Climate controls the distribution of a widespread invasive species: implications for future range expansion

W. G. McDOWELL*, A. J. BENSON[†] AND J. E. BYERS*

Odum School of Ecology, University of Georgia, Athens, GA, U.S.A.
[†]*United States Geological Survey, Southeast Ecological Science Center, Gainesville, FL, U.S.A.*

SUMMARY

1. Two dominant drivers of species distributions are climate and habitat, both of which are changing rapidly. Understanding the relative importance of variables that can control distributions is critical, especially for invasive species that may spread rapidly and have strong effects on ecosystems.

2. Here, we examine the relative importance of climate and habitat variables in controlling the distribution of the widespread invasive freshwater clam *Corbicula fluminea,* and we model its future distribution under a suite of climate scenarios using logistic regression and maximum entropy modelling (MaxEnt).

3. Logistic regression identified climate variables as more important than habitat variables in controlling *Corbicula* distribution. MaxEnt modelling predicted *Corbicula*'s range expansion westward and northward to occupy half of the contiguous United States. By 2080, *Corbicula*'s potential range will expand 25–32%, with more than half of the continental United States being climatically suitable.

4. Our combination of multiple approaches has revealed the importance of climate over habitat in controlling *Corbicula*'s distribution and validates the climate-only MaxEnt model, which can readily examine the consequences of future climate projections.

5. Given the strong influence of climate variables on *Corbicula*'s distribution, as well as *Corbicula*'s ability to disperse quickly and over long distances, *Corbicula* is poised to expand into New England and the northern Midwest of the United States. Thus, the direct effects of climate change will probably be compounded by the addition of *Corbicula* and its own influences on ecosystem function.

Keywords: climate change, distribution modelling, invasive species, model comparisons, non-native species

Introduction

Changes in climate are already affecting species distributions in ecosystems throughout North America (Parmesan & Yohe, 2003). Many researchers are using species distribution modelling to examine how ranges may shift. This approach has been criticised, however, because most models have not included non-climate variables, such as biotic factors such as competition and predation, and abiotic factors such as habitat attributes, that could themselves influence distributions (Woodward & Beerling, 1997; Pearson & Dawson, 2003; Jeschke & Strayer, 2008; Zarnetske *et al.* 2012). We must understand the relative importance of climate and non-climate variables in controlling present-day distributions to predict how climate changes will affect species distributions.

Making accurate ecological forecasts is a critical part of ecological research (Clark *et al.*, 2001). Climate changes are not occurring in isolation, but rather are coupled with other stressors such as habitat loss and land-use changes, eutrophication and invasive species. To understand the overall impact of these changes, we must also examine their interactions. Large changes in the distribution and abundance of aquatic organisms are predicted to occur over the next several decades (Wrona *et al.*, 2006). Understanding changes in distribution is especially important for invasive species, because they often are not yet at equilibrium, represent a significant component of aquatic communities and can interact strongly with native species (Cohen & Carlton, 1998; Mack *et al.*, 2000; Byers *et al.*, 2002). The conceptual basis for understanding how climate change could alter

Correspondence: William G. McDowell, Odum School of Ecology, 140 E Green St, Athens, GA 30602, U.S.A. E-mail: wgmcd@uga.edu

distributions and impacts of invasive species is well developed (Hellmann *et al.*, 2008; Rahel & Olden, 2008; Strayer, 2010), but few specific predictions have been made. Important topics to consider include quantifying how an invasive species may begin to spread with warming climate (Hellmann *et al.*, 2008) and assessing how these changes in distribution will affect aquatic ecosystems. The development of a more detailed understanding of the response of invasive species to climate change is thus critical to managing their impact and limiting their further spread.

Freshwater molluscs are an ideal group on which to focus because they provide important ecosystem services such as filtration and nutrient uptake (Strayer *et al.*, 1999; Vaughn & Hakenkamp, 2001). Also, invasive freshwater molluscs have had major impacts on aquatic ecosystems, including the New Zealand mud snail *Potamopyrgus antipodarum* that dominates nutrient cycling in western United States streams (Hall, Tank & Dybdahl, 2003), zebra mussels (*Dreissena polymorpha*) that remove phytoplankton and increase water transparency in the Great Lakes and the Hudson River (Caraco *et al.*, 1997; Hecky *et al.*, 2004) and the golden mussel *Limnoperna fortunei* that increases transparency and decreases primary productivity in South American lakes and reservoirs (Boltovskoy *et al.*, 2009).

