Predicting global habitat suitability for *Corbicula fluminea* using species distribution models: the importance of different environmental datasets.

Gama, M.¹, Crespo, D.², Dolbeth, M.^{2,3} and Anastácio, P.¹

1- MARE – Marine and Environmental Sciences Centre, Departamento de Paisagem Ambiente e Ordenamento, Escola de Ciências e Tecnologia, Universidade de Évora, Portugal

2 - CFE – Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Portugal

3 - CESAM & Department of Biology, University of Aveiro, Portugal

Corresponding author: ana.gama@ci.uc.pt, telephone: +351 266 745334/745385, fax: +351 266 745395

Abstract

Niche-based models (NBMs) are increasingly being used to predict the biological distribution of species, as well as the importance of different environmental variables on their habitat adequability. Here, we investigate the reliability of these models in predicting habitat suitability for *Corbicula fluminea*, an important freshwater bivalve invasive species. In order to determine the influence of topographic versus climatic variables, three datasets were used: 1) *CorbiculaTOPO* with topographic variables (altitude, slope and a compound topographical index); 2) *CorbiculaMIX*, combining climatic (annual mean temperature, mean temperature of warmest quarter, mean temperature of coldest quarter and annual precipitation) and topographic variables and 3) *CorbiculaCLIM* with only the climatic variables. Nine different types of models, implemented in BIOMOD2, were used and an ensemble of NBMs was built. We aimed to know how climatic suitability for these invaders changes when using different datasets of environmental variables; if the predictive reliability is similar between datasets; and which environmental variables better explain habitat adequability. Model performance was very similar between *CorbiculaMIX* and *CorbiculaCLIM*. *CorbiculaTOPO* was the dataset with the least accurate predictions. Mean temperature of the coldest quarter and altitude were the variables that influenced *C. fluminea* distribution the most. The use of an ensemble of predictions allowed us to clearly identify areas with potential to be invaded by the bivalve, in which records are not yet detected. This information can be used in management, to implement measures to delay or prevent invasions, as well as for the identification of the environmental variables that favor that invasive potential.

Keywords: *Corbicula fluminea*, BIOMOD2, climatic variables, topographic variables.

1 - Introduction

The species *Corbicula fluminea* (Müller, 1744) is considered one of the most efficient freshwater invaders worldwide and is currently found in the 5 continents (Ituarte, 1994; Mouthon, 2001; Ciutti and Cappelletti, 2009; Bódis *et al.*, 2012; Früh *et al.*, 2012; Lucy *et al.*, 2012). This bivalve species is normally found in well oxygenated waters (Belanger, 1981) and is intolerant to high salinity values (McMahon, 1999), low pH (Lucy *et al.*, 2012) and low calcium concentrations (Karatayev *et al.*, 2007). It tolerates a wide range of water temperatures (Lucy *et al.*, 2012) and prefers sand and silt sediments with higher organic matter content (Sousa *et al.*, 2008a). The occurrence of pelagic and benthic life stages enables *C. fluminea* to spread over long distances by ballast water transport in boats (Paunović *et al.*, 2007; Bódis *et al.*, 2012), by currents, and locally by diffusion due to sand extraction (Belz *et al.*, 2012). Passive transport by aquatic birds between water bodies was also observed (Darrigran, 2002; McMahon, 2002). For all these reasons this particular species has a high invasive potential.

Distribution models may be used to assess the spreading potential of invasive species allowing scientists to prioritize biodiversity conservation plans and mitigate invasion impacts. Despite the increasing area of distribution of *C. fluminea* in a small amount of time, little is known about the geographical and climatic characteristics of the areas that are potentially being invaded. Ensemble distribution models have been used recently to explore these issues for other species (Capinha *et al.*, 2011). To account for uncertainty in single Niche Based Models, scientists have been recently joining single models in ensemble analysis combining individual predictions in consensus predictions (Capinha and Anastácio, 2011).

This study models the worldwide distribution of *C. fluminea* using an ensemble of Niche-based models (NBMs) of nine modeling techniques that are implemented in

the BIOMOD modelling framework (Thuiller, 2003). The effects of three environmental datasets on the accuracy of the models were explored. We aimed to answer the following questions: i) how will the climatic suitability change for these invaders by using different environmental variables datasets; ii) is the predictive reliability similar between tested datasets; iii) which are the environmental variables that better explain habitat adequability?

