of human disturbance.

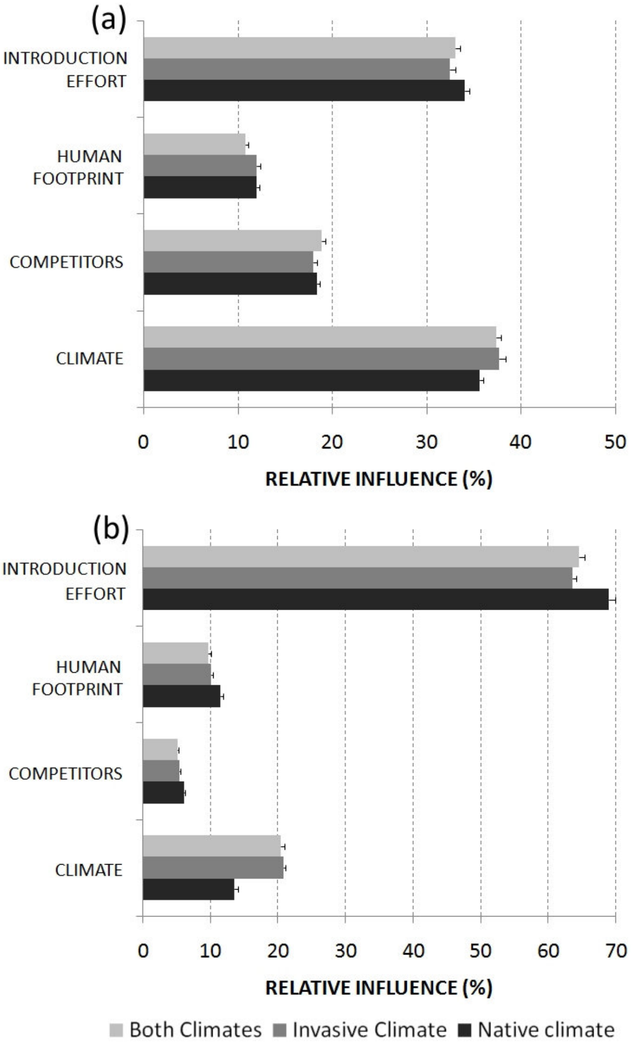


Figure 2 - Mean relative influence of explanatory variables for (a) the Signal crayfish and (b) the Red swamp crayfish. Grey shades of the bars indicate the type of climatic suitability model used in model calibration based on: the species native range (dark grey); the invaded areas outside the study area (mid grey); and these two ranges combined (light grey). Mean values were calculated from a set of 100 replicate models that were built with each dataset.

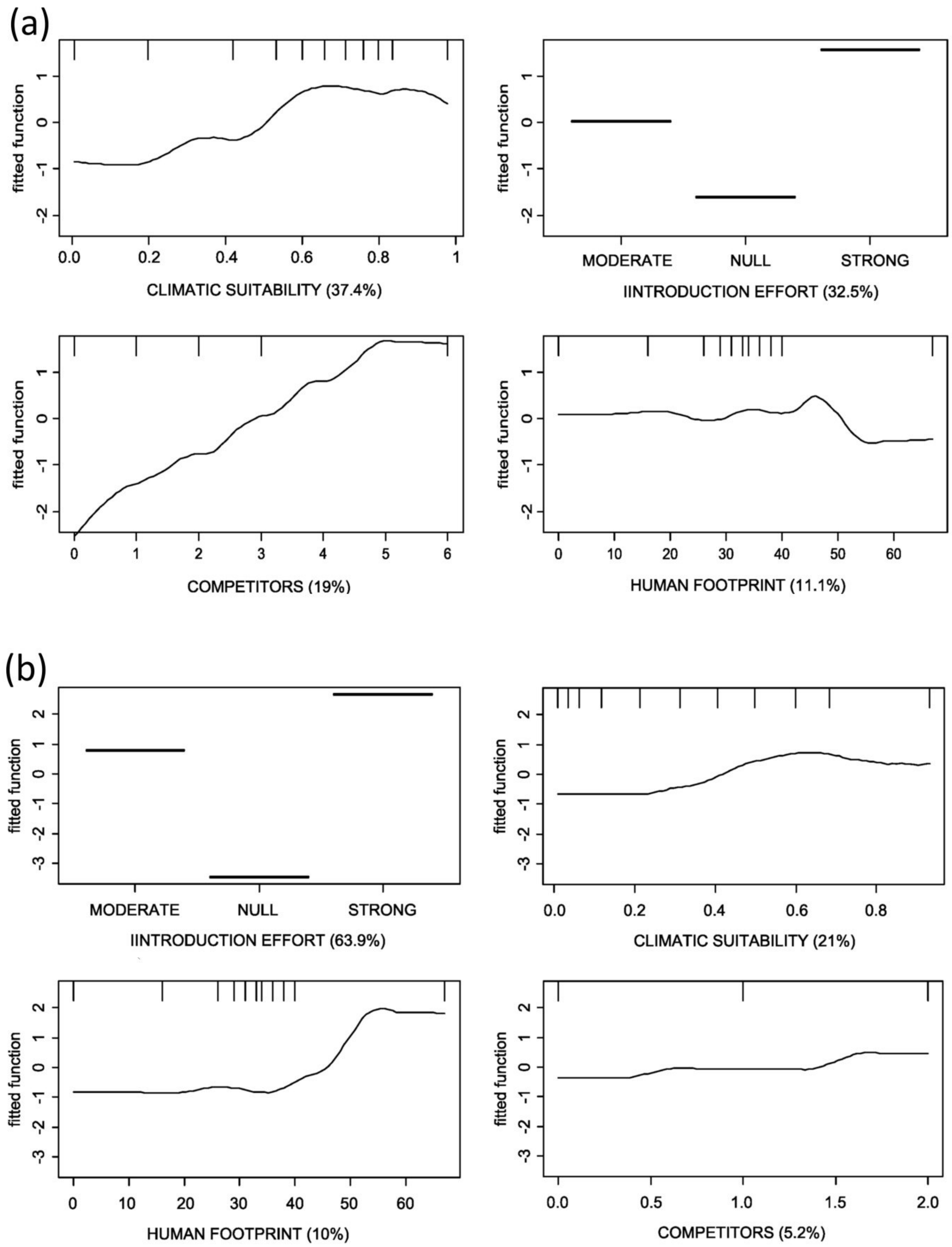


Figure 3 - Smoothed partial dependence plots of one representative boosted regression tree model for (a) the Signal crayfish and (b) the Red swamp crayfish. Climatic suitability models were calibrated with species occurrences from native ranges and invaded regions other than the study area. The relative contribution of each predictor in this specific model is shown in parentheses.

Discussion

The results of our analyses demonstrate that the current distribution of the Signal crayfish and the Red swamp crayfish in Europe can be largely predicted by an interaction of country-level introduction effort, suitability of climatic conditions and co-occurrence with competitors (the latter only for the Signal crayfish). Complementarily, they also highlight a reduced predictive power of human disturbance for both invaders and an almost null contribution of co-occurrence with competitors for the Red swamp crayfish.

Drivers of colonisation success

The high relative importance of country-level estimates of propagule pressure in determining current patterns of distribution is relatively unsurprising for these invaders. Being of commercial interest, these species were actively imported and intentionally introduced into the wild in several European countries while, in many others, they raised little commercial interest or their introduction into the wild was simply prohibited. For instance, the Signal crayfish is widely spread over Sweden, where vigorous campaigns of introduction have taken place since the 1960s (Henttonen and Hunner 1999), but only one invasive population is known for Norway (Johnsen et al. 2007), a bordering country which banned the importation of live-specimens of this species before any wild population was known (Holdich and Pöckl 2005). However, it is interesting to identify marked differences in the importance of this factor between the two invaders – noticeably higher for the Red swamp crayfish. We believe that differences in residence time and the extent of areas receiving propagules are the most likely explanations for this result. These attributes largely mediate the degree of distributional equilibrium in invasive ranges and consequently the relative importance of other factors shaping the realised distribution (e.g. Wilson et al. 2007). In the case of these two invaders, the Signal crayfish had a longer and much wider dispersal in the study area (Gherardi 2006; Souty-Grosset et al. 2006; Holdich et al. 2009) which also, considering the much lower relative importance of propagule pressure in its models, suggests that its current invasive distribution is closer to equilibrium than that of the Red Swamp crayfish.

Our results also highlight the high influence of climate in shaping the current distribution of the two invaders. Climate, either directly or indirectly, affects the suitability of abiotic conditions and it is not surprising that it plays a large role in shaping invasive distributions (e.g. Roura-Pascual et al. 2011). However, we found limited support for the 'climate matching' and related invasion hypotheses which predict that invasions take place in environmental conditions for which the invaders are pre-adapted (e.g. the 'habitat filtering' hypothesis; Catford et al. 2009). This is because the models calibrated with climatic data from other invaded areas were consistently more useful in predicting the species' European distribution than those based on data from native ranges – particularly for the Red swamp crayfish. This result also aligns with previous findings that these invaders occupy new climatic conditions (i.e. shifted their climatic niche) in new ranges worldwide (Capinha et al. 2011; Larson and Olden 2012) and thus suggest caution when using native-based correlative models for predicting climatic suitability for invaders. Several non-mutually exclusive ecological or evolutionary mechanisms may be behind this capacity. Ecological mechanisms include the absence or reduction of natural enemies in the new range (i.e. the 'enemy release' and 'enemy reduction' hypotheses; Colautti et al. 2004) or simply that these newly occupied conditions are not found in native regions. That said, more information about the ecological constraints of native populations would be necessary in order to test these possibilities. Rapid evolutionary adaptation to the new conditions is another possible mechanism given that these invaders have relatively short generation times and the variability of the European genetic pool is favoured by the importation of large numbers of specimens from distinct origins (Gherardi 2006; Souty-Grosset et al. 2006). However, the instant success of non-European specimens of Red swamp crayfish in colonising separate European regions with new climatic conditions (e.g. England, The Netherlands, Germany or Switzerland; Henttonen and Huner 1999) seems to refute the necessity of this mechanism in the invasion of this species.

Despite the great amount of evidence of strong biotic interactions taking place with other crayfish species occurring in Europe (e.g. reproductive interference or competition for food and shelter; Lindqvist and Huner 1999; Westman et al. 2002; Gherardi 2006), we found no evidence of biotic resistance as an impediment to the establishment of these two invaders. Two possibly cooperating mechanisms may account for this result. First, the superior competitive ability of these invaders (Lindqvist and Huner 1999; Westman et al. 2002; Gherardi 2006) which could be innate due to the specific traits of the invaders or have been promoted by their release from other biotic constraints (i.e. the 'evolution of increased competitive ability' hypothesis; Blossey and Notzold 1995). Second, these invaders are also vectors and partially resistant to the 'crayfish plague' (*Aphanomyces astaci*), a disease that is deadly to European crayfish species (Lindqvist and Huner 1999) – a mechanism clearly fitting within the 'enemy of my enemy' hypothesis (Colautti et al. 2004). Somewhat surprisingly, we also found that the occurrence of the Signal crayfish is strongly associated with the presence of known competitors. We hypothesise that this could be a geographic signal of propagule pressure since this invader was continuously introduced as a replacement for native species – with which it still co-occurs in many places (Gutiérrez-Yurrita et al. 1999; Skurdal et al. 1999) – and also because it shares some of the pathways of introduction of other exotic competitors (e.g. harvesting in the wild; Souty-Grosset et al. 2006; Holdich et al. 2009).

The occurrence of these two invaders was also found to be only marginally correlated with the human footprint index, suggesting that human disturbance has a reduced contribution in shaping their distribution. Nonetheless, this influence becomes stronger in highly disturbed areas, which seem to affect the Signal crayfish negatively and the Red swamp crayfish positively. Despite 'human disturbance' sometimes being referred as a potentially important factor for these invaders (e.g. Lindqvist and Huner 1999), there is little empirical evidence with which to compare our results. In addition, given the relatively low importance of this factor, it is also difficult to exclude potential confounding effects of propagule pressure. This is because these two invaders have been highly stocked for harvesting in natural to semi-natural areas across Europe. However, the Red swamp crayfish is also common as an aquarium pet or in garden ponds which may explain why it has a higher probability of occurrence in highly urbanised (i.e. disturbed) areas (Henttonen and Huner 1999). These pathways of introduction could be behind the associations found but more detailed data would be necessary to test this possibility.

Implications for management

The invasion of these two decapods in Europe is the subject of great conservation concern. They have had strong negative impacts on the already declining native crayfish populations (Holdich et al. 2009) and are also associated with reductions in local biodiversity (e.g. Correia and Anastácio 2008; Cruz et al. 2008). It is thus worrying that our results point to a widening of their ranges in the future. Although also dependent on propagule pressure, the lack of restraining factors other than abiotic conditions, supported by our results, suggests that the distribution of these decapods may increasingly become a spatial representation of their climatic niche. Projections of climatic suitability using the entire range of distribution of these species have found that many suitable areas of Europe remain unoccupied (Capinha et al. 2011; Capinha and Anastácio 2011). In addition, conservationists should also be aware that these predictions may underestimate the actual potential distribution of these invaders due to their apparent ability to occupy climatic conditions not experienced in their native range. Under this context, and since eradication has thus far proven difficult (Gherardi et al. 2011), we suggest future efforts should focus primarily on impeding new introductions. Given the large country-related component found on the invasion pattern of these species, we believe that actions at country-level, such as effective legislation, can be of particular importance in this regard.

Future directions

Studies using geographical data to infer drivers of invasion can yield important and relevant information, however they also suffer from a series of shortcomings and uncertainties. In our study, in particular, the relatively coarse resolution of the data used (50 × 50 km and country-level) may overlook important aspects of the invasion process. This is even more relevant because identified relationships can be scale-dependent (Fridley et al. 2007). Although our results clearly suggest that propagule pressure and climatic suitability are the most important drivers of the colonisation success of these invaders, finer-grained data would be useful to analyse the consistency of these results at varying spatial scales. In addition, future analyses could also greatly benefit from using predictors more directly related to the tested factors. For instance, the human footprint index is focused on terrestrial ecosystems and, despite being unquestionably correlated, it does not provide a direct representation of human disturbance in the freshwater ecosystems which the studied species inhabit. The potential availability of better predictors in the future could be of special interest in further separating causal from co-varying factors, a difficulty often found in studies assessing drivers of invasion (Lockwood et al. 2007) and which we also encountered in our study (e.g. between propagule pressure and human disturbance).

There are also several untested but well-known invasion hypotheses that are not contradicted by our results (e.g. the ‘enemy release’ hypothesis or the ‘evolution of increased competitive ability’ hypothesis). It was not possible to test these in this study due to issues of either data availability or the requirement for completely different analytical approaches. However, we encourage future work on performing such assessments. We further suggest that, when plausible, the implications of future findings should be discussed in terms of the invader’s ability to establish under new environmental conditions – particularly in the case of the Red swamp crayfish. Despite we here found little support to the ‘climate matching’ hypothesis, this seems to be the exception rather than the rule in biological invasions (Wiens and Graham 2005; Peterson 2011). Thus, understanding the mechanisms that confer certain invaders the ability to shift their climatic niches is of paramount importance in advancing predictions of future invasions (Pearman et al. 2008).

Finally, it would be also of interest to evaluate the relative importance of the tested factors on future stages of invasion. Factors related to the species’ introduction (e.g. propagule pressure) account for large relative importance in emerging invasions while distributions close to equilibrium mostly reflect the drivers of potential distributions (Wilson et al. 2007). Thus, multi-stage analyses would help provide more robust estimates of the overall importance of distinct drivers of colonisation success or failure.

Conclusions

We found evidence that, at a continental scale, the patterns of colonisation of the Signal crayfish and the Red swamp crayfish in Europe are primarily determined by a combined influence of the level of introduction effort made in each country and the abiotic conditions of the receiving areas. Additionally, we also found a reduced predictive power of the 'human disturbance', 'climate matching' and 'biotic resistance' hypotheses. These results are particularly useful for anticipating future range expansions of these species. Accordingly, and in the absence of measures impeding their dispersal, we predict that these species will increasingly occupy the many climatically suitable areas still available in Europe. This interplay between climatic suitability and propagule pressure also emphasises the need to account for multi-causality when evaluating colonisation success. Instead of testing the isolated role of particular factors, we have highlighted the need for future research in biological invasions to focus more on disentangling the relative roles of multiple, and often interacting, factors which dictate invasion success or failure. The use of integrative approaches, such as the ones we used in this research, may prove particularly useful to that end.

Acknowledgments

This study was financed by FEDER funds through the "Programa Operacional de Factores de Competitividade – COMPETE" and by national funds through "FCT – Fundação para a Ciência e Tecnologia" on the scope of the project DID (Dispersal of Invasive Decapoda) (PTDC/BIA-BEC/105182/2008). C. Capinha holds a PhD grant from FCT (SFRH / BD / 41129 / 2007). The study is a contribution to the FP7-226852 European project SCALES and the Consolider Montes CSD2008-00040 granted by the Spanish Ministry of Education and Science (MEC).

Page intentionally left blank

CHAPTER 4

Assessing the environmental requirements of invaders using ensembles of distribution models

This chapter was published in:

Capinha C, Anastácio P (2011) Assessing the environmental requirements of invaders using ensembles of distribution models. Diversity and Distributions 17:13-24.

ASSESSING THE ENVIRONMENTAL REQUIREMENTS OF INVADERS USING ENSEMBLES OF DISTRIBUTION MODELS

César Capinha and Pedro Anastácio

IMAR, Centro de Mar e Ambiente c/o Departamento de Paisagem, Ambiente e
Ordenamento, Universidade de Évora, Rua Romão Ramalho, n.º 59, 7000-
671 Évora, Portugal.

Abstract

Aim: Understanding the environmental factors determining the establishment of invasive populations is a crucial issue in the study of biological invasions. By taking into account the uncertainty of predictions, ensembles of niche based models can provide useful information. Therefore, we explored the use of consensus techniques to generate a quantitative description of the environmental conditions favouring the establishment of four problematic invasive decapods: *Cherax destructor*, *Eriocheir sinensis*, *Pacifastacus leniusculus* and *Procambarus clarkii*.

Location: Iberian Peninsula

Methods: We collected both native and invasive distribution records from multiple sources. From these data, we modelled the potential distribution of the four decapod species using eight correlative models comprising regression, classification and machine learning methods. The relative influence of the environmental variables in single models was averaged to achieve a consensus contribution of the variables. Ecological requirements were investigated by means of consensus suitability curves, a spatial analysis procedure that shows the variation of consensus suitability along the gradients of environmental variables.

Results: The predictive accuracy of single models ranged from fair to very good. Still, variability between predictions was high. Similarly, the influence of each variable in different models was also uneven. Consensus analysis identified the variables related to temperature as highly influential for all invaders. Consensus suitability curves show that *C. destructor* and *P. clarkii*, have reduced suitability in colder areas whereas the suitability for *P. leniusculus* is greatly reduced in warmer areas. The distance to the ocean was highly influential in *E. sinensis* models, with suitability showing an exponential decay as distance increased.

Main conclusions: We show that the information about the species-environment relationships obtained from niche based models is highly dependent on the characteristics of the models used. In this context we demonstrate that ensembles of models and consensus approaches can be used to identify such relationships while also allowing to assess the uncertainty of the achieved knowledge.

Keywords: *Biological invasions, consensus modelling, environmental suitability, environmental gradients, decapods, potential distribution*

Introduction

Invasions are one of the main drivers of biodiversity change in inland waters worldwide and the trend is increasing (Millennium Ecosystem Assessment, 2005). The inland waters of the Iberian Peninsula (IP) are no exception as a large number of invasives is already established (García-Berthou et al. 2007). This is of particular conservation concern mostly because of the existence of a high number of endemic (Doadrio 2001; Sanchez-Fernandez et al. 2008) and threatened species (IUCN 2009) in Iberian inland waters.

We focused our work on four invasive species of the inland waters of the IP: *Cherax destructor* (Yabby), *Eriocheir sinensis* (Chinese mitten crab), *Pacifastacus leniusculus* (Signal crayfish) and *Procambarus clarkii* (Red swamp crayfish). These decapods are known to be responsible for severe negative impacts in invaded areas such as the dissemination of the crayfish plague (Gutierrez-Yurrita et al. 1999), competition with native species (Nyström 1999; Beatty 2006; Gilbey et al. 2008), habitat and ecosystem changes (Rudnick et al. 2003; Geiger et al. 2005), negative effects on amphibian populations (Nyström 1999; Cruz et al. 2008) and on economic activities such as damaging fishing nets (Dittel and Epifanio 2009) and rice fields (Anastácio et al. 2005). *P. leniusculus* and *P. clarkii* are of commercial interest as a food item and both were legally introduced into the IP in the 1970s. They were successfully established and currently present wide ranging populations (Holdich et al. 2010). Similarly, *C. destructor* was imported for harvesting in 1983 and later introduced into the wild in Zaragoza (Bolea 1996). In spite of being considered established, this species presents a much narrower range than the previous two, being confined to localised areas in the Aragón and Navarra Spanish communities (Souty-Grosset et al. 2006). Unlike the other three species, the introduction of *E. sinensis* into the IP was most certainly unintentional, possibly by ballast waters which is its main dispersal vector (Dittel and Epifanio, 2009). This crab was first recorded in the late 1980s in the lower reaches of the Tagus River (Cabral and Costa 1999) and is currently also established in the Guadalquivir River in Seville (Cuesta et al. 2006). It was also observed in the Minho River basin but failed to establish (Sousa et al. 2008).

Despite the relatively long period since the first introductions and the numerous negative impacts of these species, the areas that could potentially be colonized ('potential distribution') in the IP are currently unknown. Moreover, it is still unclear how the environmental characteristics of the IP influence the suitability to these invaders. Niche-based models (NBM) have been regularly used to explore these issues. In general NBM consist of a correlative model calibrated with a given species distribution dataset. This model is then used to predict environmental suitability values and less often to shed light on the causal drivers of species distributions (e.g. Usio et al. 2006; Ficetola et al. 2009; Roura-Pascal et al. 2009). Despite the potential of NBM in this type of analysis a limitation to its use is the level of uncertainty. Previous studies have shown that the sources of variability in NBM are diverse, ranging from the type of correlative model used (Segurado and Araújo 2004; Elith et al. 2006; Tsoar et al. 2007; Roura-Pascual et al. 2009), its parameterisation (Hartley et al. 2006) or data characteristics (Dormann et al. 2008; Graham et al. 2008; Capinha et al. 2011). To account for uncertainty in NBM, the combination of single models in ensemble frameworks is supported (Araújo and New 2007). Further, individual models can be combined into consensus predictions using different methods such as model averaging or sum of classified models. Ensemble and consensus approaches have been shown to be a useful framework for assessing the predictive uncertainty of NBM (Buisson et al. 2011; Grenouillet et al. 2011), and even improve predictive accuracy (Marmion et al. 2009; Grenouillet et al. 2011).

In this context, the use of ensembles of NBM is also promising for examining the causal drivers of the invaders ability to establish since these can potentially provide similar improvements to those found in predicting species distributions. Nonetheless, owing to their multiple origins, the use of ensembles of models to expand the knowledge about the species ecological preferences is less straightforward than in individual NBM. Different modelling options often provide outputs in model-specific formats (e.g. decision trees and regression parameters) which hamper the use of common analysis methods or consensus techniques. The use and development of methods able to extract information from models independently of the type of algorithm used is a viable option but still, the number of studies using ensembles of NBM to gain ecological insights is limited.

In this study, we use ensembles of NBM to expand our knowledge of the ecological preferences and potential distribution of these four decapods invaders in the IP. Specifically our main objectives are: i) to use ensemble and consensus approaches to model the potential distribution of the four decapod species in IP; ii) to identify the environmental factors that best explain these distributions; iii) to explore species-environment relationships using a procedure that plots the variation of environmental suitability along the gradients of the environmental factors in the IP; iv) to use all these results to compare the observed distribution of the species in IP with the potential areas that could be colonized in order to assess the risks of invasion in a conservation perspective.

Methods

Study area

The study area corresponds to the IP. The IP is located in the extreme southwest of Europe, comprising approximately 581 000 km^2 (Figure 1). It is one of the most mountainous regions of Europe. The Iberian climate is largely heterogeneous mostly because of a combination of several maritime influences, continentality and complex orography. Still, precipitation presents a mainly eastward and southward decreasing gradient, whereas temperatures tend to decrease northward (Font 2000).

Data sources

Niche shifts can happen when invasive species colonise new areas (Broennimann and Guisan 2008). Under this prospect, an invader can occupy different conditions from those previously recorded and NBM predictions can fail to predict these shifts (Pearman et al. 2008). Given that niche conservatism is a fundamental assumption at the basis of NBM (Peterson and Vieglais 2001; Wiens and Graham 2005) the representation of the species fundamental niche should be as complete as possible (Broennimann and Guisan 2008). Thus we conducted an extensive compilation of worldwide native and invasive occurrence data of these four species in published literature, museum collections and at the global biodiversity information facility (GBIF; http://data.gbif.org). To be consistent with the spatial resolution of the environmental variables, only records providing a spatial accuracy equal or higher to one square kilometre were included. A total of 147 records were collected for *C. destructor* (105 native; 42 invasive), 306 (133 native; 173 invasive) for *E. sinensis*, 541 (224 native; 317 invasive) for *P. leniusculus* and 624 (219 native; 405 invasive) for *P. clarkii* (see Figure 1 in the Appendix C).

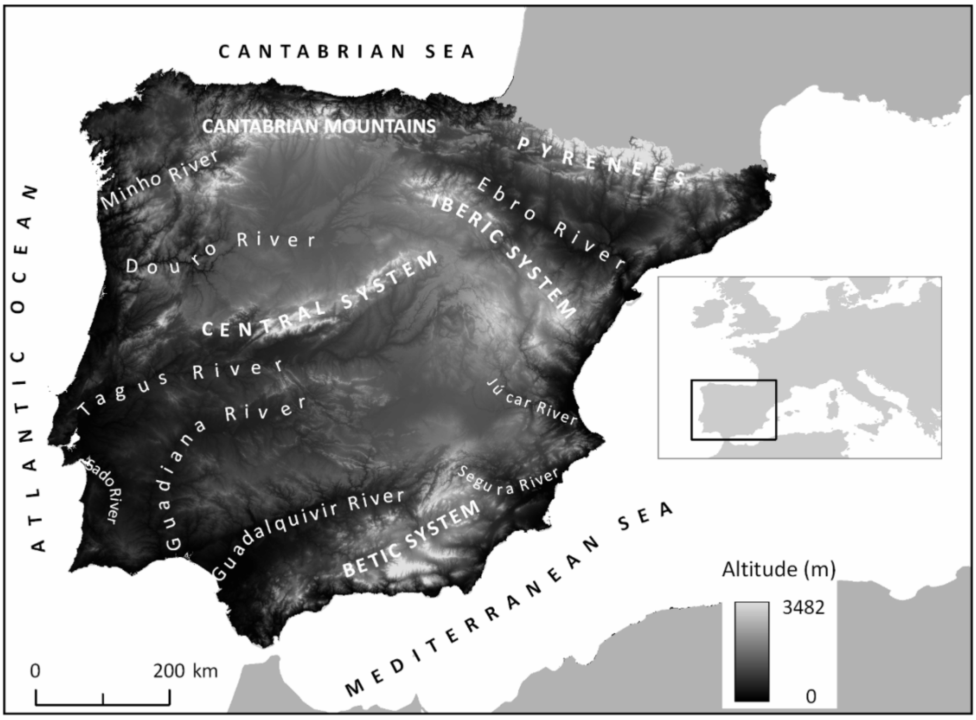


Figure 1 – General physiography of the Iberian Peninsula.

