BIODIVERSITY RESEARCH

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, 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, the 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 *Procambarus 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 the assessment of the uncertainty of the achieved knowledge.

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

*Correspondence: César Capinha, 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.
E-mail: capinha@uevora.pt

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 (IUCN, 2009) species 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 & Epifanio, 2009) and rice fields (Anastácio *et al.*, 2005). *Pacifastacus leniusculus* and *Procambarus 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 the currently present wide-ranging populations (Holdich *et al.*, 2009). 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 localized 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 & Epifanio, 2009). This crab was first recorded in the late 1980s in the lower reaches of the Tagus River (Cabral & 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-Pascual *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 & Araújo, 2004; Elith *et al.*, 2006; Tsoar *et al.*, 2007; Roura-Pascual *et al.*, 2009), its parameterization (Hartley *et al.*, 2006) or data characteristics (Dormann *et al.*, 2008; Graham *et al.*, 2008; Capinha *et al.*, 2010). To account for uncertainty in NBM, the combination of single models in ensemble frameworks is supported (Araújo & 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.*, 2010; Grenouillet *et al.*, 2010), and even improve predictive accuracy (Marmion *et al.*, 2009; Grenouillet *et al.*, 2010).

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: (1) to use ensemble and consensus approaches to model the potential distribution of the four decapod species in IP; (2) to identify the environmental factors that best explain these distributions; (3) 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; (4) to use all these results to compare the observed distribution of the species in IP with the potential areas that could be colonized 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 south-west of Europe, comprising approximately 581,000 km^2 (Fig. 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 colonize new areas (Broennimann & 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 & Vieglais, 2001; Wiens & Graham, 2005), the representation of the species fundamental niche should be as complete as possible (Broennimann & Guisan, 2008). Thus, we conducted an extensive compilation of worldwide native and invasive occurrence data of these four species in published literature,
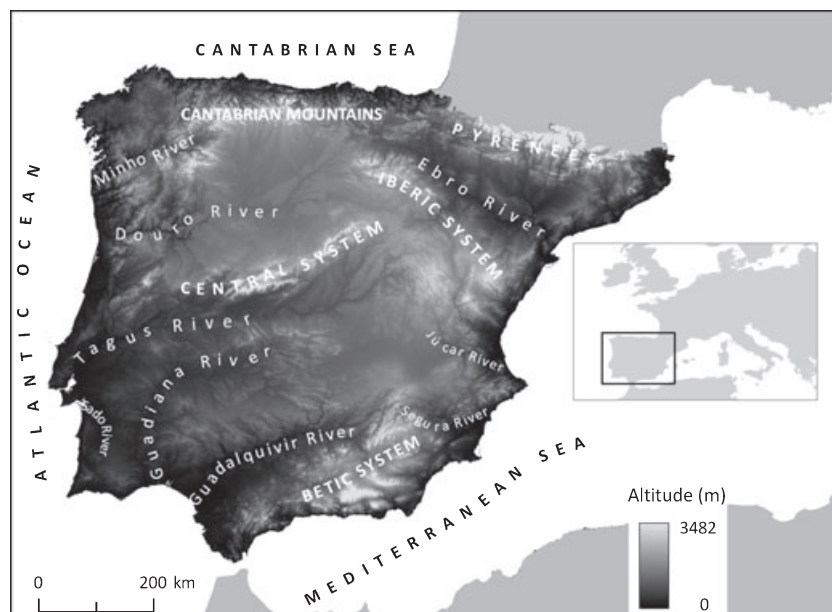
Figure 1 General physiography of the Iberian Peninsula.

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 to or higher than 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 *Procambarus clarkii* (see Fig. S1 in Supporting Information).

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 topographical index (*cti*). Despite being indirect predictors, altitude and slope were included as 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 to *E. sinensis* since this is a catadromous crab (i.e. needs to reproduce in brackish or salt waters). The compound topographical 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 [generalized linear models (GLM), generalized 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 generalized boosted regression models (GBM)].