2 - Methods

2.1 - Data sources

An extensive compilation of worldwide native and invasive records (Fig.1) (412 presences) was obtained from literature. To describe the environmental conditions, we selected up to 4 different climatic variables, with a cell resolution of 5 arc-min from the worldclim datasets (<http://www.worldclim.org/> 1.4 datasets; Hijmans *et al.*, 2005, accessed May, 2014). The climatic variables included were Annual Mean Temperature (AnnMeanTemp), Mean Temperature of Warmest Quarter (MeanTempWarmQt), Mean Temperature of Coldest Quarter (MeanTempColdQt), and Annual Precipitation (AnnPrecip), chosen from a set of 19 different climatic variables. These were chosen in order to determine the influence of temperature and precipitation patterns in the determination of habitat suitability. It was considered that the selection of these particular climatic variables were ecologically meaningful for *C. fluminea*. In addition, several topographic variables were also added to the model to allow the representation of landscape characteristics. The chosen variables were altitude and slope and a compound topographical index was also calculated. Altitude (alt) and slope give us an indication of other environmental factors such as stream size and velocity, known to influence, for example, oxygen availability (Hynes, 1970). Altitude was obtained from

the WORLDCLIM datasets. The compound topographical index (CTI) was calculated based on Cooley (2013) and was previously used for freshwater species (Capinha and Anastácio, 2011). The CTI (Moore et al., 1991) or "soil wetness" transformation can be used to model aspects of hydrologic systems and is strongly correlated with soil moisture, providing indirect information on water availability. The value of each cell in the CTI output raster is the value in a flow accumulation raster for the corresponding DEM (Digital Elevation Model), meaning that higher CTI values represent drainage depressions and lower values represent ridges. The topographical variables' slope and CTI were spatially modeled in ArcGis (version 10), using the digital elevation model of the United States Geological Survey (USGS) HYDRO1k geographical dataset (Verdin and Jenson 1996). Posteriorly, these data was resampled to match the remaining variables in R. To avoid collinearity among variables, Pearson correlation coefficients were calculated for all pairs of predictors and the environmental variables included were the ones not highly correlated ($r < 0.83$). Several published papers refer distinct r values to define high correlations, for example $r < 0.80$ (Gallardo and Aldridge, 2013), $r < 0.85$ (Capinha and Anastácio, 2011), $r < 0.80$ (Capinha et al., 2012) and some studies indicate that variables with $-0.70 < r < 0.70$ should be excluded (Domisch et al., 2013 following Green, 1979. In this study only one pair of pairwised variables had an r outside this range being the remaining lower than 0.7. We chose to keep it because we believe that the variables were considered relevant for the distribution of *C. fluminea*.

To determine the influence of topographic versus climatic variables, three ensemble models with increasing complexity (with respect to the number of environmental variables selected) were created: - *Corbicula*TOPO, with alt, slope and CTI; - *Corbicula*CLIM, with AnnMeanTemp, MeanTempWarmQt, MeanTempColdQt

75 and AnnPrecip; and *CorbiculaMIX*, with AnnMeanTemp, MeanTempWarmQt,
76 MeanTempColdQt and AnnPrecip, alt, slope and CTL.

2.2 - Ensemble distribution models

To model the potential distribution of *C. fluminea*, 9 different algorithms in the BIOMOD2 package (Thuiller, 2009) were used, in R software (version: 2.14.0; R Development Core Team, 2011). Algorithms were SRE (surface range envelop or BIOCLIM), CTA (classification tree analysis), RF (random forest for classification and regression), MARS (multivariate adaptive regression spline), FDA (flexible discriminant analysis), GLM (generalized linear models), GAM (generalized additive models), GBM (generalized boosted regression models) and ANN (artificial neural networks). Species occurrence data, both from native and invasive records, were coupled with pseudo-absences data generated worldwide (randomly) because this option has been previously used with more powerful results in predicting potential distributions (Capinha *et al.*, 2011) as opposite to using real absences which may lead to underestimation issues. Each of the datasets had species occurrence records coupled with an equal number of independently generated pseudo-absences. Models were built using 80% of the coupled occurrences and pseudo-absence data and the remaining 20% were withheld for evaluating predictions. In the BIOMOD2 package one can determine or evaluate the variables importance through a randomization method (Thuiller, 2009). We assumed that the most important variables were the ones with a relative importance above the mean of the predictor variables within each subset. True skill statistics (TSS) was used as a measure of model performance. TSS takes into account both omission and commission errors and ranges also from -1 to 1, not being affected by prevalence as KAPPA (Allouche *et al.*, 2006). Values ranging from 0.2 to 0.5 were considered poor, from 0.6 to 0.8 useful, and values larger than 0.8 were good to excellent (as in Coetzee *et al.*, 2009). We used the area under the receiver operating characteristic curve (AUC) as an alternative measure of accuracy. Prediction accuracy is considered to be similar to