To describe the environmental conditions we selected 10 worldwide environmental variables, with a cell resolution of one square kilometre. From the WORLDCLIM v1.4 datasets (Hijmans et al. 2005), we included six climatic variables: annual mean temperature (*amtemp*); mean temperature of warmest quarter (*mtempwq*); mean temperature of coldest quarter (*mtempcq*); annual precipitation (*anpre*); precipitation of wettest quarter (*prewq*) and precipitation of driest quarter (*predq*). The choice of these variables was based on the fact that climatic variability is known to be of high importance for several physiological attributes of these species, such as regulating their body temperatures. To represent the landscape characteristics four physiographic variables were also included: altitude (*alt*); slope (*slope*); distance to ocean (*disto*) and a compound topographic index (*cti*). Despite being indirect predictors altitude and slope were included since they usually correlate with other environmental characteristics that are known to influence the distribution of these decapods such as stream velocity and size. Distance to ocean was included as a surrogate of water salinity. This information is particularly relevant for *E. sinensis* since this is a catadromous crab (i.e. needs to reproduce in brackish or salt waters). The compound topographic index indicates the landscape’s capacity to retain water by relating slope and area contributing to flow. These four physiographic variables were spatially modelled in GIS using an improved version of the SRTM 90m digital elevation data (Jarvis et al. 2008), after its bicubic resampling to a one square kilometre resolution. To avoid potential redundancy between environmental variables we calculated the Pearson correlation coefficients for all pairs of predictors. It appears that the predictors used were not highly correlated ($|r|<0.85$).

Distribution models

To model the potential distribution of the four species in the IP we used eight different correlative algorithms implemented in the BIOMOD package (Thuiller et al. 2009) for R-CRAN (R Development Core Team, 2009). These included three regression algorithms [generalised linear models (GLM), generalised additive models (GAM) and multivariate adaptive regression splines (MARS)], two classification methods [mixture discriminant analysis (MDA) and classification tree analysis (CTA)] and three machine learning methods [artificial neural networks (ANN), random forest for classification and regression (RF) and generalised boosted regression models (GBM)].

We coupled species occurrence data with pseudo-absences randomly generated worldwide since this option has previously shown to provide good results in predicting potential distributions of these invaders (Capinha et al. 2011). To maximise the pseudo-absences' representation of the available environmental conditions and to avoid unbalanced sample effects (Hosmer and Lemeshow 1989), we created 10 datasets for each species. Each of these datasets had species occurrence records coupled with an equal number of independently drawn pseudo-absences.

To reduce the uncertainty associated with the parameterisation of different NBM algorithms, several parameter values for each algorithm were tested. Models were built using 80% of the occurrences and pseudo-absences of each of the 10 datasets for each species. The remaining 20% were withheld for evaluating predictions using the area under the receiver operating characteristic curve (AUC) (Hanley and McNeil 1982). Each of these replicate models was projected to the IP and the final projection for each modelling algorithm was achieved by averaging the 10 replicate models from the parameterization achieving higher mean AUC values (Table 1 in the Appendix C). The consensus models were then calculated by combining the eight final projections for each invader through a weighted average, being the weights given according to its average AUC as described by Eqn 1:

$$WAi = \frac{\Sigma i(AUCji \times pji)}{\Sigma i(AUCji)}$$

(Eqn 1)

where *WAi* is the weighed prediction for the *i*th invader and pj_i is the probability of presence for the *i*th invader and the *j*th single algorithm. The uncertainty of the resulting ensembles was also assessed by calculating and mapping the standard deviation between the predictions of each of the eight models composing the ensembles.

Relative importance of environmental variables

The BIOMOD package used for modelling species distribution allows the evaluation of variables' importance through a variable randomisation method (Thuiller et al. 2009). The process starts by creating a standard prediction using all environmental variables. Then, BIOMOD produces a set of new predictions, each having one of the independent variables randomised. The evaluation of variable's importance uses the correlation scores between the standard prediction and each of the new predictions. If the correlation score is high, then the difference between the two models is small and, therefore, the relative importance of the randomised variable is low. The calculation of the relative importance is made by subtracting to 1 the correlation score obtained. Thus these values will range from 0 (lowest importance) to 1 (highest importance). This method is particularly valuable because of its independence of the modelling algorithms used. To obtain the relative importance of each of the environmental variables, we performed this calculation for each of the 10 replicate models of the statistical algorithms and we then averaged the correlation scores across the 10 replicate models. The final importance value of each variable in the ensemble corresponded to the average influence across the eight algorithms weighted by the AUC values of each algorithm. To assess the variability between algorithms, the standard deviation was also calculated. The most important predictor variables were defined as those with a relative importance above the mean value of the 10 predictor variables.

Quantitative description of the consensus suitability

Besides identifying the most important variables in explaining the distribution of the decapods studied we also wanted to gain a deeper understanding about the spatial patterns of the consensus suitability projected for the IP. To assess the relationships between environmental variables and probability of occurrence in individual models one can make use of available methods for extracting response curves (e.g. Elith et al. 2005). These responses can then be combined to provide consensus responses, in a process similar to consensus prediction. Still, since these methods do not take into account the interaction among variables (i.e. they rely on setting to constant values all variables except the variable of interest), the environmental suitability along the gradients of these variables will most likely differ from any of the individual responses. This reduces our ability to assess the ecological validity of consensus suitability and hinder from exploring the spatial reasoning of the predictions obtained for the study area. In this context we developed a method that provides a quantitative description of the variation of consensus suitability along the gradient of any environmental variable of interest. Here we used this method to describe the variation of environmental suitability along the gradients of each of the most important predictors in the IP. The procedure used was conducted using Ilwis 3.7 Open GIS (http://52north.org/) automated by the Visual

DialogScript programming language (Commercial Research Ltd.) and its implementation starts by reclassifying each of the environmental variables into equal interval raster maps. For example, the continuous values of mean annual temperature were grouped into multiple classes using class amplitudes of 2ºC (e.g. 0–2ºC; >2–4ºC; >4–6ºC…). We chose class amplitudes by counterbalancing computer processing time with the number of analyses performed along the variable's gradient. The next step was to create new maps, each containing only the portion of the consensus suitability map overlapping each of these classes (Figure 2). Finally, for each of these extracted suitability maps, a set of descriptive statistics was calculated (first quartile, median and third quartile). To obtain more representative values, we only extracted these statistics for classes represented by a minimum of 1000 cells in the variables's raster map. The adoption of this criterion, in most cases, reduced the amplitude of the analysed variable's ranges, since their range limits commonly comprise very small areas. Once this process was concluded we obtained a statistical description of the suitability values along a classified gradient of any environmental variable of interest. These statistics were then plotted using the mean value of each of the variable's classes as the *x*-axis and the obtained descriptive statistics in the *y*-axis.

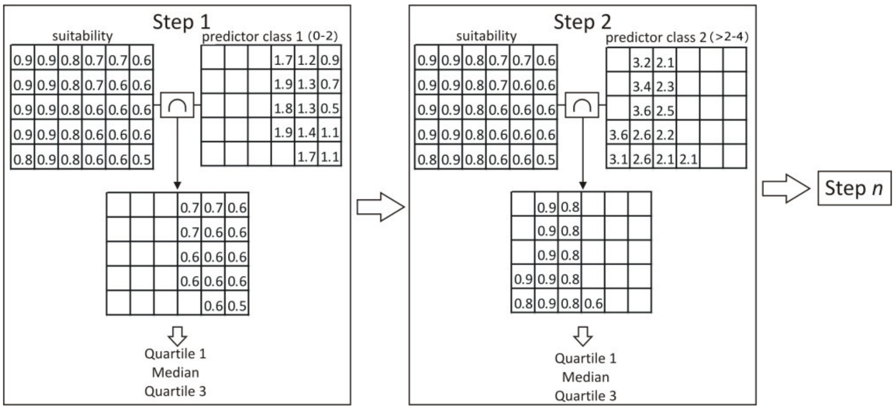


Figure 2 – Schematic representation of the framework used to build the consensus suitability curves. The consensus suitability map is spatially intersected with areas having a predefined range of values of the variable under analysis. The process is shown for two classes of one variable; class one ranging from zero to two and class two from higher than two to four. From these intersections *n* fractions of the consensus suitability map are obtained. For each of these a set of descriptive statistics is then calculated (median, first and third quartiles).

Results

Predictive performance

Table 1 shows the mean AUC value for each algorithm, ranging from fair to very good accuracies. Models for *C. destructor* gather most of the least accurate predictions while also showing the highest variability between accuracies of distinct algorithms (mean AUC = 0.84; standard deviation = ± 0.029). Predictions for *E. sinensis* are also dominated by fair accuracies but still slightly more accurate (mean AUC values = 0.88 ± 0.026) than those of *C. destructor*. Models for both *P. leniusculus* and *P. clarkii* show, in general, good accuracies and lower variability in their predictive performance (mean AUC values = 0.91 ± 0.023 and mean AUC values = 0.91 ± 0.02 respectively).

Table 1 - Area under the receiver operation characteristics curve (AUC) and standard deviation (SD) values for each of the algorithms used for predicting distribution of the four decapod species. Evaluated algorithms include: artificial neural networks (ANN); classification tree analysis (CTA); generalized additive models (GAM); generalized boosted regression models (GBM); generalized linear models (GLM); multivariate adaptive regression splines (MARS); mixture discriminant analysis (MDA) and random forest for classification and regression (RF).

	C. destructor		***E. sinenis***		***P. leniusculus***		***P. clarkii***	
	AUC	SD	AUC	SD	AUC	SD	AUC	SD
ANN	0.83	0.04	0.87	0.02	0.88	0.03	0.91	0.02
CTA	0.82	0.03	0.87	0.02	0.89	0.01	0.90	0.03
GAM	0.86	0.02	0.91	0.02	0.87	0.03	0.92	0.02
GBM	0.86	0.02	0.90	0.03	0.95	0.02	0.92	0.01
GLM	0.80	0.04	0.84	0.04	0.92	0.03	0.88	0.02
MARS	0.84	0.02	0.88	0.02	0.93	0.03	0.91	0.02
MDA	0.85	0.03	0.88	0.03	0.94	0.03	0.92	0.02
RF	0.85	0.03	0.87	0.03	0.93	0.01	0.92	0.02

Relative importance of environmental variables

For all four species only three environmental variables out of ten have values of relative importance higher than the mean importance value (Figure 3). Large standard deviation values show that a large variability can occur in the influence that a variable has in predictions resulting from different algorithms. This occurs for all four species and due to their statistical dependence is higher for variables achieving high mean influence. Factors related to temperature are the most influential for all species especially for *P. leniusculus* and *P. clarkii* (Figures 3c and 3d respectively). The only noticeable exception is the distance to the ocean in *E. sinensis* models, which appears to drive strongly its distribution (Figure 3b).

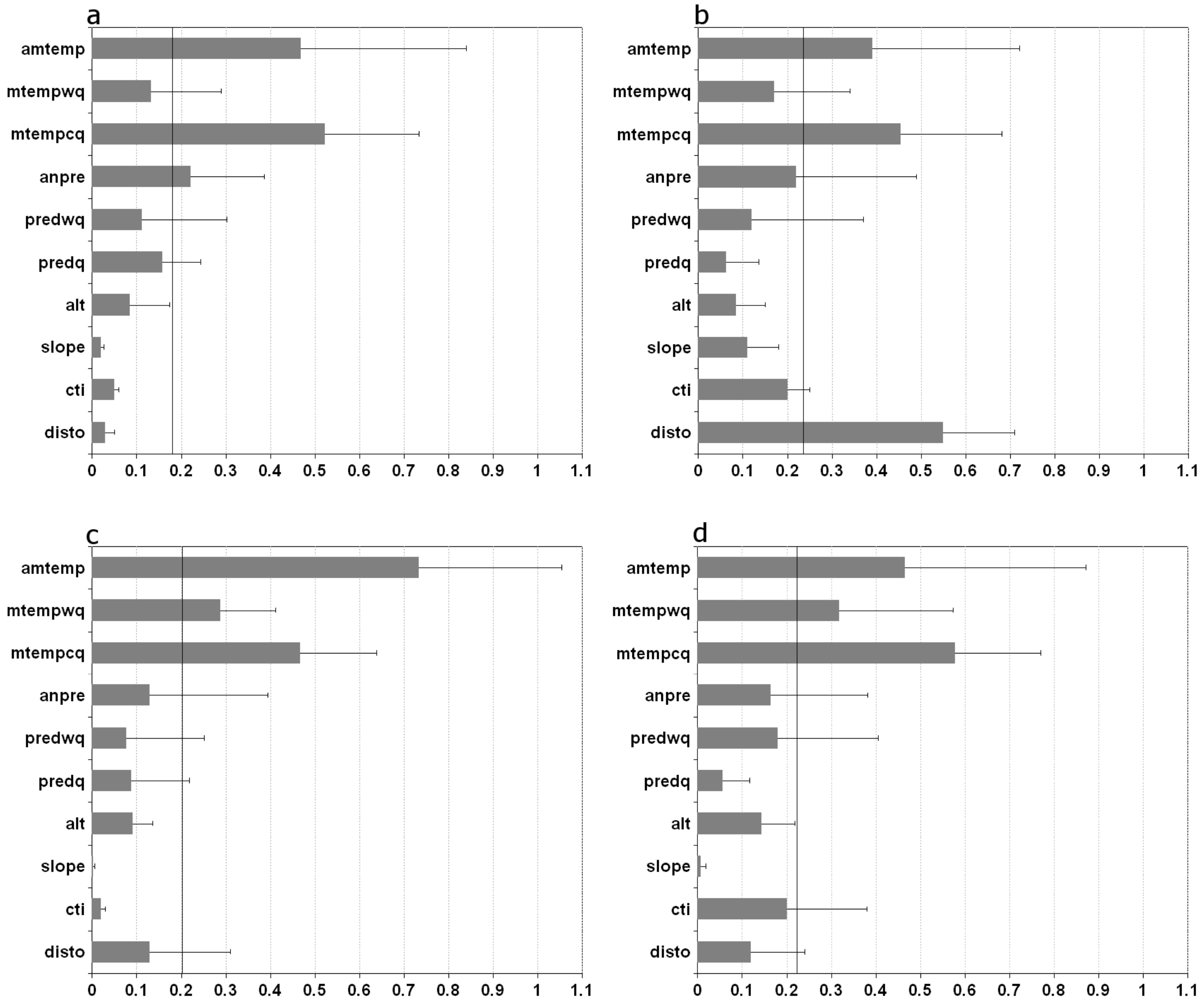


Figure 3 – Relative importance of the environmental variables used to predict the distributions of (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii*. Bars refer to the mean value of relative importance obtained from eight different modelling algorithms and whisker to the standard deviation. Continuous black line represents the mean relative importance value across the 10 predictors.

Environmental suitability within the IP

The suitability plots for *C. destructor, E. sinensis* and *P. clarkii* show that these invaders are better adapted to areas of warmer temperature in the IP (Figure 4). Still, they respond differently depending on the thermal variables and position on the thermal gradient. *P. clarkii* is the species showing wider thermal tolerance. The suitability for this invader steeply increases in areas having mean temperatures above 0ºC in winter, a mean annual temperature above 6ºC and mean summer temperatures above 13ºC (Figures 4j and 4k). *C. destructor*, however, shows a lower suitability in areas with low temperatures, with suitability values increasing only for areas above 12ºC in annual mean temperature and above 4ºC in winter (Figures 4a and 4b). Similarly, higher suitability for *E. sinensis* is also associated with warmer conditions. Still, this species is primarily conditioned by distance to ocean, and the suitability plot for this variable shows an exponential decay of suitability as distance increases (Figure 4d). Distinctively from the previous three invaders, *P. leniusculus* is the species with the highest ability to cope with low temperatures and, unlike the three other species, areas with warm temperatures do not seem suitable for that species whatever the season (Figures 4g-i).

For *C. destructor* the geographic areas with the highest suitability are mostly located in the southwest of the IP, on the Cantabrian coast, the Ebro valley and along the eastern littoral regions (Figure 5a). Still, variability between statistical algorithms is high for some areas, particularly in the southwest of the IP (Figure 6a). Areas of higher suitability for *E. sinensis* are mostly grouped into two geographical regions: (1) the lower Tagus and Sado basins and (2) the lower Guadalquivir River basin, each within approximately 100 km of river mouths (Figure 5b). Other smaller areas are also suitable for this species, mostly in the western and eastern littoral areas of the peninsula. Overall, the variability in the predictions for this species is lower than for the other three invaders. Higher uncertainty values are confined to small areas that are mostly concentrated on the east coast of the IP (Figure 6b). The decrease of temperatures northward in the IP and the preference of *P. leniusculus* for colder areas clearly determines a north/south partition in the potential distribution of this invader (Figure 5c). Still, the ensemble of this species presents wide areas of high variability between single predictions (Figure 6c). *P. clarkii* is the species showing wider invasiveness potential. Suitable conditions are found across most of the IP except at higher altitudes (Figure 5d). The uncertainty map for this species also identifies some areas of higher predictive variability between single predictions, especially in the upper sections of the Douro and Tagus basins (Figure 6d).

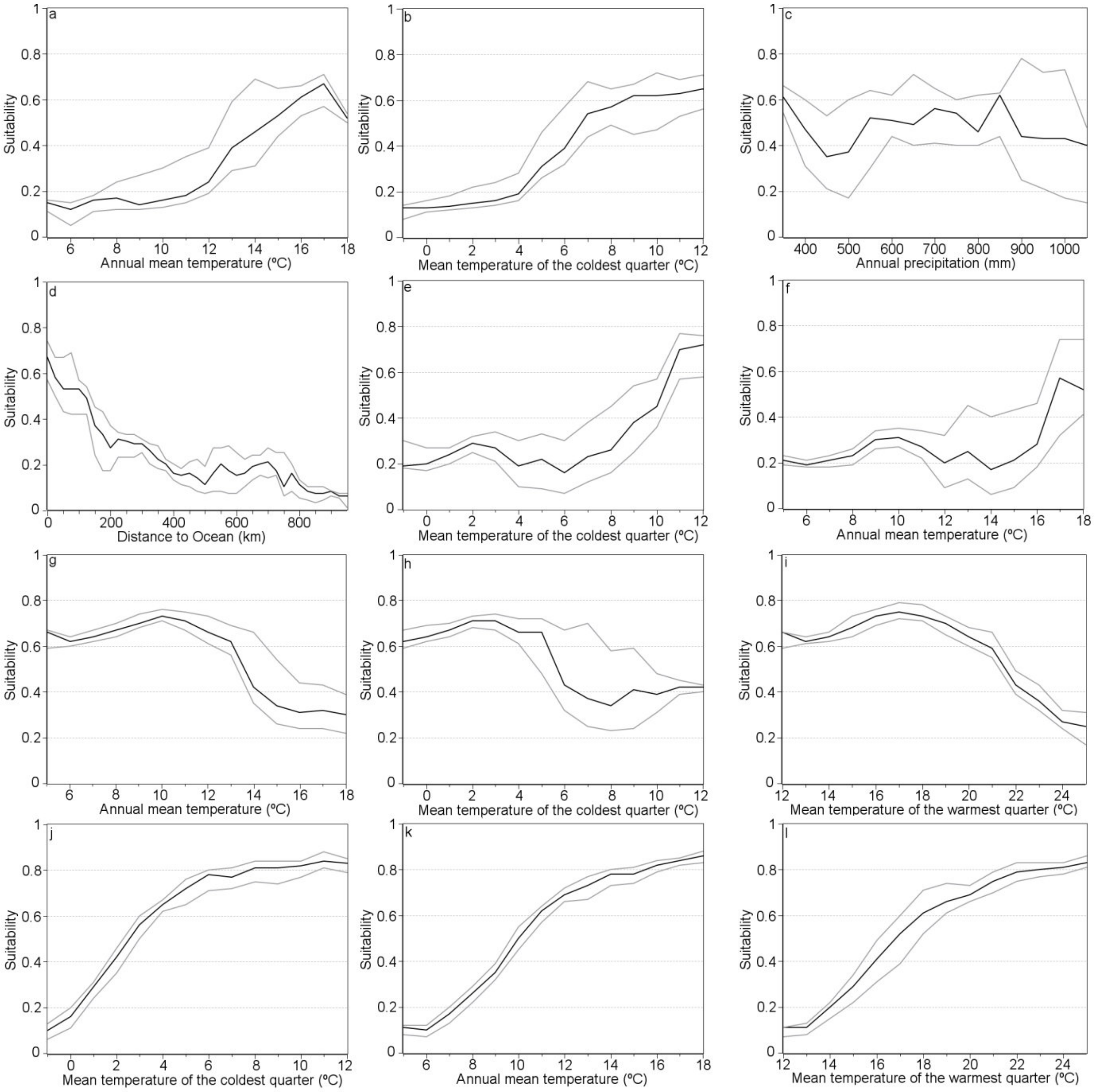


Figure 4 – Suitability variation along the gradients of the most important environmental variables for predicting the distributions of (a-c) *Cherax destructor*, (d-f) *Eriocheir sinensis*, (g-i), *Pacifastacus leniusculus* and (j-l) *Procambarus clarkii*. Grey lines represent first and third quartile and black line the median.

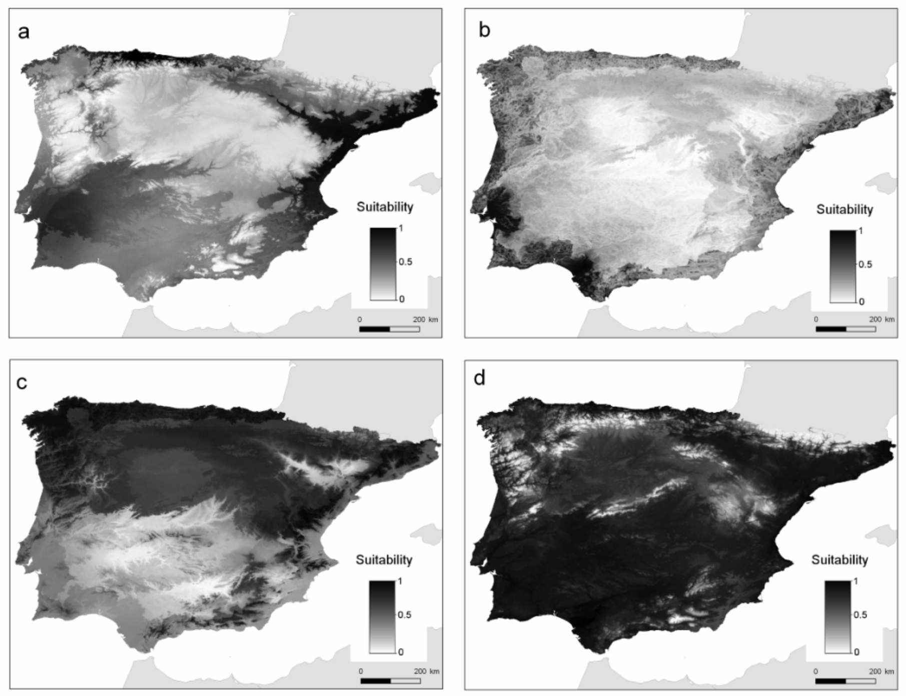


Figure 5 – Consensus potential distribution models obtained by a weighted average of the single models for (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* in the Iberian Peninsula.

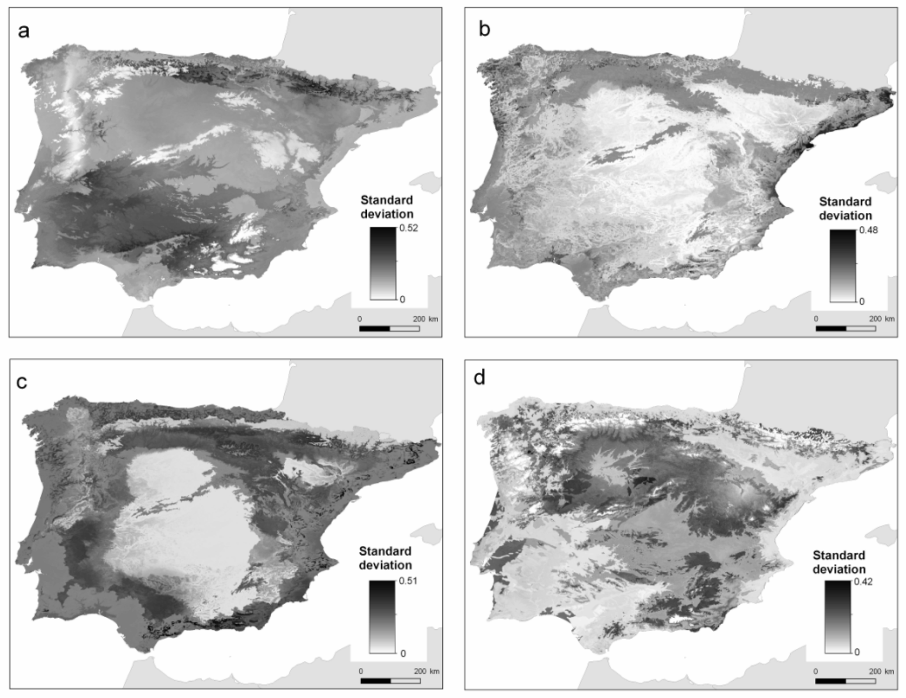


Figure 6 – Variability among the single-models included in the ensembles for (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* in the Iberian Peninsula.

Discussion

Species environmental requirements

In this work, we sought to explore the environmental requirements of four invasive decapods and analyse how the IP suited them. The strong predictive performance of some single models for these invaders supports well the accuracy of the species-environment relationships they found. Still, previous studies have shown that correlative models can fail to detect the correct relationships between the species and the environment (Austin et al. 2006). Furthermore, the identification of correct species/environment relationships is also complicated by the difficulties of comparisons with physiological data (but see Austin et al. 2009). NBM rely on information obtained at the population level and in natural conditions while physiological studies are mostly performed at the individual level, comprising isolated phases of their lifecycles and by maintaining non-studied environmental factors at fixed values (e.g. Nakata et al. 2002; Paglianti and Gherardi 2004). Our results demonstrate that even among models achieving good accuracy variability between models can be high (e.g. for *P. leniusculus*). In this context, the use of ensembles of NBM for examining species/environment relationships is useful since it allows assessing the degree of uncertainty of the achieved knowledge.

To explore the environmental requirements of our four invasive decapods, we first identified the environmental factors which explain the most their distribution. Results showed that a large variability can occur in the influence that each variable has for different algorithms. Given this uncertainty the use of a consensus approach seems more adequate than relying on any single algorithm evaluation. Our results showed that temperature variables, particularly annual mean temperature and mean temperature in winter, highly influence the distributions of all species. The only exception was the distance to the ocean for *E. sinensis*, a catadromous crab. The overall importance of temperature in predicting these invaders' distributions is unsurprising. Besides determining body temperature, temperature variation is also highly influential in sensitive life history processes of these decapods such as growth, mating and recruitment (Nyström 2002; Dittel and Epifanio 2009). This high influence of temperature is also of special concern under the threat of climate warming (IPCC 2007).

Changes in both the spatial and temporal thermal patterns will most certainly cause changes in the ranges of these invaders and, as such, further research regarding this matter is warranted.

The consensus suitability curves presented in this work explored how suitability changes along the gradients of the environmental factors in the IP. This information provides a deeper insight on the spatial patterns of the consensus suitability projected for this area. For example, *C. destructor* and *P. clarkii*, which are two warm water crayfish, present rather different invasive abilities for the IP. The consensus model for *P. clarkii* shows a much wider range, with suitable conditions being found for most of the peninsula, whereas for *C. destructor* most of the northern IP seems unsuitable. A combined analysis of the suitability curves' shapes for these two species provides a detailed explanation for this dissimilarity (i.e. lower suitability in colder areas to *C. destructor* than *P. clarkii*).

In the same manner, despite *E. sinensis* also benefiting from warmer temperatures, the consensus prediction shows that the most suitable regions are almost entirely restricted to the first 100 km from the coast. The higher relative importance of the distance to ocean and the exponential decay of suitability as distance increases provide a clear explanation for this spatial pattern. This spatial confinement is also supported by the current knowledge about the occurrence of this species in the IP. Cabral and Costa (1999) found the most upstream record of this invader in the Tagus River to be only 80 km from river mouth, the same distance found for the majority of the invasive population found in the Guadalquivir (García de Lomas, pers. comm.). Similarly, the north/south partition shown in the *P. leniusculus* prediction can be easily understood by its inability to support the warmer temperatures found in the south of the peninsula.

Invasive ability in the IP

Holdich et al. (2009) provided the most up-to-date description of distributional data for crayfish in Europe. Despite using a much coarser resolution (50 × 50 km) they presented a pattern for *P. leniusculus* in the IP showing the bulk of invasion in the Iberic, Central and Cantabric mountainous systems and the Douro basin. The suitable areas predicted by our study are highly consistent with this description. However, the fact that our potential distribution map shows large extents of suitable areas that are still unoccupied is of concern. Among them, the northwest of the peninsula might be the one of highest concern because of its wide extent and contiguity with established populations. Moreover, the natural spread to some of these areas is being noticed (Costa et al. 2009). As such, in the absence of control measures this region will probably also be part of the species' invasive range in the IP in the coming years.