Of the common aquatic invaders in the contiguous United States, the Asian clam *Corbicula fluminea* (hereafter *Corbicula*) is the most widely distributed (McMahon & Bogan, 2001). Since it was first introduced to Washington State in 1938, *Corbicula* spread eastward in the 1950s and reached the southeastern United States in the 1960s (McMahon & Bogan, 2001). *Corbicula* has invaded rivers throughout the United States, from southern New York to Florida on the east coast of the United States, west to Texas on the Gulf Coast and throughout the Midwest and West Coast states (McMahon & Bogan, 2001). *Corbicula* can reach densities of several thousand individuals per square metre (Cohen *et al.*, 1984; Phelps, 1994; Cataldo & Boltovsky, 1999; Sousa *et al.*, 2008b) and can strongly affect river ecosystems. Documented effects include decreasing phytoplankton concentrations, reducing rates of organic matter processing and reducing the abundance of submerged aquatic vegetation through indirect effects (Cohen *et al.*, 1984; Lauritsen, 1986; Phelps, 1994; Hakenkamp & Palmer, 1998).

The status of *Corbicula* as an invader that disperses rapidly, presumably aided by some means of human-mediated transport, makes it especially capable of invading new habitats as they become suitable. Therefore, it is critical to understand how climate changes may alter the ability of *Corbicula* to persist in new habitats. This study uses two different modelling approaches to determine the variables that control the distribution of *Corbicula*. First, to compare the relative importance of climate and habitat as predictors of *Corbicula* presence and absence, we constructed logistic regression models that examine all combinations of the predictor variables. Second, we created a maximum entropy (MaxEnt) model that only uses data on *Corbicula* presence and produces a continuous, spatially explicit prediction of suitability. To create this continuous prediction, the predictive variables must be spatially continuous, and therefore, the model is limited to climate-related predictive variables. Results of model are then used to assess probably range expansions of *Corbicula* in the United States over the coming decades.

Methods

Modelling approaches

We modelled suitability of locations within the continental United States for *Corbicula* using a logistic regression and a MaxEnt model. The logistic regression uses a general linear model to determine the suitability of a location using known presence and absence locations. A logistic regression is ideal for modelling binary data, such as presence and absence, as it has a higher power than the analysis of transformed data (Warton & Hui, 2011). The logistic regression also allowed us to include habitat variables, such as water chemistry and substrate variables that were only available as discrete data points, as well as discretised climate variables. By including both climate and habitat variables, we could test the importance of habitat and climate factors alone and in combination. Logistic regression, however, is only able to model habitat suitability for individual points; sampling these points is labour intensive because true presence and absence points are required, and the number of sites included is much lower than for the MaxEnt model.

We also created a MaxEnt model, which is a machine-learning approach that uses presence-only data and spatially explicit data layers of predictive environmental variables to model species distributions (Elith *et al.*, 2006). MaxEnt modelling outperforms other commonly used species distribution models (e.g. GARP, BIOCLIM), because it works well with incomplete data, small sample sizes and gaps, and therefore is well suited to modelling work with invasive species (Jarnevich *et al.*, 2006; Kumar *et al.*, 2009).

MaxEnt has two major advantages over logistic regression. First, MaxEnt is able to use presence-only records, allowing a much larger set of known *Corbicula* locations to be used. Second, with the availability of spatially explicit predictive climatic variables across a region, MaxEnt directly produces a spatially explicit output that describes how suitable an area is for *Corbicula*. Logistic regression, in contrast, makes predictions for individual points based on a broad range of site-specific habitat information, but spatial gaps in data availability limit its ability to produce a continuous spatial output. Because we elected to create a continuous MaxEnt model using climate variables, we cannot include the habitat data because continuous habitat data are generally not available. Thus, unlike logistic regression, our MaxEnt model cannot readily allow comparisons of the relative importance of habitat and climate variables. By combining logistic regression and MaxEnt, however, by combining logistic regression and MaxEnt, we were able to identify habitat variables that influence *Corbicula*'s distribution that could not be included in the MaxEnt model, quantify the amount of variability in *Corbicula*'s distribution that models looking only at climate factors on *Corbicula* can explain and create spatially explicit predictions about the potential distribution of *Corbicula* under future climate scenarios.