random for AUC values lower than 0.5, poor, for values in the range 0.5–0.7, fair in the range 0.7–0.9, and excellent for values greater than 0.9 (Swets, 1988). To determine if there were differences in TSS and AUC model accuracy between CorbiculaTOPO, CorbiculaCLIM and CorbiculaMIX subsets a Friedman test was applied, followed by pairwise comparisons by a Wilcoxon signed-ranks test, using a Bonferroni correction. The Friedman test is a non-parametric statistical test similar to the parametric repeated measures ANOVA used to detect differences in treatments across multiple test attempts (Zar, 1999). The sensitivity-specificity approach (<http://r-forge.r-project.org/projects/biomod/>) was used to reclassify the resulting continuous maps in binary maps (unsuitable and suitable areas). In Liu *et al.*, (2005) it was shown that this method is one of the most reliable for choosing a reclassification threshold. In this modelling context, sensitivity is the proportion of observed presences that are predicted as such and therefore is a measure of omission errors whereas specificity is the proportion of observed absences that are predicted as such, quantifying commission errors.

3 - Results

3.1 - Predictive model performance

For the AUC, *CorbiculaTOPO* gathered the least accurate predictions among the different environmental datasets (Table 1). The AUC values for the other two environmental datasets were similar and displayed, in general, good accuracies. When we analyze TSS scores, the variables that best performed were the ones used in *CorbiculaCLIM*, that is, only the climatic variables, followed by *CorbiculaMIX*, combining both climatic and topographic variables (Table1). The set of variables of *CorbiculaTOPO* were the ones that predicted less accurately the distribution of *C. fluminea*, similarly to results obtained for AUC. For the TSS, the variability associated with the predictions (given by the standard deviation) was similar to the ones predicted by AUC. In general, models had good agreement between observed data and model predictions, especially when AUC is used (Table 1).

There was a statistically significant difference in both TSS and AUC model accuracy values depending on which type of subset was evaluated (Friedman test on AUC, $\chi^2(2) = 16.76$, $p < 0.001$ and Friedman test on TSS $\chi^2(2) = 14.97$, $p = 0.001$). *Post hoc* analysis using Wilcoxon signed-rank tests were conducted, with a Bonferroni correction (alpha was set to 0.05/3). There were no significant differences between *CorbiculaCLIM* and *CorbiculaMIX* for both AUC and TSS ($Z = -0.447$, $p = 0.655$ and $Z = -0.447$, $p = 0.952$ respectively). *CorbiculaTOPO* was significantly different from both *CorbiculaCLIM* ($Z = -2.670$, $p = 0.008$ for AUC; $Z = -2.666$, $p = 0.008$ for TSS) and *CorbiculaMIX* ($Z = -2.680$, $p = 0.007$ for AUC; $Z = -2.660$, $p = 0.008$ for TSS) with lower model accuracy scores.

3.2 - Relative importance of environmental variables

For the *CorbiculaMIX* environmental dataset, AnnMeanTemp, MeanTempColdQt and Alt had higher relative importance than the mean (Fig. 2). For the other two datasets only one of the variables tested (1/3 for *CorbiculaTOPO* and 1/4 for *CorbiculaCLIM*) was more important than the mean. For the topographic variables, altitude was always the most important variable in the two datasets where it was used (*CorbiculaMIX* and *CorbiculaTOPO*). As to the climatic variables, in the *CorbiculaMIX* dataset, AnnMeanTemp and MeanTempColdQt, had similar importance (approximately 0.31 for both), whereas for the *CorbiculaCLIM* dataset, MeanTempColdQt was the only climatic variable with higher importance than the mean (mean value of 0.53). This variable had also higher importance than the correspondent variables in the *CorbiculaMIX* dataset (mean values of 0.53 vs 0.31).

3.3 - Ensemble modeling

One of the main features of the BIOMOD2 software is the ability of combining predictions made in single models in an ensemble. Sensitivity and specificity in the ensemble output provides the percentage of presences and absences correctly predicted, respectively. The datasets containing the climatic and topographic variables combined (*CorbiculaMIX*) or with the climatic variables only (*CorbiculaCLIM*), were better for predicting species habitat adequability than the dataset with the topographic variables alone (*CorbiculaTOPO*) (Table 2). In these two datasets, habitat adequability variation was very similar, between 0.9 and 98.7 for the *CorbiculaMIX* dataset and 0.8 and 97.2 for *CorbiculaCLIM* (Supplementary Material – Fig. 1). Sensitivity was the highest for the *CorbiculaMIX* dataset, meaning that the percentage of the correctly predicted

presences was the highest (Table 2). On the contrary, the specificity (i.e., the % of absences correctly predicted) was the lowest, but still extremely high in this subset.