P. clarkii, remains unrecorded only in a reduced number of areas, of which the two wider ones are the northwest of the IP and the Pyrenees (Holdich et al. 2009). Our consensus distribution map for this species shows that the Pyrenees comprise large extents of unsuitable areas and, as such, the species absence shown by the authors might be explained by a lower suitability of these areas. However, our consensus prediction show that in the northwest of the IP, especially the lowlands, many areas are highly suitable and, similarly to *P. leniusculus*, in the absence of control measures, these will probably become part of the *P. clarkii* invasive range.

According to the distribution data provided by Souty-Grosset et al. (2006), *C. destructor* presents a small invasive range in the IP located in the upper Ebro basin. Our consensus prediction for this species shows that this region is dominated by moderate suitability values. These suboptimal conditions might help explain the reduced colonisation rate of *C. destructor* in the IP compared with the previous two crayfish species. The initial areas of introduction of both *P. leniusculus* and *P. clarkii* (upper central and southwest Spain respectively) took place in areas predicted as highly suitable, which certainly favoured their naturalisation and subsequent natural spread. Still, it is worth noticing that the wider areas offering high environmental suitability for *C. destructor* in the IP are also located in the Ebro basin, downstream of its current invasive range.

The ensemble for *E. sinensis* shows that the two invasive populations currently known in the IP (Tagus estuary and Guadalquivir) are located in areas with high environmental suitability. The close proximity of wide highly suitable areas such as the Doñana National Park could be of particular concern in the coming years. This large wetland is one of the most important protected areas in Europe and might be severely disrupted if the establishment of this crab occurs.

Methodological issues and future directions

Ensembles of predictions allowed us to clearly identify areas for which environmental suitability is less certain such as the southwest of the IP for *C. destructor* or the upper sections of the Douro and Tagus basins for *P. clarkii*. Future refinement of the modelling framework (e.g. by including additional variables) might provide a better understanding of how environmental variability influences suitability in these areas. Despite the usefulness of the information provided by consensus suitability curves, some limitations should be recognised. These curves correspond to a quantitative description of the NBM predictions and are thereby based on the same assumptions and have the same limitations (e.g. not taking into account the existence of dispersal constrains or biotic interactions shaping the species distributions; see Hirzel and Le Lay, 2008 for a recent review). Moreover, consensus suitability curves describe the function fitted for the variable under analysis while also incorporating the influence of the other variables, their coefficients and the method used for combining predictions (i.e. the full model). Thus, these curves should not be detached from the study area for which the predictions are projected. This is because new areas might provide different environmental combinations (i.e. the factors interacting with the environmental variable under study might differ) and thereby consensus suitability values may differ.

Conclusions

Finding the combination of environmental conditions that determines the suitability to invasive populations is a crucial issue in the study of biological invasions. Our results have shown that the use of distinct correlative methods in NBM produces distinct outcomes. This variability occurs both in the predicted distributions and in the results aiming to provide insights on species-environment relationships. These findings support previous studies suggesting the use of consensus predictions to examine the uncertainty of distribution models for invaders. We further support the use of consensus of NBM results to increase the knowledge regarding the relationships between the environment and the invaders ability to establish. The uncertainty contained in this information should always be clearly presented and is fundamental to understand the limits of the obtained knowledge.

Using this framework we were able to identify the environmental factors responsible for determining the potential distribution of four invasive decapods in the IP. We found that factors related to temperature are the most influential for all four invaders, being distance to the ocean the only exception in the *E. sinensis* models. Areas with low temperatures have reduced suitability to *P. clarkii* and *C. destructor*, while for *P. leniusculus* it is mostly in the warmer areas that environmental suitability is low. Our predictions indicate that there are still some uninvaded areas of the IP that possess adequate environmental conditions for these invaders. While the expansion of the *P. leniusculus* and *P. clarkii* invasive ranges in IP will most certainly continue to many of these areas, efforts should be placed to avoid the establishment of *C*. *destructor* and *E*. *sinensis* outside their current ranges since these are in close proximity to wide suitable areas some of which of high conservation importance.

Acknowledgements

We thank the three anonymous reviewers who provided helpful comments and suggestions on earlier drafts of this manuscript. We also thank the many researchers and institutions that provided species distribution data. This research work benefited from the DID (Dispersal of Invasive Decapoda) (PTDC/BIA-BEC/105182/2008) research project co-financed by the Portuguese Fundação para a Ciência e Tecnologia (FCT), and the European Union programs COMPETE, QREN and FEDER. C. Capinha holds a PhD grant from FCT (SFRH / BD / 41129 / 2007).

Page intentionally left blank

CHAPTER 5

Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability

This chapter was published in:

Capinha C, Anastácio P, Tenedório JA (2012) Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability. Biological Invasions. [DOI: 10.1007/s10530-012-0187-z]

PREDICTING THE IMPACT OF CLIMATE CHANGE ON THE INVASIVE DECAPODS OF THE IBERIAN INLAND WATERS: AN ASSESSMENT OF RELIABILITY

César Capinha[1]; Pedro Anastácio[1]; José António Tenedório[2]

[1]IMAR, Centro de Mar e Ambiente c/o Departamento de Paisagem, Ambiente e Ordenamento, Universidade de Évora, Rua Romão Ramalho, n.º 59, 7000-671 Évora, Portugal.

[2] e-GEO, Centro de Estudos de Geografia e Planeamento Regional, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas, Lisboa, Portugal.

Abstract

In an effort to predict the impact of climate change on the distribution of existing invasive species, niche-based models (NBMs) are being increasingly used to make forecasts. Here, we investigate the reliability of these models in predicting future climatic suitability for 4 invasive decapods of the Iberian Peninsula: *Cherax destructor*, *Eriocheir sinensis*, *Pacifastacus leniusculus* and *Procambarus clarkii*. From an ensemble of forecasts generated by 5 distinct algorithms (generalized linear models, artificial neural networks, support vector machines, random forests and alternating decision trees), we calculated consensus predictions for current conditions and 3 future time periods (2030, 2050 and 2080) under low and high scenarios of greenhouse gas emissions. Three criteria were examined to infer the robustness of the forecasts: ability to predict current distributions, inter-model variability and degree of environmental extrapolation. Our results indicate an overall decline in climatic suitability for the 4 invaders as time progresses. However, we also identified highly distinct levels of predictive uncertainty among species. Good indicators of reliability were found for *Procambarus clarkii* and *Pacifastacus leniusculus*, whereas the predictions for *C. destructor* showed low predictive performance, low inter-model agreement and wide areas of environmental extrapolation. For *E. sinensis*, the models also showed high variability with respect to areas projected to lose climatic suitability. Overall, our results highlight the need to consider and evaluate multiple sources of uncertainty when using NBM predictions for invaders under current and future conditions.

Keywords: *Climate change; Decapods; Ensemble forecasting; Environmental extrapolation; Iberian Peninsula; Uncertainty*

Introduction

The climate has been rapidly changing in recent decades (Vose et al. 2005; IPCC 2007), and simulations of future conditions suggest that surface temperatures may increase and precipitation patterns may change (IPCC 2007; Rahmstorf et al. 2007). These modifications are likely to have a substantial effect on key components of invasions (Dukes and Mooney 1999; Galil et al. 2007; Hellmann et al. 2008; Rahel and Olden 2008; Walther et al. 2009), including the distribution of existing invasive species (Hellmann et al. 2008; Rahel and Olden 2008). For example, it is generally expected that cold-temperature constraints will decrease at higher-latitude or upper-elevation range limits, whereas warm-temperature constraints will increase at lower-latitude or lower-elevation range limits (Hellmann et al. 2008).

The Iberian Peninsula, like many other areas of the world, is seriously impacted by invasive species (Drake 2009). Of these species, few are as successful as the crayfish *Procambarus clarkii* (Red swamp crayfish) and *Pacifastacus leniusculus* (Signal crayfish). These 2 species are native to North America and were introduced to the Iberian Peninsula in the 1970s. Since then, they have become widespread (Gutiérrez-Yurrita et al. 1999; Souty-Grosset et al. 2006) and have had a large impact on ecosystems and human activities, which have been classified as 'positive' or 'negative' by affected stakeholders. For example, the harvesting and angling of these species have become popular activities in some regions, providing extra income to practitioners (Gutiérrez-Yurrita et al. 1999). In contrast, other individuals have suffered major losses as a result of the grazing of their crops by crayfish, particularly *Procambarus clarkii* (Gutiérrez-Yurrita et al. 1999; Anastácio et al. 2005). A similar dichotomy has also been observed in conservation. For example, these invaders are a known cause for the decline of several freshwater species (Ilhéu et al. 2007; Arce and Alonso 2011), whereas they have also become a new and rich food source for a number of vertebrate predators, most notably the Eurasian otter (*Lutra lutra*) (Clavero et al. 2003; Arce and Alonso 2011).

In addition to these 2 decapods, *Cherax destructor* (Yabby) and *Eriocheir sinensis* (Chinese mitten crab) have also been established in Iberian inland waters for some decades. *E. sinensis* is a catadromous crab that is native to eastern Asia, which was first detected in the Tagus estuary of the Iberian Peninsula, and is thought to have been established since the late 1980s (Cabral and Costa 1999). At present, this species is also known to occur in the Guadalquivir River (Garcia-de-Lomas et al. 2010). To date, no significant impact by *E. sinensis* has been recorded in the Iberian Peninsula; however, this species has caused a number of negative impacts at other invaded areas, such as reducing the integrity of stream banks and levees due to its burrowing activity (Rudnick et al. 2003) or damaging fishing nets (Dittel and Epifanio 2009). *C. destructor* is a crayfish native to Eastern and Central Australia. This species was introduced to some sites in the north of the Iberian Peninsula during the 1980s and, regardless of eradication efforts, a few populations remain (Camargo and Gonzalo 2009, cited in Holdich et al. 2009). Like *E. sinensi*s, the potential impacts of this species on the Iberian Peninsula remain relatively unknown. However, recent studies in other areas of the world suggest that this species might have serious negative impacts on invaded freshwater ecosystems, particularly on native crayfish (Beatty 2006; Lynas et al. 2007).

In a previous study, our research group found that the potential distribution of these 4 invaders is strongly correlated with climate, particularly thermal factors (Capinha and Anastácio 2011). We also found that, *C. destructor*, *E. sinensis* and *Procambarus clarkii* are positively associated to warmer conditions in the Iberian Peninsula. We therefore hypothesized that future climate warming might increase the climatic suitability of the Iberian Peninsula to these invaders. Inversely, for *Pacifastacus leniusculus* we anticipated a reduction of suitable areas, since this species is associated with the colder areas of the Iberian Peninsula (Capinha and Anastácio 2011). Assessing changes on the potential distribution of *Pacifastacus leniusculus* and *Procambarus clarkii* is of particular importance, given the large number of impacts that these species have in the study area. Information is also of interest for *C. destructor* and *E. sinensis*, since an increase in the climatic suitability of the Iberian Peninsula might promote a widening of their ranges.

In recent years, niche-based models (NBMs) (Guisan and Zimmermann 2000; Guisan and Thuiller 2005) have become popular tools to forecast climatic suitability for non-indigenous species (e.g. Beaumont et al. 2009b; Ficetola et al. 2009; Gallagher et al. 2010b). These models relate species occurrences with climate data, to implicitly capture the species-environment relationships responsible for shaping species distributions. Such correlations have a degree of uncertainty, since they assume that species are in equilibrium with the environment, which is often not the case for invaders (Guisan and Thuiller 2005; Elith and Leathwick 2009; Elith et al. 2010). In addition, when developing predictions for new areas or time periods, these models also rely on the conservatism of species-environment relationships (i.e. niche conservatism; Peterson 2003; also see Pearman et al. 2008 for a review). Even if these ecological assumptions are met, the application of NBMs across space and time is also subject to several methodological uncertainties. These include the variability arising from (1) the use of distinct statistical models (Dormann et al. 2008; Diniz-Filho et al. 2009; Buisson et al. 2010); (2) the way that these models are parameterised (Hartley 2006); (3) the quantity, quality and adequacy of dependent and independent data (Dormann et al. 2008; Diniz-Filho et al. 2009; Buisson et al. 2010; Austin and Van Niel 2011); (4) and/or the uncertainty already found in predictions of general circulation models and scenarios of greenhouse gas emissions (Beaumont et al. 2008; Dormann et al. 2008; Diniz-Filho et al. 2009; Buisson et al. 2010). In addition, it should also be taken into account that the potential existence of new environmental conditions in the prediction layers (hereafter termed ‘environmental extrapolation’) is also an important, but often neglected, source of uncertainty (Thuiller et al. 2004; Dormann 2007; Elith and Leathwick 2009; Fitzpatrick and Hargrove 2009). This possibility raises both ecological and methodological uncertainties, since the relationships between the species and these new environments are not known; therefore, the behaviour of models under these circumstances is likely to also differ substantially (Pearson et al. 2006; Elith and Graham 2009).

Given these challenges, NBM predictions should be accompanied with an assessment of their uncertainty. While ecological uncertainties are usually difficult to evaluate a priori (Guisan and Thuiller 2005; Pearman et al. 2008; Elith and Leathwick 2009), 2 main approaches have been used to address methodological uncertainties: (1) using a single modelling option that is considered to provide ‘optimal’ results (e.g. Phillips et al. 2006; Peterson et al. 2007) or (2) building an ensemble of distinct but plausible models and analysing the range of the results (Araújo and New 2007). In general, the choice for the second option is supported on the lack of agreement on what are the overall ‘best’ modelling options (e.g. Elith et al. 2006; Peterson et al. 2007; Phillips 2008; VanDerWal et al. 2009; Capinha et al. 2011). Consequently, ensembles of NBM have been used in a large number of applications, such as predicting the current distribution of rare species (e.g. Le Lay et al. 2010), predicting the potential distribution of invasive species (e.g. Roura-Pascual et al. 2009; Capinha and Anastácio 2011) or forecasting climatic suitability (e.g. Diniz-Filho et al. 2010).

Here, we built an ensemble of NBMs to predict the climatic suitability of the Iberian Peninsula for *C. destructor*, *E. sinensis*, *Pacifastacus leniusculus* and *Procambarus clarkii* in 3 future time periods (2030, 2050 and 2070) under scenarios of low and high greenhouse gas emissions. Since our main interest was to increase the reliability of our forecasts, we specifically tested the effects of 2 well-known sources of uncertainty of NBMs during the analysis of our target species: (1) the use of distinct statistical models (Capinha and Anastácio 2011) and (2) the strategy used for extracting pseudo-absence data (Capinha et al. 2011). In addition, we also applied a simple method to spatially delimit and measure the level of environmental extrapolation that arises when predictions are made for new conditions. We used the results from these procedures to address the 2 following questions: (1) how will the climatic suitability of the Iberian Peninsula change for these invaders in the future, and (2) are the levels of predictive reliability equal among all 4 species?

Methods

Species data

Many invasive species are not in equilibrium with the environment in their new ranges (Guisan and Thuiller 2005; Elith and Leathwick 2009). Thus, to better estimate the fundamental niche of such species, recent research suggests that, if available, both native and invasive distributions should be used when building models (Broennimann and Guisan 2008; Beaumont et al. 2009a; Capinha et al. 2011). Therefore, we compiled both native and invasive occurrence records of our target species from multiple sources, including the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/), museum collections, published literature and several field surveys that we have previously conducted in the Iberian Peninsula. From these sources, we selected data that provided a spatial precision equal to or higher than 5 km. To avoid the duplication of information and given the spatial resolution of our climate data (see below), we only included 1 occurrence record per 10 × 10-km grid cell. This process resulted in a total of 128 (93 native and 35 invasive) occurrence records for *C. destructor*, 233 (98 native and 135 invasive) for *E. sinensis*, 437 (171 native and 266 invasive) for *Pacifastacus leniusculus* and 512 (195 native and 317 invasive) for *Procambarus clarkii* (see Figure 1 to 4 in the Appendix D for species ranges and occurrence data).

Climate data

We used downscaled climate model data from the Intergovernmental Panel on Climate Change (IPCC) fourth assessment, which was made available by Jones et al. (2009). These datasets provide both current conditions and predictions of future conditions on the monthly averages of maximum and minimum temperatures, monthly rainfall and the number of rainy days using a 5 arc-minutes spatial resolution (≈10 km). We collected this information for current distribution conditions (1950–2000), and for 3 future time periods, i.e. 2021–2040 (hereafter denoted as the 2030 period), 2041–2060 (denoted as the 2050 period) and 2071–2090 (denoted as the 2080 period) with respect to 2 possible alternative emission scenarios indicated by the IPCC Special Report on Emissions Scenarios (IPCC-SRES): B1 (low greenhouse gases emissions scenario) and A2 (high greenhouse gases emissions scenario) (Nakicenovic and Swart 2000). Datasets of future climates were extracted from 2 coupled general circulation models, the CSIRO Mark 3.0 (Australia's Commonwealth Scientific and Industrial Research Organisation) and the CNRM-CM3 (Centre National de Recherches Météorologiques). The climatic predictions of these 2 CGCMs were averaged into a single prediction, from which a set of 6 non-redundant climatic variables (pairwise Pearson correlations ≤ ± 0.8) were calculated for both current and future time periods: mean diurnal range (ºC); mean maximum temperature of the warmest month (ºC); mean minimum temperature of the coldest month (ºC); mean yearly total rainfall (mm); mean maximum rainfall in the driest month (mm); and maximum number of rainy days in a month (days). We considered that this selection of climatic variables was ecologically meaningful for the selected target species since climate, and especially thermal conditions, appear to have a major influence on their distributions (Souty-Grosset et al. 2006; Lynas et al. 2007; Dittel and Epifanio 2009; Sahlin et al. 2010; Capinha and Anastácio 2011). We also assumed that these land-based variables are representative of climatic conditions found in inland waters (where our species occur), since the conditions in these 2 systems are roughly correlated (Caissie 2006; Sabater et al. 2009).

NBMs

To forecast the future potential ranges of the 4 selected invasive decapods, we used 5 distinct correlative algorithms implemented in the Weka data mining software (v 3.6.2) (Hall et al. 2009), i.e. generalised linear models (GLM; McCullagh and Nelder 1989), multilayer perceptron artificial neural networks (MP-ANN; Sarle 1994), support vector machines (SVM; Vapnik 1995), random forests (RF; Breiman 2001) and alternating decision trees (ADT; Freund and Mason 1999; Pfahringer et al. 2001). With the exception of ADT, these models have been widely used in NBMs, and their predictive performance has been previously tested by a number of studies (e.g. Drake et al. 2006; Elith et al. 2006; Pearson et al. 2006; see Text 1 in the Appendix D for more details).

To calibrate and validate the NBMs, we created datasets containing current climatic conditions transmitted by the species occurrences and pseudo-absences. Pseudo-absences were used as a substitute for real absences since information on the species absence or locations in which invasions have failed is almost non-existent for our target species. Despite allowing for the use of discrimination models that are considered to be more accurate than models based on presence-only data (Brotons et al. 2004; Elith et al. 2006), these records are a known cause of uncertainty in NBMs. This uncertainty is mostly related to the delimitation of the areas used to sample these records (VanDerWal et al. 2009; Capinha et al. 2011). To account for some of this uncertainty, we tested 2 major types of extraction methods in this study. The first consisted of a random sample of worldwide climatic conditions, excluding areas that contained species occurrences. This extraction method has previously been shown to perform well with these invaders (Capinha and Anastácio 2011; Capinha et al. 2011); however, it potentially samples many areas where these species have never dispersed to, but that may in fact be suitable (see the discussion in Capinha et al. 2011). Hence, we tested a second strategy with the aim to increase the number of sampling locations that receive propagules of these species (i.e. areas close to where the species is established), but where they have not been recorded. By this, we assumed that these areas are representative of unsuitable conditions. This was done by jointly sampling areas where species do not occur inside the species native range (defined here as the convex-hull of its occurrence records), with the outer areas surrounding the species native range limit. Taking into consideration the high active dispersal ability of our target species (Souty-Grosset et al. 2006; Dittel and Epifanio 2009; Bernardo et al. 2011), we tested incrementally increasing distances from the species native range at intervals of 200 km, with a maximum distance of 1000 km, to delimit the outer area surrounding the species range (following VanDerWal et al. 2009). Because of the catadromous life cycle of *E. sinensis*, we removed from the sampling areas of this species all locations exceeding the maximum distance to the coast found on its occurrence dataset (≈1100 km).

In total, 6 groups of calibration datasets were prepared for each species: 1 using pseudo-absences extracted worldwide and 5 using pseudo-absences extracted in the vicinity of the species native range. Each group comprised 20 replicate calibration datasets. These datasets included the current climatic conditions transmitted by the species occurrences, and an equal number of climatic conditions transmitted by an independently drawn set of pseudo-absences. The use of replicate models with an equal prevalence of occurrences and pseudo-absences allowed the overall representativeness of the conditions sampled by pseudo-absences to be increased, without inducing predictive trends towards a more prevalent type of response. Before the final predictions were made, all the models were validated using a stratified 'leave-one-out' cross-validation procedure. This procedure consists of using 1 occurrence and 1 pseudo-absence previously removed from the model calibration for comparison with the predictions. The process is repeated so that each pair of records is used once as the validation data. These results were then used to calculate the area-under-the-curve of the receiver-operating characteristic (AUC) (Hanley and McNeil 1982) for both native and invasive ranges by using the results of pairs with both native and invasive occurrences, respectively.

Since, overall, we obtained equivalent or higher predictive performances when using local pseudo-absences sampled as far as 800 km away from the species native range (see results), we used the models calibrated from this data to make predictions under current and future climate conditions. Final climatic suitability predictions were made by obtaining an average across the replicate predictions. For this purpose, we initially excluded all predictions with poor accuracy (AUC < 0.8; cf. Thuiller et al. 2005) at discriminating either native or invasive occurrences from pseudo-absences. Next, we rescaled the remaining predictions into a 0 to 1 range, in an attempt to reduce differences in the calibration of probabilities between different statistical algorithms. Finally, the models were combined using a weighted average, with the weight of each model being given by the product of one from its mean AUC score. A schematic representation of the methodological framework is available in Figure 5 in the Appendix D.

Analysis of the predictions

The consensus predictions were spatially projected using a 10 × 10-km cell resolution. To quantitatively describe the magnitude of the changes in suitability of future climates, for each species we calculated the percentage of total area of the Iberian Peninsula that would be climatically suitable. To obtain these measurements, we classified the consensus predictions by using a probability threshold of 0.5 (≥0.5 = suitable), which corresponds to the presence/pseudo-absence prevalence used for model calibration (Liu et al. 2005).

We also evaluated the variability associated with the use of multiple correlative models. This was achieved by calculating the standard deviation among the predictions of the replicate models used to generate each of the consensus predictions. We also evaluated the amount of environmental extrapolation in each of the forecasts. To identify the areas where predictions were extrapolated, we spatially delimited all areas where the value of at least 1 of the predictors was outside of its range on the calibration datasets (transmitted either by occurrence or pseudo-absence records). We also measured the magnitude of this difference by calculating a simple measure of distance between the conditions found in the calibration sets and those found in each area of extrapolation. This calculation is given by equation 1:

$$Ex = \sum_i ((|V_i - L_i|) \times 100 \ / A_i) \qquad \text{(eq. 1)}$$

where *Ex* is the magnitude of environmental extrapolation, *Vi* is the value of the *i*th variable falling outside of its range in the calibration set, *Li* is the value of the closest range boundary of that variable in the calibration set, and *Ai* the amplitude of the variable in the calibration set. This measurement was calculated for all map cells of each scenario and time period, and it corresponds to the summation of the percentage of extrapolation (i.e. the prediction of conditions beyond the calibration ranges) that occurred for each environmental variable. The percentages are based on the amplitude of the variable in the calibration set. For example, if extrapolation occurs for only one 1 variable, a cell value higher than 100% indicates that the distance to the closest range boundary exceeds the amplitude of that variable in the calibration set.

Results

Predictive performances under current conditions

Overall, pseudo-absences that were sampled at a maximum of 200 km from the species native range provided the least accurate predictions (Figure 6 in the Appendix D). However, a clear increase in accuracy was found in models that sampled at maximum distances of 800 km, and no significant difference was found in accuracy when larger areas were sampled (Table 1; Figure 6 in the Appendix D).

Table 1 Area under the receiver operation characteristics curve (AUC) for each of the algorithms used for predicting current climatic suitability in native and invaded areas. Values refer to models using pseudo-absences extracted in areas without the species occurrence located at a maximum of 800km from its native range boundary. Evaluated algorithms include: generalised linear models (GLM); artificial neural networks (ANN); alternating decision trees (ADT); random forests (RF); support vector graphics (SVG).

	Cherax destructor				*Eriocheir sinensis*			
	Native		Invasive		Native		Invasive	
	AUC	SD	AUC	SD	AUC	SD	AUC	SD
GLM	0.76	0.02	0.79	0.01	0.81	0.01	0.86	0.01
ANN	0.81	0.01	0.83	0.02	0.83	0.01	0.89	0.01
ADT	0.81	0.02	0.85	0.02	0.83	0.01	0.89	0.01
RF	0.81	0.01	0.84	0.01	0.86	0.01	0.93	0.01
SVG	0.79	0.02	0.82	0.02	0.83	0.01	0.88	0.01

	Pacifastacus leniusculus				*Procambarus clarkii*			
	Native		Invasive		Native		Invasive	
	AUC	SD	AUC	SD	AUC	SD	AUC	SD
GLM	0.84	0.01	0.86	0.01	0.85	0.01	0.93	0.01
ANN	0.84	0.01	0.86	0.01	0.85	0.01	0.92	0.01
ADT	0.86	0.01	0.88	0.01	0.91	0.01	0.94	0.01
RF	0.89	0.01	0.92	0.01	0.90	0.01	0.96	0.01
SVG	0.82	0.01	0.85	0.01	0.88	0.01	0.92	0.01

Our models also showed consistently higher accuracies when discriminating invasive occurrences from pseudo-absences extracted in the vicinity of the species native range. This difference ranges from a mean increase of 0.01 to 0.08 in AUC values and is found irrespective of species (Figure 6 in the Appendix D). Focusing specifically on the results of the models used for building the consensus predictions (Table 1), we found a consistent pattern of higher accuracy for the 2 tree-based algorithms. RF almost consistently provided the best performing models (mean AUC = 0.89; standard deviation = ± 0.05), followed by ADT (mean AUC = 0.87 ± 0.04). In comparison, the 3 remaining correlative methods provided lower mean scores, with SVG being the least accurate algorithm (mean AUC = 0.85 ± 0.04). In addition, all the models achieved the highest mean performance for *Procambarus clarkii* (mean AUC for native occurrences = 0.88 ± 0.03; mean AUC for invasive occurrences = 0.93 ± 0.02), followed by *Pacifastacus leniusculus* (mean AUC for native occurrences = 0.85 ± 0.03; mean AUC for invasive occurrences = 0.87 ±0.03), *E. sinensis* (mean AUC for native occurrences = 0.83 ± 0.02; mean AUC for invasive occurrences = 0.89 ± 0.03), and finally *C. destructor*, which had the lowest accuracies (mean AUC for native occurrences = 0.80 ± 0.02; mean AUC for invasive occurrences = 0.83 ± 0.02). In general, the variability of the predictive performances of each statistical model was low (max AUC variation = 0.02), which supports the independence of their accuracy scores from the environments sampled by each replicate sample of pseudo-absences.

Climatic suitability in the future

Overall, the consensus predictions show a decline in climatically suitable areas for the 4 invaders as climate change progresses (Figure 1). The invader with the highest loss of climatically suitable areas is *Pacifastacus leniusculus* (Figure 1c). As time progresses, the potential distribution of this species shows a continuous loss of suitability, mostly in the southern areas of the Iberian Peninsula (Figure 2c). Under the most extreme scenario for climatic change (i.e. A2 in 2080), suitable areas for this invader are projected to be less than a third of those under present-day conditions (Figure 1c). These areas are expected to be mostly confined to the northern and northwestern fringes of the Iberian Peninsula, and a few higher altitude regions. Projections for *Procambarus clarkii* show that this is the least affected species by climate change (Fig 1d). Future projections for this invader show moderate changes from its current potential distribution (Figure 2d), with the decline in suitable areas only being of note for the 2080 period (Figure 1d). Most of the areas projected to lose climatic suitability for this species are found on the interior of the Iberian Peninsula

(Figure 2d). A reduction in climatic suitability is also predicted to occur for *C. destructor* (Figure 1a). Major changes also occur mostly in the interior area of the Iberian Peninsula, where this species is projected to lose some suitable areas with increasing levels of climatic change (Figure 2a). Projections of climatic suitability for *E. sinensis* also show noticeable changes in future climates, especially in relation to the loss of suitable areas along the Southern Atlantic and Mediterranean coasts of the Peninsula (Figure 2b).