Data sources

Corbicula locations. Data on known locations of *Corbicula* came from two primary sources: the Environmental Protection Agency's (EPA) Environmental Monitoring and Assessment Program (EMAP) and the United States Geological Survey's Nonindigenous Aquatic Species (NAS) Database. During the EMAP surveys, a total of 1392 stream sites in the coterminous United States were sampled by EPA staff scientists and contractors with identical methods during summer months in 2000–2004, with most sampling occurring in 2003 and 2004. These data are freely available for ecological modelling (Paulsen *et al.*, 2008). The logistic regression can only use the points from the EMAP surveys, as it requires true presence and absence points. For the MaxEnt model, we combined the presence locations from the EPA EMAP surveys and the USGS NAS Database, because unlike logistic regression, a MaxEnt model does not require true absence points. This allowed us to use the locations from the USGS NAS Database, despite the lack of known absence points associated with them, increasing the overall number of *Corbicula* locations in the model by an order of magnitude. Many locations within the

USGS NAS Database are clustered within small geographic areas. To prevent giving too much weight in the MaxEnt model to small areas with many known *Corbicula* locations, the list of known points was converted to a raster with the same cell size as the predictive climate variables using ArcCatalog (ESRI, Redlands, CA, U.S.A.). This raster was then converted back to a list of points. By doing this, we ensured that no single cell had more than one location to it, reducing the total number of *Corbicula* locations from over 3500 to approximately 2800.

Predictive variables. During EMAP sampling, a variety of data were collected, including water quality, habitat and substrate measurements (Table S1). These habitat variables were only used for the logistic regression, as they are not available for points in the USGS NAS Database, nor are these variables available in a continuous, spatially explicit data layer. Climate data were used as predictive variables in both the logistic regression and the MaxEnt modelling. We used BioClim variables, a set of seventeen climate variables that provide a continuous data layer of climatic variables that are interpolated between weather stations. BioClim variables focus on aspects that could control species distributions, such as annual temperature, timing and amount of precipitation, and extreme conditions such as high temperature in the warmest month and low temperature in the coldest month (Hijmans *et al.*, 2005) (Table S2). Because the logistic regression models specific points, rather than a continuous geographic area, BioClim variable values were extracted to the sampling points from the EMAP surveys.

Multicollinearity between explanatory variables can lead to inaccurate model parameterisation and the exclusion of significant predictive variables (Graham, 2003). To avoid multicollinearity among the predictive variables, pairwise Pearson's correlation coefficients were calculated for all combinations of the variables, and one variable from any pair of variables with a Pearson's correlation coefficient greater than 0.7 ($r^2 > 0.49$) was removed from consideration (Moore & McCabe, 1993). We selected the variable to remove based on biological mechanisms. For example, calcium was highly correlated with conductivity ($\rho = 0.94$); from this pair, conductivity was removed, due to the documented importance of calcium concentrations for other invasive shell-building molluscs, such as the zebra mussel (Hinks & Mackie, 1991; Whittier *et al.*, 2008).