The probability data were transformed into suitable-unsuitable maps (Fig. 3). Once again, the percentage of area suitable for the species was the lowest for the *Corbicula* *TOPO* dataset and similar for the other 2 datasets, which had similar areas (6.2 vs. 7.2 respectively, Table 2, Fig. 3). Both dataset predictions shared 4.9% of common adequate areas of the continental earth.

4 - Discussion

We explored the influence of using different environmental datasets to model *C. fluminea* potential distribution. Despite the good accuracy of the model's predictions, there was some variability amongst individual models, which has also been reported in other works (Capinha and Anastácio, 2011). The climatic environmental variables led to more accurate modeling both when used alone as well as when coupled with the topographic variables, as the results of the Friedman test indicate. In contrast, the topographic variable datasets used alone led to fair accuracy predictions. Therefore, for *C. fluminea*, coupling climatic and topographic variables, or using only the climatic variables seems to be the best modeling approach.

Altitude was always one of the most important variables in the two datasets where it was used. Altitude may be inversely correlated with stream velocity (Leopold, 1953), influencing particle size and consequently sand or silt deposition (Vannote *et al.*, 1980). As such, locations with higher altitude may not be adequate for the establishment of the species due to the decrease of suitable habitats (Vannote *et al.*, 1980). Additionally, dispersion to headwater reaches may be difficult due to geographical anthropogenic or natural barriers (as dams). Nonetheless it has been suggested that *C. fluminea* may travel up to 3m per day in-stream (Voelz *et al.*, 1998) and also longitudinal dispersal with fish may be possible. Filaments may facilitate passive transport of small individuals (Prezant and Chalermwat, 1984) and Martinho *et al.*, (2008) found *Corbicula* individuals in gut contents of fish species such as *Platichthys flesus* and *Solea solea*. Passive transportation between water bodies on the feet or feathers of aquatic birds was also observed (McMahon, 2002), however dispersal to higher altitudes is still improbable due to the unlikely presence of suitable habitats.

In terms of climatic variables, those related to lower temperature values seemed to play an important part in habitat adequability. Temperature is an important variable for several physiological processes (Mouthon, 2001; Ortmann and Grieshaber, 2003; Sousa *et al.*, 2008b) and it influences habitat adequability in several freshwater species (Gallardo and Aldridge, 2013). Despite tolerating a wide range of water temperatures (from 2 to 35-37°C – Lucy *et al.*, 2012), *C. fluminea* fails to survive at temperatures below 2°C (Mattice, 1977). Winter ice formation also limits the distribution and impacts of *C. fluminea* in northern latitudes (Simard *et al.*, 2012). Nonetheless, in recent years, reports of the spread of the species further north into Canada and Ireland were published (Kirkendale and Clare, 2008; Caffrey *et al.* 2011). These reports suggest increased tolerance to low water temperatures or thermal pollution (Mattice, 1977; Simard *et al.*, 2012).

As our results suggest, *Corbicula*TOPO had the lowest accuracy values (TSS and AUC) being those significantly different from the other two subsets. This in conjunction with the highest specificity observed (Table 2) allowed us to conclude that this subset was the one that performed the worst. It is also our understanding that the model with the best sensitivity should be considered the best in determining the presence of *Corbicula* (*Corbicula*MIX). Sensitivity indicates how responsive the model is to the condition being tested for, meaning how well it correctly indicates a positive result when the positive condition is true. Specificity represents the absence of commission errors showing how specific the model is to the condition being tested for, or how well it correctly indicates a negative result when the condition really is absent. Interpretation of commission error rates is more complex than omission error rates due to the fact that information documenting species' absences is difficult to obtain with

accuracy and the species may truly be present at many sites where the evaluation dataset indicates absence, which will inflate commission error estimates (Peterson *et al.*, 2012).