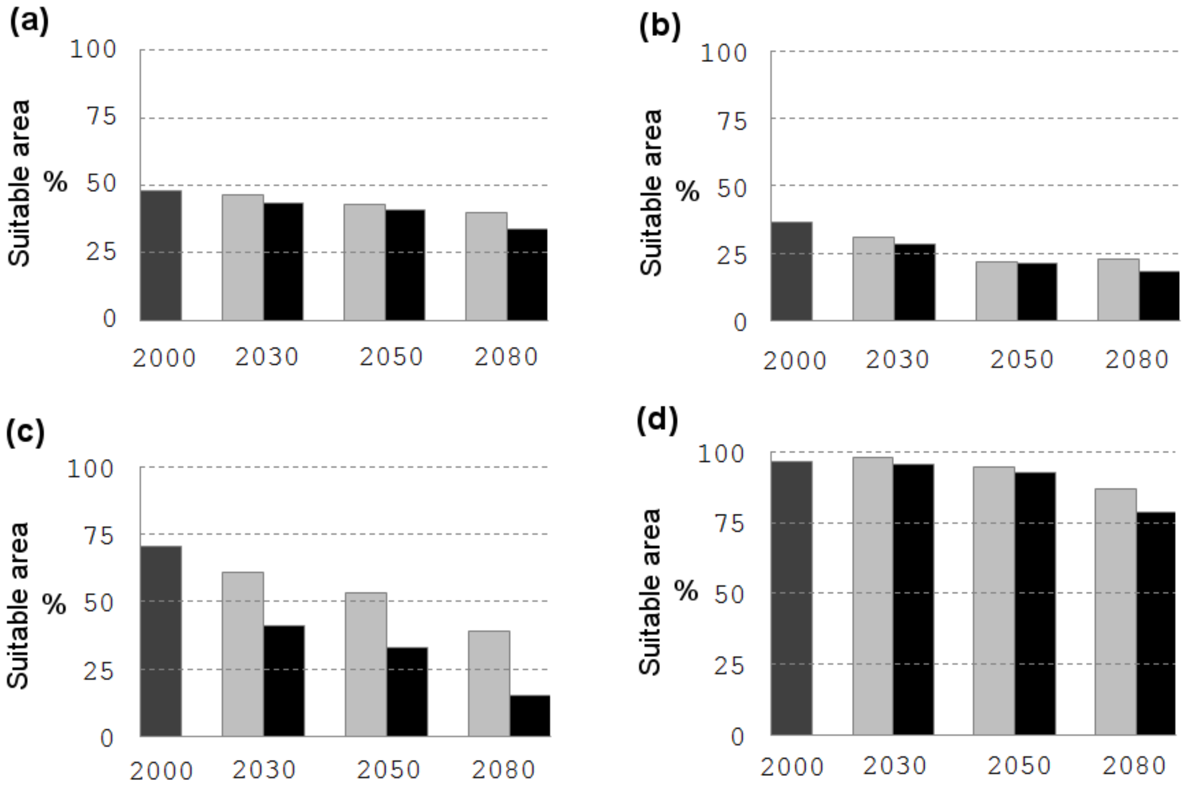


Figure 1 - Percentage of the Iberian Peninsula climatically suitable for a) *Cherax destructor*, b) *Eriocheir sinensis*, c) *Pacifastacus leniusculus* and d) *Procambarus clarkii*. Predictions are made for current climatic conditions (2000; dark grey) and three future time periods (2030, 2050 and 2080) under a low (B1; light grey) and high (A2; black) greenhouse gases emission scenario. Suitable areas were defined as those achieving a climatic suitability equal or higher than 0.5.

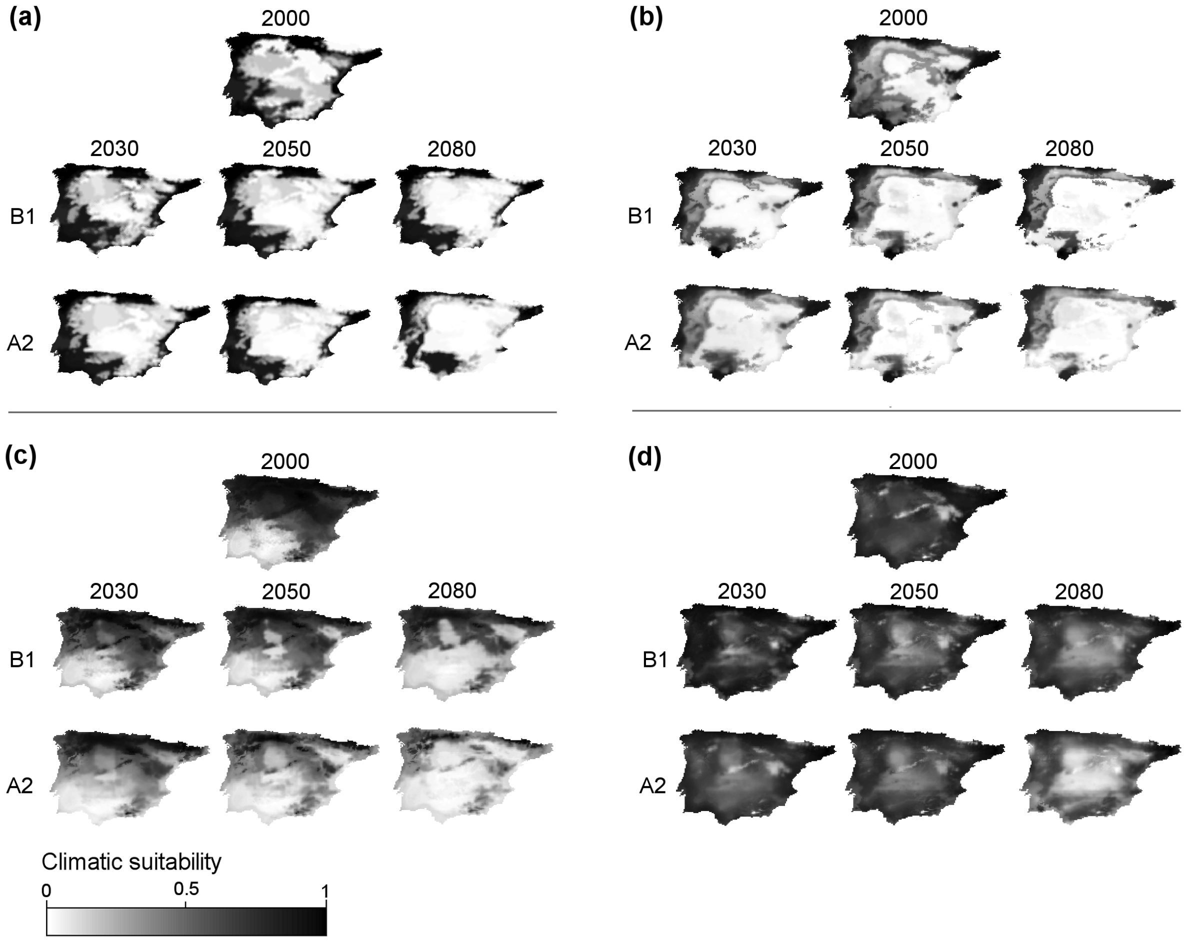


Figure 2 - Climatically suitable areas of the Iberian Peninsula for a) *Cherax destructor*, b) *Eriocheir sinensis*, c) *Pacifastacus leniusculus* and d) *Procambarus clarkii*. Predictions are made for current climate (2000) and three future time periods (2030, 2050 and 2080) under a low (B1) and high (A2) greenhouse gases emission scenario.

Predictive variability and environmental extrapolation

Out of the 4 invaders, the highest variability in generating consensus predictions was obtained for *C. destructor* (Figure 3a). This high variability arose in many areas, with the north and northwest of the Iberian Peninsula consistently being the major areas of model divergence. The forecasts for *E. sinensis* also showed several areas with medium to high inter-model divergence. The highest variability was mostly concentrated along the Southern Atlantic and Mediterranean coasts of the Iberian Peninsula (Figure 3b). For *Pacifastacus leniusculus*, variability among models was generally low. Nonetheless, some areas of higher disagreement were found in the northern and higher altitude regions (Figure 3c). The models for *Procambarus clarkii* also showed relatively low levels of inter-model variability, with areas of more noticeable variability being mostly confined to the west and southwest coastal areas of the Iberian Peninsula (Figure 3d).

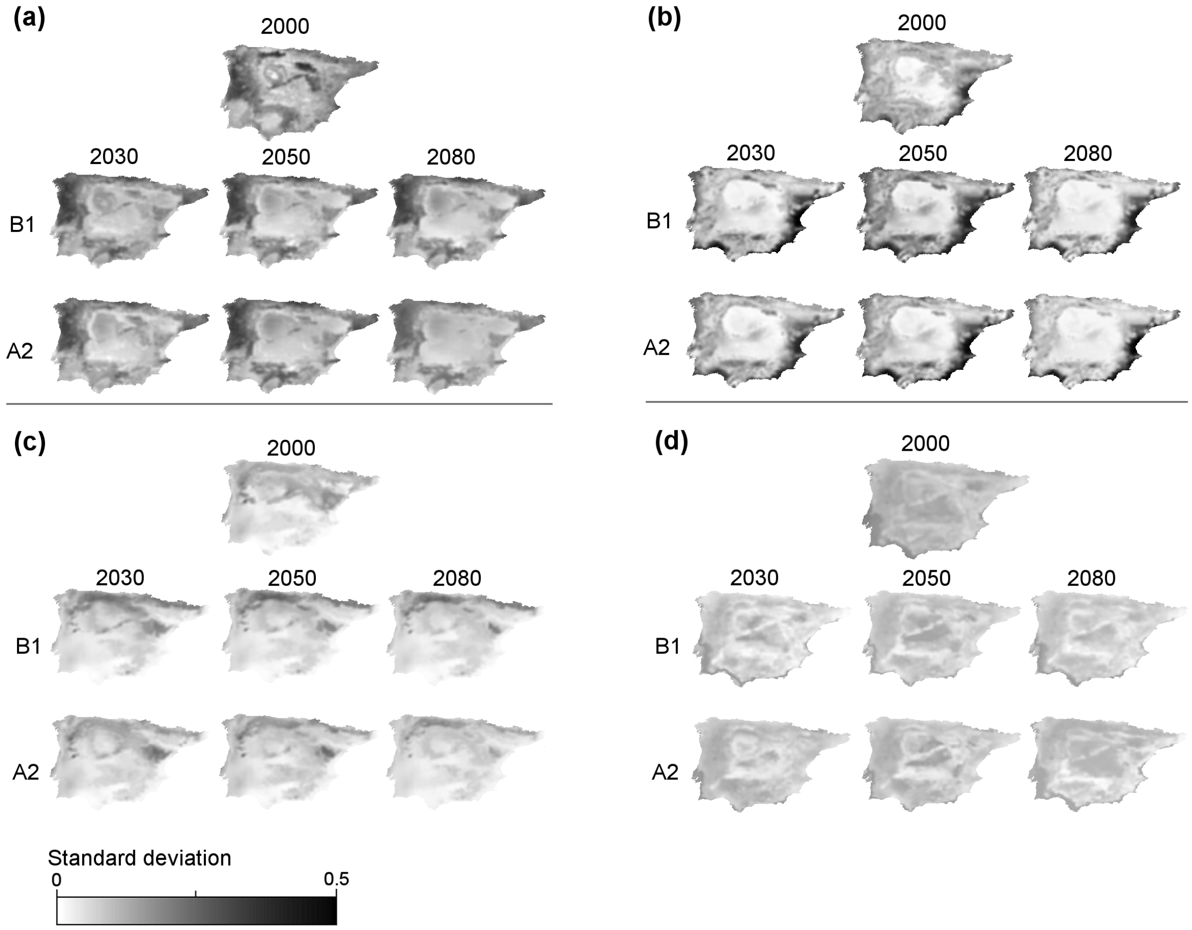


Figure 3 - Standard deviation among the single-models included in the consensus predictions for (a) *Cherax destructor*, b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii*. Predictions are made for current climate (2000) and three future time periods (2030, 2050 and 2080) under a low (B1) and high (A2) greenhouse gases emission scenario.

Areas of extrapolation occurred in the majority of the forecasts; however, their extension was highly distinct for different species and scenarios. These areas are extremely reduced or absent in the forecasts for *Procambarus clarkii*, whereas they range from approximately 11% to 51% of the total area in the predictions for *C. destructor* (see Table 1 in the Appendix D). In the models of *Pacifastacus leniusculus*, extrapolation also occurred across wide areas (from 18% to 46% of the study area), whereas for *E. sinensis* these areas were relatively smaller (2 to 23%). The magnitude of extrapolation was also variable. *Pacifastacus leniusculus* and *E. sinensis* exhibited the highest degree of extrapolation, whereas this parameter was generally lower for *C. destructor*, and was insignificant for *Procambarus clarkii* (Figure 4). Excluding the models for *Procambarus clarkii*, which were practically free from environmental extrapolation, the magnitude and geographical extent of areas where this type of extrapolation occurs increased with increasing changes in future climates for all other species.

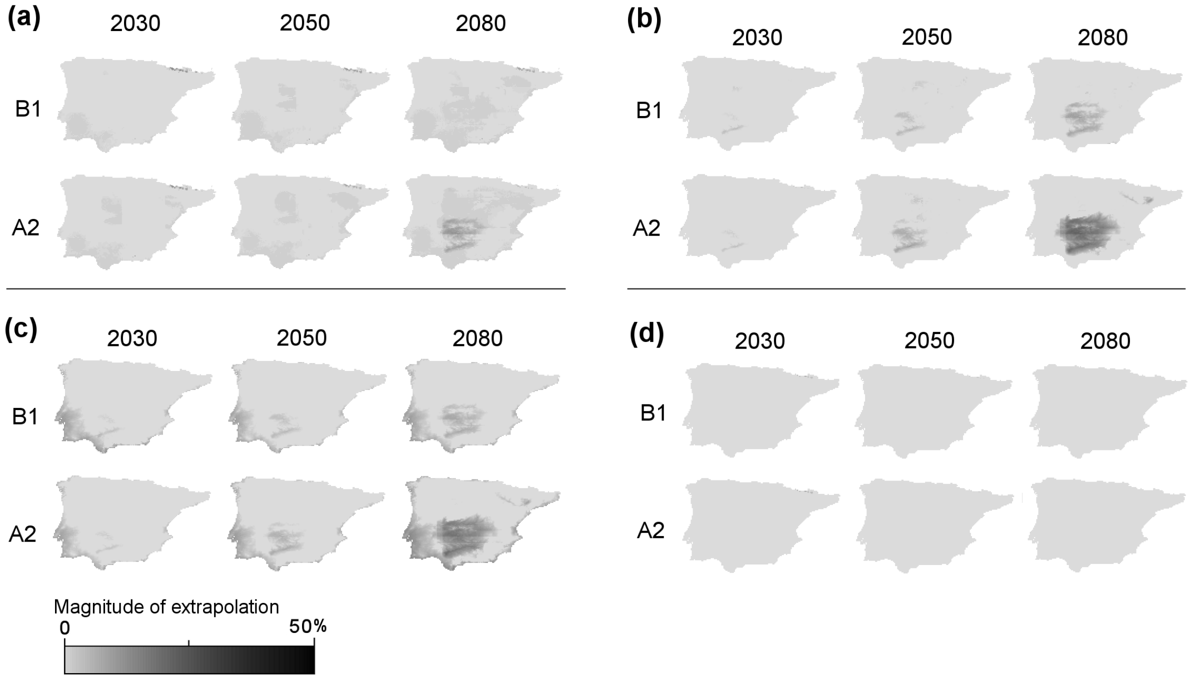


Figure 4 - Environmental extrapolation in each climatic scenario for (a) *Cherax destructor*, b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii*. Predictions are made for current climate (2000) and three future time periods (2030, 2050 and 2080) under a low (B1) and high (A2) greenhouse gases emission scenario

Discussion

Given the accelerating rate and magnitude of climate change worldwide (IPCC 2007), the need to forecast the impacts of these changes on the distribution of existing invasive species is of obvious importance. Despite other modelling approaches being available (e.g. Kearney and Porter 2009), NBMs have become the primary reference tool in forecasting these impacts because of the increased accessibility to modelling algorithms and sampling data, together with the relatively straightforward process of model building. However, previous research using NBMs to forecast future potential ranges of invaders has rarely been accompanied by explicit assessments of the uncertainty of their predictions, despite the large number of sources of uncertainty that are involved (but see Roura-Pascual et al. 2009). Our results highlight that the levels of reliability of these forecasts can be highly distinct among species and climate change projections.

How reliable are our predictions?

We found that differences in predictive performance were dependent on the statistical algorithm. RF and ADT consistently showed the highest predictive ability, followed by ANN, GLM, and finally SVG, which delivered the poorest mean accuracy. Comparative studies have shown that statistical algorithms able to generate complex responses provide a better fit to current species distributions (Elith et al. 2006; Mateo et al. 2010). However, it has been argued that this complexity arises at the expense of generalisation (Araújo and Rahbek 2006), resulting in a decline in the ability of these models to predict for new areas (e.g. invasions) or the impact of climate change. A more informative interpretation of predictive performances is to observe the patterns of performance of multiple models among species. Interestingly, our species showed a consistent pattern of distinct performances, with all models performing better for *Procambarus clarkii,* followed by *Pacifastacus leniusculus* and *E. sinensis*, and finally *C. destructor*, which received the lowest accuracies. This pattern is similar to the overall variability found in the forecasts. These species differences may be related to multiple factors, such as the degree of equilibrium with the environment, species characteristics and/or the quality and quantity of the distribution data (Elith and Leathwick 2009; Newbold 2010; Grenouillet et al. 2011) and a generalised low fit across multiple modelling techniques together with a high predictive variability, as found for *C. destructor*, indicates that the processes determining species distributions are not well represented by the sampled data.

A large number of studies have analysed the relationships between the ecological characteristics of species and the ability of models to predict current distributions (see Grenouillet et al. 2011 and references therein). These studies found that, overall, species with narrower niche breadth ('specialist' species) and narrower ranges provide the highest accuracies. However, our results contradict these findings, since the best performing species (*Procambarus clarkii*) was the most widespread invader. Yet, its physiological tolerances are similar to those of *C. destructor* (Nyström 2002), the least widespread of our species, which received the least accurate predictions. It is not clear what could have caused this pattern, but, when specifically focusing on these 2 species, we hypothesise that it could be related to (1) the degree of equilibrium of the invaders within the sampled environments or (2) data bias. For example, *Procambarus clarkii* has been deliberately introduced to many countries worldwide, and is currently invasive on all continents with the exception of Australia and Antarctica. In contrast, *C. destructor*, which has received a lower level of introduction effort worldwide, is only found in Australia and in some localised regions of Europe. Hence, the higher equilibrium of *Procambarus clarkii* with environmental conditions worldwide allows a better description of its fundamental niche, and consequently more robust predictions of its potential distribution. Furthermore, the degree of dispersal of all 4 target species in this study was also related to the availability of occurrence data. Accurate geographical data reporting either the native or invasive distribution of *C. destructor* at the spatial scale of this work was particularly difficult to find. In fact, for some relatively wide areas of its native range, we were not able to collect any occurrence data. This factor may also underlie the lower level of predictive reliability found for this species, since it may have caused an underestimation of the species climatic niche and also because it potentially increased the chances of sampling pseudo-absences in areas that are, in fact, suitable for its existence. Pseudo-absences are a well known source of uncertainty in NBMs (VanDerWal et al. 2009; Capinha et al. 2011), and, in the context of model evaluation, records that are extracted from suitable locations might cause an increase in Type I statistical errors (false positives), leading to a reduction in the discrimination values calculated by AUC. We also suspect that these records are the cause for the higher ability of models in discriminating invasive ranges compared to native ranges based on pseudo-absences extracted from the surrounding area of species' native occurrences. This is because these pseudo-absences and native occurrences are inherently spatially autocorrelated, which reduces the ability of models to differentiate between the two. These implications concur with the high influence of pseudo-absences found in previous NBM exercises and, despite a large number of studies having already assessed their effects, we believe that guidelines about the best ways to generate these models under different circumstances are still lacking.

Previous studies have highlighted the risks of predicting species distributions into new environmental spaces (Thuiller et al. 2004; Elith and Leathwick 2009; Fitzpatrick and Hargrove 2009). However, it is still not general practice to provide spatially explicit assessments about this source of uncertainty. The method we employed here is a simple but practical way of performing such assessment. We were able to geographically delimit the areas where climatic conditions were beyond the ranges used for model calibration, as well as quantify their dissimilarities. A more thorough evaluation should also test for new environmental combinations (see Fitzpatrick and Hargrove 2009; Elith et al. 2010). However, we consider that the strategy used here could be useful as a preliminary description of the problem. Our analysis highlighted remarkable differences on the spatial extent and magnitude of environmental extrapolation among species and future climate projections. Wider geographical extents of environmental extrapolation were found for *C. destructor*; however, *E. sinensis* and *Pacifastacus leniusculus* achieved higher dissimilarities to current conditions in some areas. *Procambarus clarkii* was virtually free of this source of uncertainty. In addition, the level of extrapolation was higher when the projected changes in the climate of the Peninsula were more extreme. These results highlight that, even in modelling exercises developed under the same methodological framework, the level of predictive uncertainty may be extremely distinct between species and/or scenarios of climate change. It is also interesting to verify that, despite known divergences on how models behave under extrapolation (Pearson et al. 2006; Elith and Graham 2009), a visual comparison between the maps of variability among predictions and those of environmental extrapolation reveals that the 2 factors are not related. This observation therefore indicates that maps of variability among predictions, which are being increasingly used to tackle uncertainty in NBM exercises, provide little or no signal about the uncertainty arising from environmental extrapolation.

Overall, we found distinct indicators of reliability in the predictions for the 4 sampled species, with *Procambarus clarkii* and *C. destructor* being at opposite ends and *Pacifastacus leniusculus* and *E. sinensis* falling in between. Nonetheless, this evaluation refers only to sources of uncertainty that might be assessed in models, specifically methodological issues. Ultimately, ecological uncertainty is much more difficult to evaluate, and would require different approaches, such as experimental work. For example, 1 major concern is the possibility of species shifting their climate niches because of the release of historical and geographical constraints or because of evolutionary changes (Pearman et al. 2008; Lavergne and Molofsky 2007). In fact, the species evaluated in the current study are known to have established under new environmental conditions in their invaded ranges (Capinha et al. 2011), and the possibility to continue doing so should be expected. This ability would inevitably cause the ranges predicted in this study to be an underestimate. Hence, we acted cautiously, and calibrated our models using occurrences from both native and adventive ranges, aiming to provide a better estimate of the species fundamental niche. However, this procedure does not guarantee that the entire range of conditions in which our species are able to persist have been captured, particularly for those with lower levels of introduction effort worldwide, such as *C. destructor*.

Potential distributions under future climate change

Contrary to our initial expectations, the potential distributions of *Procambarus clarkii*, *E. sinensis* and *C. destructor* are projected to contract in the future. Today, the areas least suitable for *Procambarus clarkii* are exclusively found in the higher altitude mountains of the Iberian Peninsula, which are characterized by low temperatures and high precipitation levels. However, as time progresses, some currently optimal areas are also projected to become less suitable for this species, particularly in the interior of the Iberian Peninsula where the maximum extremes of higher temperature and dryness are projected to occur. Despite this loss of suitability, the warmer and dryer areas of the Peninsula are shown to remain potentially suitable to this invader, even under the most severe climate change scenario that we simulated here (A2 in 2080). *C. destructor*, which is also a warm-water crayfish, shows a similar pattern of range change. However, in contrast to that predicted for *Procambarus clarkii*, most of the interior of the Peninsula is shown to become clearly unsuitable in the mid to long-term. We are uncertain about the future unsuitability of this area, as most studies suggest that the two species have similar tolerances to warm temperatures (Nyström 2002). In addition, the high levels of predictive uncertainty found for *C. destructor* (see above) suggest that a refinement of the modelling framework should be conducted in order to obtain higher confidence about the impact that climate change will have in its potential distribution.

For *E. sinensis*, our forecasts show that the majority of the north and northwest of the Peninsula will remain climatically suitable in the future, but an overall loss of suitability is expected to occur in southern areas. However, the models also showed a high predictive variability in most of these southern areas. In fact, no population of this species currently exists in regions with similar climatic conditions to those projected for these southern areas (e.g. in the Mediterranean basin). However, it is unclear if the reason for this absence is due to climatic constraints, or other factors, such as dispersal limitation. Thus, while the continued suitability of the northern areas seems reliable, the loss projected to occur in the south is more uncertain and should be treated with caution.

Pacifastacus leniusculus is the species with the highest loss of climatic suitability in the future. This decrease is consistent with what we expected for this invader. *Pacifastacus leniusculus* is a cold-water crayfish that is able to survive in cold climates up to a latitude of 66°N in Scandinavia (Holdich 2002). However, this species has a low tolerance to very high temperatures, a characteristic that is often linked to its inability to establish in some regions (e.g. Sahlin et al. 2010), including most of Southern Iberia (Capinha and Anastácio 2011). This relationship is also apparent in our predictions, which show the current climate in the south of the Iberian Peninsula as unsuitable for this crayfish. In both of the climatic scenarios that we tested here, these ‘southern climates’ are expected to shift north towards the Cantabric Sea (see Figure 7 in the Appendix D) which explains the severe range contraction predicted for this species.

Conclusions

As climate change progresses, the Iberian Peninsula is projected to lose climatically suitable habitats for the 4 invaders studied here. However, the reliability of the forecasts was highly variable among species. These differences support that the interpretation of NBM predictions of future potential ranges for invaders is not a straightforward process (Elith and Leathwick 2009; Elith et al. 2010). Ecologists should therefore make an effort to assess the uncertainties of their models, and, while some of these may be very difficult to quantify (e.g. niche shifts or biotic processes), others (mostly those related to methodological issues) can be evaluated by using relatively straightforward procedures. Measuring the agreement among predictions obtained from multiple statistical methods proved to be a useful approach under this context. For our 4 target species, we also found that environmental extrapolation was unrelated to the level of agreement between predictions. Predicting beyond the ranges of calibration data is an important source of uncertainty and, thus, we recommend that forecasts of climatic niches should be accompanied with maps of environmental extrapolation. Finally, we also highlighted that the differences found in the reliability of our forecasts might be related to multiple factors and, thus, we encourage future studies to investigate the conditions leading to more reliable predictions.

ACKNOWLEDGEMENTS

We thank the many researchers and institutions that provided species distribution data. We are also grateful to the associate Editor and 2 anonymous reviewers for helpful comments on the manuscript. This research work benefited from the DID (Dispersal of Invasive Decapoda) (PTDC/BIA-BEC/105182/2008) research project co-financed by the Portuguese Fundação para a Ciência e

Tecnologia (FCT), and the European Union programs COMPETE, QREN and FEDER. C.C. holds a PhD grant from FCT (SFRH/BD/41129/2007).

Page intentionally left blank

CHAPTER 6

General discussion and conclusions

1. General discussion and future perspectives

Overall, our results corroborate the strong predictive power of abiotic factors –particularly climate– in shaping the potential distribution of invasive species. Niche based models (NBM) based on the characteristics of physical environments were able to successfully predict the invasive ranges of the four invaders worldwide. Moreover, the high predictive power of abiotic factors was maintained at distinct spatial resolutions (i.e. 1 × 1 km, 10 × 10 km and 50 × 50 km grid cells) and also when the relative importance of non-abiotic factors was tested simultaneously. This consistent pattern of accuracy is supportive of the use of NBM based on environmental conditions for building risk assessments for invasive species (Elith and Leathwick 2009; Jiménez-Valverde et al. 2011; Peterson 2011; Petitpierre et al. 2012). However, we also identified both methodological and ecological/evolutionary sources of uncertainty, which can greatly reduce the model's predictive accuracy or contribute for an increase of their uncertainty.