For the logistic regression, 16 habitat variables were retained following the removal of collinear pairs

(Table S1). For climate variables, which were used in both logistic regression and MaxEnt modelling, four of the original 17 remained following the removal of collinear variables (Table S1). No habitat variables were collinear with climate variables. The four climate variables that remained were annual precipitation, precipitation in the warmest quarter, precipitation in the driest quarter and minimum temperature in the coldest month. Maximum temperature in the warmest month was initially considered as a predictive variable, as mass mortality of *Corbicula* has been observed at high temperatures (Cohen *et al.*, 1984; Ilarri *et al.*, 2011) and a warming climate could increase temperatures beyond *Corbicula*'s maximum temperature of 35 °C (McMahon & Bogan, 2001). However, in initial model runs that included maximum temperature in the warmest month, there was a positive correlation between maximum temperature in the warmest month and *Corbicula* presence. *Corbicula* is sensitive to high water temperatures (McMahon & Bogan, 2001), which can lead to mass mortality events (Cohen *et al.*, 1984; Ilarri *et al.*, 2011), and there is no clear mechanism as to why higher maximum temperatures would lead to an increased likelihood of *Corbicula* presence. More importantly, its inclusion precluded the inclusion of a related, collinear variable, namely minimum temperature in the coldest month, which has a strong effect on *Corbicula* presence and solid mechanistic underpinnings (Werner & Rothhaupt, 2008). Thus, maximum temperature in the warmest month was the collinear temperature variable we excluded from the model.

Comparing habitat and climate variables: logistic regression

To identify which individual variables should be included, we conducted a stepwise regression in R (R Development Core Team, 2008) and evaluated candidate models with AICc, a modified version of Akaike's Information Criterion (AIC) (Akaike, 1973) that is adjusted to account for a finite sample size (Burnham & Anderson, 2002). This model initially considered the 16 habitat variables and the four climate variables remaining following the removal of collinear variables as described above. From the results of the stepwise regression, a list of candidate models was created using methods described in Burnham & Anderson (2002). Using model weights, variables that appeared in a total of 0.95 candidate models by weight were identified and were classified as either 'Habitat' (from the EPA EMAP surveys) or 'Climate' (from BioClim variables).

To determine the relative importance of climate and habitat variables in controlling *Corbicula*'s presence or absence, three models were created: a model with only habitat variables, a model with only climate variables and a model with both climate and habitat variables. These three models were compared using AICc, model weights and pseudo-R^2, calculated for the three different models.

Spatially explicit, climate-only model: maximum entropy model

We created the MaxEnt model with MaxEnt software (version 3.3.3e, Phillips, Anderson & Schapire, 2006) using the 2800 known locations of *Corbicula* from our data sources. A MaxEnt model does not use known absence points, instead it generates pseudo-absences from a defined spatial area. This approach has been criticised, especially when it includes areas that have not been sampled or novel habitats (Peterson, Papes & Eaton, 2007). However, given the extremely wide distribution of *Corbicula* within the United States already and because the points from the EPA EMAP surveys are part of a systematic, nationwide survey of streams and rivers, creating pseudo-absences from the entire continental United States, avoids these issues (Phillips, 2007). Using predictive variables input as a continuous ASCII grid, the model fits the distribution of a species over the area of the grid. A MaxEnt model uses several different methods, including linear, nonlinear, hinge and threshold functions, to fit the distribution of known points the best possible way (Elith *et al.*, 2006). MaxEnt creates a model from the predictive variables; model results range from zero to one and are interpreted using *a priori* selected thresholds based on varying levels of permissiveness. For example, the 'minimum training presence' threshold is typically the lowest threshold and therefore the most permissive threshold. Using this threshold, the model output does not need to be very close to one in order for the area to be deemed climatically suitable, and thus it will classify a much larger area as suitable for *Corbicula* than other, more stringent (higher) thresholds. We selected three different thresholds to interpret the MaxEnt model, ranging from liberal (equate entropy of thresholded and original distributions) to conservative (equal test sensitivity and specificity).

Using current climatic data, a MaxEnt model was created using the four independent BioClim variables. If a variable did not contribute significantly to the model fit, as measured by per cent contribution to the overall model fit and by permutation importance (a measure of

how much the model is changed by random variation in a given variable), that variable was removed from the model, and the more parsimonious model was rerun. Cross-validation was performed on five replicate Max-Ent models, using 80% of the known *Corbicula* locations to train the model and the remaining 20% to test the model. Test data were sampled without replacement so that each of the 2800 *Corbicula* locations was used in the test set once and the training set four times. To prevent over-fitting of the response curves of predictive variables, a regularisation multiplier of 10 was used. This prevented the inclusion of complicated higher-order polynomial functions that, despite fitting the data better, produced response curves that were not biologically meaningful.