The two best ensemble models, the one with climatic data only and the one with climatic and topographic data, correctly predicted regions initially invaded by the species, for example, in Europe (Pérez-Quintero, 2008) or North America (Counts, 1986), as well as in areas recently invaded, such as the UK or Ireland (Elliott and Ermgassen, 2011; Lucy *et al.*, 2012). For the two environmental datasets, suitable-unsuitable areas obtained from ensemble predictions showed similar occurrence areas (6.2 and 7.2% of total continental area respectively) and both datasets shared 4.9% of common continental adequate areas. From these ensemble predictions we may observe that there are several areas with high habitat suitability for *Corbicula fluminea*, which have not yet been invaded.

The possibility of *C. fluminea* invasion raises a number of concerns since it is known to promote several economic problems, for example impacting industrial facilities and water treatments plants (Rosa *et al.*, 2011). Asian clams can also outcompete native species, such as Unionidae (Cooper *et al.*, 2005, Sousa *et al.*, 2008b), although a few authors observed no conclusive evidence of this (Strayer, 1999). The appearance of this species in South African ecosystems would be an example of a high risk situation since *C. fluminea* could competitively outcome native species, already at risk (Graf and Cummings, 2007; Bogan, 2008). In the south of the African continent or in the southern part of Australia, habitat adequability is high but no species record has been detected yet. Nonetheless, in both regions, closely related species (*C. africana* in Africa and *C. australis* in Australia; Byrne *et al.*, 2000) have been recorded, which may indicate that the ecological niche of *C. fluminea* may be already occupied. In Africa, reports identified *C. fluminea* in Morocco in 2008 and 2011 (Clavero *et al.*, 2012), and

C. africana (Krauss, 1848) (Korniushin, 2004) and *C. fluminalis* (Müller, 1774) (Korniushin, 2004) in South Africa.

However, one should consider that species identification within *Corbicula* species has raised several taxonomic concerns. For example, some authors consider *C. fluminea* and *C. fluminalis* as morphotypes, and the Global Invasive Species Database (GISD) consider them as synonymous (Korniushin 2004; Paunović *et al.*, 2008). Currently, the existence of several discrepancies in the descriptions and classification of members of the genera may lead to underestimations of the species invasive range. The records used in this work belonged to specimens identified as *C. fluminea* only. However, the occurrence of similar species within areas considered suitable for *C. fluminea* by our habitat adequability map indicates that environmental characteristics may be prone for the establishment of the species.

5 - Conclusions

Our study shows that South African and Australian environments seem particularly prone to invasion by *C. fluminea*. Therefore, a careful monitoring of the freshwater habitats in these regions is advisable to allow early detection and prevent *C. fluminea* invasion. The results from our models may support the establishment of management measures to delay or prevent invasions, as well as to identify the environmental variables that favor the invasive potential of *C. fluminea*. Ensemble Species Distribution Models can provide important insights into the threat posed by invasive species, nonetheless as our work suggests, importance should be paid in the selection of environmental variables used to model species distribution.

Acknowledgments

We would like to thank to the many researchers and institutions that provided species distribution data. We are also grateful to Paula Mendes and César Capinha for the valuable help with spatial analysis tools. This study was financed by FEDER funds through the “Programa Operacional de Factores de Competitividade – COMPETE” and by national funds through “FCT – Fundação para a Ciência e Tecnologia” within the scope of the project BIOCHANGED - Impact of biodiversity evenness changes in the ecosystem functioning of estuaries (PTDC/MAR/111901/2009) and of the project Pest (C/MAR/UI0284/2011). The PhD grant SFRH/BD/80252/2011 (D. Crespo) and post-doc grant SFRH/BPD/41117/2007 (M. Dolbeth) supported by FCT are also acknowledged.

References

- Allouche O, Tsoar A, Kamon R (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223-1232.
- Belanger SE (1991). The effect of dissolved oxygen, sediment, and sewage treatment plant discharges upon growth, survival and density of Asiatic clams. *Hydrobiologia*, 218, 113-126.
- Belz CE, Darrigran G, Mäder Netto OS, Boeger WA, Ribeiro PJ (2012). Analysis of Four Dispersion Vectors in Inland Waters: The Case of the Invading Bivalves in South America. *Journal of Shellfish Research*, 31(3), 777-784.
- Byrne M, Phelps H, Church T, Adair V, Selvakumaraswamy P, Potts J (2000). Reproduction and development of the freshwater clam *Corbicula australis* in southeast Australia. *Hydrobiologia*, 418, 185–197.
- Bódis E, Sipkay C, Tóth B, Oertel N, Nosek J, Hornung E (2012) Spatial and temporal variation in biomass and size structure of *Corbicula fluminea* in Danube River catchment, Hungary. *Biologia*, 67(4), 739-750.
- Bogan AE (2008). Global diversity of freshwater mussels (Mollusca, Bivalvia) in freshwater. *Freshwater Animal Diversity Assessment Developments in Hydrobiology*, 198, 139-147.
- Caffrey JM, Evers S, Millane M, Moran H (2011). Current status of Ireland's newest invasive species - the Asian clam *Corbicula fluminea* (Müller, 1774). *Aquatic Invasions*, 6, 291-299.
- Capinha C, Anastácio P (2011). Assessing the environmental requirements of invaders using ensembles of distribution models. *Diversity and Distributions*, 17, 13-24.