Of particular relevance was the consistent finding that, in Europe, the Red swamp crayfish invades environmental conditions markedly distinct from the ones it occupies in its native range, a result also recently corroborated by Larson and Olden (2012). Niche shifts are a well know challenge for predicting biological invasions (Broennimann et al. 2007; Pearman et al. 2008; Gallagher et al. 2010a) and for the case of the Red swamp crayfish it remains unclear what can be behind the observed changes. At the most basic level niche shifts correspond to modifications of the species realized niche (ecological origin) and/or its fundamental niche (evolutionary origin) however, a more precise knowledge of the mechanisms causing these changes is needed for attaining more reliable predictions of environmental suitability under global change (Pearman et al. 2008; Wiens et al. 2010; Hoffmann and Sgrò 2011). Questions such as if certain characteristics of the species or environmental conditions favour the occurrence of niche shifts remain unanswered (Pearman et al. 2008). Thus, the identification of distinct levels of niche conservatism found here can be highly valuable information for future work addressing these questions. Moreover, future research specifically exploring the mechanisms behind the niche shift observed for the Red swamp crayfish would also be desirable.

Besides ecological and evolutionary uncertainty we also identified multiple sources of methodological uncertainty in our predictions. Selection of calibration data is one of the most basic steps in NBM. However, we found that some of the differing selection strategies commonly used in the literature can yield markedly distinct predictions and that, for some cases, these perform close to what would be expected at random. For minimizing the uncertainty caused by potential niche shifts our results agreed with previous research advocating the simultaneous use of native and invasive distribution data for model calibration (Broennimann and Guisan 2008; Beaumont et al., 2009a). Complementary, we found that in the absence of better information, pseudo-absences should aim at increasing the representativeness of the conditions for which predictions will be made, even if this represents sampling areas for which the degree of propagule pressure is not known (e.g. worldwide or in the surrounding of native ranges). An alternative approach has been advocated by Elith et al. (2010) which consist on making the model's response constant outside of the range of the training data (i.e. 'clamping'). Future work should aim at testing the advantages and limitations of one method over the other as well as their complementarities. Regardless of the method used, we also found that future analyses on the uncertainty of NBM predictions benefit from the analysis of maps of environmental extrapolation. This is because this source of uncertainty is not captured by maps of inter-model variability which are the most commonly used method for evaluating predictive uncertainty.

Several other methodological factors showed to have clear implications on the reliability of the predictions. The clearest of these referred to the dissimilarities among predictions from distinct statistical algorithms, but these were also found among distinct model parameterizations or initial conditions (i.e. dependent and independent calibration data). In accordance with previous research, we found that ensemble modelling and consensus methods are a valuable framework to deal with the distinct outcomes (e.g. Araújo and New 2007; Marmion et al. 2009). In this thesis we further developed this approach for assessing species-environment relationships allowing to gain ecological insight and/or to validate the rationale of the predictions. Using this framework we identified temperature as the most important factor – among a group of multiple abiotic factors – in shaping the potential distribution of the three crayfish species in the Iberian Peninsula. Differences in their potential distributions were found to be mainly originated by distinct responses along thermal gradients. The complementary use of maps of inter-model variability showed to be also of importance because we found high variability even among predictions with good discrimination accuracy. By being spatially explicit, this measure allowed identifying the areas least consensual among predictions. A very clear example of the importance of this type of evaluation was found in the predictions of future climatic suitability of the Iberian Peninsula for the Chinese mitten crab, since the areas with higher model disagreement were precisely those projected to lose climatic suitability to the species.

Taken as a whole, our results give support to the combined use of distinct techniques for evaluating the reliability of NBM predictions. Traditional evaluation approaches have almost entirely relied on the ability of models to discriminate between the species current occurrences and absences (or pseudo-absences). However, this approach has been criticized when models aim at predicting across space and/or time because known species distributions –and inherently the validation data– do not reflect the fundamental niches, but instead the realized niches (Jiménez-Valverde et al. 2008). Thus, alternative or complementary approaches should be considered for predictions across space and/or time. Accordingly, we complemented traditional evaluation techniques with the approaches discussed above and found that better estimates of predictive reliability can be obtained.

Our analysis also highlighted a high invasion potential for the four invasive decapods studied. Working at macro-scales we identified a large number of environmentally suitable areas worldwide, the majority of which remaining unoccupied. Similar results were found for the Iberian Peninsula concerning the three least spread invaders (Chinese mitten crab; Signal crayfish and Yabby). These results were also recently supported by independent studies working with the Red swamp crayfish and the Signal crayfish (Liu et al. 2011a; Larson and Olden 2012). It is also expected that climatic changes will cause modifications on the potential ranges of the four invaders. We here focused on the Iberian Peninsula and found that the magnitude of the expected changes varies among species but also that a decline in the extent of suitable areas is expected for all four species as time progresses. These patterns are location specific and with the exception of the Red swamp crayfish (Liu et al. 2011a) the implications of climate change for these invaders at the global scale remain undetermined, thus future work on this matter would be valuable.

The importance of estimates of environmental suitability was also emphasized by our results which suggested that, in Europe, the current distribution of the Red swamp crayfish and the Signal crayfish is largely associated with country-level propagule pressure and climatic suitability of receiving areas. Other well known invasion mechanisms such as biotic resistance or human disturbance showed a reduced or null contribution to the process. Given data limitations we were only able to perform this assessment for Europe and using coarse-grained data. It is thus not possible to know if this pattern can be extrapolated to other invaded regions of the globe or at finer resolutions; however, it clearly indicates a strong invasive ability for the two species. If more detailed data arises in the future we suggest performing similar analyses for new areas and at multiple spatial resolutions.

Our results also allowed a deeper insight about the environmental conditions potentially determining the successful establishment of each invader. From the multiple abiotic factors tested, thermal factors consistently showed to be of high importance. The Red swamp crayfish and the Yabby evidenced a higher preference for warmer conditions than the Signal crayfish, a species more tolerant to cold. However, distinct tolerances to low temperatures were found between the Red swamp crayfish and the Yabby. Proximity to coastal waters was also an expected requirement of the Chinese mitten crab. Future work could use this sort of information (i.e. correlative) for comparison and/or integration with data originated from physiological experiments, potentially yielding a deeper knowledge about the invader's fundamental niches (e.g. Monahan 2009).

Knowledge about current and future potential ranges of the four invasive decapods is of specific interest for conservation. The patterns of climatic suitability expressed by our models indicate a large overlap of suitable areas and biodiversity hotspots worldwide (Myers et al. 2000). Although some of these hotspots may be largely representing species immune to the negative effects of these invaders (e.g. birds) others comprise a large number of endemic taxa potentially affected (e.g. Mediterranean amphibians; Cruz et al. 2008; Ficetola et al. 2011). This thesis did not specifically aim at estimating the impacts of the invaders in nature. However, future work could further extend our analyses to obtain more precise estimates of potential biodiversity loss or for supporting conservation actions. For instance, given the large amount of evidence that North American invasive crayfish are a primary driver of decline of European crayfish fauna (Holdich et al. 2009), a certainly valuable future research would be to estimate current and future overlap of potential ranges between these two groups of species.

2. Conclusions

Overall, our results agree with the general perception that the predictive ability of invasion ecology in general, and niche based models for invasive species in particular, have increased in recent years (Davis 2009; Elith and Leathwick 2009; Petitpierre et al. 2012). Using correlative models we were able to predict the current invasive ranges of four problematic decapods at multiple spatial scales. In addition, it was also found that a thorough evaluation of multiple sources of uncertainty can be useful for investigating the reliability of these predictions across space and/or time. However, our results also highlighted multiple methodological issues that should continue being examined in order to obtain more reliable predictions. Among these are even basic aspects of model building such as the selection of calibration data or statistical model. Future research may particularly benefit from the increasing availability of large-scale datasets of species distributions (e.g. the Global Biodiversity Information Facility). We suggest using these data for revaluating long-standing questions such as the identification of 'ideal' statistical algorithms, desirable levels of model complexity or data requirements. Answering questions such as these would be certainly valuable for increasing the predictive ability (and overall reliability) of invasion ecology.

Besides methodological sources of uncertainty, it is also important to examine potential ecological and/or evolutionary changes. Similarly to previous studies (e.g. Broennimann et al. 2007; Gallagher et al. 2010a) we found that invaders are able to shift their niches (i.e. occupy environmental conditions distinct from those occupied previously). Much of these shifts can be originated from changes on the invaders realized niches (Perman et al. 2008; Rödder et al. 2009) thus emphasizing the importance of an increase in the integration of biotic factors and dispersal barriers into the analyses. Similarly, a closer integration with evolutionary models and phylogenetics may improve the anticipation of changes on fundamental niches (Perman et al. 2008).

Our analyses also indicate that in the short-term the four decapods studied will continue to expand their ranges worldwide. In the Iberian Peninsula this may also occur, but in the long-term it is expected an overall hampering effect of climate change. In fact, climate and particularly thermal conditions showed a strong correlation with the distribution of these species. Thus, we suggest that estimates of potential distribution –and associated impacts– of these invaders in new regions should consider the fast changes predicted for future climates (IPCC 2007).

Page intentionally left blank

LIST OF REFERENCES

Aldridge DC, Elliott P, Moggridge GD (2005) Microencapsulated BioBullets for the control of biofouling Zebra mussels. *Environmental Science & Technology* 40:975-979

Ahlroth P, Alatalo RV, Holopainen A, et al. (2003) Founder population size and number of source populations enhance colonization success in waterstriders. *Oecologia* 137:617-620

Anastácio P, Nielsen S, Marques J (1999) CRISP (crayfish and rice integrated system of production): 2. Modelling crayfish (*Procambarus clarkii*) population dynamics. *Ecological Modelling* 123:5-16

Anastácio PM, Parente VS, Correia AM (2005) Crayfish effects on seeds and seedlings: identification and quantification of damage. *Freshwater biology* 50:697-704

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

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22:42-47

Araújo MB, Rahbek C (2006) How does climate change affect biodiversity? *Science* 313:1396

Araújo MB, Whittaker RJ, Ladle RJ, et al. (2005) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography* 14:529-538

Arce JA, Alonso F (2011) Factors related to the presence of the *Austropotamobius pallipes* (Lereboullet, 1858) species complex in calcareous mountin rivers in central Spain. *Knowledge and Management of Aquatic Ecosystems*:25

Austin MP, Belbin L, Meyers JA, et al. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory. *Ecological Modelling* 199:197-216

Austin MP, Smith TM, Van Niel KP, et al. (2009) Physiological responses and statistical models of the environmental niche: a comparative study of two co-occurring Eucalyptus species. *Journal of Ecology* 97:496-507

Austin MP, Van Niel KP (2011) Improving species distribution models for climate change studies: variable selection and scale. *Journal of Biogeography* 38:1-8

Bahn V, McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography* 16:733-742

Beatty SJ (2006) The diet and trophic positions of translocated, sympatric populations of *Cherax destructor* and *Cherax cainii* in the Hutt River, Western Australia: evidence of resource overlap. *Marine and Freshwater Research* 57:825-835

Beaumont LJ, Gallagher RV, Thuiller W, et al. (2009a) Different climatic envelopes among invasive populations may lead to underestimations of current and future biological invasions. *Diversity and Distributions* 15:409-420

Beaumont LJ, Gallagher RV, Downey PO, et al. (2009b) Modelling the impact of *Hieracium* spp. on protected areas in Australia under future climates. *Ecography* 32:757-764

Beaumont LJ, Hughes L, Pitman AJ (2008) Why is the choice of future climate scenarios for species distribution modelling important? *Ecology Letters* 11:1135-1146

Bernardo JM, Costa AM, Bruxelas S, et al. (2011) Dispersal and coexistence of two non-native crayfish species (*Pacifastacus leniusculus* and *Procambarus clarkii*) in NE Portugal over a 10-year period. *Knowledge and Management of Aquatic Ecosystems* 401:e28

Berry ZC, Wevill K, Curran TJ (2011) The invasive weed *Lantana camara* increases fire risk in dry rainforest by altering fuel beds. *Weed Research* 51:525-533

Blossey B, Notzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants: a hypothesis. Journal of Ecology 83:887-889

Bolea L (1996) Primera cita de *Cherax destructor* (Crustacea: Decapoda: Parastacidae) en Europa. *Boletín de la Sociedad Entomológica Aragonesa* 14:49-51

Bondar CA, Bottriell K, Zeron K, et al. (2005) Does trophic position of the omnivorous signal crayfish (*Pacifastacus leniusculus*) in a stream food web vary with life history stage or density? *Canadian Journal of Fisheries and Aquatic Sciences* 62:2632-2639

Breiman L (2001) Random forests. *Machine learning* 45:5-32

Broennimann O, Guisan A (2008) Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters* 4:585-589

Broennimann O, Treier UA, Müller-Schärer H, et al. (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters* 10:701-709

Brotons L, Thuiller W, Araújo MB, et al. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437-448

Buisson L, Thuiller W, Casajus N, et al. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology* 16:1145-1157

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag

Cabral H, Costa M (1999) On the occurrence of the Chinese mitten crab, *Eriocheir sinensis*. Portugal (Decapoda, Brachyura). *Crustaceana* 72:55-58

Caissie D (2006) The thermal regime of rivers: a review. *Freshwater biology* 51:1389-1406

Capinha C, Anastácio P (2011) Assessing the environmental requirements of invaders using ensembles of distribution models. *Diversity and Distributions* 17:13-24

Capinha C, Broton L, Anastácio P (unpublished) Geographic variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish

Capinha C, Leung B, Anastácio P (2011) Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. *Ecography* 34:448-459

Cassey P, Blackburn TM, Sol D, et al. (2004) Global patterns of introduction effort and establishment success in birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271:405-408

Catford JA, Jansson R, Nilsson C (2009) Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Diversity and Distributions* 15:22-40

Chefaoui RM, Lobo JM (2008) Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210:478-486

Chiron F, Shirley S, Kark S (2009) Human-related processes drive the richness of exotic birds in Europe. *Proceedings of the Royal Society B: Biological Sciences* 276:47

Clavero M, Brotons L, Pons P, et al. (2009) Prominent role of invasive species in avian biodiversity loss. *Biological Conservation* 142:2043-2049

Clavero M, Hermoso V, Brotons L, et al. (2010) Natural, human and spatial constraints to expanding populations of otters in the Iberian Peninsula. *Journal of Biogeography* 37:2345-2357

Clavero M, Prenda J, Delibes M (2003) Trophic diversity of the otter (*Lutra lutra* L.) in temperate and Mediterranean freshwater habitats. *Journal of Biogeography* 30:761-769

Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20:37-46

Cohen AN, Carlton JT (1997) Transoceanic transport mechanisms: The introduction of the Chinese mitten crab *Eriocheir sinensis* to California. *Pacific Science* 51:1-11

Colautti RI (2005) Are characteristics of introduced salmonid fishes biased by propagule pressure? *Canadian Journal of Fisheries and Aquatic Sciences* 62:950-959

Colautti RI, Ricciardi A, Grigorovich IA, et al. (2004) Is invasion success explained by the enemy release hypothesis? *Ecology Letters* 7:721-733

Colwell RK, Rangel TF (2009) Hutchinson's duality: The once and future niche. *Proceedings of the National Academy of Sciences* 106:19651-19658

Correia AM (2003) Food choice by the introduced crayfish *Procambarus clarkii*. *Annales Zoologici Fennici* 40:517-528

Correia AM, Anastácio PM (2008) Shifts in aquatic macroinvertebrate biodiversity associated with the presence and size of an alien crayfish. *Ecological Research* 23:729-734

Correia AM, Ferreira Ó (1995) Burrowing behavior of the introduced red swamp crayfish *Procambarus clarkii* (Decapoda: Cambaridae) in Portugal. *Journal of Crustacean Biology* 15:248-257

Costa AM, Bruxelas S, Bernardo J, Teixeira A (2009) Colonization of Rio Maçãs (North Portugal) by two exotic crayfish *Pacifastacus leniusculus* and *Procambarus clarkii*. In: Bioleif (ed) *World Conference on Biological Invasions and Ecosystem Functioning*. Candeias Artes Gráficas, Oporto

Côté IM, Reynolds JD (2002) Predictive ecology to the rescue? *Science* 298:1181-1182

Craig JF (1992) Human-induced changes in the composition of fish communities in the African Great Lakes. *Reviews in Fish Biology and Fisheries* 2:93-124

Cruz M, Segurado P, Sousa M, et al. (2008) Collapse of the amphibian community of the Paul do Boquilobo Natural Reserve (central Portugal) after the arrival of the exotic American crayfish *Procambarus clarkii*. *The Herpetological Journal* 18:197-204

Cuesta J, González-Ortegón E, Rodríguez A, et al. (2006) The decapod custacean community of the Guadalquivir Estuary (SW Spain): Seasonal and inter-year changes in community structure. *Hydrobiologia* 557:85-95

Cumberlidge N (2009) Freshwater Crabs and Shrimps (Crustacea: Decapoda) of the Nile Basin. In: Dumont HJ (ed) *The Nile*. Springer, pp. 547-561.

Davis MA (2009) *Invasion Biology*. Oxford University Press, Oxford

Davis MA (2011) Invasion biology. In: Simberloff D and Rejmánek M (eds) *Encyclopedia of biological invasions*. University of California Press, London, pp. 364-369

DeRivera CE, Ruiz GM, Hines AH, et al. (2005) Biotic resistance to invasion: native predator limits abundance and distribution of an introduced crab. *Ecology* 86:3364-3376

DeWalt SJ, Denslow JS, Ickes K (2004) Natural-enemy release facilitates habitat expansion of the invasive tropical shrub *Clidemia hirta*. *Ecology* 85:471-483

Diniz-Filho JAF, Mauricio Bini L, Fernando Rangel T, et al. (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* 32:897-906

Diniz-Filho JAF, Nabout JC, Bini LM, et al. (2010) Ensemble forecasting shifts in climatically suitable areas for *Tropidacris cristata* (Orthoptera: Acridoidea: Romaleidae). *Insect Conservation and Diversity* 3:213-221

Dittel AI, Epifanio CE (2009) Invasion biology of the Chinese mitten crab *Eriochier sinensis*: A brief review. *Journal of Experimental Marine Biology and Ecology* 374:79-92

Doadrio I (2001) *Atlas y Libro Rojo de los Peces Continentales de España*. Dirección General de Conservación de la Naturaleza and Museo Nacional de Ciencias Naturales, Madrid

Dormann C (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology* 8:387-397

Dormann C, McPherson J, Araújo M, et al. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609-628

Dormann CF, Purschke O, Marquez JRG, et al. (2008) Components of uncertainty in species distribution analysis: A case study of the Great Grey Shrike. *Ecology* 89:3371-3386

Drake JA (2009) *Handbook of alien species in Europe*. Springer, Dordrecht

Drake JM, Lodge DM (2006) Allee effects, propagule pressure and the probability of establishment: risk analysis for biological invasions. *Biological Invasions* 8:365-375

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

Dukes J, Mooney H (1999) Does global change increase the success of biological invaders? *Trends in Ecology and Evolution* 14:135-139

Dunn JC, McClymont HE, Christmas M, et al. (2009) Competition and parasitism in the native White Clawed Crayfish *Austropotamobius pallipes* and the invasive Signal Crayfish *Pacifastacus leniusculus* in the UK. *Biological Invasions* 11:315-324

Elith J, Ferrier S, Huettmann F, et al. (2005) The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling* 186:280-289

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

Elith J, Graham CH, Anderson RP, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151

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

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

Elith J, Leathwick J, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802-813

Elton CS (1958) *The ecology of invasions by animals and plants*. Methuen, London

Essl F, Dullinger S, Rabitsch W, et al. (2011) Socioeconomic legacy yields an invasion debt. *Proceedings of the National Academy of Sciences* 108:203-207

Ficetola GF, Siesa ME, Manenti R, et al. (2011) Early assessment of the impact of alien species: differential consequences of an invasive crayfish on adult and larval amphibians. *Diversity and Distributions* 17:1141-1151

Ficetola GF, Thuiller W, Miaud C (2007) Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. *Diversity and Distributions* 13:476-485

Ficetola GF, Thuiller W, Padoa-Schioppa E (2009) From introduction to the establishment of alien species: bioclimatic differences between presence and reproduction localities in the slider turtle. *Diversity and Distributions* 15:108-116

Fitzpatrick M, Hargrove W (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation* 18:2255-2261

Fitzpatrick MC, Weltzin JF, Sanders NJ, et al. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* 16:24-33

Font I (2000) *Climatología de España y Portugal*. Ediciones Universidad Salamanca, Salamanca

Freund Y, Mason L (1999) The alternating decision tree learning algorithm. *Proceeding of the Sixteenth International Conference on Machine Learning*. Bled, pp. 124-133

Fridley JD, Stachowicz JJ, Naeem S, et al. (2007) The invasion paradox: reconciling pattern and process in species invasions. *Ecology* 88:3-17

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: A statistical view of boosting. *The annals of statistics* 28:337-374

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *The annals of statistics* 29:1189-1232

Friedman JH, Meulman JJ (2003) Multiple additive regression trees with application in epidemiology. *Statistics in medicine* 22:1365-1381

Galil BS, Nehring S, Panov V (2007) Waterways as Invasion Highways – Impact of Climate Change and Globalization. In: Nentwig W (ed) *Biological Invasions*. Springer Berlin Heidelberg, pp. 59-74

Gallagher RV, Beaumont LJ, Hughes L, et al. (2010a) Evidence for climatic niche and biome shifts between native and novel ranges in plant species introduced to Australia. *Journal of Ecology* 98:790-799

Gallagher R, Hughes L, Leishman M, et al. (2010b) Predicted impact of exotic vines on an endangered ecological community under future climate change. *Biological Invasions* 12:4049-4063

García-Berthou E, Boix D, Clavero M (2007) Non-indigenous animal species naturalized in Iberian inland waters. In: Gherardi F (ed) Biological *invaders in inland waters: Profiles, distribution, and threats*. Springer, Dordrecht, pp. 123-140

Garcia-de-Lomas J, Dana E, López-Santiago J, et al. (2010) Management of the Chinese mitten crab, *Eriocheir sinensis* (H. Milne Edwards, 1853) in the Guadalquivir Estuary (Southern Spain). *Aquatic Invasions* 5:323-330

Geiger W, Alcorlo P, Baltan, et al. (2005) Impact of an introduced Crustacean on the trophic webs of Mediterranean wetlands. *Biological Invasions* 7:49-73

Gherardi F (2006) Crayfish invading Europe: the case study of *Procambarus clarkii*. *Marine and Freshwater Behaviour and Physiology* 39:175-191

Gherardi F (2007) Biological invasions in inland waters: an overview. In: Gherardi F (ed*) Biological invaders in inland waters: Profiles, distribution, and threats*. Springer Netherlands, pp. 3-25

Gherardi F, Aquiloni L, Diéguez-Uribeondo J, et al. (2011) Managing invasive crayfish: is there a hope? *Aquatic Sciences - Research Across Boundaries* 73:185-200

Gherardi F, Cioni A (2004) Agonism and interference competition in freshwater decapods. *Behaviour*:1297-1324

Gilbey V, Attrill MJ, Coleman RA (2008) Juvenile Chinese mitten crabs (*Eriocheir sinensis*) in the Thames estuary: distribution, movement and possible interactions with the native crab *Carcinus maenas*. *Biological Invasions* 10:67-77

Gollasch S (2006) NOBANIS – Invasive Alien Species Fact Sheet – *Eriocheir sinensis*. Online Database of the North European and Baltic Network on Invasive Alien Species – NOBANIS [Available at: www.nobanis.org; accessed 12/04/2010]

Graham CH, Elith J, Hijmans RJ, et al. (2008) The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* 45:239-247

Grenouillet G, Buisson L, Casajus N, et al. (2011) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34:9-17

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

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

Guisan A, Zimmermann NE, Elith J, et al. (2007) What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics?. *Ecological Monographs* 77:615-630

Gurevitch J, Fox GA, Wardle GM, et al. (2011) Emergent insights from the synthesis of conceptual frameworks for biological invasions. *Ecology Letters* 14:407-418

Gutiérrez-Yurrita PJ, Martinez JM, Bravo-Utrera MA, et al. (1999) The status of crayfish populations in Spain and Portugal. In: Gherardi F and Holdich DM (eds) *Crayfish in Europe as alien species. How to make the best of a bad situation*. AA Balkema, Rotterdam, pp. 161-192

Hall M, Frank E, Holmes G, et al. (2009) The WEKA data mining software: an update. *SIGKDD Explorer Newsletter* 11:10-18

Hanley J, McNeil B (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29-36

Hartley S, Harris R, Lester PJ (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters* 9:1068-1079

Hashim A (2010) Occurrence of the chinese mitten crab *Eriocheir sinensis* (H. Milne Edwards) in South Iraq. *Mesopotamian Journal of Marine Science* 25:31-36

Hayes K, Barry S (2008) Are there any consistent predictors of invasion success? *Biological Invasions* 10:483-506

Heinimaa S, Pursiainen M (2008) Signal crayfish *Pacifastacus leniusculus* at northerly latitudes: A search for the distribution limits. *Freshwater Crayfish* 16:37-41

Hellmann J, Byers J, Bierwagen B, et al. (2008) Five potential consequences of climate change for invasive species. *Conservation Biology* 22:534-543

Henttonen P, Hunner JV (1999) The introduction of alien species of crayfish in Europe: A historical introduction. In: Gherardi F and Holdich DM (eds) *Crayfish in Europe as alien species: How to make the best of a bad situation?* AA Balkema, Rotterdam, pp. 13-22

Hijmans RJ, Cameron SE, Parra JL, et al. (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

Hirzel AH, Le Lay G (2008) Habitat suitability modelling and niche theory. *Journal of Applied Ecology* 45:1372-1381

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature* 470:479-485

Hogger JB (1986) Aspects of the introduction of "signal crayfish", *Pacifastacus leniusculus* (Dana), into the southern United Kingdom. 1. Growth and survival. *Aquaculture* 58:27-44

Holdich DM (2002) Distribution of crayfish in Europe and some adjoining countries. *Bulletin Français de la Pêche et de la Pisciculture* 367:611-650

Holdich DM, Pöckl M (2005) Does legislation work in protecting vulnerable species? Proceeding of CRAYNET Innsbruck conference 2004. *Bulletin Français de la Pêche et de la Pisciculture* 376-377:809-827

Holdich D, Pöckl M (2007) Invasive crustaceans in European inland waters In: Gherardi F (ed*) Biological invaders in inland waters: Profiles, distribution, and threats*. Springer Netherlands, pp. 29-75

Holdich D, Reynolds J, Souty-Grosset C, et al. (2009) A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. *Knowledge and Management of Aquatic Ecosystems* 394-395:e46

Hosmer DW, Lemeshow S (1989) Applied logistic regression. Wiley, New York

Hulme PE, Pyšek P, Nentwig W, et al. (2009) Will threat of biological invasions unite the European Union. *Science* 324:40-41

Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:415-442

Ilhéu M, Bernardo J, Fernandes S (2007) Predation of invasive crayfish on aquatic vertebrates: the effect of *Procambarus clarkii* on fish assemblages in Mediterranean temporary streams. In: Gherardi F (ed) *Biological invaders in inland waters: Profiles, distribution, and threats*. Springer Netherlands, pp. 543-558

IPCC (2007) Climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, et al. (eds) *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.

