

Research Article

Modelling of the potential distribution of *Limnoperna fortunei* (Dunker, 1857) on a global scale

Mônica de Cássia Souza Campos^{1*}, André Felipe Alves de Andrade¹, Bárbara Kunzmann¹, Danielle Diniz Galvão¹, Fabiano Alcísio Silva¹, Antônio Valadao Cardoso¹, Marcela David Carvalho² and Helen Regina Mota²

¹CETEC SENAI Technological Centre – Av. José Cândido da Silveira, 2000 - Belo Horizonte/MG, 31170-000 Brazil

²The Electrical Power Company of Minas Gerais – CEMIG, Av. Barbacena, 1200, Santo Agostinho, Belo Horizonte/MG, 30190-131 Brazil

E-mail: mosouza@fiemg.com.br (CMSC), andrefaandrade@gmail.com (AFAA), bgkunzmann@gmail.com (BK), daniellebioufmg@gmail.com (DDG), fabianobiologia2@gmail.com (FAS), avcardoso2007@gmail.com (AVC), marcela.david@cemig.com.br (MCD), helen.mota@cemig.com.br (HRM)

*Corresponding author

Received: 20 November 2013 / Accepted: 25 July 2014 / Published online: 6 August 2014

Handling editor: Vadim Panov

Abstract

Predictive modelling of species' distributions is an important tool in biogeography, evolution, ecology, conservation, and invasive-species management. In this study we applied four different algorithms: Mahalanobis Distance, Domain, GARP and MAXENT, using them to predict the potential distribution of *Limnoperna fortunei*, a freshwater mussel native to Southeast Asia and a major fouling pest of water supply systems in Hong Kong, Japan, and South America. For model input, we compiled native and invaded occurrence data from Asia (71 points) and South America (248 points) from the literature and BIOCLIM's environmental layers related to air temperature and precipitation. To evaluate model quality we used different "training" and "test" data sets. On the Mahalanobis Distance and Domain algorithms, three sets of training data were used: 1) Asia points; 2) South America points; 3) Asia and South America points. For MAXENT the combinations were: 1) South America points (25% test data/75% training data); 2) Asia points (25% test data/75% training data); 3) South America training data/Asia test data; 4) Asia training data/ South America test data; 5) Asia + South America points (25% test data/75% training data). Comparing the responses of the four types of algorithms used, it was found that MAXENT was the most conservative model (i.e. it produced a smaller area of suitable habitats) followed in order by GARP, Domain and Mahalanobis Distance, which proved to be the widest. In general, the best results corresponded to models in which the points of occurrence covered a greater environmental variability (Asia+South America 25% test data/75% training data). They showed better performance for predicting correctly the occurrence of regions already known to host the species. An *ensemble* map was produced based on the best scenarios for each algorithm. This tool performed well in assessing the potential global distribution of *L. fortunei* even though it was generated from climatic macro variables without the use of locale-specific abiotic variables, which are more difficult to obtain.

Key words: invasive species, golden mussel, spatial distribution models, niche modelling

Introduction

Species Distribution Models (SDMs) have the objective of characterizing the ecological niche of a species and projecting it within geographical space. The maps of potential distribution thus generated are useful, for example, in forecasting the capacity of invasion of exotic species (Rödger et al. 2009). According to Kluza and McNyset (2005) the modelling of a spatially explicit ecological niche depends on the supposition that the ecological niche of a species (in the Grinnell 1917 sense) offers a stable restriction on its

geographical distribution and that this current distribution contains sufficient information for understanding the restriction (Peterson 2003). In addition to the environmental conditions, other factors also can influence the distribution of a species, for example, biotic relationships and its capacity of dispersion - either by means of its own movements or the dispersion of propagules by external agents, as this element determines which parts of the world are accessible to the individuals of the species (Soberón 2007). In the modelling of the potential spatial distribution, nevertheless, the potential niche for the species

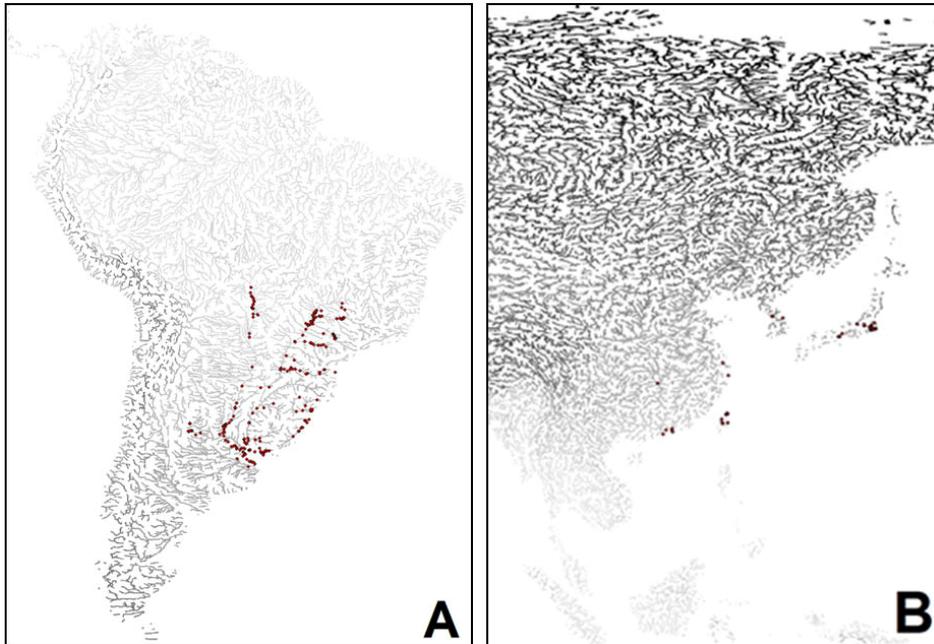


Figure 1. Points of occurrence of *Limnoperna fortunei*. Occurrence data were taken from a database created by Minas Gerais Technological Centre – (CETEC 2011) A) South America; B) Asia

of interest is modelled and is represented by the whole space corresponding to areas that prove to be environmentally suitable for its establishment.

For the development of models of ecological niches (e.g. SDMs), there is a fundamental need to understand the natural history of the individual being studied. In this case, the species is the freshwater bivalve invasive mollusc *Limnoperna fortunei* (Dunker, 1857), popularly known as the golden mussel. Its natural habitat is the Chinese rivers and streams of Southeast Asia. In recent years, as a consequence of the increase in international trade and the great flow of craft, this mollusc has been expanding its distribution into various parts of the world. On the Asian continent, in 1965, it invaded the waters of Hong Kong reaching Japan, Taiwan, Cambodia, Indonesia, Korea, Laos, Thailand and Vietnam (Morton 1996; Ricciardi 1998) and in 1991, it also reached the waters of South America in the estuary of the Río de la Plata river, in the Argentine (Pastorino et al. 1993). From then on, it swiftly extended its distribution over the South American continent (Figure 1A).