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

Capinha C, Anastácio P, Tenedório JA (2012). Predicting the impact of climate change on the invasive decapods of the Iberian inland waters: an assessment of reliability. *Biological Invasions*, 14, 1737-1751.

Ciutti F, Cappelletti C (2009). First record of *Corbicula fluminalis* (Müller, 1774) in Lake Garda (Italy), living in sympatry with *Corbicula fluminea* (Müller, 1774). *Journal of Limnology*, 68, 162-165.

Clavero M, Araújo R, Calzada J, Delibes M, Fernández N, Gutiérrez-Expósito C, Revilla E, Román J (2012). The first invasive bivalve in African freshwaters: invasion portrait and management options. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 22, 277-280.

Coetzee BWT, Roberston MP, Erasmus BFN, van Rensburg BJ, Thuiller W (2009). Ensemble models predict important bird areas in southern Africa will become less effective for conserving endemic birds under climate change. *Global Ecology and Biogeography*, 18, 701-710.

Cooley SW (2013). GIS4Geomorphology: <http://www.gis4geomorphology.com> (Accessed May, 2014)

Cooper NL, Bidwell JR, Cherry DS (2005). Potential effects of Asian clam (*Corbicula fluminea*) die-offs on native freshwater mussels (Unionidae) II: porewater ammonia. *Journal of the North American Benthological Society*, 24(2), 381-394.

Counts CL (1986). The zoogeography and history of invasion of the United States by *Corbicula fluminea* (Bivalvia: Corbiculidae). *American Malacological Bulletin Special Edition*, 2, 7-39.

Darrigran G (2002). Potential impact of filter-feeding invaders on temperate inland freshwater environments. *Biological Invasions*, 4, 145-156.

Domisch S, Kuemmelen M, Jähnig SC., Haase P. (2013). Choice of study area and predictors affect habitat suitability projections, but not the performance of species distribution models of stream biota. *Ecological Modelling*, 257, 1-10

Elliott P, Ermgassen SE (2008). The Asian clam (*Corbicula fluminea*) in the River Thames, London, England. *Aquatic Invasions*, 3, 54-60.

Früh D, Stoll S, Haase P (2012). Physico-chemical variables determining the invasion risk of freshwater habitats by alien mollusks and crustaceans. *Ecology and Evolution*, 2(11), 2843-2853.

Gallardo B, Aldridge DC (2013). Evaluating the combined threat of climate change and biological invasions on endangered species. *Biological Conservation* 160, 225-233.

Graf DL, Cummings KS (2007). Review of the systematics and global diversity of freshwater mussel species (Bivalvia: Unionoida). *Journal of Molluscan Studies*, 73(4), 291-314.

Green, R., 1979. Sampling Design and Statistical Methods for Environmental Biologists. Wiley, New York.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Hynes HBN (1970). The ecology of running waters. Univ. Toronto Press. 555 pp.

Ituarte CF (1994). *Corbicula* and *Neocorbicula* (Bivalvia: Corbiculidae) in the Paraná, Uruguay, and Río de la Plata Basins. *Nautilus*, 107, 129-135.