IUCN (2009) *Red List of Threatened Species* [Available at: http://www.iucnredlist.org; accessed 12/05/2010]

Jarvis, A., Reuter, H.I., Nelson, A. & Guevara E. (2008) *Hole-filled seamless SRTM data V4*, International Centre for Tropical Agriculture (CIAT), [Available at: http://srtm.csi.cgiar.org; accessed 06/09/2009]

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

Jiménez-Valverde A, Peterson A, Soberón J, et al. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13:2785-2797

Johnsen S, Taugbøl T, Andersen O, et al. (2007) The first record of the non-indigenous signal crayfish *Pacifastacus leniusculus* in Norway*. Biological Invasions* 9:939-941

Jones PG, Thornton PK, Heinke J (2009) Generating characteristic daily weather data using downscaled climate model data from the IPCC's Fourth Assessment pp. 19 [Available at: https://hc.box.net/shared/f2gk053td8; accessed 15/02/2011]

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution* 16:199-204

Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution* 17:164-170

Kearney M, Phillips BL, Tracy CR, et al. (2008) Modelling species distributions without using species distributions: the cane toad in Australia under current and future climates. *Ecography* 31:423-434

Kearney M, Porter W (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* 12:334-350

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics*:159-174

Larson ER, Olden JD (2008) Do schools and golf courses represent emerging pathways for crayfish invasions. *Aquatic Invasions* 3:465-468

Larson E. Olden J (2012) Using avatar species to model the potential distribution of emerging invaders. *Global Ecology and Biogeography* DOI: 10.1111/j.1466-8238.2012.00758.x.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences* 104:3883

Lek S, Guégan JF (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120:65-73

Le Lay G, Engler R, Franc E, et al. (2010) Prospective sampling based on model ensembles improves the detection of rare species. *Ecography* 33:1015-1027

Leung B, Mandrak NE (2007) The risk of establishment of aquatic invasive species: joining invasibility and propagule pressure. *Proceedings of the Royal Society B: Biological Sciences* 274:2603-2609

Lindqvist OV, Huner J (1999) Life history characteristics of crayfish: What makes some of them good colonizers. In: Gherardi F and Holdich DM (eds) *Crayfish in Europe as alien species. How to make the best of a bad situation*. AA Balkema, Rotterdam, pp. 23-30

Liu C, Berry PM, Dawson TP, et al. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385-393

Liu X, Guo Z, Ke Z, et al. (2011a) Increasing potential risk of a global aquatic invader in Europe in contrast to other continents under future climate change. *PloS one* 6:e18429

Liu C, White M, Newell G (2011b) Measuring and comparing the accuracy of species distribution models with presence–absence data. *Ecography* 34:232-243

Lobo JM (2008) More complex distribution models or more representative data? *Biodiversity Informatics* 5:14-19

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution 20:223-228

Lockwood JL, Cassey P, Blackburn TM (2009) The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions* 15:904-910

Lockwood JL, Hoopes MF, Marchetti MP (2007) *Invasion ecology*. Blackwell Publishing Malden

Loo SE, Nally RM, Lake P (2007) Forecasting New Zealand mudsnail invasion range: model comparisons using native and invaded ranges. *Ecological Applications* 17:181-189

Lounibos LP (2011) Human disease vectors. In: Simberloff D and Rejmánek M (eds) *Encyclopedia of biological invasions*. University of California Press, London, pp. 150-154

Lowe S, Browne M, Boudjelas S (2000) 100 of the world’s worst invasive alien species. A selection from the global invasive species database. Invasive Species Specialist Group, Auckland, New Zealand [Available at: www.issg.org/booklet.pdf; accessed 03/01/2010]

Lynas J, Lindhjem PA, Storey A, et al. (2004) Is the Yabby, *Cherax destructor* (Parastacidae) in Western Australia an ecological threat? *Freshwater Crayfish* 14:37-44

Lynas J, Storey A, Knott B (2007) Introduction and spread of crayfish (Parastacidae) in Western Australia and their potential to displace indigenous species. In: Gherardi F (ed) *Biological invaders in inland waters: Profiles, distribution, and threats*. Springer Netherlands, pp. 577-596

Marmion M, Parviainen M, Luoto M, et al. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions* 15:59-69

Mateo RG, Croat TB, Felicísimo ÁM, et al. (2010) Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions* 16:84-94

Mau-Crimmins TM, Schussman HR, Geiger EL (2006) Can the invaded range of a species be predicted sufficiently using only native-range data?: Lehmann lovegrass (*Eragrostis lehmanniana*) in the southwestern United States. *Ecological Modelling* 193:736-746

McCullagh P, Nelder J (1989) Generalized Linear Models. Chapman and Hall/CRC

McGeoch MA, Butchart SHM, Spear D, et al. (2010) Global indicators of biological invasion: species numbers, biodiversity impact and policy responses. *Diversity and Distributions* 16:95-108

Meyer J-Y, Medeiros AC (2011) Melastomes. In: Simberloff D and Rejmánek M (eds) *pedia of biological invasions*. University of California Press Ltd., London, pp. 458-462

Myers N, Mittermeier RA, Mittermeier CG, et al. (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853-858

Millennium Ecosystem Assessment, (2005) *Ecosystems and human well-being: biodiversity synthesis*, Island Press, Washington DC.

Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* 25:693-712

Monahan WB (2009) A mechanistic niche model for measuring species' distributional responses to seasonal temperature gradients. *PloS one* 4:e7921

Morrissy N (1994) Yabby frequently asked questions. Department of Fisheries of the Government of Western Australia [Available at: http://www.fish.wa.gov.au/docs/aq/aq001/index.php; accessed 20/03/2012]

Muñoz AR, Real R (2006) Assessing the potential range expansion of the exotic monk parakeet in Spain. *Diversity and Distributions* 12:656-665

Nakata K, Hamano T, Hayashi KI, et al. (2002) Lethal limits of high temperature for two crayfishes, the native species *Cambaroides japonicus* and the alien species *Pacifastacus leniusculus* in Japan. *Fisheries Science* 68:763-767

Nakicenovic N, Swart R (2000) *Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge

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

Nyström P (1999) Ecological impact of introduced and native crayfish on freshwater communities: European perspectives. In: Gherardi F and Holdich DM (eds) *Crayfish in Europe as alien species. How to make the best of a bad situation*. AA Balkema, Rotterdam, pp.63-85

Nyström P (2002) Ecology. In: Holdich DM (ed) *Biology of freshwater crayfish*. Blackwell Science, Oxford, pp. 192–235

Nuñez MA, Moretti A, Simberloff D (2011) Propagule pressure hypothesis not supported by an 80-year experiment on woody species invasion. *Oikos* 120:1311–1316

Özesmi SL, Tan CO, Özesmi U (2006) Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling* 195:83-93

Paglianti A, Gherardi F (2004) Combined effects of temperature and diet on growth and survival of young-of-year crayfish: A comparison between indigenous and invasive species. *Journal of Crustacean Biology* 24:140-148

Panning A (1939) The Chinese mitten crab. Smithsonian Institute Annual Report. pp. 361-375

Pearce JL, Boyce MS (2006) Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43:405-412

Pearman PB, Guisan A, Broennimann O, et al. (2008) Niche dynamics in space and time. *Trends in Ecology and Evolution* 23:149-158

Pearson DE, McKelvey KS, Ruggiero LF (2000) Non-target effects of an introduced biological control agent on deer mouse ecology. *Oecologia* 122:121-128

Pearson R, Thuiller W, Araújo M, et al. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography* 33:1704-1711

Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78:419-433

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

Peterson AT, Papes M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550-560

Peterson AT, Vieglais DA (2001) Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *Bioscience* 51:363-371

Petitpierre B, Kueffer C, Broennimann O, et al. (2012) Climatic Niche Shifts Are Rare Among Terrestrial Plant Invaders. *Science* 335:1344-1348

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

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231-259

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161-175

Phillips SJ, Dudík M, Elith J, et al. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19:181-197

Pimentel D, McNair S, Janecka J, et al. (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems & Environment* 84:1-20

Pfahringer B, Holmes G, Kirkby R (2001) Optimizing the Induction of Alternating Decision Trees. In: Cheung D, Williams G and Li Q (eds) *Advances in Knowledge Discovery and Data Mining*. Springer Berlin / Heidelberg, pp. 477-487

Pyšek P, Jarošík V, Hulme PE, et al. (2010) Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences* 107:12157

R Development Core Team (2009) *R: a language and environment for statistical computing, version 2.9.0*. R Foundation for Statistical Computing, Vienna, Austria.

R Development Core Team (2011) *R: a language and environment for statistical computing, version 2.12.1*. R Foundation for Statistical Computing, Vienna, Austria.

Rahel F, Olden J (2008) Assessing the effects of climate change on aquatic invasive species. *Conservation Biology* 22:521-533

Rahmstorf S, Cazenave A, Church JA, et al. (2007) Recent climate observations compared to projections. *Science* 316:709

Régnier C, Fontaine B, Bouchet P (2009) Not knowing, not recording, not listing: numerous unnoticed mollusk extinctions. *Conservation Biology* 23:1214-1221

Ricciardi A (2007) Are Modern Biological Invasions an unprecedented form of global change? *Conservation Biology* 21:329-336

Ricciardi A (2011) Crustaceans (other). In: Simberloff D and Rejmánek M (eds) Encyclopedia of biological invasions. University of California Press, London, pp. 135-137

Ridgeway G (2010) *gbm: Generalized Boosted Regression Models*. R Package Version 1.6-3.1. [Available at: http://cran.r-project.org/web/packages /gbm/index.html; accessed 20/02/2011]

Robin C, Capron G, Desprez-Loustau ML (2001) Root infection by *Phytophthora cinnamomi* in seedlings of three oak species. *Plant Pathology* 50:708-716

Rödder D, Schmidtlein S, Veith M, et al. (2009) Alien Invasive Slider Turtle in Unpredicted Habitat: A Matter of Niche Shift or of Predictors Studied? *PloS one* 4:e7843

Roura-Pascual N, Brotons L, Peterson AT, et al. (2009) Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions* 11:1017-1031

Roura-Pascual N, Hui C, Ikeda T, et al. (2011) Relative roles of climatic suitability and anthropogenic influence in determining the pattern of spread in a global invader. *Proceedings of the National Academy of Sciences*, 108:220-225

Rudnick DA, Hieb K, Grimmer KF, et al. (2003) Patterns and processes of biological invasion: The chinese mitten crab in San Francisco Bay. *Basic and Applied Ecology* 4:249-262

Sabater S, Feio MJ, Graça MAS, et al. (2009) Chapter 4 - The Iberian Rivers. In: Tockner K, Uehlinger U and Robinson CT (eds) *Rivers of Europe*. Academic Press, London, pp. 113-149

Sahlin U, Smith HG, Edsman L, et al. (2010) Time to establishment success for introduced signal crayfish in Sweden – a statistical evaluation when success is partially known. *Journal of Applied Ecology* 47:1044-1052

Sanchez-Fernandez D, Bilton DT, Abellan P, et al. (2008) Are the endemic water beetles of the Iberian Peninsula and the Balearic Islands effectively protected? *Biological Conservation* 141:1612-1627

Sanderson EW, Jaiteh M, Levy MA, et al. (2002) The human footprint and the last of the wild. *BioScience* 52:891-904

Santika T (2011) Assessing the effect of prevalence on the predictive performance of species distribution models using simulated data. *Global Ecology and Biogeography* 20:181-192

Sarle W (1994) Neural Networks and Statistical Models. *Proceedings of the Nineteenth Annual SAS Users Groups International Conference*. SAS Institute Inc., Cary, pp. 1538–1550

Segurado P, Araújo MB (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31:1555-1568

Semple GP, Rouse DB, McLain KR (1995) *Cherax destructor*, *C. tenuimanus* and *C. quadricarinatus* (Decapoda: Parastacidae): a comparative review of biological traits relating to aquaculture potential. *Freshwater Crayfish* 8:495-503

Simberloff D (2009) The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* 40:81-102

Skurdal J, Taugbøl T, Burba A, et al. (1999) Crayfish introductions in the Nordic and Baltic countries. In: Gherardi F and Holdich DM (eds) *Crayfish in Europe as alien species. How to make the best of a bad situation*. AA Balkema, Rotterdam, pp. 193-219

Soberón J, Nakamura M (2009) Niches and distributional areas: concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences* 106:19644-19650

Sousa R, Dias SC, Guilhermino L, et al. (2008) Minho River tidal freshwater wetlands: threats to faunal biodiversity. *Aquatic Biology* 3:237-250

Souty-Grosset C, Holdich DM, Noël PY, et al. (2006) *Atlas of crayfish in Europe*. Muséum National d'Histoire Naturelle, Paris

Stachowicz JJ, Tilman D (2005) Species invasions and the relationships between species diversity, community saturation, and ecosystem functioning. In: Sax DF, Stachowicz JJ and Gaines SD (eds) *Species Invasions: Insights into Ecology, Evolution, and Biogeography*. Sinauer Associates, Sunderland, pp. 41-64

Steiner FM, Schlick-Steiner BC, VanDerWal J, et al. (2008) Combined modelling of distribution and niche in invasion biology: a case study of two invasive Tetramorium ant species. *Diversity and Distributions* 14:538-545

Stohlgren TJ, Ma P, Kumar S, et al. (2010) Ensemble habitat mapping of invasive plant species. *Risk Analysis* 30:224-235

Thuiller W, Brotons L, Araújo MB, et al. (2004) Effects of restricting environmental range of data to project current and future species distributions. *Ecography* 27:165-172

Thuiller W, Lafourcade B, Engler R, et al. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32:369-373

Thuiller W, Richardson D, Pysek P, et al. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* 11:2234-2250

Thuiller W, Richardson DM, Rouget M, et al. (2006) Interactions between environment, species traits, and human uses describe patterns of plant invasions. *Ecology* 87:1755-1769

Tsoar A, Allouche O, Steinitz O, et al. (2007) A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions* 13:397-405

Usio N, Nakajima H, Kamiyama R, et al. (2006) Predicting the distribution of invasive crayfish (*Pacifastacus leniusculus*) in a Kusiro Moor marsh (Japan) using classification and regression trees. *Ecological Research* 21:271-277

Vanderploeg HA, Liebig JR, Carmichael WW, et al. (2001) Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Canadian Journal of Fisheries and Aquatic Sciences* 58:1208-1221

VanDerWal J, Shoo LP, Graham C, et al. (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* 220:589-594

Van Kleunen M, Dawson W, Schlaepfer D, et al. (2010) Are invaders different? A conceptual framework of comparative approaches for assessing determinants of invasiveness. *Ecology Letters* 13:947-958

Vapnik VN (1995) *The nature of statistical learning theory*. Springer-Verlag New York, Inc.

Veilleux É, De Lafontaine Y (2007) Biological synopsis of the Chinese mitten crab (*Eriocheir sinensis*). Canadian Manuscript Report of Fisheries and Aquatic Sciences 2812:1-45

Verdin K, Jenson S (1996) Development of continental scale DEMs and extraction of hydrographic features. *Proceedings of the Third Conference on GIS and Environmental Modeling*. Santa Fe, pp. 21-26

VonHolle B, Simberloff D (2005) Ecological resistance to biological invasion overwhelmed by propagule pressure. *Ecology* 86:3212-3218

Vose R, Easterling D, Gleason B (2005) Maximum and minimum temperature trends for the globe: An update through 2004. *Geophysical Research Letters* 32:L23822

Walther G, Roques A, Hulme P, et al. (2009) Alien species in a warmer world: risks and opportunities. *Trends in Ecology and Evolution* 24:686-693

Welk E, Schubert K, Hoffmann MH (2002) Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in North America. *Diversity and Distributions* 8:219-233

Westman K, Savolainen R, Julkunen M (2002) Replacement of the native crayfish *Astacus astacus* by the introduced species *Pacifastacus leniusculus* in a small, enclosed Finnish lake: a 30-year study. *Ecography* 25:53-73

Wiens JJ, Ackerly DD, Allen AP, et al. (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13:1310-1324

Wiens JJ, Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics* 36:519-539

Wiles GJ, Bart J, Beck RE, et al. (2003) Impacts of the Brown Tree Snake: patterns of decline and species persistence in Guam's avifauna. *Conservation Biology* 17:1350-1360

Williamson MH (1996) *Biological invasions*. Chapman & Hall, London

Williamson M, Fitter A (1996) The varying success of invaders. *Ecology* 77:1661-1666

Wilson JRU, Richardson DM, Rouget M, et al. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions* 13:11-22

Wisz M, Guisan A (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC ecology 9:8

Wisz MS, Hijmans R, Li J, et al. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14:763-773

Withnall F (2000) Biology of Yabbies (*Cherax destructor*). Aquaculture notes. Department of Natural Resources and Environment, Melbourne, pp. 4

Witten I, Frank E (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, London

Zimmermann NE, Edwards Jr TC, Graham CH, et al. (2010) New trends in species distribution modelling. *Ecography* 33:985-989

Page intentionally left blank

APPENDIX

APPENDIX A (CHAPTER2)

Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets

Table 1 - Number of records in calibration datasets for each species using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA). Occurrence records in NRO and IRO models correspond to 80% of total records collected for each species in its respective range. NIO models include an equal number of native and invasive occurrences corresponding each to the minimum available of one of the ranges. These occurrences were coupled with a same number of pseudo-absences for each pseudo-absence type. For maximizing their representativeness we made 20 replicate datasets for each combination, each containing an independently drawn set of pseudo-absences.

Model	***C. destructor***	***E. sinensis***	***P. leniusculus***	***P. clarkii***
NRO-NRA	82 (native)	81 (native)	100 (native)	138 (native)
NRO-NIA	82 (native)	81 (native)	100 (native)	138 (native)
NRO-WRA	82 (native)	81 (native)	100 (native)	138 (native)
NIO-NRA	82 (41 native + 41 invasive)	162 (81 native + 81 invasive)	200 (100 native + 100 invasive)	276 (138 native + 138 invasive)
NIO-NIA	82 (41 native + 41 invasive)	162 (81 native + 81 invasive)	200 (100 native + 100 invasive)	276 (138 native + 138 invasive)
NIO-WRA	82 (41 native + 41 invasive)	162 (81 native + 81 invasive)	200 (100 native + 100 invasive)	276 (138 native + 138 invasive)
IRO-NRA	41 (invasive)	155 (invasive)	352 (invasive)	340 (invasive)
IRO-NIA	41 (invasive)	155 (invasive)	352 (invasive)	340 (invasive)
IRO-WRA	41 (invasive)	155 (invasive)	352 (invasive)	340 (invasive)

Table 2 - Best mean root mean squared errors (RMSE) its standard deviation (SD) and number of nodes in the artificial neural network providing it during the training sessions. Number of nodes was based on the number of calibration cases (*n*) in each calibration dataset type. Models were built using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA).

		NRO			**NIO**			**IRO**		
Species	**Pseudo-absence**	**Best mean**	**SD**	**Nodes in**	**Best mean**	**SD**	**Nodes in**	**Best mean**	**SD**	**Nodes in**
C. destructor	NRA	0.49	0.11	0	0.44	0.07	(*n*/10)/2	0.16	0.04	0
	NIA	0.47	0.1	(*n*/10)/2	0.49	0.08	(*n*/10)/2	0.43	0.06	*n*/10
	WRA	0.30	0.07	(*n*/10)/2	0.30	0.54	*n*/10	0.34	0.07	*n*/10
E. sinensis	NRA	0.43	0.09	(*n*/10)/2	0.34	0.05	(*n*/10)/2	0.15	0.04	0
	NIA	0.34	0.08	(*n*/10)/2	0.36	0.06	(*n*/10)/2	0.23	0.05	(*n*/10)/2
	WRA	0.23	0.05	*n*/10	0.29	0.04	(*n*/10)/2	0.24	0.06	*n*/10
P. leniusculus	NRA	0.44	0.09	*n*/10	0.38	0.06	0	0.19	0.05	(*n*/10)/2
	NIA	0.38	0.06	(*n*/10)/2	0.37	0.06	(*n*/10)/2	0.31	0.05	*n*/10
	WRA	0.35	0.09	*n*/10	0.27	0.07	(*n*/10)/2	0.27	0.06	0
P. clarkii	NRA	0.42	0.11	*n*/10	0.38	0.11	(*n*/10)/2	0.18	0.05	*n*/10
	NIA	0.3	0.05	*n*/10	0.39	0.07	(*n*/10)/2	0.28	0.03	(*n*/10)/2
	WRA	0.22	0.03	(*n*/10)/2	0.32	0.04	*n*/10	0.30	0.04	(*n*/10)/2

Table 3 - Number of records used for evaluating the predictions. Numbers of occurrences correspond to 20% of total records collected for each species in its respective range. These records were coupled with an equal number of worldwide random pseudo-absences. For maximizing the representativeness of pseudo-absences we made 10 replicates of each validation dataset, each containing an independently drawn set of pseudo-absences.

Species	**Range**	**Validation occurrences**	**Validation pseudo-absences**
C. destructor	Native	21	21
	Invasive	10	10
E. sinensis	Native	20	20
	Invasive	39	39
P. leniusculus	Native	25	25
	Invasive	88	88
P. clarkii	Native	35	35
	Invasive	85	85

Table 4 - Standard deviation of validation results of kappa statistic (*k*) and area under the curve of receiver-operating characteristic (ROC-AUC) for native ranges using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA).

		NRO		**NIO**		**IRO**	
Species	Pseudo-absences	SD (*k*)	SD (ROC-AUC)	SD (*k*)	SD (ROC-AUC)	SD (*k*)	SD (ROC-AUC)
C. destructor	NRA	0.066	0.061	0.072	0.060	0.050	0.061
	NIA	0.062	0.053	0.052	0.057	0.056	0.042
	WRA	0.031	0.029	0.031	0.028	0.029	0.021
E. sinensis	NRA	0.033	0.031	0.031	0.030	0.041	0.030
	NIA	0.029	0.035	0.045	0.032	0.042	0.046
	WRA	0.022	0.017	0.020	0.011	0.024	0.023
P. leniusculus	NRA	0.031	0.026	0.021	0.030	0.039	0.035
	NIA	0.055	0.032	0.041	0.034	0.061	0.047
	WRA	0.028	0.015	0.022	0.022	0.020	0.013
P. clarkii	NRA	0.022	0.020	0.037	0.042	0.009	0.011
	NIA	0.029	0.021	0.034	0.033	0.022	0.018
	WRA	0.019	0.011	0.022	0.014	0.028	0.024

Table 5 - Standard deviation of validation results of kappa statistic (*k*) and area under the curve of receiver-operating characteristic (ROC-AUC) for invasive ranges using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRA), native and invasive ranges pseudo-absences random extraction (NIA) and the common spatially worldwide random pseudo-absences (WRA).

		NRO		**NIO**		**IRO**	
Species	Pseudo-absences	*k* SD	ROC-AUC SD	*k* SD	ROC-AUC SD	*k* SD	ROC-AUC SD
C. destructor	NRA	0.041	0.036	0.055	0.034	0.031	0.041
	NIA	0.042	0.041	0.050	0.038	0.031	0.033
	WRA	0.033	0.029	0.027	0.027	0.020	0.021
E. sinensis	NRA	0.042	0.036	0.028	0.028	0.034	0.032
	NIA	0.029	0.032	0.022	0.024	0.034	0.030
	WRA	0.025	0.017	0.028	0.022	0.023	0.019
P. leniusculus	NRA	0.036	0.033	0.039	0.042	0.035	0.036
	NIA	0.036	0.038	0.032	0.034	0.022	0.022
	WRA	0.031	0.029	0.026	0.023	0.022	0.019
P. clarkii	NRA	0.026	0.028	0.031	0.031	0.043	0.044
	NIA	0.036	0.037	0.022	0.025	0.032	0.029
	WRA	0.042	0.038	0.024	0.026	0.021	0.018

Table 6 - Standard deviation of validation results of kappa statistic (*k*) and area under the curve of receiver-operating characteristic (ROC-AUC) of a weighted average of all single-models *WA(all)*, a weighted average of all single-models using native range occurrences *WA(NIO)*, a weighted average of all single-models using native and invasive occurrences *WA(NIO)*, a weighted average of all single-models with invasive range occurrences *WA(IRO)*, a weighted average of all single-models using native range pseudo-absences *WA(NRA)*, a weighted average of all single-models using native and invasive ranges pseudo-absences random extraction *WA(NIA)* and a weighted average of all single-models using worldwide random pseudo-absences *WA(WRA)*.

		WA(all)		**WA(NRO)**		**WA(NIO)**		**WA(IRO)**		**WA(NRA)**		**WA(NIA)**		**WA(WRA)**	
Species	**Range**	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC	*k*	ROC-AUC
C. destructor	Native	0.58	0.72	0.71	0.81	0.57	0.69	0.47	0.57	0	0.15	0.17	0.32	0.9	0.98
	Invaded	0.61	0.73	0.53	0.64	0.66	0.8	0.64	0.74	0.23	0.52	0.27	0.4	0.79	0.96
E. sinensis	Native	0.64	0.73	0.75	0.87	0.65	0.79	0.34	0.38	0.12	0.49	0.38	0.55	0.62	0.88
	Invaded	0.73	0.85	0.64	0.74	0.79	0.89	0.79	0.9	0.7	0.89	0.71	0.82	0.86	0.97
P. leniusculus	Native	0.61	0.73	0.75	0.85	0.59	0.73	0.27	0.56	0.11	0.55	0.53	0.61	0.74	0.94
	Invaded	0.73	0.84	0.58	0.7	0.74	0.88	0.74	0.92	0.77	0.95	0.63	0.76	0.87	0.97
P. clarkii	Native	0.62	0.77	0.8	0.91	0.61	0.7	0.27	0.32	0	0.34	0.12	0.24	0.84	0.98
	Invaded	0.69	0.84	0.55	0.67	0.73	0.88	0.76	0.92	0.67	0.85	0.67	0.79	0.79	0.94

Table 7 - Ranks of variable importance measured using the connection weights approach (Olden et al. 2004). Models were built using native range occurrences (NRO), native and invasive occurrences (NIO) and invaded range occurrences (IRO), coupled with within native range pseudo-absences random extraction (NRO), native and invasive ranges pseudo-absences random extraction (NIA), and the common spatially worldwide random pseudo-absences (WRA).

		Independent variables									
		alt	amte	anpr	dist	fros	maxtw	mintcm	pred	prew	trang
NRO - NRA	*C. destructor*	4	5	2	-	3	6	1	8	9	7
	E. sinensis	2	5	3	4	1	7	6	8	9	10
	P. leniusculus	5	3	2	-	4	8	1	7	6	9
	P. clarkii	5	6	8	-	1	2	4	7	3	9
NRO - NIA	*C. destructor*	5	2	4	-	1	6	3	8	9	7
	E. sinensis	1	4	7	2	5	9	3	6	10	8
	P. leniusculus	9	4	3	-	1	8	2	5	6	7
	P. clarkii	6	2	8	-	3	4	1	5	7	9
NRO - NRA	*C. destructor*	7	3	5	-	4	1	2	8	6	9
	E. sinensis	7	6	5	1	3	4	2	8	9	10
	P. leniusculus	8	5	2	-	4	3	1	9	6	7
	P. clarkii	9	2	5	-	6	1	3	4	8	7
NIO - NRA	*C. destructor*	7	2	4	-	3	1	6	5	9	8
	E. sinensis	1	3	5	6	2	9	4	8	7	10
	P. leniusculus	4	1	5	-	2	6	3	8	9	7
	P. clarkii	5	4	6	-	3	1	7	8	2	9
NIO - NIA	*C. destructor*	6	2	4	-	3	9	1	7	8	5
	E. sinensis	3	5	4	1	2	7	6	8	10	9
	P. leniusculus	8	1	6	-	2	7	3	4	5	9
	P. clarkii	6	3	4	-	2	5	1	9	8	7
NIO - WRA	*C. destructor*	8	2	6	-	5	3	1	7	4	9
	E. sinensis	6	4	7	1	2	3	5	8	10	9
	P. leniusculus	9	4	3	-	5	1	2	8	7	6
	P. clarkii	9	2	4	-	5	3	1	6	8	7
IRO - NRA	*C. destructor*	3	5	6	-	1	4	2	7	9	8
	E. sinensis	1	3	6	4	2	8	5	10	9	7
	P. leniusculus	4	3	5	-	2	7	1	6	9	8
	P. clarkii	4	7	6	-	3	1	2	9	8	5
IRO - NIA	*C. destructor*	4	1	8	-	3	5	2	6	9	7
	E. sinensis	3	4	6	1	2	7	5	8	10	9
	P. leniusculus	5	1	7	-	4	6	2	3	8	9
	P. clarkii	4	3	6	-	2	7	1	9	5	8
IRO - WRA	*C. destructor*	6	4	7	-	3	2	1	5	8	9
	E. sinensis	9	4	3	2	5	1	6	10	8	7
	P. leniusculus	7	5	4	-	3	2	1	6	9	8
	P. clarkii	9	3	2	-	7	4	1	5	8	6

References

Olden JD, Joy MK, Death RG (2004) An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* 178: 389-397

APPENDIX B (CHAPTER3)

Geographic variability in propagule pressure and climatic suitability explain the European distribution of two highly invasive crayfish

Table 1 List of crayfish species which have co-occurring European populations with the Red swamp crayfish or the Signal crayfish and are known competitors for resources or space. Listed references represent only a subset of studies referring to these biotic interactions.