We applied the MaxEnt model created using current climate conditions and using the threshold values described above, to three projections of climate in the year 2080 corresponding to different emission scenarios: A2a, which includes large human population growth, large amounts of energy use and slow technological change; A1b, which includes lower population growth, high energy use and a mix of fossil fuel and other energy sources; and finally the B2a scenario, which includes medium population growth and a mix of energy sources that are consistent with usage today (Intergovernmental Panel on Climate Change (IPCC), 2007). These scenarios provide a range of projected temperature increases by 2080, ranging from 2.4 C (B2a) to 3.4 C (A2a) (Intergovernmental Panel on Climate Change (IPCC), 2007). Although this is not the full range of projected temperature rise predicted by other models, the selected climate data have a consistent, simple, straightforward methodology for spatial downscaling (Delta Method) and use the same source model (Canadian Centre for Climate Modelling and Analysis) while varying only human population and energy sources.

Results

Comparing habitat and climate variables: logistic regression

Stepwise regression identified nine variables that appeared in more than 95% of candidate models, by weight. Six were habitat variables (water depth, ammonium, pH, per cent fine gravel, total suspended solids and per cent agriculture in the catchment), and three were climate variables (minimum temperature in the coldest month, annual precipitation and precipitation in the driest quarter) (Table 1). When comparing the

Table 1 Variables selected in stepwise regression of *Corbicula* presence. Habitat variables are from the EPA EMAP surveys, and climate variables are from the BioClim data set. All variables in this list appeared in more than 0.95 of models, by weight. The effect that an increase in the variable has on the likelihood of *Corbicula* presence is indicated in parentheses

Habitat		Climate
Depth (+)	Per cent agriculture in catchment (+)	Min. temperature coldest month (+)
pH (+)	Per cent fine gravel (+)	Annual precipitation (+)
Ammonium (−)	Total suspended solids (−)	Precipitation driest quarter (+)

relative importance of the six habitat variables and three climate variables, the combined model, which included all nine variables, had the best fit for *Corbicula* presence and absence from the EPA EMAP data, as measured by AIC, area under the curve (AUC) and pseudo-R^2 (Fig. 1, Table 2). However, the climate model performed nearly as well, whereas the habitat-only model lagged behind both of the other two candidate models in all measures of fit (Table 2).

Spatially explicit, climate-only model: maximum entropy model

The final MaxEnt model included three climatic variables: minimum temperature in the coldest month, precipitation in the warmest quarter and precipitation in the driest quarter (Table S2). Initial runs of the MaxEnt model indicated that annual precipitation, one of the four *a priori* climate variables initially considered (Table S2), was not a significant predictor of *Corbicula* presence or absence. Annual precipitation was removed from the model, and a final model was run. From the final model run, the replicates had an average AUC of 0.762 (out of one) for the test data, with a standard

Table 2 Comparisons between the three model types used in the logistic regression: habitat-only (EPA EMAP data), climate data-only (BioClim variables) and the combined model. By all measures, the combined model performs best. However, the climate-only model does a reasonable job modelling *Corbicula* presence and absence, particularly as measured by AIC and pseudo-R^2. The climate model provides the majority of the fit of the combined model despite having only half the number of variables retained in the habitat model

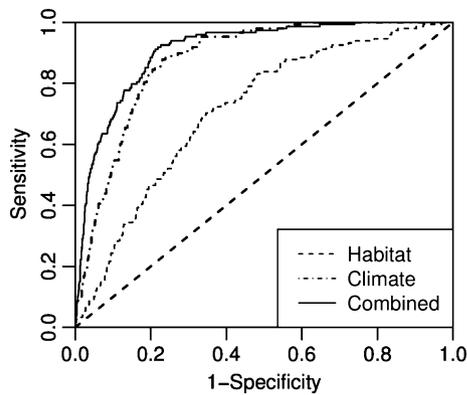|	K	AIC	ΔAIC	Weight	Pseudo-R^2
Combined	9	695	0	1	0.41
Climate-only	3	769	74	9.00E-17	0.32
Habitat-only	6	988	293	2.00E-64	0.079