We coupled species occurrence data with pseudo-absences randomly generated worldwide, because this option has previously shown to provide good results in predicting potential distributions of these invaders (Capinha *et al.*, 2010). To maximize the pseudo-absences' representation of the available environmental conditions and to avoid unbalanced sample effects (Hosmer & 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 parameterization 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 & 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 S1). 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:

$$\text{WA}_i = \frac{\sum i(\text{AUC}_{ji} \times P_{ji})}{\sum i(\text{AUC}_{ji})} \qquad (1)$$

where WA_i is the weighed prediction for the ith invader, and P_{ji} is the probability of presence for the ith invader and the jth 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 randomization 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 randomized. 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 randomized 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 asses 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, because 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., UK), 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 2°C (e.g. 0–2; > 2–4; > 4–6°C…). 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 (Fig. 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' raster map. The adoption of this criterion, in most cases, reduced the amplitude of the analysed variable's ranges, as 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.

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
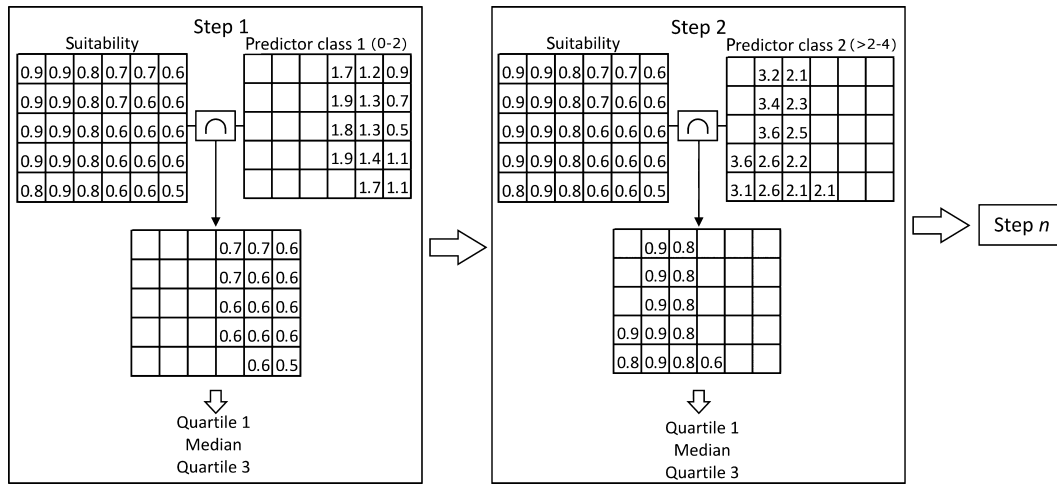
Figure 2 Schematic representation of the framework used to build the consensus suitability curves. The consensus suitability map is spatially intersected with areas having a predefined range of values of the variable under analysis. The process is shown for two classes of one variable; class one ranging 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).

Table 1 Area under the receiver operating characteristics curve (AUC) and standard deviation (SD) values for each of the algorithms used for predicting distribution of the four decapod species.

	Cherax destructor		*Eriocheir sinensis*		*Pacifastacus leniusculus*		*Procambarus 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

Evaluated algorithms include the following: 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).

both *P. leniusculus* and *Procambarus 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).

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 (Fig. 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 most of the environmental variables for all four species. Factors related to temperature are the most influential for all species especially for *P. leniusculus* and *Procambarus clarkii* (Fig. 3c,d respectively). The only noticeable exception is the distance to the ocean in *E. sinensis* models, which appears to drive strongly its distribution (Fig. 3b).

Environmental suitability within the IP

The suitability curves for *C. destructor*, *E. sinensis* and *Procambarus clarkii* show that these invaders are better adapted to areas of warmer temperature in the IP (Fig. 4). Still, they respond differently depending on the thermal variables and position on the thermal gradient. *Procambarus clarkii* is the species showing wider thermal tolerance. The suitability for this invader steeply increases in areas having mean temperatures above 0°C in winter, a mean annual temperature above 6°C and mean summer temperatures above 13°C (Fig. 4j,k). *Cherax destructor*, however, shows a lower suitability in areas with low temperatures, with suitability values increasing only for areas above 12°C in annual mean temperature and above 4°C in winter (Fig. 4a,b). 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 (Fig. 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 (Fig. 4g–i).