	Red swamp crayfish	**Signal crayfish**
Red swamp crayfish (*Procambarus clarkii* Girard)		Alonso and Martínez (2006)
Spiny cheek crayfish (*Orconectes limosus* Rafinesque)	-	Hudina et al. (2011)
Noble crayfish (*Astacus astacus* Linnaeus)	-	Bäck (1995)
White-clawed crayfish (*Austropotamobius pallipes* Lereboullet)	Gherardi and Cioni (2004)	Dunn et al. (2009)
Stone crayfish *(Austropotamobius torrentium* Shrank)	-	Vorburger and Ribi (1999)
Narrow-clawed crayfish (*Astacus leptodactylus* Eschscholtz)	-	Holdich et al. (1995)

References

Alonso F, Martínez R (2006) Shelter competition between two invasive crayfish species: a laboratory study. *Bulletin Français de la Pêche et de la Pisciculture.* 380-381: 1121-1132

Bäck BS (1995) Replacement of the native crayfish *Astacus astacus* by the introduced species *Pacifastacus leniusculus* in a Swedish lake: possible causes and mechanisms. *Freshwater biology* 33: 291-304

Dunn J, McClymont H, Christmas M et al. (2009) Competition and parasitism in the native White Clawed Crayfish *Austropotamobius pallipes* and the invasive Signal Crayfish *Pacifastacus leniusculus* in the UK. *Biological Invasions* 11: 315-324

Gherardi F, Cioni A (2004) Agonism and interference competition in freshwater decapods. *Behaviour*, 141: 1297-1324

Holdich DM, Reader JP, Rogers WD et al. (1995) Interactions between three species of crayfish (*Austropotamobius pallipes*, *Astacus leptodactylus* and *Pacifastacus leniusculus*). *Freshwater Crayfish* 10: 46-56

Hudina S, Galić N, Roessink I, et al.(2011) Competitive interactions between co-occurring invaders: identifying asymmetries between two invasive crayfish species. *Biological Invasions*, 13: 1791-1803

Vorburger C, Ribi G (1999) Aggression and competition for shelter between a native and an introduced crayfish in Europe. *Freshwater biology* 42: 111-119

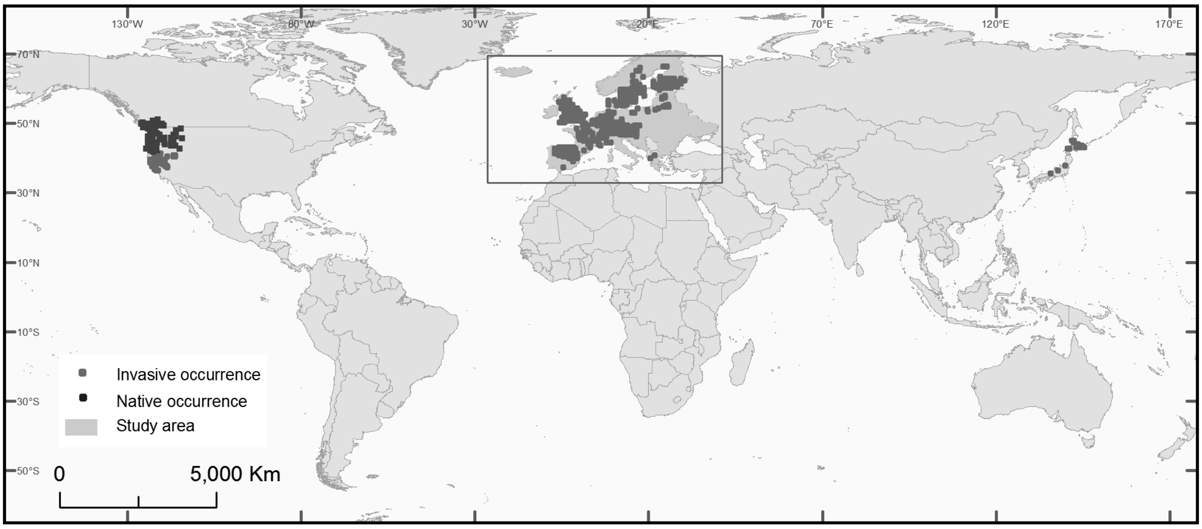


Figure 1 - Worldwide occurrences for the Signal crayfish

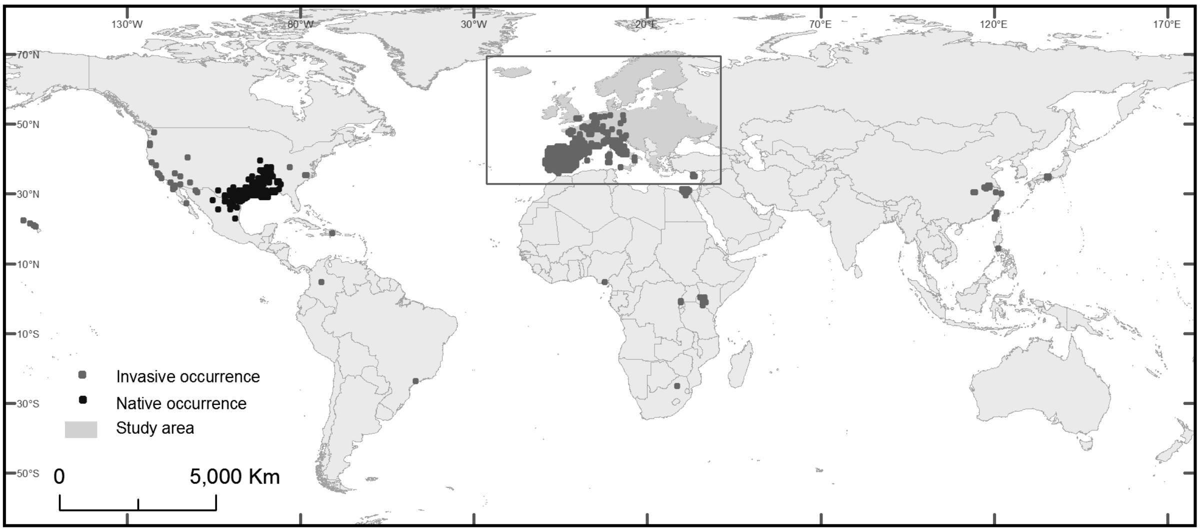


Figure 2 - Worldwide occurrences for the Red swamp crayfish

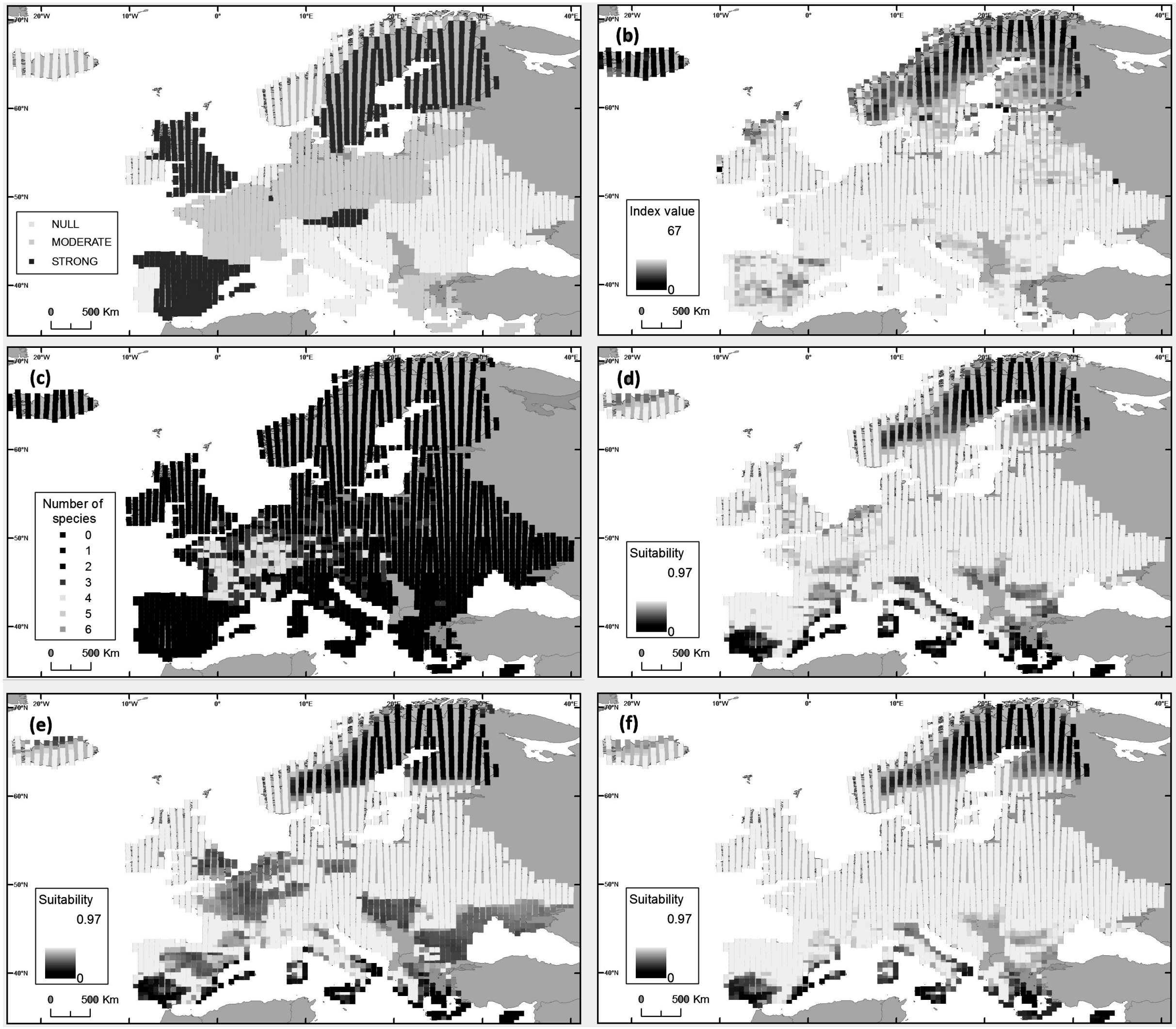


Figure 3 - Candidate explanatory variables used for predicting the current European distribution of the Signal crayfish: (a) country-level introduction effort; (b) mean human footprint; (c) richness of competing crayfish species; (d) climatic suitability based on the climatic conditions occupied in the native range; (e) climatic suitability based on the climatic conditions occupied in invaded areas outside Europe and (f) climatic suitability based on the climatic conditions occupied in the native range and in invaded areas outside Europe

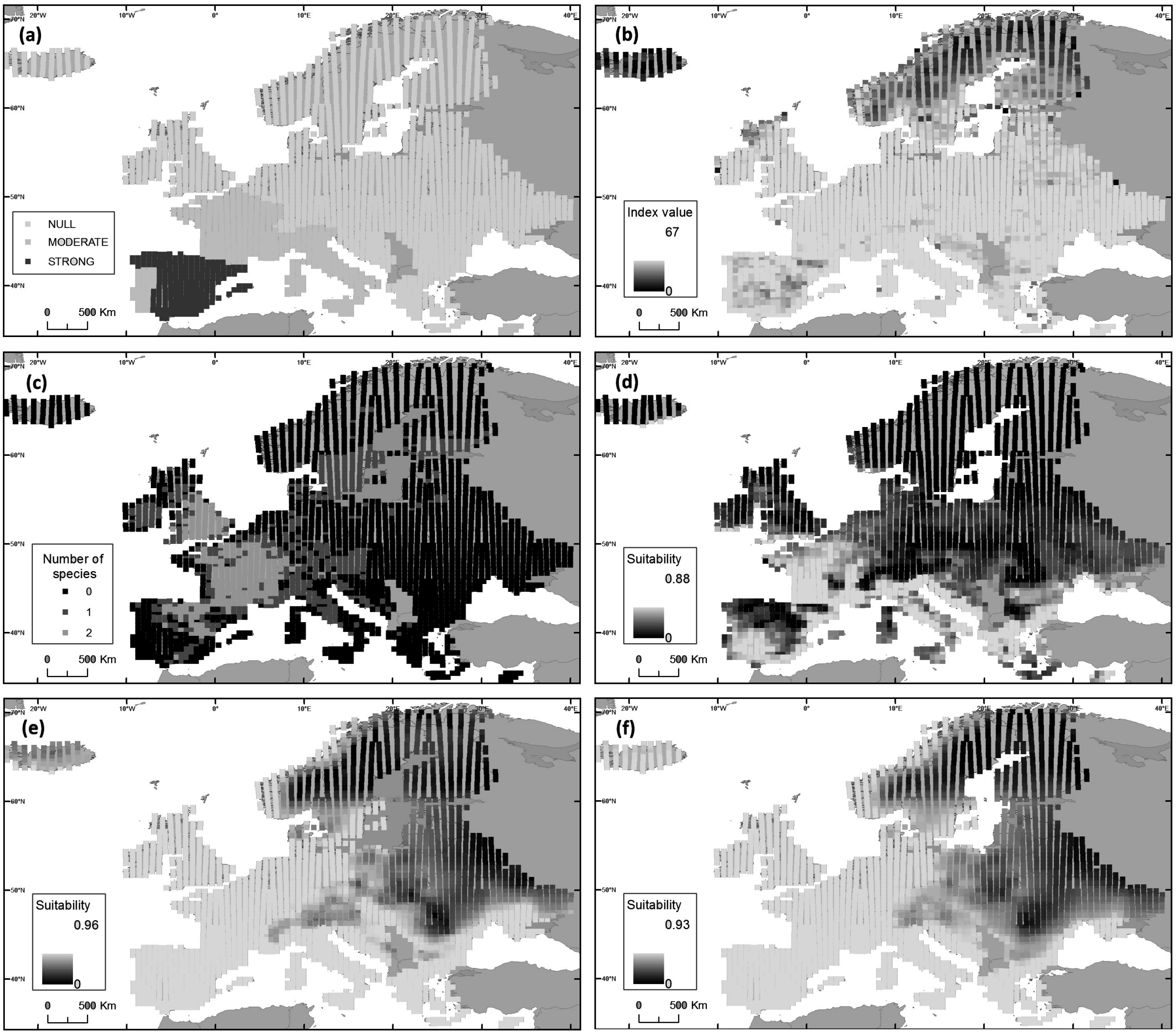


Figure 4 - Candidate explanatory variables used for predicting the current European distribution of the Red swamp crayfish: (a) country-level introduction effort; (b) mean human footprint; (c) richness of competing crayfish species; (d) climatic suitability based on the climatic conditions occupied in the native range; (e) climatic suitability based on the climatic conditions occupied in invaded areas outside Europe and (f) climatic suitability based on the climatic conditions occupied in the native range and in invaded areas outside Europe

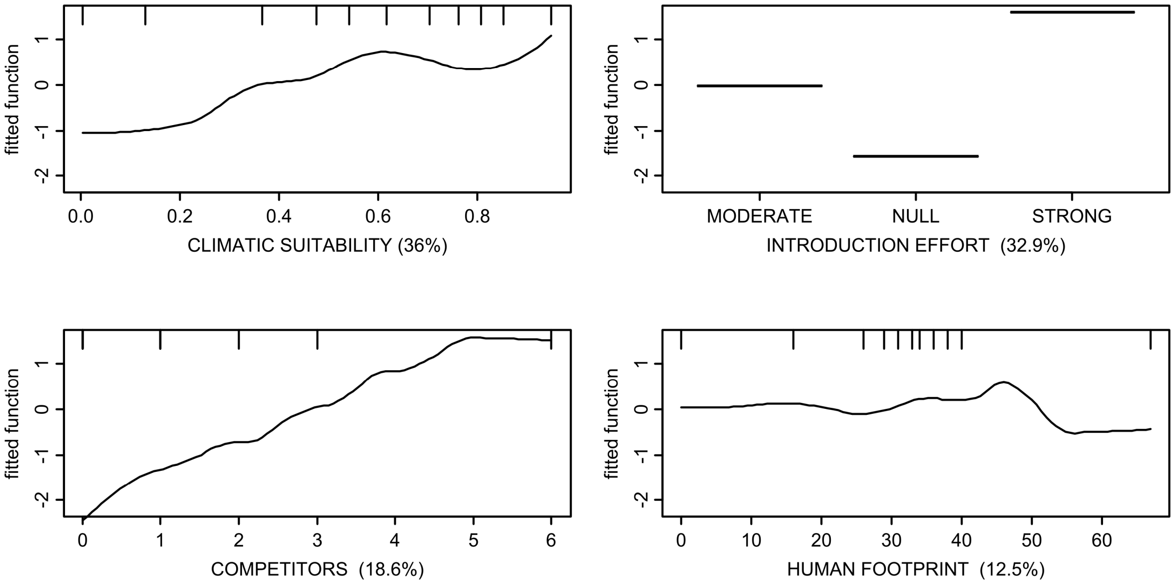


Figure 5 - Smoothed partial dependence plots of one representative boosted regression trees model for the Signal crayfish using a climatic suitability model based on the species native range. The relative contribution of each predictor in this specific model is show in parenthesis

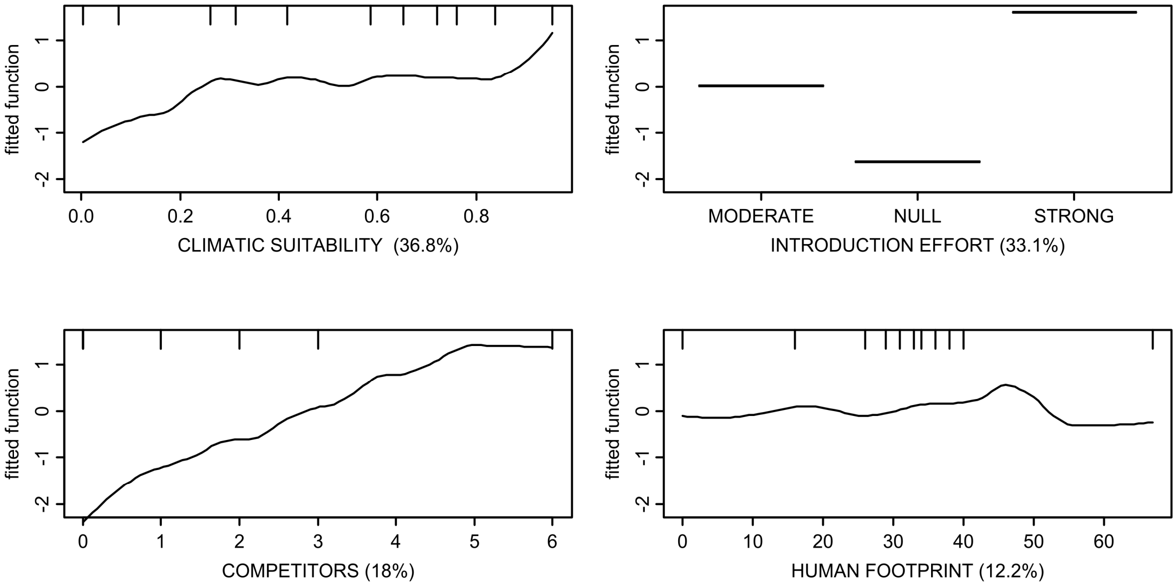


Figure 6 - Smoothed partial dependence plots of one representative boosted regression trees model for the Signal crayfish using climatic suitability based on the species invasive range outside the study area. The relative contribution of each predictor in this specific model is show in parenthesis

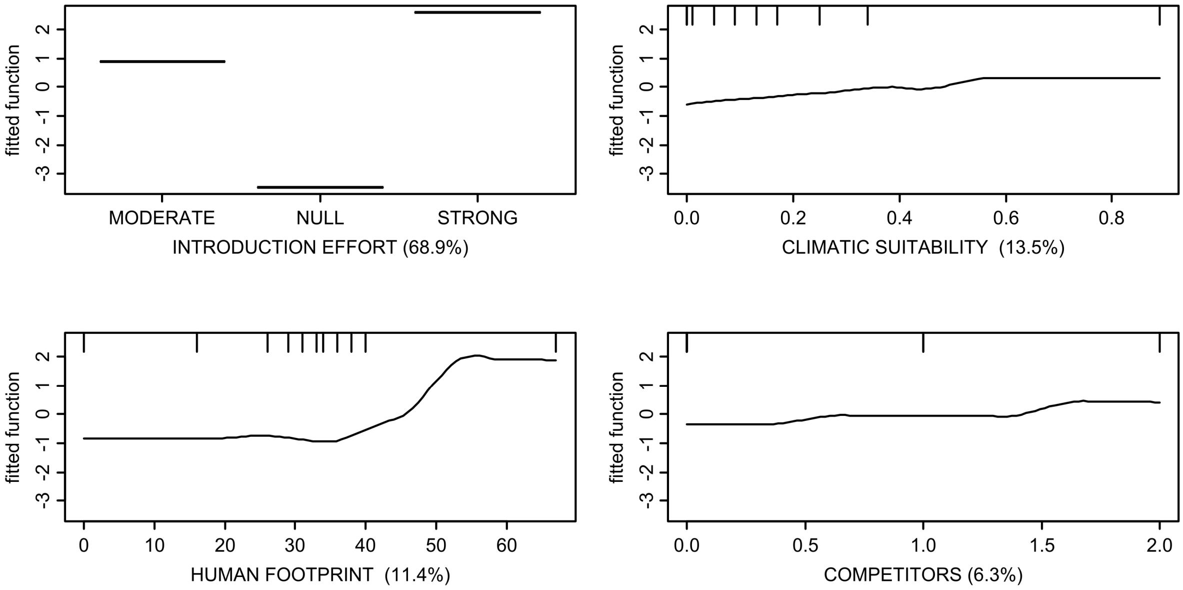


Figure 7 - Smoothed partial dependence plots of one representative boosted regression trees model for the Red swamp crayfish using a climatic suitability model based on the species native range. The relative contribution of each predictor in this specific model is show in parenthesis

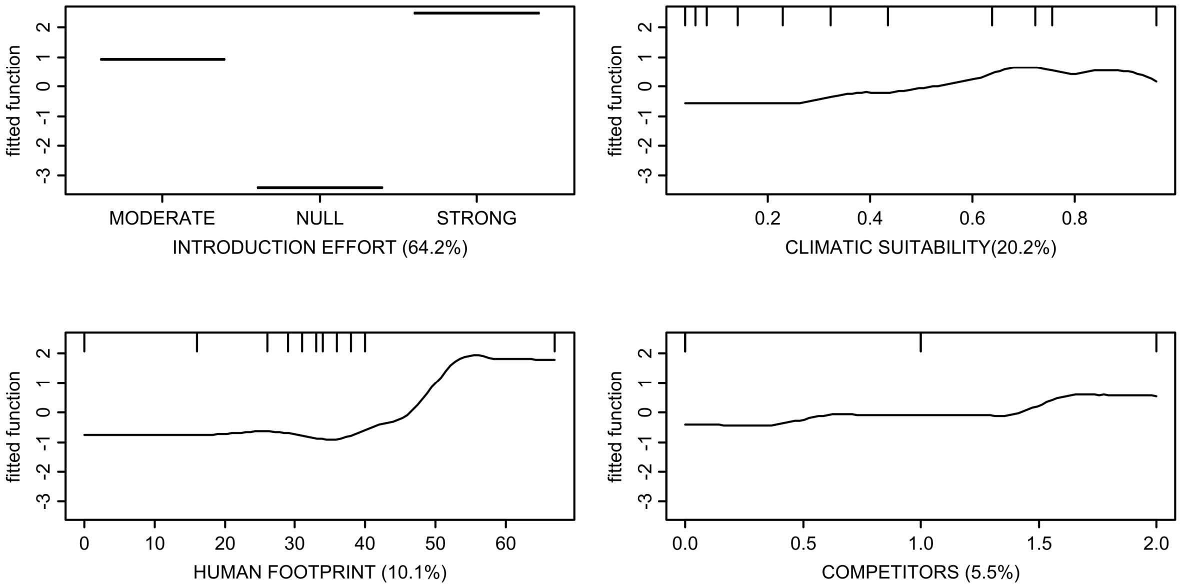


Figure 8 - Smoothed partial dependence plots of one representative boosted regression trees model for the Red swamp crayfish using climatic suitability based on the species invasive range outside the study area. The relative contribution of each predictor in this specific model is show in parenthesis

APPENDIX C (CHAPTER4)

Assessing the environmental requirements of invaders using ensembles of distribution models

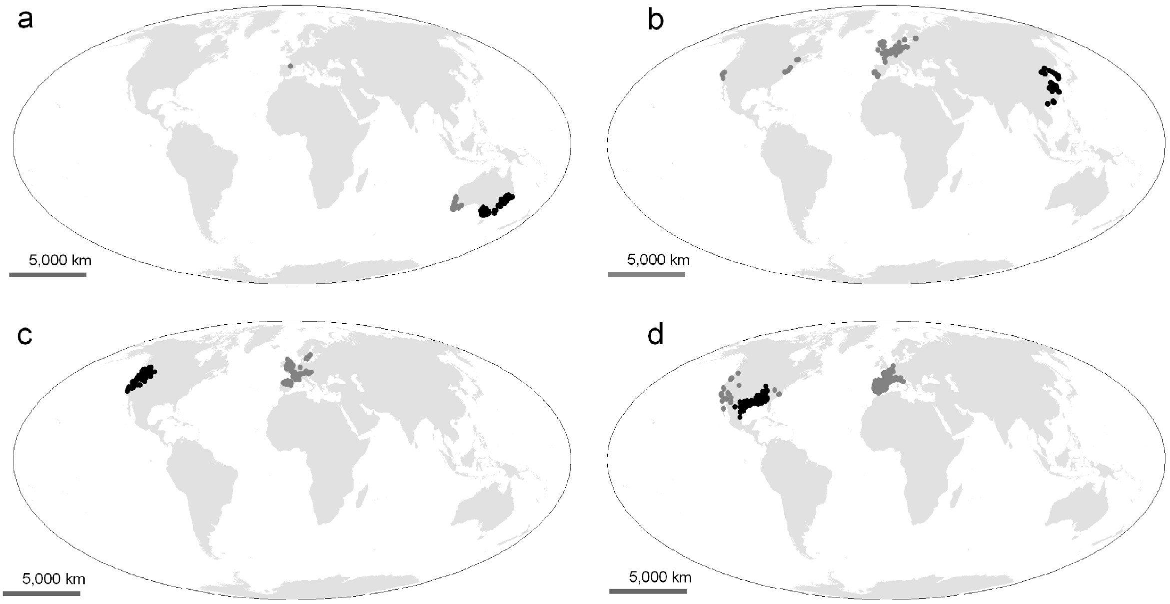


Figure 1 - Native (black circles) and invasive occurrences (grey circles) of (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* used for model calibration

Table 1 - Parameterization values selected (x) for predicting the distribution of the four decapod species using: artificial neural networks (ANN); classification tree analysis (CTA); generalized additive models (GAM); generalized boosted regression models (GBM) and generalized linear models (GLM). Tested values were chosen by taking into account the default values used by the BIOMOD package (Thuiller et al. 2009) and making them vary at regular intervals. Parameters not mentioned were set to default values. For the multivariate adaptive regression splines, mixture discriminant analysis and random forest for classification and regression algorithms we only tested the default parameter values. When distinct parameterizations provided equal predictive performance the most parsimonious model was selected.

Model	**Parameter**	**Value**	**Best performing parametrization**			
			C.	*E.*	*P.*	*P.*
ANN	Number of cross-validation episodes	1				
		5	x			
		10		x	x	x
		15				
CTA	Number of cross-validation episodes	20				
		40	x			
		60		x		
		80			x	x
		100				
GAM	Degree of smoothing of the spline function	3	x	x		
		4			x	
		5				x
		6				
GBM	Maximum number of trees	1000				
		2000	x	x	x	
		3000				x
		4000				
		5000				
		6000				
GLM	Complexity of the terms	linear	x	x		
		polynomial			x	
		quadratic				x

References

Thuiller W, Lafourcade B, Engler R et al. (2009) BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* 32:369–373

APPENDIX D (CHAPTER 5)

Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability

Table 1 - Percentage of the study area where at least one climatic variable falls outside its range in the species calibration data. Predictions were made for three future time periods (2030, 2050 and 2080) under low (B1) and high (A2) greenhouse gases emission scenarios.