- Karatayev AY, Boltovskoy D, Padilla DK, Burlakova LE (2007). The invasive bivalves *Dreissena polymorpha* and *Limnoperna fortunei*: parallels, contrasts, potential spread and invasion impacts. *Journal of Shellfish Research*, 26, 205-213.
- Kirkendale L, Clare J (2008). The Asiatic clam (*Corbicula fluminea*) rediscovered on Vancouver Island. *The Victoria Naturalist*, 65, 12-16.
- Korniushin AV (2004). A revision of some Asian and African freshwater clams assigned to *Corbicula fluminalis* (Müller, 1774) (Mollusca: Bivalvia: Corbiculidae), with a review of anatomical characters and reproductive features based on museum collections. *Hydrobiologia*, 529, 251-270.
- Leopold LB (1953). Downstream change of velocity in rivers. *American Journal of Science*, 251(8), 606-624.
- Liu C, Berry PM, Dawson TP, Pearson RG (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28, 385-393.
- Lucy FE, Karatayev A, Burlakova L (2012). Prediction for the spread, population density, and impacts of *Corbicula fluminea* in Ireland. *Aquatic invasions*, 7, 465-474.
- Martinho F, Leitão R, Neto JM, Cabral H, Lagardère F, Pardal MA, (2008) .Estuarine colonization, population structure and nursery functioning for 0-group sea bass (*Dicentrarchus labrax*), flounder (*Platichthys flesus*) and sole (*Solea solea*) in a mesotidal temperate estuary. *Journal of Applied Ichthyology*, 24, 229-237.
- Mattice JS (1977). Interactions of *Corbicula* sp. with power plants. In: Britton JC (ed) *First Int. Corbicula Symp.* Fort Worth, Texas, USA, pp 119-13.
- McMahon RF (2002). Evolutionary and physiological adaptations of aquatic invasive animals: r selection versus resistance. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 1235-1244.

McMahon RF (1999). Invasive characteristics of the freshwater bivalve, *Corbicula fluminea*. In: Claudi R, Leach JH (eds), Nonindigenous Freshwater Organisms: Vectors, Biology and Impacts. Lewis Publishers, CRC Press, London, UK, pp 305-313.

McMahon RF (2002). Evolutionary and physiological adaptations of aquatic invasive animals: r selection versus resistance. Canadian Journal of Fisheries and Aquatic Sciences, 59, 1235-1244.

Moore ID, Grayson RB, Ladson AR (1991). Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications. Hydrological Processes, 5, 3-30.

Mouthon J (2001). Life cycle and populations dynamics of the Asian clam *Corbicula fluminea* (Bivalvia: Corbiculidae) in the Saone River at Lyon (France). Hydrobiologia, 452, 109-119.

Ortmann C, Grieshaber MK (2003). Energy metabolism and valve closure behaviour in the Asian clam *Corbicula fluminea*. The Journal of Experimental Biology, 206, 4167-4178.

Paunović M, Csányi B, Knežević S, Simić V, Nenadić D, Jakovčev-Todorović D, Stojanović B, Cakić P (2007). Distribution of Asian clams *Corbicula fluminea* (Müller, 1774) and *C. fluminalis* (Müller, 1774) in Serbia. Aquatic Invasions, 2, 99-106.

Pérez-Quintero JC (2008). Revision of the distribution of *Corbicula fluminea* (Müller, 1744) in the Iberian Peninsula. Aquatic Invasions, 3(3), 355-358.

Peterson AT, Soberón J, Anderson RP, Pearson RG, Martínez-Meyer E, Nakamura M, Araújo M.B (2012). Ecological Niches and Geographic Distributions: A Modeling Perspective. Princeton University Press, Princeton. 328pp.

Prezant RS, Chalermwat,K (1984). Flotation of the Bivalve *Corbicula fluminea* as a Means of Dispersal. Science, 225 (4669), 1491-1493. doi: 10.1126/science.225.4669.1491

Swets J (1988). Measuring the accuracy of diagnostic systems. Science, 240, 1285-1293.

Sousa R, Antunes C, Guilhermino L (2008a) Ecology of the invasive Asian clam *Corbicula fluminea* (Müller, 1774) in aquatic ecosystems: an overview. Annales de Limnologie - International Journal of Limnology, 44, 85-94.

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

Strayer DL (1999). Effects of alien species on freshwater mollusks in North America. Journal of the North American Benthological Society, 18, 74-98.

Simard A, Paquet A, Jutras C, et al. (2012). North American range extension of the invasive Asian clam in a St. Lawrence River power station thermal plume. Aquatic Invasions, 7,81-89.

Rosa IC, Pereira JL, Gomes J, Saraiva PM, Gonçalves F, Costa R (2011). The Asian clam *Corbicula fluminea* in the European freshwater-dependent industry: A latent threat or a friendly enemy? Ecological Economics, 70, 1805-1813.

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

Thuiller W (2009) BIOMOD – a platform for ensemble forecasting of species distributions. Ecography, 32, 369-373.

431 Vannote RL, Minshal GW, Cummins KW, Seddel JR, Gushing E (1980). The
432 river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences, 37, 130-
433 137.

434 Verdin K, Jenson S (1996). "Development of continental scale DEMs and
435 extraction of hydrographic features", proceedings of the Third Conference on GIS and
436 Environmental Modeling, Santa Fe, New Mexico, NCGIA.