Morton (1977) defined *L. fortunei* as a species adapted for colonization over a wide range of aquatic environments, possessing various important

features for a successful invader, such as a cycle of short duration, capacity of rapid growth and high fertility, in addition to wide physiological tolerance to several abiotic factors that frequently limit other aquatic invertebrates (Oliveira et al. 2011). Its dispersion is closely associated to activity of human origin and may also utilize natural mechanisms (Darrigran and Damborenea 2006). In the larval planktonic phase natural dispersion occurs by means of water currents, but in the adult and juvenile stages it can also occur by fixing to substrates.

L. fortunei lives on average three years and reaches between 3 and 4 cm when adult. It is generally found in locations with oxygenated water, although it is capable of surviving successfully in saline waters up to 3 psu, lakes, humid areas and other water courses. However, this species is capable of tolerating concentrations ranging from distilled water to solutions containing 20% seawater. It colonizes waterbody margins and bottom substrates in densities that range from 1 to 150,000 ind/m².

The invasion of the golden mussel can cause extensive negative impacts for the environment, and also economic loss, both caused basically by extensive population growth. Change in the diet

of native predators, establishment of competition for space and food, change in the transparency of the water bodies containing denser populations of the mussel, creation of new microhabitats and the shift of native species are some of the ecological impacts related to this invasive bivalve (Darrigran et al. 1998; Penchaszadeh et al. 2000; Darrigran 2002; Mansur et al. 2003; Mansur et al. 2004; Darrigran and Damborenea 2005; Sylvester et al. 2007; Sardiña et al. 2008; Darrigran and Damborenea 2011). Economically, especially in industries that use raw water, e.g. the hydroelectric sector, the species can result in considerable losses by embedding firmly in layers on various submerged surfaces, such as wood, rock, plastic and even glass (Faria et al. 2006). The resulting biofouling blocks pipelines and produces head loss and stoppages of the systems for maintenance (Darrigran et al. 2007). Moreover the infesting of grids and other component structures of hydro-power plants increases the frequency of cleaning operations, which results in stoppages of machines and reduction of power generation. Other sectors can also be impacted such as water supply, agriculture, fluvial transport and aquaculture operations.

Frequently the detection of an invasive species occurs when the level of infestation is already quite advanced. Thus the modelling of the potential distribution and the risk maps generated represent essential tools for directing focussed efforts on prevention, control and reduction of the impacts caused by the invasion of *L. fortunei*.

Objectives

The idea was to compare the forecasting performance of four algorithms commonly utilized to generate potential distribution maps: two models that utilize environmental distances as metrics: Mahalanobis Distance, Domain; the genetic algorithm, GARP and an algorithm based on the principle of maximum entropy MAXENT. In this way it was intended to forecast the dispersion of the *L. fortunei* species on a global scale.

Material and methods

Collection of occurrence data

Part of the *L. fortunei* occurrence data were taken from a database created by researchers of the Minas Gerais Technological Centre – (CETEC 2012) compiled by consulting the scientific literature or generated by previous monitoring programmes (Campos et al. 2012). Altogether, 319

georeferenced records of the occurrence of the organisms were utilized; 71 points of occurrence were in Asia and 248 points in South America. These points are illustrated in Figure 1.

Obtaining environmental data - layers

The climatic layers utilized to run the distribution model were taken from the Worldclim database (Global Climate Data <http://www.worldclim.org/bioclim>) eight being initially selected from the total related to temperature and rainfall: Annual Average Temperature, Minimum Temperature of the Coldest Month, Average Temperature of the Coldest Quarter, Average Temperature of the Warmest Quarter, Rainfall of the Wettest Month, Rainfall of the Driest Month, Rainfall of the Wettest Quarter, Rainfall of the Driest Quarter. The layers with a high degree of covariance between them (> 0.7) were excluded and only one of them was chosen to represent the others. Thus only four layers were selected for the simulations: Annual Average Temperature, Minimum Temperature of the Coldest Month, Rainfall of the Wettest Month and Rainfall of the Driest Month.

Algorithms

The following algorithms were utilized to generate the distribution maps:

Mahalanobis Distance and Domain

These two algorithms use different metrics, being derivations of the Euclidian distance and are available in an “open source” platform denominated OpenModeller (<http://openmodeller.cria.org.br>).

The Mahalanobis Distance algorithm is structured about the existence of an optimum ecological point, defined by the construction of a centroid for all the points of occurrence in the whole ecological space. The smaller the distance, the greater is the similarity between regions, and the greater the probability of the species being present. The Mahalanobis Distance produces an envelope in the form of an ellipse around the “optimum” within the ecological space. When the algorithm is applied to a species potential distribution model, the main conditions of a group of habitat variables are typically compared with each other.

The Domain, as different from the Mahalanobis Distance, is not based on a centroid, but utilizes the Gower distance, and because of this has little influence on the sampling bias. In the case of

this algorithm, there are various envelopes around the point.

- Maximum distance in relation to the reference environmental space: 0.1 (above this the conditions are considered unsuitable for the presence of the species);
- “Nearest ‘n’ points”: 1 (which means that the distance was measured to the nearest point, while if it were 0 the environmental distance would be measured in relation to the average of all the points of occurrence).

GARP

GARP (Genetic Algorithm for Rule Set Production) – This is a much used algorithm, based on artificial intelligence and works by combining groups of rules with the intention of generating a more precise forecast in the region considered (Stockwell and Noble 1992). The rules represent a multivariate group of relationships between points of occurrence of the species and environmental variables. The algorithm utilizes bioclimatic, atomic rules and regression logistics (Stockwell and Peters 1999).

The algorithms utilized above are inserted in the OpenModeller Desktop 1.1.0 Platform, and to run the models, 50% of the points of occurrence were considered for training, 20 models altogether being run with a convergence limit of 0.01 and 400 iterations were made.

Three classes of models for each algorithm were generated according to the origin of the set of occurrence data utilized for the simulations, totalling 9 models generated:

- Class 1 “all the points” models: three models using all 319 points;
- Class 2 “Asia points” models: three models using only the 71 Asia points;
- Class 3 “South America points” models: three models using only the 248 South America points.

MAXENT

MAXENT is software based on the principle of maximum entropy for modelling of the species’ habitats. This algorithm requires the input of a set of layers or environmental variables (such as rainfall, altitude, etc.), as well as a set of georeferenced occurrence locations, to produce a distribution model of the species in question (Phillips et al. 2006; Elith et al. 2010). The MAXENT algorithm estimates the geographical distribution of the species looking for the

probability of distribution of maximum entropy (that is, more spread out, in other words, approximating more a uniform distribution), subject to a set of restrictions that represent the incomplete information on the desired distribution (Phillips et al. 2006). The 3.3.3k version was utilized. In this programme, the information available on the distribution of the species is presented as a set of variables of real value, denominated “features” and the restrictions are the expected value of each “feature” which should coincide with its empirical average, the average value of a set of points of the sample collected from the distribution of the species (Phillips et al. 2006). Among the possibilities present in the MAXENT, ‘linear features’, and ‘quadratic features’ were utilized, as according to Phillips et al. (2006), the joint use of the two results in the variance of the environmental variable approaching the observed value.