For *C. destructor*, the geographical areas with the highest suitability are mostly in the south-west of the IP, on the Cantabrian coast, the Ebro valley and along the eastern littoral regions (Fig. 5a). Still, variability between statistical algorithms
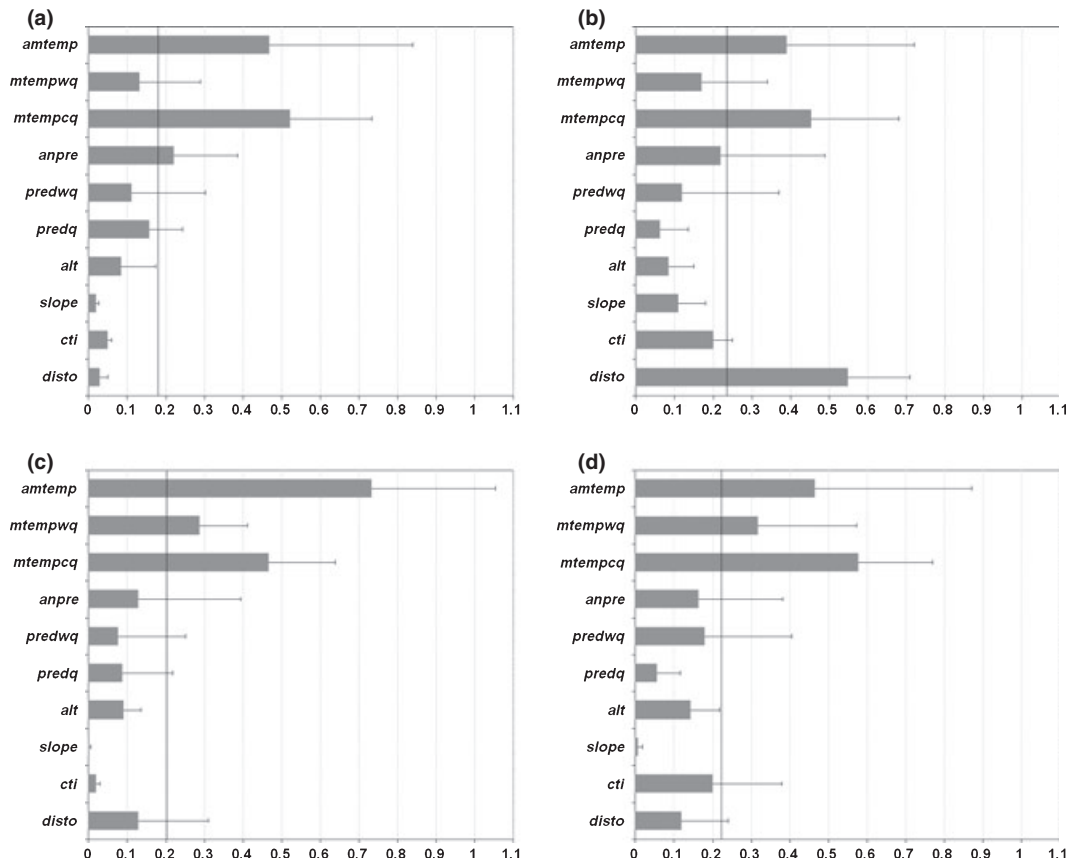
Figure 3 Relative importance of the environmental variables used to predict the distributions of (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii*. Bars refer to the mean value of relative importance obtained from eight different modelling algorithms and whisker to the standard deviation. Continuous black line represents the mean relative importance value across the 10 predictors.

is high for some areas, particularly in the south-west of the IP (Fig. 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 *c.* 100 km of river mouths (Fig. 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 (Fig. 6b). The decrease in 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 (Fig. 5c). Still, the ensemble of this species presents wide areas of high variability between single predictions (Fig. 6c). *Procambarus clarkii* is the species showing wider invasiveness potential. Suitable conditions are found across most of the IP except at higher altitudes (Fig. 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 (Fig. 6d).