437 Voelz NJ, McArthur JV, Rader RB (1998). Upstream mobility of the asiatic
438 clam *Corbicula fluminea*: Identifying potential dispersal agents. Journal of Freshwater
439 Ecology. 13(1), 39-45.

440 Zar JH (1999). Biostatistical Analysis. Prentice Hall, New Jersey, 663 pp.

441 Figure legends

442 Fig. 1 – Records of *C. fluminea* presence worldwide, used to model the current
443 distribution of the species.

444 Fig. 2 - Mean value of relative importance of the environmental variables used to
445 predict the distribution of *Corbicula fluminea* obtained from nine different modeling
446 algorithms with corresponding standard deviation values. Grey line - mean relative
447 importance. *CorbiculaMIX*, *CorbiculaCLIM* and *CorbiculaTOPO* correspond to
448 different environmental datasets.

449 Fig.3 - Binary maps (suitable-unsuitable areas) of predicted worldwide distribution of
450 *C. fluminea* in the three different environmental datasets: *CorbiculaTOPO*,
451 *CorbiculaCLIM* and *CorbiculaMIX*. Black means suitable

Figure1

0 3,700 7,400 14,800 Kilometers

A horizontal scale bar with four major tick marks. The first tick mark is at the left end, labeled '0'. The second tick mark is labeled '3,700'. The third tick mark is labeled '7,400'. The fourth tick mark is at the right end, labeled '14,800 Kilometers'. There are also smaller, unlabeled tick marks between the major ones.

Figure2

CorbiculaTOPO

CorbiculaCLIM

CorbiculaMIX

Figure3

Table1

Table 1 – Evaluation of *C. fluminea* distribution models (mean values $\pm$ S.D.) by the area under the receiver operating characteristics curve (AUC) and by the true skill statistic (TSS), using 3 different environmental datasets: *CorbiculaMIX* (climatic and topographic datasets), *CorbiculaCLIM* (climatic) and *CorbiculaTOPO* (topographic). The models were: SRE (surface range envelop or BIOCLIM), CTA (classification tree analysis), RF (random forest for classification and regression), MARS (multivariate adaptive regression spline), FDA (Flexible discriminant analysis), GLM (generalized linear models), GAM (generalized additive models), GBM (generalized boosted regression models) and ANN (artificial neural networks).

Algorithms		<i>CorbiculaTOPO</i>		<i>CorbiculaCLIM</i>		<i>CorbiculaMIX</i>	
AUC	SRE	0.70	0.01	0.81	0.01	0.78	0.01
	CTA	0.75	0.07	0.92	0.08	0.92	0.15
	RF	0.92	0.15	0.98	0.04	0.98	0.07
	MARS	0.91	0.03	0.94	0.04	0.96	0.02
	FDA	0.87	0.07	0.94	0.02	0.94	0.03
	GLM	0.85	0.05	0.93	0.03	0.93	0.01
	GBM	0.89	0.00	0.97	0.02	0.97	0.03
	GAM	0.90	0.03	0.96	0.02	0.96	0.01
	ANN	0.86	0.05	0.91	0.06	0.91	0.06
	mean	0.84	0.06	0.92	0.04	0.92	0.04
TSS	SRE	0.31	0.02	0.62	0.06	0.65	0.04
	CTA	0.58	0.10	0.82	0.07	0.87	0.01
	RF	0.80	0.20	0.93	0.06	0.92	0.05
	MARS	0.61	0.05	0.80	0.03	0.79	0.03
	FDA	0.61	0.04	0.78	0.03	0.78	0.04
	GLM	0.62	0.08	0.76	0.06	0.76	0.03
	GBM	0.66	0.05	0.86	0.04	0.85	0.02
	GAM	0.62	0.05	0.85	0.03	0.85	0.05
	ANN	0.63	0.07	0.78	0.01	0.76	0.08
	mean	0.60	0.08	0.79	0.04	0.80	0.04

Table 2 – Sensitivity and Specificity calculated for the different environmental datasets *CorbiculaTOPO*, *CorbiculaCLIM* and *CorbiculaMIX*. Percentage (%) of suitable and unsuitable terrestrial areas calculated after binary map resampling based on a specificity and sensitivity approach.

TSS	Sensitivity	Specificity	unsuitable (%)	suitable (%)
<i>CorbiculaTOPO</i>	95.57	99.35	98.1	1.9
<i>Corbicula CLIM</i>	94.60	96.50	92.8	7.2
<i>CorbiculaMIX</i>	97.78	95.30	93.8	6.2