With the purpose of evaluating the average behaviour of the MAXENT algorithm and to enable the statistical testing of the differences observed in the performances (Phillips et al. 2006), 10 replications were done for each model. In each, the occurrence data were partitioned (by random selection) into data for “training” and data for “testing”, in accordance with the following scheme:

Model Group 01. All the South America (AMS) and Asia (A) input occurrences (75% training/25% test);

Model Group 02. AMS input points (75% training/25% test);

Model Group 03. Asia input points (75% training/25% test);

Model Group 04. AMS input points and training / Asia points as test;

Model Group 05. Asia input points of entry and training/AMS points as test.

All the models to be tested fall into the category of presence-only models. Despite presence-only models having a place in modern ecology, they have potential limitations (Elith et al. 2006). According to these authors in many instances evaluation focuses on predictive performance, some known occurrences are withheld from model development and accuracy is assessed based on how well models predict the withheld data. In presence-only modelling, such withheld data are unlikely to provide a general test of model accuracy in predicting species’ distributions, because the occurrence records often have biases in both geographic and environmental space and

such biases will persist in common resampling designs. More importantly, withheld data are presence-only, which limits the options for, and power of, statistical evaluations of predictive performance.

On the other hand, usually the only reliable information on the distribution of organisms is from their recorded presence. Contrary to presence data, reliable absence data are rare and hard to obtain; confirming that a species is absent from a locality is a difficult task (Jiménez-Valverde et al. 2008) that becomes almost unaffordable in the case of the coarse resolution grid cells used in most studies.

Even knowing the limitations of the presence-only models, data on real absences were not considered in this work. *L. fortunei* is an invasive species, r strategist and with great capacity for the colonization of environments. Because of this, possible registrations of the absence of this species are still difficult to presume to be "real absences", as it is in expansion and it could colonize other localities. In this way the absences cannot be inferred with certainty. In other words, in the case of an invasive species like *L. fortunei*, the true potential range of occurrence may differ from the realized range because of dispersal limitation, competition or other factors, so that evaluating model performance is a complex task and use of observed absences may be misleading (Elith et al. 2006), for instance, true absences could be allocated in areas that are either unsuitable for the species, or that are suitable but currently do not host any population. In this way for the case of modelling the potential distribution of an invasive species the techniques that play down the importance of absence information may be better suited to estimation of the ecological and distributional potential of the species, whereas methods incorporating absence information more directly may be more suitable for estimating actual distributions of species (Jimenez-Valverde et al. 2011).

ArcGIS

The records of occurrence and their coordinates were previously analysed for consistency and treated in the software ArcGIS 9.3 of the ESRI with the purpose of eliminating the pseudo-absences generated outside the hydrography, a Buffer of 4 km being performed by means of the Spatial Analyst tool for the environmental layers of the Worldclim. The points that, even so, were not located within the rivers were manually

Table 1. Receiver Operating Characteristics ROC and Area Under the Curve AUC for the three classes of models generated by the Domain, GARP and Mahalanobis Distance algorithms.

Algorithms /Classes of models		AUC
Domain:		
Class 1	South America points	0.99
Class 2	Asia points	0.995
Class 3	All the points	0.99
GARP:		
Class 1	South America points	0.96
Class 2	Asia points	0.975
Class 3	All the points	0.95
Mahalanobis Distance:		
Class 1	South America points	0.97
Class 2	Asia points	0.97
Class 3	All the points	0.96

relocated to the nearest river. The outputs generated by the algorithms were transformed into rasters by the ArcMap tool.

Statistical

The statistics used to ascertain model quality were: Area Under the Curve (AUC) and the Receiver Operation Characteristic (ROC) calculated by the OpenModeller software itself and also by the MAXENT.

The evaluation of a model is based on the forecast performance and includes the determination of a minimum threshold of the quantitative value produced for the potential presence of a species. The sensibility of a model is defined as the proportion of true presences in relation to the total of presences predicted by the model. The specificity is defined as the ratio of true absences in relation to the total of absences predicted by the model. Thus, the Receiver Operating Characteristics (ROC) curve is obtained plotting the sensibility against 1 minus the specificity for different values of the probability threshold, generating an evaluation method of the threshold independent of the model (Manel et al. 2001). In addition, the area below the curve AUC is extensively used in species distribution modelling (SDM), characterizing the performance of the model, in all possible thresholds, based on a single value that can be used as an objective approach in comparing different models (Elith et al. 2006; Phillips et al. 2006). The AUC varies from 0 to 1, where 1 indicates high performance, while values lower than 0.5 indicate low performance (Luoto et al. 2005; Elith et al. 2006).

In spite of recent criticisms (e.g., Lobo et al. 2008), AUC can still be useful comparing models of a same species in a similar geographical space.

Table 2. Values of the Area Under the Curve (AUC) utilizing ROC. For the five classes of models generated by MAXENT.

Models	AUC
Group 01. All the input occurrences (AMS+Asia) (75% training/ 25% test)	0.970
Group 02. AMS input points (75% training/25% test)	0.979
Group 03. Asia input points (75% training/25% test)	0.992
Group 04. AMS input points and training / Asia points as test	0.979
Group 05. Asia input points and training / AMS points as test	0.994

Models with values over 0.75 are considered potentially useful (Elith 2002). In this work true records of absence were not considered so the calculation of the AUC made use of the “background” data (also called pseudo-absences) chosen uniformly and randomly from the study area (Phillips et al. 2006).

Lastly a composite map from the best predicted distributions from each of the four models was created. Using the raster calculator tool in ArcGIS, the average predicted likelihood of occurrence from the best models of the four different techniques was calculated and the standard deviation of the predicted occurrence, to provide a map highlighting where the models agree in their predictions, and where they don't.

Results

Performance analysis of the models generated by the AUC

The models generated by the **Domain**, **GARP** and **Mahalanobis Distance** algorithms presented optimum performance according to the values of AUC generated by the Openmodeller (AUC \geq 0.95) are presented in Table 1. The Class 3 Model “all the points” generated by these algorithms presented the worst performance, while all the Domain models obtained the best performance.

The models generated by the MAXENT presented AUC values over 0.97 (Table 2), also indicating good performance. Within the models generated by Maxent, those belonging to the Group 01 (“all the points”) also presented the lowest value for the AUC.

*Potential global distribution of *L. fortunei**

The potential global distribution scenarios of the invasive bivalve were distinct in accordance with the algorithm employed and the nature and size of the sample of the information on the real presence of the species (Figures 2 to 4).