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 & Gherardi, 2004). Our results demonstrate that even among models achieving good accuracy variability between models
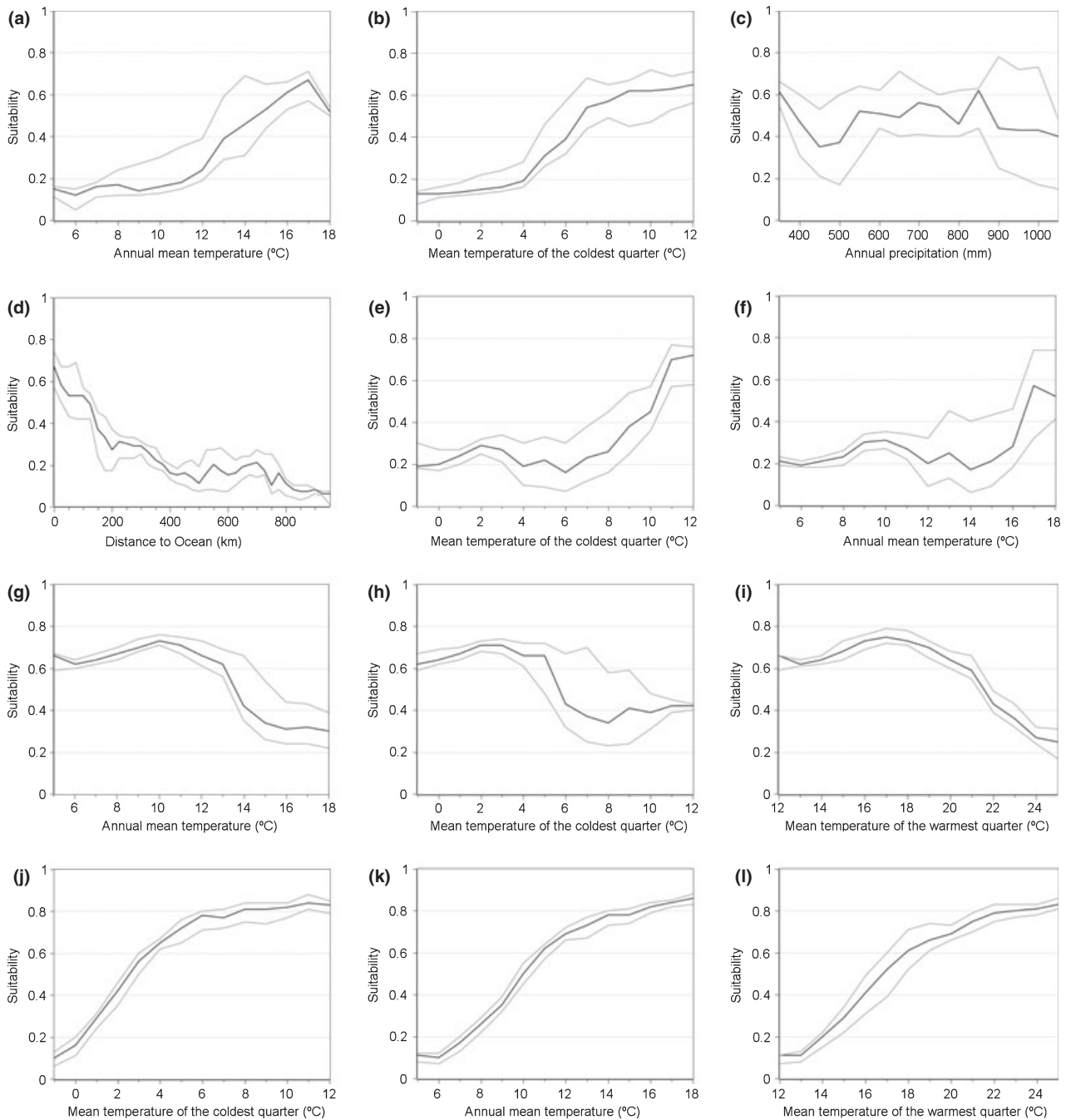
Figure 4 Suitability variation along the gradients of the most important environmental variables for predicting the distributions of (a–c) *Cherax destructor*, (d–f) *Eriocheir sinensis*, (g–i), *Pacifastacus leniusculus* and (j–l) *Procambarus clarkii*. Grey lines represent first and third quartile and black line the median.