	C. destructor		*E. sinensis*		*P. leniusculus*		*P. clarkii*	
	B1	A2	B1	A2	B1	A2	B1	A2
2030	11.2	19.6	2.6	2.1	18.4	15.2	0	0.5
2050	19.4	25.5	6.2	10.9	18.7	26.1	0	0.2
2080	38.6	50.6	13	22.8	28.3	46.2	0	0

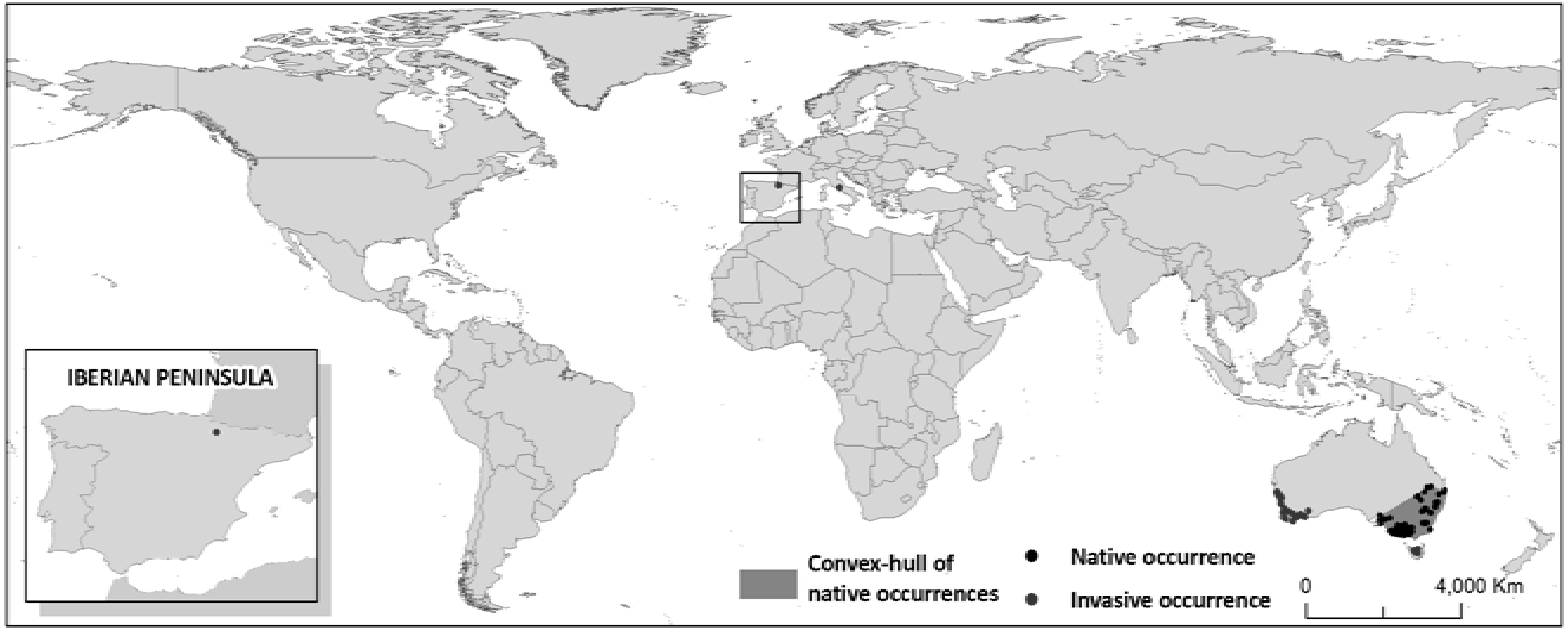


Figure 1 - Native and invasive occurrences of *Cherax destructor* used for model calibration. Native range is delimited as the convex-hull of the species native occurrences

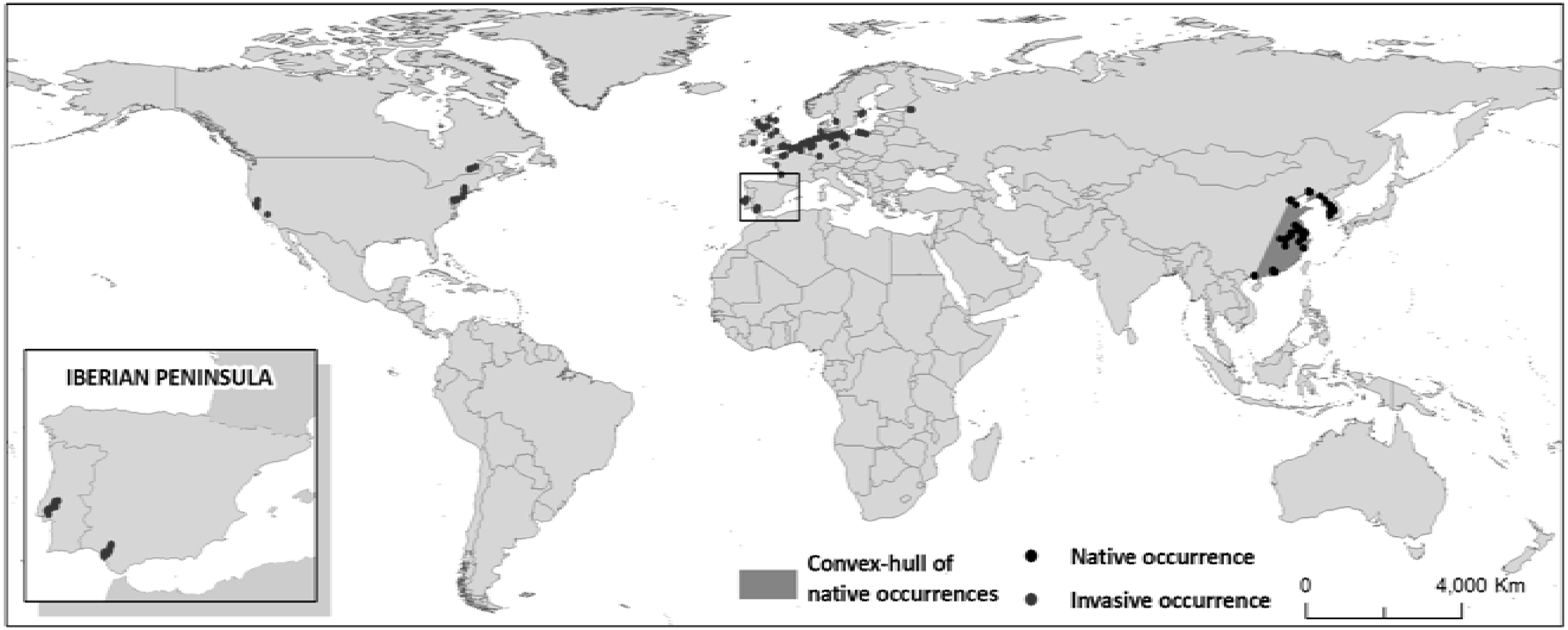


Figure 2 - Native and invasive occurrences of *Eriocheir sinenis* used for model calibration. Native range is delimited as the convex-hull of the species native occurrences

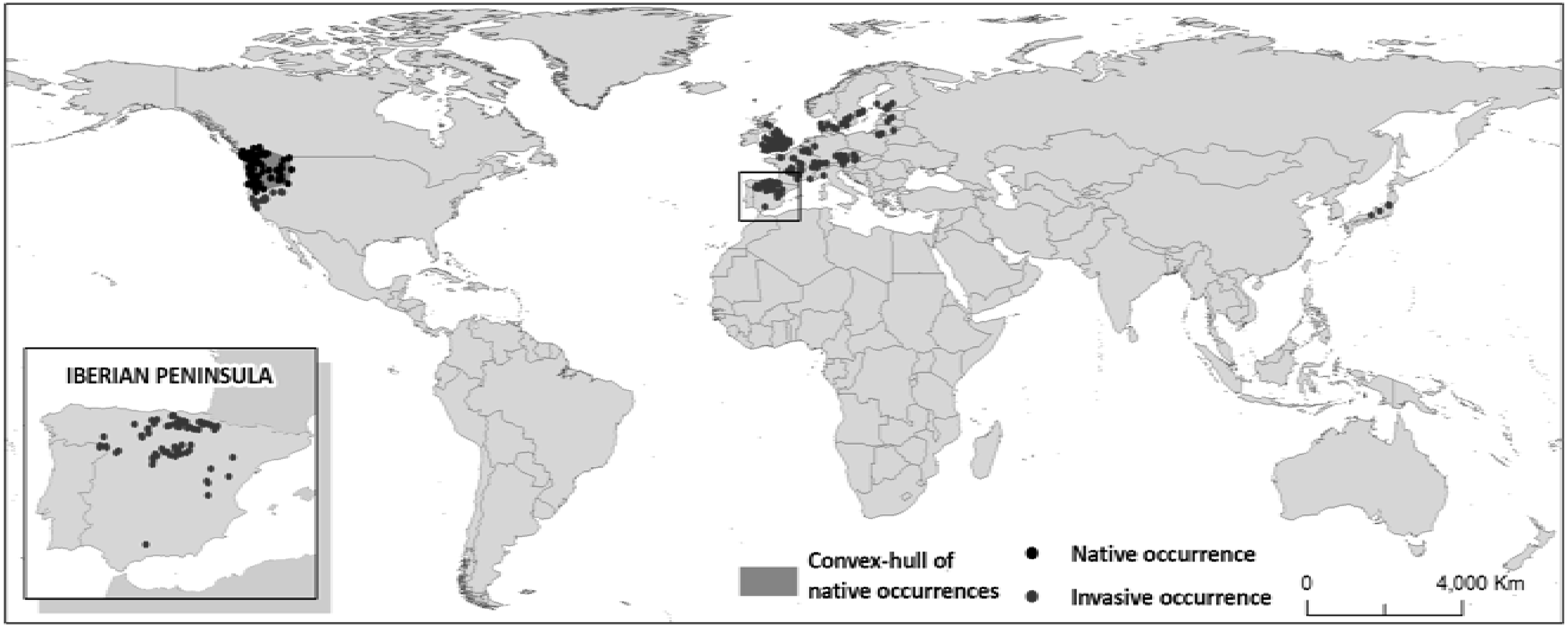


Figure 3 - Native and invasive occurrences of *Pacifastacus leniusculus* used for model calibration. Native range is delimited as the convex-hull of the species native occurrences

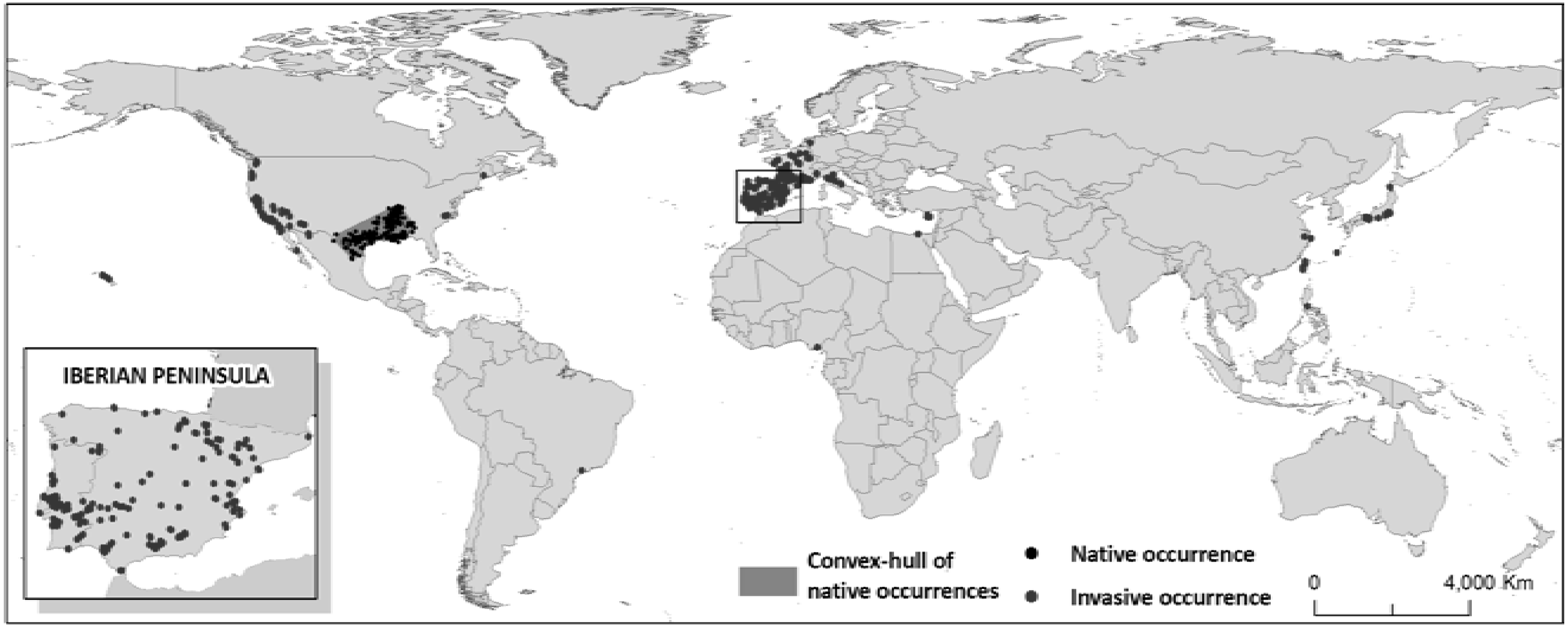


Figure 4 - Native and invasive occurrences of *Procambarus clarkii* used for model calibration. Native range is delimited as the convex-hull of the species native occurrences

Text 1 - Description of the correlative algorithms used

Generalized linear models (GLM) correspond to a flexible generalisation of the ordinary least squares regression since they allow multiple distributions to be modelled in the response variable and non-constant variance functions (Guisan et al. 2002). Here, we used binomial GLMs with a logit link function to predict the probability of species presence. Despite being widely used in NBMs, GLMs assume a linear relationship between the response and the predictors, and as such may have limited success if these relationships are complex (Elith at al. 2006). Multilayer perceptron artificial neural networks (MP-ANN) are supervised machine learning models that use a set of training data to infer functions that best describe the relationships between a response variable and a set of predictors. Due to their non-parametric nature, MP-ANNs are considered to be particularly appropriate for addressing the complexity and non-linearity of ecological data (Lek and Guégan 1999). However, the degree of flexibility of these models should be regulated to avoid overfitting the training data. Here we used MP-ANN with 1 hidden layer composed of 8 processing nodes, each with a logistic activation function. To avoid overfitted relationships, we used a stopping rule in the training process. Models were allowed to search for the best fit over a maximum of 6000 learning cycles, as long as the predictions did not exceed more than 500 consecutive cycles without improving their predictive performance. Support vector machines (SVM) are a distinct set of supervised machine learning models. These models produce classification rules by building a set of hyperplanes in the environmental space that is provided by the predictors. New data are then classified by using the hyperplane that provides the largest separation between the classes found on the response variable. For this study, the SVM model was trained using the sequential minimal optimisation algorithm (Platt 1999). Final probability estimates were obtained by fitting logistic regression models to the outputs of the SVM. Random forests (RFs) are an ensemble classifier that is based on the generation of multiple classification trees. Classification is made by choosing the class predicted by the higher number of classification trees in the forest. Here we used RF with 100 classification trees. Alternating decision trees (ADT) is another type of tree based classification method that uses boosting to combine weak classifiers based on 1-level decision trees. ADT is distinct from most decision tree methods in that instances can take multiple paths, and classification is based on the sum of the weights found along them. To our knowledge, this is the first study to use this statistical algorithm for the analysis of species distributions. Nonetheless, we selected this method on the basis of its good predictive accuracy supported by other areas of research (e.g., Zhao and Zhang 2008). In this study we ran ADT with 20 boosting interactions.

References

Elith J, Graham CH, Anderson RP, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151

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

Lek S, Guégan JF (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120:65-73

Platt JC (1999) Fast training of support vector machines using sequential minimal optimization. In: Schölkopf B, Burges CJC and Smola AJ (eds) *Advances in kernel methods - Support Vector Learning*. MIT Press, Cambridge, pp 185-208

Zhao Y, Zhang Y (2008) Comparison of decision tree methods for finding active objects. *Advances in Space Research* 41:1955-1959

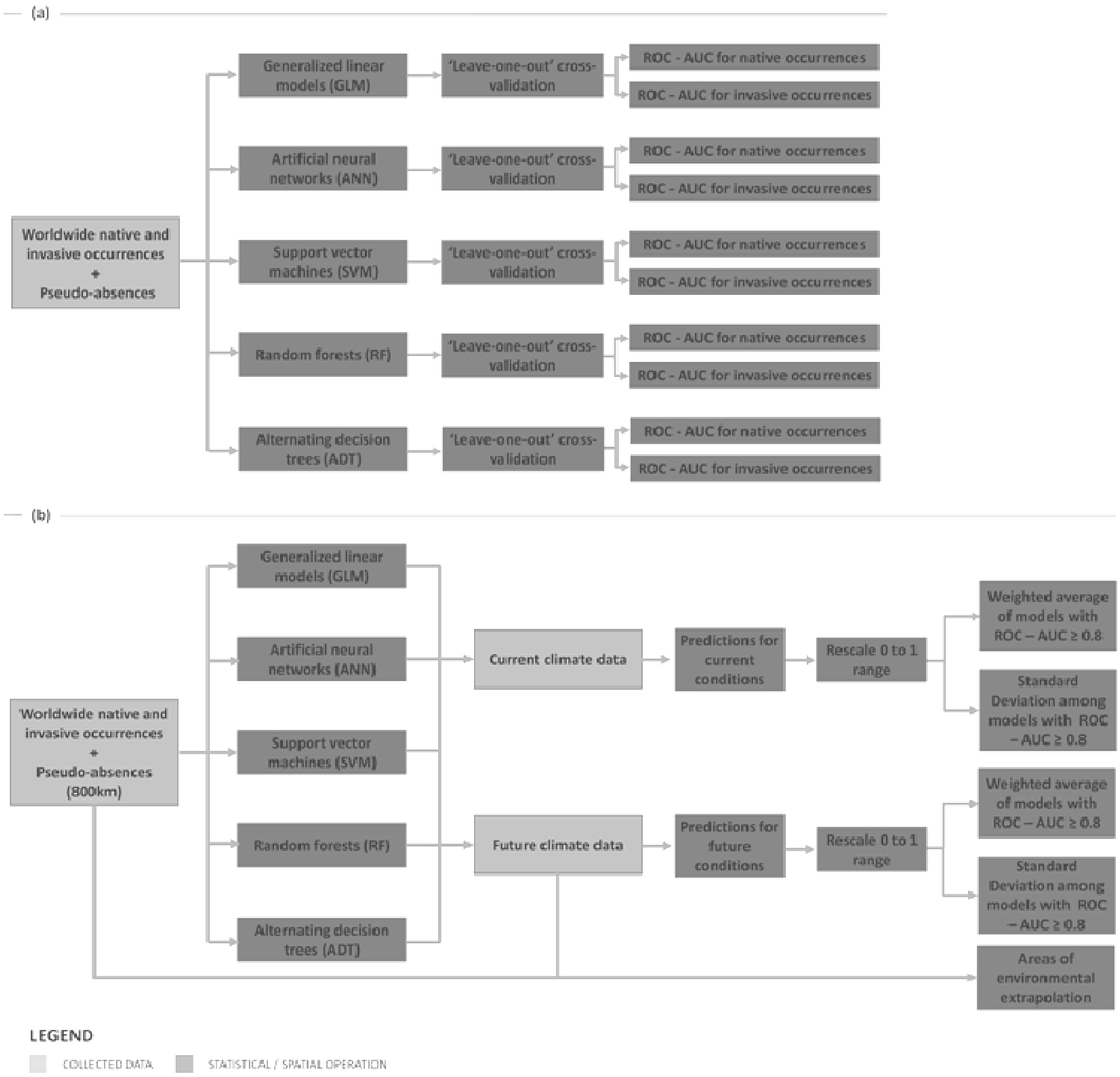


Figure 5 - Schematic representation of the methodological framework for (a) evaluating predictive performances and selecting type of pseudo-absences and, (b) predicting under current and future conditions and assessing predictive variability and environmental extrapolation

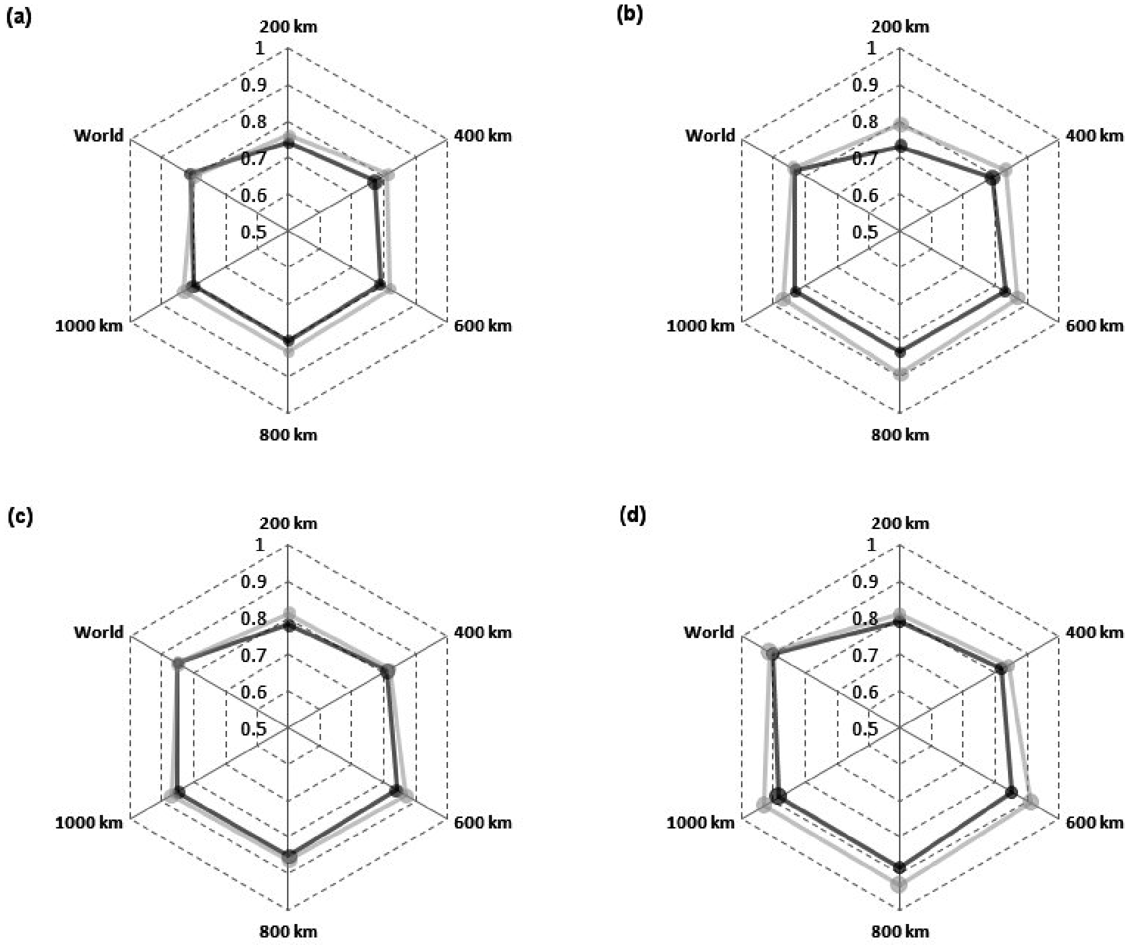


Figure 6 - Mean AUC values of all models in discriminating either native (dark grey) or invasive (light grey) occurrences of (a) *Cherax destructor*, (b) *Eriocheir sinensis*,(c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* from local pseudo absences extracted at distinct maximum distances (200, 400, 600, 800 and 1000 km) from the convex-hull of the species native occurrences. Circle size refers to standard deviation

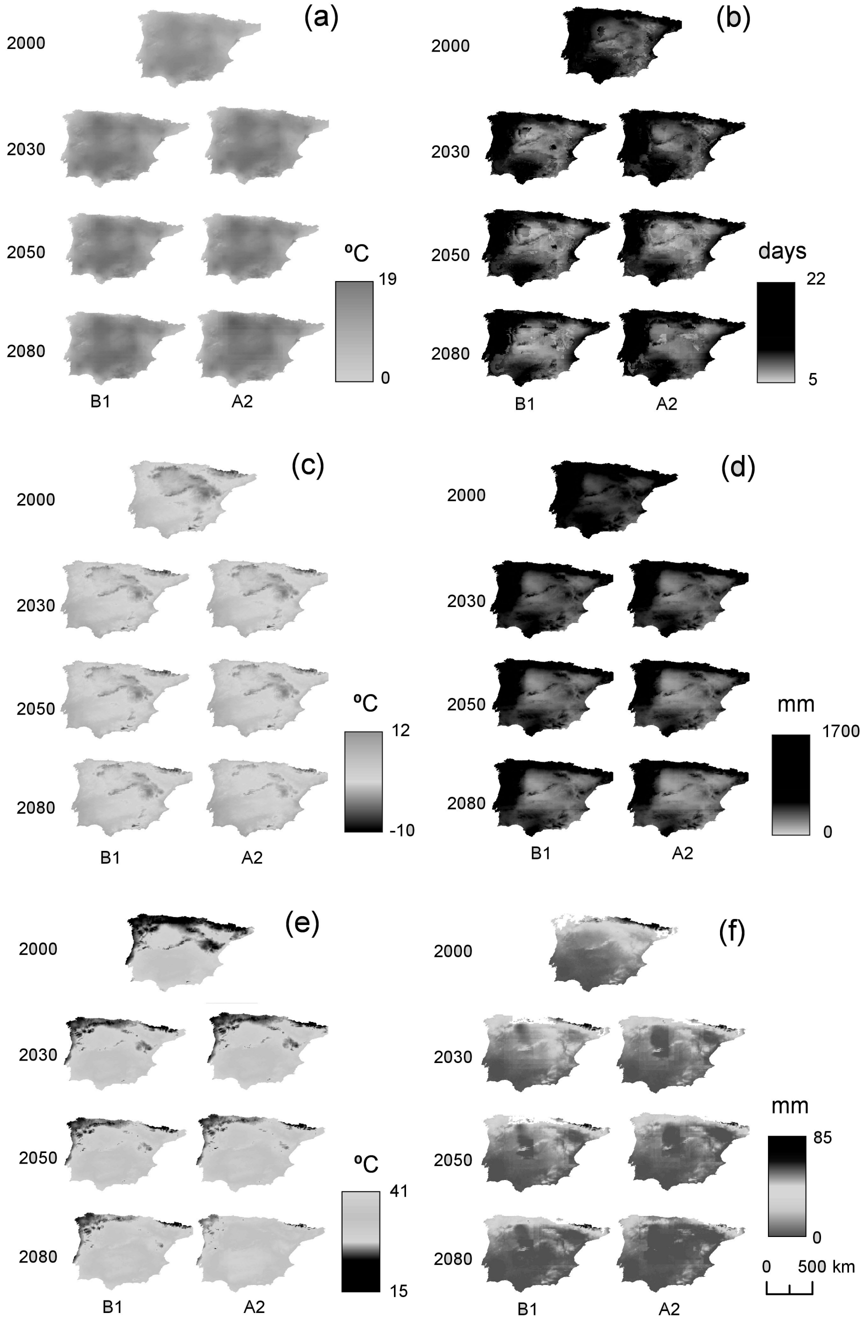


Figure 7 - Spatial distribution of (a) mean diurnal range, (b) maximum number of rainy days in a month, (c) mean minimum temperature of the coldest month, (d) mean yearly total rainfall, (e) mean maximum temperature of the warmest month and (f) mean maximum rainfall in the driest month, for both current (2000) and future time periods (2030, 2050 and 2080) under a low (B1) and high (A2) greenhouse gases emission scenario

u évora
Instituto de Investigação e
Formação Avançada - IIFA

Contactos:
Universidade de Évora
Instituto de Investigação e Formação Avançada - IIFA
Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94
7002-554 Évora | Portugal
Tel: (+351) 266 706 581
Fax: (+351) 266 744 677
email: iifa@uevora.pt