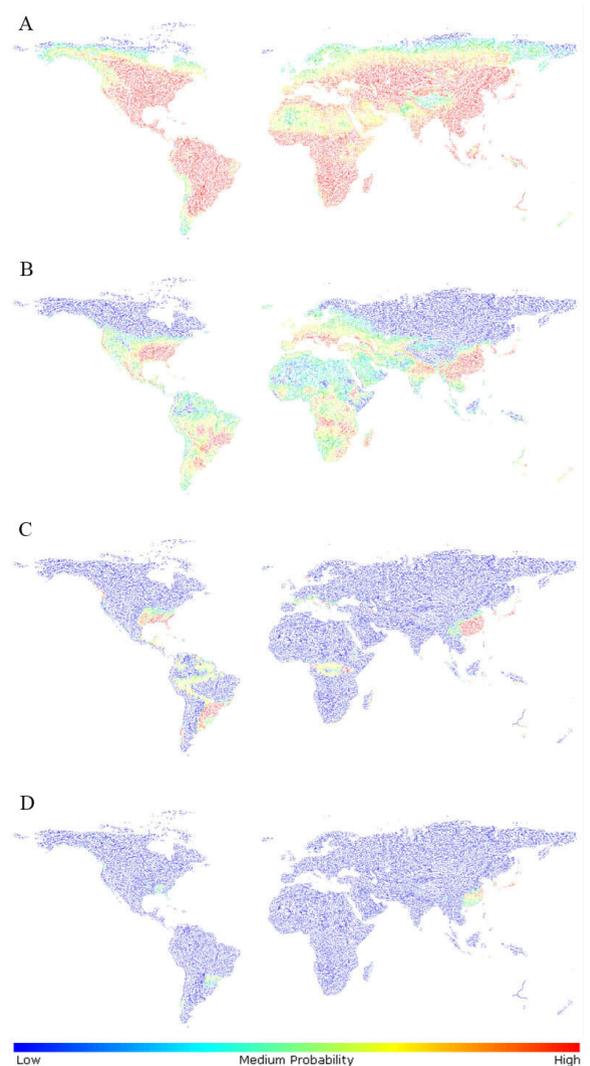


Figure 2. Potential distribution of *L. fortunei*, with records of presence from Asia, generated by the algorithms Mahalanobis Distance (A), Domain (B), GARP (C) and MAXENT (D).

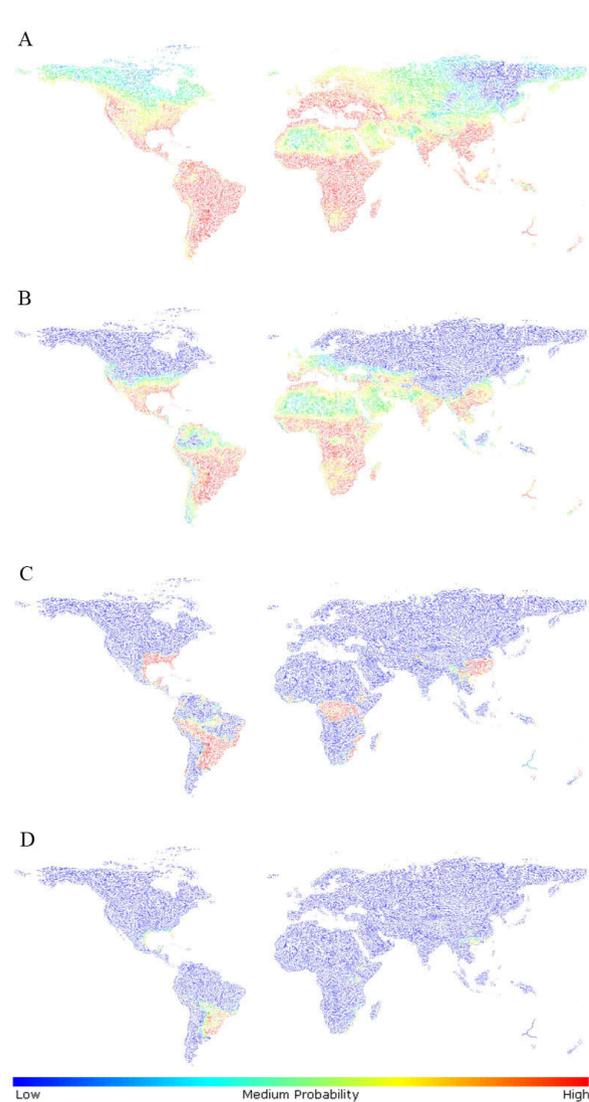


Figure 3. Potential distribution of *L. fortunei*, with records of presence from South America, generated by the algorithms Mahalanobis Distance (A), Domain (B), GARP (C) and MAXENT (D).

Mahalanobis Distance

The models generated by the Mahalanobis Distance algorithm gave an AUC between 0.96 and 0.97. In all the variations utilized regarding the input data (Figures 2A to 4A) this algorithm was, compared to the others, that which presented a great flexibility in the forecast of the potential distribution area of the invader. In all the scenarios considered, the models generated by

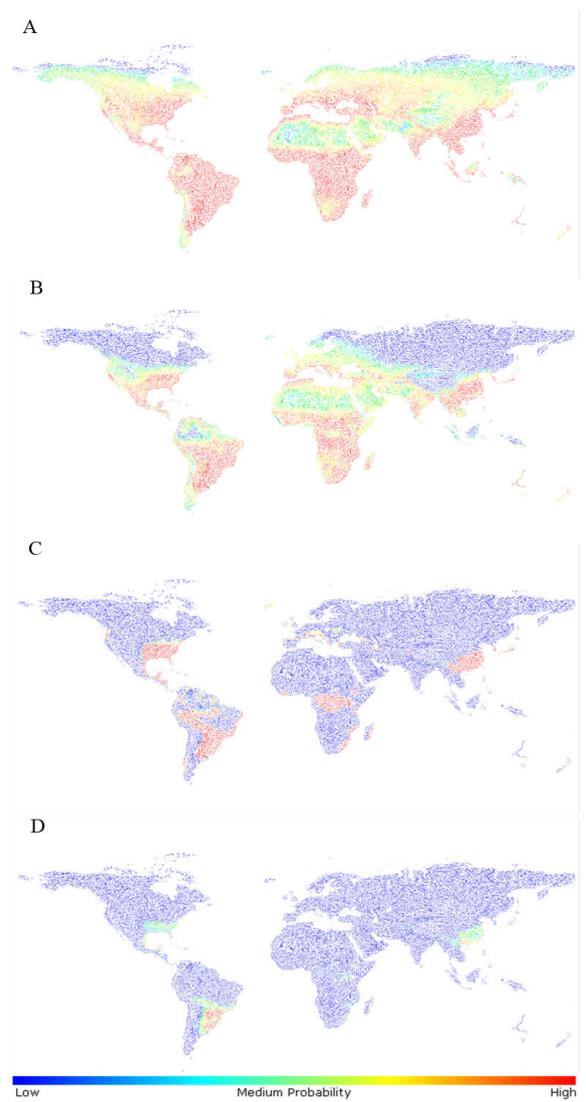


Figure 4. Potential distribution of *L. fortunei*, with records of presence from South America and Asia, generated by the algorithms Mahalanobis Distance (A), Domain (B), GARP (C) and MAXENT (D).

the Mahalanobis Distance, indicate the invasion of the bivalve with considerable levels of probability even in areas with extreme environmental conditions of temperature and/or rainfall, much in excess of its limits of tolerance (Ricciardi 1998) and, accordingly, unsuitable for its establishment.