can be high (e.g. for *P. leniusculus*). In this context, the use of ensembles of NBM for examining species/environment relationships is useful as 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 that 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'
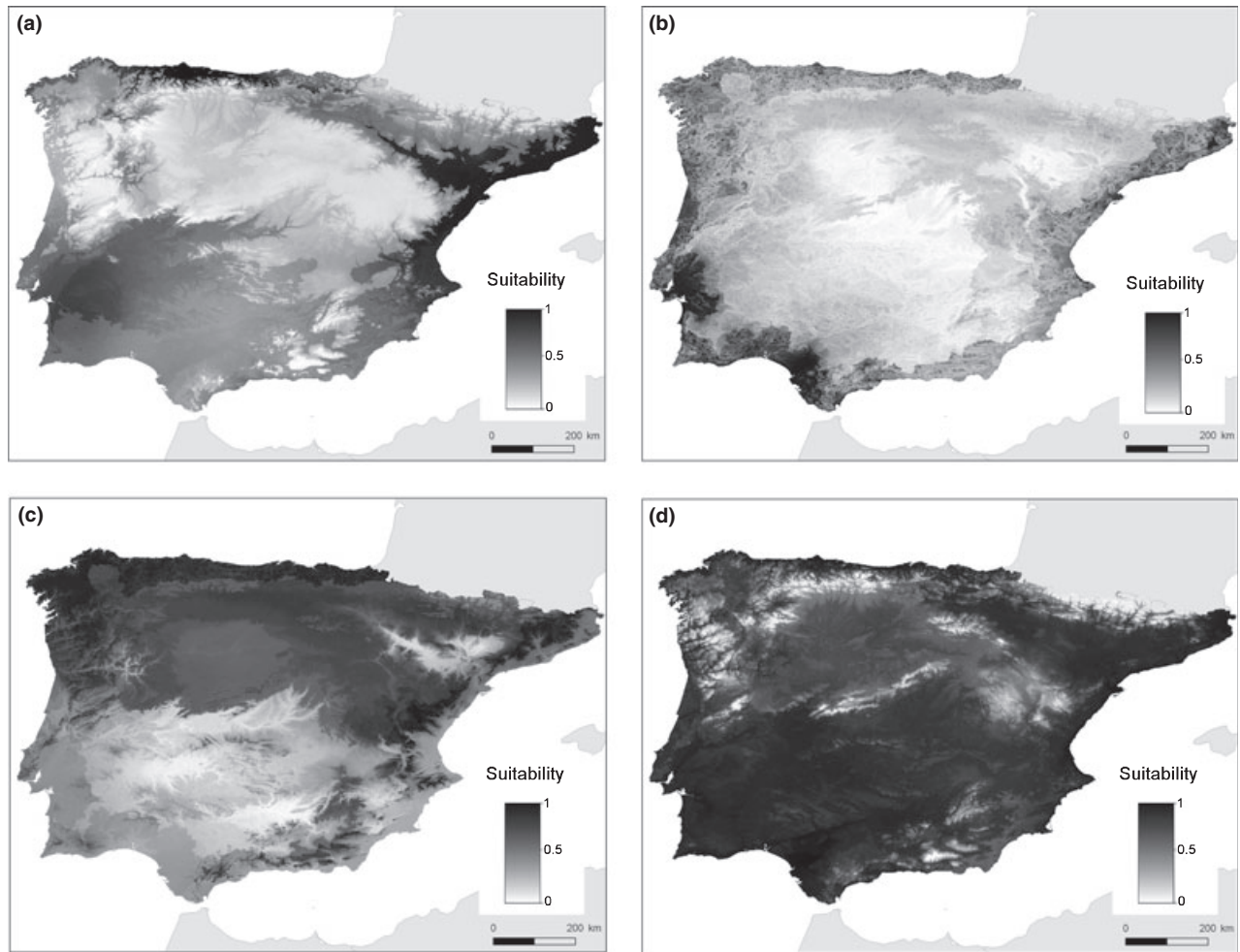
Figure 5 Consensus potential distribution models obtained by a weighted average of the single models for (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* in the Iberian Peninsula.