The model generated by Mahalanobis Distance with records of Asia (Figure 3A), presented a flexible performance and worse than the subsequent one (South America), predicting the whole of South America and Africa with a high risk of

invasion. The model generalized also the occurrence of the species in Central and North America, primarily in Mexico and the United States and Europe, and indicated an average risk of invasion in places such as Finland, Norway and Britain which climatically would not be very propitious for the establishment of the species. This model furthermore did not predict a high probability of occurrence of the mussel in areas of known presence such as China.

Comparing the performance of the algorithm Mahalanobis Distance in relation to the origin of the points of occurrence, it can be seen in Figure 4A that the model generated from the actual points of occurrence of *L. fortunei* in South America, was able to predict correctly the invasion of Asia, indicating a high probability of occurrence in China, which is its native environment (Morton 1977), in addition to Japan, South Korea (partially) and Taiwan (Ricciardi and Rasmussen 1998), countries already invaded by the mussel. The model also forecasted correctly the invasion of South America, in the Plate basin, in the River Paraná, which occurred from 1991 on through the River Plate estuary and which at present reaches the headwaters of the River Paraná (Campos et al. 2012). Moreover the model based on the records from South America demonstrates that extreme areas such as the Sahara Desert and Northern Europe possess some chance of invasion, something that makes no sense.

The model produced utilizing all the points of occurrence (Figure 4A) demonstrated an intermediate performance in relation to those preceding it. This model predicted correctly the existence of the mussel in Southeast Asia including South Korea and also in South America. Such a model did not overestimate the spatial distribution of the species such as the model that considered only the Asia points, but was more flexible in predicting the high risk of invasion in Russia than the model based on presence points in South America.

Domain

In its turn the models generated by the Domain, (Figures 2 to 4B) were less general and visibly better than those produced by the Mahalanobis Distance. Based on its values of AUC = 0.99 they can be considered models of high predictive accuracy and they predicted the expansion of *L. fortunei* with an expected logic, based on the points of occurrence utilized by it.

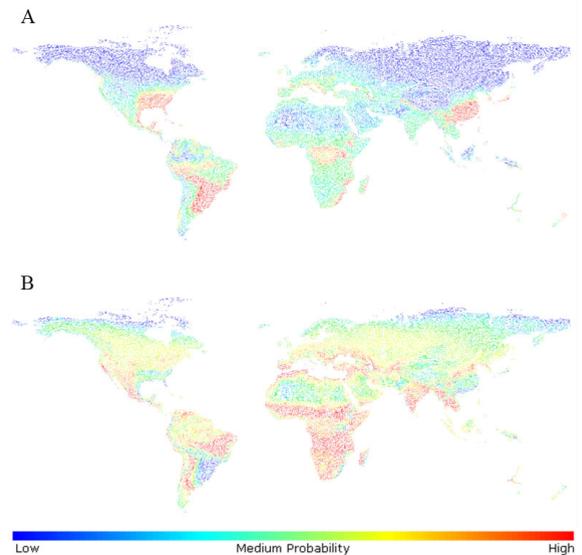


Figure 5. Maps of the averages (A) and deviations (B) of the best scenarios of potential distribution generated by the four algorithms utilized (Mahalanobis Distance, Domain, Garp and Maxent).

The model obtained by Domain utilizing only records from Asia (Figure 2B) indicated a high risk of invasion for certain regions north of the Equator such as the Southeast coast of the United States, Southern Europe and Southeast Asia where it was already expected to be predicted. It showed coherently that Canada and the North of Asia are places with very low chances of occupation by the mussel. The Domain model based on the presences of the Asiatic region indicated correctly the invasion of South America, although with less risk than in countries of the Northern Hemisphere. In being based on climatic variables of temperature and rainfall the model that utilizes records from Asia predicts areas of the USA and Mediterranean Europe with greater probability of invasion than South America, in the latter's already invaded part.

This model also predicted expansion to the Amazon basin, Central Africa and the West Coast of the USA.

In considering only the true presences of South America (Figure 3B), the model Domain predicted invasion primarily for the Southern Hemisphere and not so well for Southeast Asia, which is plausible considering the nature of the environmental data inputted, macro climatic variables with values related to the Southern Hemi-

sphere. It demonstrated a massive expansion in South America, with the exception of the axis corresponding to the Andes (cold and dry) and an equatorial area corresponding to the hot region and with the maximum average annual rainfall levels on the continent, two situations with extreme averages and distinct from the average environmental conditions related to the points of occurrence. It should be further noted that high altitudes in South America, as a determinant factor of the climate, also contribute to reduce the environmental suitability for the species as is shown by the low risk indicated for the Andean region and the Guiana Highlands. Darrigran et al. 2011 show that certain conditions connected to the concentration of solids in suspension, intermittence of water flow and salinity are some aspects that limit the dispersion of *Limnoperna* for the Andean tributaries of the Plate basin in the Argentine.

According to such a model, the mussel will still invade the South-Central region of Africa, in addition to Mexico, the south of the USA, and the Mediterranean region.

Including the environmental data of the South America points and the Asia data (Figure 4B), the model generated becomes less restrictive than the previous ones increasing its capacity of prediction for both hemispheres.

GARP

The models generated by this algorithm (Figures 2 to 4C) tended to be more conservative than those based on environmental distances, maintaining the areas with great potential of invasion by the mussel near to the points of occurrence supplied to generate the model and diminishing the degree of environmental suitability generally.

Taking into consideration only the Asia points (Figure 2D) the forecast of a high risk of invasion provided by this model was restricted in the majority of cases to areas with a subtropical climate north or south maintaining a relation with the average weather conditions of the occurrence data. It was able to predict the native region of the bivalve although it limited the area of invasion even in Southeast Asia. The model predicted correctly the invasion of South America, including the locations that have been invaded recently such as the headwaters of the River Paraná. It showed that the Southeast of the United States possesses high potential for being invaded. Generally the model made correct forecasts but it was very limited to the actual presence data.

Utilizing the South America points (Figure 3D) there was a greater expansion of the forecast of areas potentially invaded in the countries of the Southern Hemisphere in relation to the previous model and reduction of the risk forecast for the countries of the Northern Hemisphere. The model also predicted occurrence in its native territory, China, not however demonstrating the invasion in Japan.

The inclusion of all the points in the model (Figure 4D) also expands a little more its capacity of forecasting. From this simulation the layout of a probable route of dispersion can be seen, reaching the Amazon basin. This was the best of the three models generating an intermediate scenario in relation to those produced with Asia points or with South America points.

MAXENT

This was unequivocally the most restrictive algorithm of all (Figures 2 to 4D) and that implied greater adhesion between the real presence data and the data simulated. AUC values for this algorithm were between 0.970 and 0.994.

The model generated contemplating the Asia data (Figure 2D) was excessively restrictive. It predicted correctly the occurrence of the invader in Asiatic areas that are currently occupied by the species. But its general performance tends to underestimate the potential of invasion, for example, in not forecasting the expansion of the species in Brazilian territory and in not forecasting the invasion of the River Plate region. The Model based on South America data (Figure 3D) was limited to forecasting the invasion of South America and some few points located in the Asiatic region where the species is native.

Generally the analysis for the models generated by the MAXENT maintained consistency with the models generated by the other algorithms repeating the sampling bias of the spatial distribution of the records of occurrence: models with input data of AMS predicted better for the Southern Hemisphere and models generated with the Asia data predicted better for the Northern Hemisphere.

The MAXENT Model with all the occurrences, AMS and Asia (Figure 4D) was considered to be of better performance, as in addition to predicting correctly the occurrence in regions already known for the species, it was the one that expanded more the possibility of distribution of *L. fortunei*, reaching areas of Southeast Asia, a considerable part of South and Central America, in addition to the Southeast region of the United States.

Discussion

According to Jiménez-Valverde (2012) a good use of species distribution models requires a clear distinction of the differences between potential and realized distributions (Soberón 2007). While potential distribution refers to the places where a species could survive and reproduce due to the existence of suitable environmental conditions, realized distribution refers to the places where a species actually lives. In the case of the modelling of invasive species for example, forecasting potential distribution may be the most appropriate approach while the estimates of realized distribution are more indicated for conservation studies (Peterson 2006). Because of this, according to Jiménez-Valverde (2012) different strategies, including data and modelling techniques, are required for approaching one concept or the other, as are the strategies used to evaluate the models.

In a first analysis, taking into account only the AUC values obtained, all the algorithms presented a good degree of accuracy in their forecasts ($AUC > 0.95$). However, this tool proved unsatisfactory, when applied to the analysis of the maps generated

According to Lobo et al. (2010) similar AUC scores can be obtained with predictions of the distribution in geographic space very different one from the other and hence, these measures do not provide reliable estimates of SDM performance.

AUC is only truly informative when there are true instances of absence available and the objective is the estimation of the realized distribution (Jiménez-Valverde 2012). When the potential distribution is the goal of the research, the AUC is not an appropriate performance measure because the weight of commission errors is much lower than that of the omission errors. Thus in the case of potential distribution of invasive species, a form of validation suggested as partially possible by Lobo et al. (2010) would be to examine the success of the predictions of presences in spatial or temporal scenarios.

The results generated by this work are in accordance with the spatial distribution portrayed in the prior work of Kluza and Mc Nyset (2005) which used only the GARP and was based on a different climatic and environmental database from that utilized in this work. The different algorithms employed showed that even when climatic macro variables are used, such tools can be effective, above all for an approach on a

global scale. Aligned between themselves and with the preceding work, all the algorithms were capable of forecasting the invasion of South America and conversely, also indicated the establishment of the species in its native region, although with different degrees of predictive power and reliability. All the models indicate the great invasive potential that this species has, with the capacity to establish itself basically on all the continents.

The influence of the size and origin of the sample of presence data on the performance of the algorithms is clear. Probably such differentiated responses between the algorithms tested reflect the influence of the manner in which the presence data are utilized by these different techniques.

The results of this work confirm the well-known fact that the predictive performance of individual spatial distribution models varies widely among methods and species (Poulos et al. 2012).

The algorithms based on environmental distances were more flexible and generalized the area of potential distribution of the species when compared to the algorithms of the artificial intelligence type such as the MAXENT or GARP. These latter were very coincident with the input data, modelling the niche in a more restrictive manner and close to the real presence data. Within the set of methods used, those that characterise the background environment and that can differentially weight variables (Maxent and Garp) were more conservative than those that use presence data alone (Mahalanobis Distance, DOMAIN).

Mainly for the potential distribution maps generated by GARP and MAXENT, it was demonstrated that the increase of input information related to real presences can improve performance and eliminate problems of spatial autocorrelation. In the majority of cases tested, the models that included all the records of occurrence having presented better performance than those that considered data subsets. Thus the models with records of occurrence only from Asia enlarge the risks of invasion for environments of the Northern Hemisphere and underestimate the risk of invasion for the areas of the Southern Hemisphere and the replies occurring when records of occurrence only from South America are used are in the contrary sense. According to Fielding and Bell (1997), models that utilize all the available data will be, on average, better than those models based on data subsets. Consequently, if the data are few or partitioned so that the size of the training set is smaller, there is a tendency to reduce the accuracy

of the model. In the present study 319 points of occurrence of the species in the world were considered, 78% of the records coming from South America (AMS-invaded region) and only 22% from Asia (AS- native region). These results reinforce the idea of an ideal approach proposed by Jimenez-Valverde et al. (2011) for the modelling of invasive species that would take into consideration all available information coming from native and invaded regions, as well as those provided by different time slices, since it may enhance characterization of the species' fundamental niche.

According to Jiménez-Valverde et al. (2008) complex techniques may be more suitable to model the realized distribution than simple ones, which may be more appropriate to estimate the potential distribution. Those techniques that are able to establish the more complex relationships between dependent and independent variables will overfit the presence data more strongly. Unavoidably, this will result in predicted extents of occurrence that are smaller than those suggested by simpler techniques. This fact was explicit in the distribution maps generated by Maxent, that proved to be overly adjusted to the areas of presence registration (Figures 2D and 3D) representing much more realized distribution of *L. fortunei* than its potential distribution.

In the contrary sense, the models generated by the Mahalanobis Distance algorithm, generalized excessively the area potentially occupied by the species showing low capacity of discrimination regarding the risk of invasion of the different geographical regions and thus had a low degree of reliability.

Considering that the focus of this study is the modelling of a quite aggressive invasive species and with ample tolerance limits to environmental variables, it was adjudged that very conservative forecasts such as those produced by MAXENT are not appropriate for the modelling of potential distribution of an invasive species. However, excessively flexible forecasts such as those generated by the environmental envelopes estimated by the Mahalanobis Distance, are equally ineffective for the focus of this work.