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 & 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 into the spatial patterns of the consensus suitability projected for this area. For example, *C. destructor* and *Procambarus clarkii*, which are two warm water crayfish, present rather different invasive abilities for the IP. The consensus model for *Procambarus 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 *Procambarus 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 & 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.
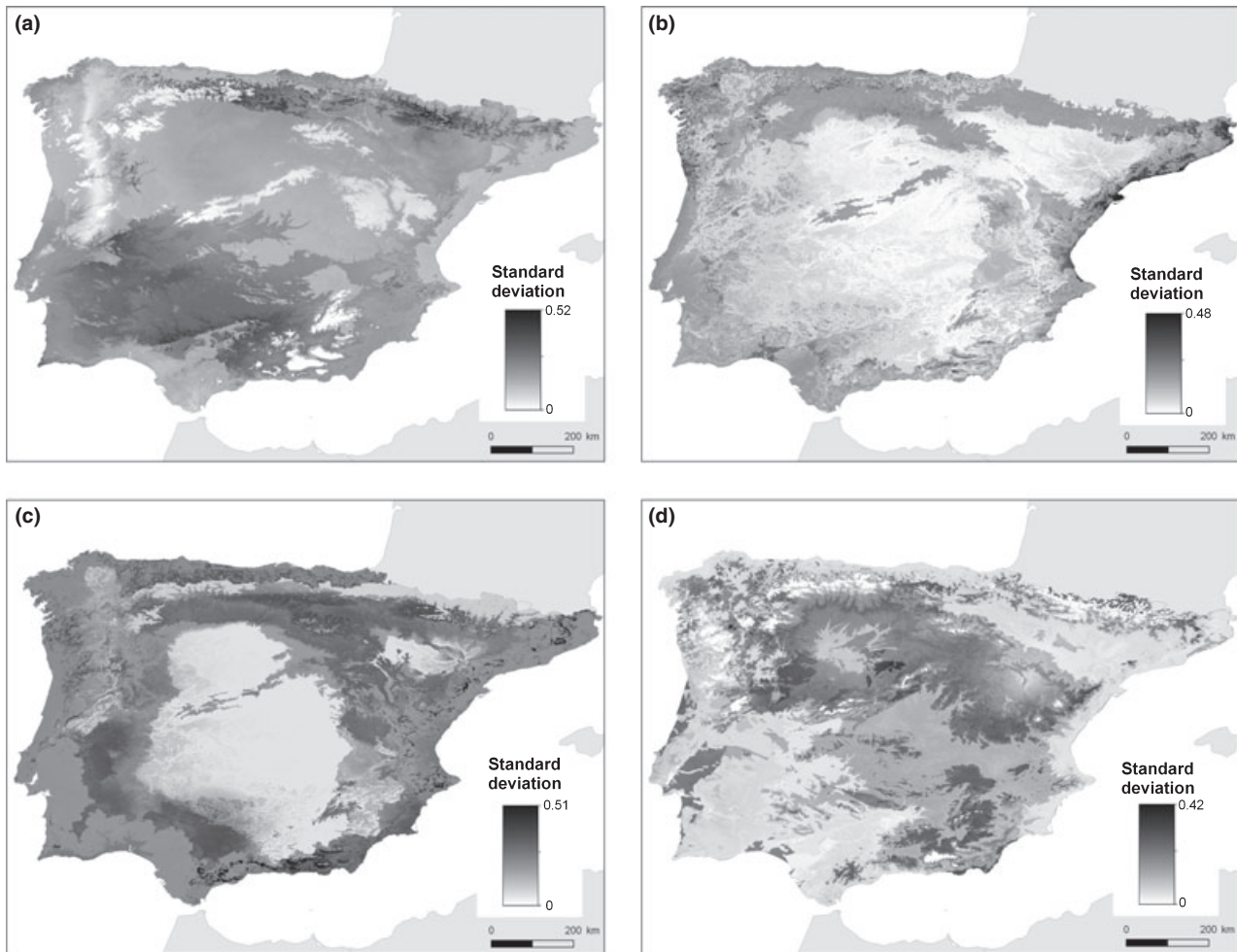
Figure 6 Variability among the single-models included in the ensembles for (a) *Cherax destructor*, (b) *Eriocheir sinensis*, (c) *Pacifastacus leniusculus* and (d) *Procambarus clarkii* in the Iberian Peninsula.

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 north-west of the peninsula might be the one of highest concerns 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.

Procambarus clarkii remains unrecorded only in a reduced number of areas, of which the two wider ones are the north-west 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 shows that in the north-west of the IP, especially the lowlands, many areas are highly suitable and, similar to *P. leniusculus*, in the absence of control measures, these will probably become part of the *Procambarus 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 colonization rate of *C. destructor* in the IP compared with the previous two crayfish species. The initial areas of introduction of both *P. leniusculus* and *Procambarus clarkii* (upper central and south-west Spain respectively) took place in areas predicted as highly suitable, which certainly favoured their naturalization 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 south-west of the IP for *C. destructor* or the upper sections of the Douro and Tagus basins for *Procambarus 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 recognized. 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 & 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 into 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 *Procambarus 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 *Procambarus 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.

REFERENCES

Anastácio, P.M., Parente, V.S. & Correia, A.M. (2005) Crayfish effects on seeds and seedlings: identification and quantification of damage. *Freshwater Biology*, **50**, 697–704.

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

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

Austin, M.P., Smith, T.M., Van Niel, K.P. & Wellington, A.B. (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.

Beatty, S.J. (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.

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

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

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

Cabral, H.N. & Costa, M.J. (1999) On the occurrence of the Chinese mitten crab, *Eriocheir sinensis*, in Portugal (Decapoda, Brachyura). *Crustaceana*, **72**, 55–58.

Capinha, C., Leung, B. & Anastácio, P. (2010) Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. *Ecography*, doi: 10.1111/j.1600-0587.2010.06369.x.

Costa, A.M., Bruxelas, S., Bernardo, J. & Teixeira, A. (2009) Colonization of Rio Maças (North Portugal) by two exotic crayfish *Pacifastacus leniusculus* and *Procambarus clarkii*. *World Conference on Biological Invasions and Ecosystem Functioning* (ed. Bioleif), p. 25. Candeias Artes Gráficas, Oporto.

Cruz, M.J., Segurado, P., Sousa, M. & Rebelo, R. (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*. *Herpetological Journal*, **18**, 197–204.

Cuesta, J., González-Ortegón, E., Rodríguez, A., Baldó, F., Vilas, C. & Drake, P. (2006) The decapod crustacean community of the Guadalquivir estuary (SW Spain): seasonal and inter-year changes in community structure. *Hydrobiologia*, **557**, 85–95.

Dittel, A.I. & Epifanio, C.E. (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.F., Purschke, O., Marquez, J.R.G., Lautenbach, S. & Schroder, B. (2008) Components of uncertainty in species distribution analysis: a case study of the Great Grey Shrike. *Ecology*, **89**, 3371–3386.

Elith, J., Ferrier, S., Huettmann, F. & Leathwick, J. (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, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Ficetola, G.F., 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.

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

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

Geiger, W., Alcorlo, P., Baltanas, A. & Montes, C. (2005) Impact of an introduced crustacean on the trophic webs of Mediterranean wetlands. *Biological Invasions*, **7**, 49–73.