For this reason the choice of the algorithm should be orientated by the modelling's objective and the characteristics of the species. In the case of invading species, with wide environmental tolerance, more generalized replies such as those given by the Domain, for instance, can trace an interesting scenario from a more preventive

perspective showing the whole potential of expansion of the invader, whereas more adjusted scenarios such as those portrayed by the MAXENT can indicate invasion hot spots for which priority actions and allocation of resources should be targeted (the realized niche). Another question related to potential distribution models refers to the fact that even in suitable conditions, the risk of invasion should take into account the accessibility to the new regions and the favourable inter specific relationships, aspects not considered by the algorithms utilized (Soberon 2007).

We established the following classification according to the capacity of generalization of the tools: Mahalanobis Distance, Domain, GARP and MAXENT. Similar evaluations of the forecasting performance of the algorithms have already been obtained by other studies: MAXENT outperforms GARP (Phillips et al. 2006) and some presence-only methods (e.g. DOMAIN, ENFA, Hirzel et al. 2002) have advantages over BIOCLIM (Loiselle et al. 2003).

Ensemble maps constructed based on the averages and standard deviations of the best scenarios of the four algorithms are presented in Figure 5. This integrating tools permits us to account for the possible limitations or biases of any one modelling technique. As a result, this tends to iron out the differences and limitations created by different techniques and to identify where all the models agree in their level of predicted occurrence, and where they vary. This provided a measure of certainty in the predicted occurrence (Rochinni et al. 2011).

Considering the *ensemble* maps, it can be seen from the average of the best forecasts of all the models, that the invasive mussel can encounter regions of great environmental suitability on all the continents, although on some areas, as the African continent, the deviations present a certain degree of variability and less reliability.

It deserves special mention the fact that all the models predicted the invasion of the Southeast part of the United States, and the region around the Great Lakes which was the target of invasion by the mussel zebra (*Dreissena polymorpha* (Pallas, 1771) - a dreissenidae bivalve with ecological characteristics very similar to the golden mussel. *L. fortunei* has been indicated in the literature as an invasive species more aggressive than *D. polymorpha* and that because of this could occupy the southern region of North America (Karatayev et al. 2007) not invaded by the first. The global models generated confirm this

invasion in the Southeast of the USA, however they also reveal the importance of temperature as a determinant factor of the area of expansion of *L. fortunei*, indicating low potential for its establishment in regions with extreme temperatures (northern reaches, e.g. Finland and Norway).

Considering that attachment to vessels is by far the most important dispersion mechanism of *Limnoperna fortunei*, Boltovskoy et al. (2006) suggested that the Amazon, Orinoco and Magdalena basins in South America are under high risk of invasion by this mussel, especially through their estuarine gateways. The *ensemble* map of the forecasts foresees high risk of invasion for the Amazon and Magdalena Basins, however for the Orinoco region, the variability of the responses of the algorithms was high and high environmental suitability for the invader in this region was not confirmed.

The average of all models also shows that the species currently restricted to the Plate basin on the South American continent, may encounter favourable environmental conditions for spreading into a substantial portion of this continent with the already expected impacts on the neotropical biota.

Acknowledgements

This work was financed by CEMIG - *Companhia de Energia de Minas Gerais* (The Power Utility Company of the State of Minas Gerais) and the ANEEL - The National Electrical Power Agency, and benefited from help with scholarships for scientific initiation from the CNPq (the National Research Council). We especially thank to anonymous reviewers and the scientific editor for their constructive comments on the manuscript.

References

- Boltovskoy D, Correa N, Cataldo D, Sylvester F (2006) Dispersion and ecological impact of the invasive freshwater bivalve *Limnoperna fortunei* in the Rio de la Plata watershed and beyond. *Biological Invasions* 8(4): 947–963, <http://dx.doi.org/10.1007/s10530-005-5107-z>
- Campos MCS, Lanzer R, Castro PT (2012) Hydrological stress as a limiting factor of the invasion of *Limnoperna fortunei* (Dunker, 1857) in the Upper Paraná River (Brazil). *Acta Limnologica Brasiliensia* 24(1): 64–82, <http://dx.doi.org/10.1590/S2179-975X2012005000027>
- CETEC (2012) Fundação Centro Tecnológico de Minas Gerais (CETEC) - Centro de Bioengenharia de Espécies Invasoras de Hidrelétricas. <http://base.cbeih.org/view/Limnoperna+fortunei> (Accessed 23 March 2012)
- Darrigran G (2002) Potential impact of filter-feeding invaders on temperate inland freshwater environments. *Biological Invasions* 4: 145–156, <http://dx.doi.org/10.1023/A:1020521811416>
- Darrigran GA, Damborenea MC (2005) A bioinvasion history in South America. *Limnoperna fortunei* (Dunker, 1857), the golden mussel. *American Malacological Bulletin* 20: 105–112
- Darrigran GA, Damborenea MC (eds) (2006) Bio-invasion del mejillón dorado en el continente americano. Edulp, La Plata Argentina, 220 pp
- Darrigran GA, Damborenea MC (2011) Ecosystem Engineering Impact of *Limnoperna fortunei* in South America. *Zoological Science* 28(1): 1–7, <http://dx.doi.org/10.2108/zsj.28.1>
- Darrigran GA, Martin SM, Gullo B, Armendariz L (1998) Macroinvertebrates associated with *Limnoperna fortunei* (Dunker, 1857) (Bivalvia, Mytilidae) in Río de la Plata, Argentina. *Hydrobiologia* 367: 223–230, <http://dx.doi.org/10.1023/A:1003244603854>
- Darrigran G; Damborenea C, Greco N (2007). Freshwater invasive bivalves in man-made environments: A case study of larvae biology of *Limnoperna fortunei* in a Hydroelectric Power Plant in South America. *AMBIO* 36(7):575–579, [http://dx.doi.org/10.1579/0044-7447\(2007\)36\[575:AEFPAP\]2.0.CO;2](http://dx.doi.org/10.1579/0044-7447(2007)36[575:AEFPAP]2.0.CO;2)
- Elith J (2002) Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In: Ferson S, Burgman M (eds), *Quantitative Methods for Conservation Biology*. New York: Springer-Verlag, pp 39–58
- Elith J, Graham CH, Anderson RP, Dudi'k M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick J R, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J. McC, Peterson AT, Phillips SJ, Richardson KS, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151, <http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* (1): 330–342
- Faria EA, Branco JRT, Campos MCS, Oliveira MD, Rolla ME (2006) Estudo das características antiincrustantes de materiais. *Revista Escola de Minas* 59(2): 233–238, <http://dx.doi.org/10.1590/S0370-44672006000200014>
- Fielding AH, Bell JFA (1997) Review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1): 38–49, <http://dx.doi.org/10.1017/S0376892997000088>
- Hawkins BA, Diniz-Filho JAF, Bini LM, De Marco P, Blackburn TM (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography* 30: 375–384, <http://dx.doi.org/10.1111/j.0906-7590.2007.05117.x>
- Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027–2036, [http://dx.doi.org/10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)
- 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(6): 885–890, <http://dx.doi.org/10.1111/j.1472-4642.2008.00496.x>
- Jiménez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragón P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13(12): 2785–2797, <http://dx.doi.org/10.1007/s10530-011-9963-4>
- Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography* 21(4): 498–507, <http://dx.doi.org/10.1111/j.1466-8238.2011.00683.x>
- Karatayev A, Boltovskoy D, Padilla DK, Burlakova LE (2007) The invasive bivalves *Dreissena polymorpha* and *Limnoperna fortunei*: Parallels, contrasts, potential spread and invasion. *Journal of Shellfish Research* 26(1): 205–213, [http://dx.doi.org/10.2983/0730-8000\(2007\)26\[205:TIBDPA\]2.0.CO;2](http://dx.doi.org/10.2983/0730-8000(2007)26[205:TIBDPA]2.0.CO;2)
- Kluza DA, McNyset M (2005) Ecological niche modeling of aquatic invasive species. *Aquatic Invaders* 16(1): 1–7

- Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151, <http://dx.doi.org/10.1111/j.1466-8238.2007.00358.x>
- Lobo JM, Jiménez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33(1): 103–114, <http://dx.doi.org/10.1111/j.1600-0587.2009.06039.x>
- Loiselle BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, Williams PH (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology* 17: 1591–1600, <http://dx.doi.org/10.1111/j.1523-1739.2003.00233.x>
- Luoto M, Poyry J, Heikkinen RK, Saarinen K (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography* 14: 575–584, <http://dx.doi.org/10.1111/j.1466-822X.2005.00186.x>
- Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38: 921–931, <http://dx.doi.org/10.1046/j.1365-2664.2001.00647.x>
- Mansur MCD, Santos CP, Darrigran G, Heydrich I, Callil CT, Cardoso FR (2003) Primeiros dados quali-quantitativos do mexilhão-dourado, *Limnoperna fortunei* (Dunker), no Delta do Jacuí, no Lago Guaíba e na Laguna dos Patos, Rio Grande do Sul, Brasil e alguns aspectos de sua invasão no novo ambiente. *Revista Brasileira de Zoologia* 20(1): 75–84, <http://dx.doi.org/10.1590/S0101-81752003000100009>
- Mansur MCD, Quevedo CB, Santos CP, Callil CT (2004) Prováveis vias de introdução de *Limnoperna fortunei* (Dunker, 1857) (Mollusca, Bivalvia, Mytilidae) na bacia da laguna dos Patos, Rio Grande do Sul e novos registros de invasão no Brasil, pelas bacias do alto Paraná e Paraguai. In: Silva JSV, Souza RCCL (eds), Água de Lastro e Bioinvasão. Interciência, Rio de Janeiro, Brasil, pp 33–38
- Morton B (1977) The population dynamics of *Limnoperna fortunei* (Dunker 1857) (Bivalvia: Mytilacea) in Plover Cove reservoir, Hong Kong. *Malacologia* 16: 165–182
- Morton B (1996) The aquatic nuisance species problem: a global perspective and review. In: D'Itri FM (ed), Zebra Mussels and Other Aquatic Nuisance Species. Ann Arbor Press, Chelsea, MI, USA, pp 1–54
- Oliveira MD, Calheiros DF, Jacobi CM, Hamilton SK (2011) Abiotic factors controlling the establishment and abundance of the invasive golden mussel *Limnoperna fortunei*. *Biological Invasions* 13: 717–729, <http://dx.doi.org/10.1007/s10530-010-9862-0>
- Pastorino G, Darrigran G, Martin S, Lunaschi L (1993) *Limnoperna fortunei* (Dunker, 1857) (Mytilidae), nuevo bivalvo em águas del Río de La Plata. *Neotropica* 39(101–102): 171
- Penchaszadeh P, Darrigran G, Angulo C, Averbuj A, Brignoccoli N, Brögger M, Dogliotti A, Pirez N (2000) Predation on the invasive freshwater mussel *Limnoperna fortunei* (Dunker, 1857) (Mytilidae) by the fish *Leporinus obtusidens* Valenciennes, 1846 (Anostomidae) in the Río de la Plata, Argentina. *Journal Shellfish Research* 19(1): 229–231
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78: 419–433, <http://dx.doi.org/10.1086/378926>
- Peterson AT (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* 3: 59–72
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259, <http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026>
- Poulos H, Chernoff B, Fuller P, Butman D (2012) Ensemble forecasting of potential habitat for three invasive fishes. *Aquatic Invasions* 7: 59–72, <http://dx.doi.org/10.3391/ai.2012.7.1.007>
- Ricciardi A (1998) Global range expansion of the Asian mussel *Limnoperna fortunei* (Mytilidae): another fouling threat to freshwater systems. *Biofouling* 13: 97–106, <http://dx.doi.org/10.1080/08927019809378374>
- Ricciardi A, Rasmussen JB (1998) Predicting the identity and impact of future biological invaders: a priority for aquatic resource management. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1759–1765, <http://dx.doi.org/10.1139/f98-066>
- Rocchini D, Hortal J, Lengyel S, Lobo JM, Jimenez-Valverde A, Ricotta C, Bacaro G, Chiarucci A (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography* 35(2): 211–226, <http://dx.doi.org/10.1177/0309133311399491>
- Rödger D, Schmidtlein S, Veith M, Lotters S (2009) Alien Invasive Slider Turtle in Unpredicted Habitat: A Matter of Niche Shift or of Predictors Studied? *Plos One* 4(11): 1–9, <http://dx.doi.org/10.1371/journal.pone.0007843>
- Sardiña P, Cataldo D, Boltovskoy D (2008) The effects of the invasive mussel, *Limnoperna fortunei*, on associated fauna in South American freshwaters: importance of physical structure and food supply. *Fundamental and Applied Limnology* 173: 135–144, <http://dx.doi.org/10.1127/1863-9135/2008/0173-0135>
- Soberon J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10: 1115–1123, <http://dx.doi.org/10.1111/j.1461-0248.2007.01107.x>
- Stockwell DRB, Noble IR (1992) Induction of sets of rules from animal distribution data: A robust and informative method of analysis. *Mathematics and Computers in Simulation* 33: 385–390, [http://dx.doi.org/10.1016/0378-4754\(92\)90126-2](http://dx.doi.org/10.1016/0378-4754(92)90126-2)
- Stockwell D, Peters ED (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *Lecture Notes in Computer Science* 13: 143–158
- Sylvester F, Boltovskoy D, Cataldo D (2007) The invasive bivalve *Limnoperna fortunei* enhances benthic invertebrate densities in South American floodplain Rivers. *Hydrobiologia* 589: 15–27, <http://dx.doi.org/10.1007/s10750-007-0708-4>