Gilbey, V., Attrill, M.J. & Coleman, R.A. (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.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A. & Gro, N.P.S.W. (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. & Lek, S. (2010) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, doi: 10.1111/j.1600-0587.2010.06152.x.

Gutierrez-Yurrita, P.J., Martinez, J.M., Bravo-Utrera, M.A., Montes, C., Ilhéu, M. & Bernardo, J.M. (1999) The status of crayfish populations in Spain and Portugal. *Crayfish in Europe as Alien Species*, **11**, 161–192.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (Roc) curve. *Radiology*, **143**, 29–36.

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

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

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

Holdich, D.M., Reynolds, J.D., Souty-Grosset, C. & Sibley, P.J. (2009) A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. *Knowledge and Management of Aquatic Ecosystems*, **(394–395)**, 11.

Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression*. Wiley, New York.

IPCC (2007) Climate change 2007: the physical science basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller), p. 996. Cambridge University Press, Cambridge, UK.

IUCN (2009) *Red list of threatened species*. Available at: http://www.iucnredlist.org (accessed 12 May 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 6 September 2009).

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in

predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.

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

Nakata, K., Hamano, T., Hayashi, K.I. & Kawai, T. (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.

Nyström, P. (1999) Ecological impact of introduced and native crayfish on freshwater communities: European perspectives. *Crayfish in Europe as Alien Species*, **11**, 63–85.

Nyström, P. (2002) Ecology. *Biology of freshwater crayfish* (ed. by D.M. Holdich), pp. 192–235, Blackwell Science, Oxford.

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.

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

Peterson, A.T. & Vieglais, D.A. (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience*, **51**, 363–371.

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

Roura-Pascual, N., Brotons, L., Peterson, A.T. & Thuiller, W. (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.

Rudnick, D.A., Hieb, K., Grimmer, K.F. & Resh, V.H. (2003) Patterns and processes of biological invasion: the Chinese mitten crab in San Francisco Bay. *Basic and Applied Ecology*, **4**, 249–262.

Sanchez-Fernandez, D., Bilton, D.T., Abellan, P., Ribera, I., Vezasco, J. & Millan, A. (2008) Are the endemic water beetles of the Iberian Peninsula and the Balearic Islands effectively protected? *Biological Conservation*, **141**, 1612–1627.

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

Sousa, R., Dias, S.C., Guilhermino, L. & Antunes, C. (2008) Minho River tidal freshwater wetlands: threats to faunal biodiversity. *Aquatic Biology*, **3**, 237–250.

Souty-Grosset, C., Holdich, D.M., Noël, P.Y., Reynolds, J.D. & Haffner, P. (2006) *Atlas of Crayfish in Europe*. Muséum national d'Histoire naturelle, Paris.

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

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

Usio, N., Nakajima, H., Kamiyama, R., Wakana, I., Hiruta, S. & Takamura, N. (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.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Native and invasive occurrence records.

Table S1 Parameterization values tested for predicting the invaders distributions.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

César Capinha is a PhD candidate at the 'IMAR – Centre for Sea and Environment' in the University of Évora, Portugal. Its research interest focuses on predictive tools aiming to understand distribution ranges and dispersal of both native and alien species. In his PhD, he is working with four problematic invasive decapods in the inland waters of Iberian Peninsula. This research subject also led him to develop interests in the fields of freshwater ecology and management of invasive species.

Pedro Anastácio is Professor Auxiliar, at the University of Évora, Portugal and a Senior Researcher at the 'IMAR – Centre for Sea and Environment'. He lectures Ecological Modelling, Quantitative Methods in Ecology and more recently Invasive Species Impact and Control. Its main current research interests are modelling freshwater invasive species dispersion and impact and also wetland ecology. Most of its research deals with the ecology of water-associated organisms, namely, freshwater crayfish, beach amphipods, bivalves, amphibians and waterfowl.

Author contributions: C.C. and P.A. conceived the ideas; C.C. collected the data; C.C. and P.A. analysed the data; and C.C. led the writing.

Editor: Wilfried Thuiller