Fig. 1 Receiver operating curve (ROC) for the three candidate models in the logistic regression: habitat-only (EPA EMAP data, dashed line), climate-only (BioClim data, dot and dashed line) and the combined model (solid line). Sensitivity, the proportion of points that the model identifies as suitable that are truly suitable is on the *y*-axis. On the *x*-axis, 1 – specificity is plotted, which is a measure of the false-positive rate. The climate-only model outperformed the habitat-only model (area under curve = 0.890 and 0.713, respectively). The combined model performed best (AUC = 0.917), but was only slightly better than the climate-only model at predicting *Corbicula*'s presence. A perfect model would have an AUC of 1. The expected results from a random relationship are included in the 1:1 line and would have an AUC of 0.5 (large dashed line).

deviation of 0.009 (Fig. 2). Of the three variables included in the final model, minimum temperature in the coldest month was the most important by per cent contribution (58.4%) and permutation importance (55), followed by precipitation in the driest quarter (27.9% contribution and 26.1 permutation importance) and finally by precipitation in the warmest quarter (13.7% contribution and 18.9 permutation importance).

Comparing the climate variables identified in the logistic model and the MaxEnt model, precipitation in the driest quarter and minimum temperature in the coldest month were identified in both models as important controllers of *Corbicula* distribution. The MaxEnt model identified precipitation in the warmest quarter as an important variable, whereas annual precipitation was retained in the logistic regression. Precipitation in the warmest quarter was a candidate variable for the logistic regression, but was not identified as a significant predictor of *Corbicula*'s distribution. Likewise, annual precipitation was a candidate variable in the MaxEnt model, but was not a significant predictor of *Corbicula*'s distribution using that modelling approach.

The MaxEnt model suggests that *Corbicula* is limited to southern New York and the southern extent of the Great Lakes under current climate conditions, leaving much of the Midwest climatically unsuitable (Fig. 3a). Under all three future climate scenarios, the range is
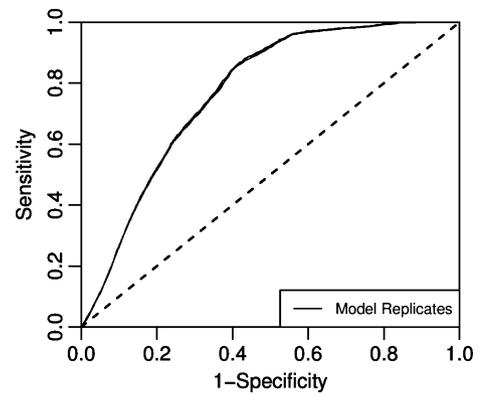

Fig. 2 Receiver operating curve (ROC) for the five replicates of the maximum entropy model using the three selected BioClim variables: minimum temperature in the coldest month, precipitation in the driest quarter and precipitation in the warmest quarter. The average area under curve was 0.762 and was very consistent across the five model replicates.

predicted to expand throughout New England, the Midwest and into the Great Lakes by 2080, or potentially sooner (Fig. 3b,c,d). The exact range varies between the three climate scenarios, but the overall range increase is fairly consistent. Using the intermediate threshold (10% training presence) presented in Fig. 3, *Corbicula*'s potential range is predicted to increase from 37% of the United States by area to 46, 49 or 48.5% under the B2A, A1B and A2A climate scenarios, respectively. This increase in range represents a total areal increase of 25–32% in the distribution of *Corbicula*, and this occurs despite slight range retractions in Louisiana and southern Florida under the A1B climate scenario (Fig. 3c) and nearly all of Florida in the A2A climate scenario (Fig. 3d).

Discussion

The distribution of the widespread invasive freshwater clam *Corbicula fluminea* is controlled both by habitat variables, such as substrate and pH, and by climate variables, indicating the potential interaction between invasive species and climate change, as predicted by many studies (Hellmann *et al.*, 2008; Rahel & Olden, 2008; Strayer, 2010). Within our extensive data set, climate variables are more important than habitat in driving the distribution of *Corbicula*. This supports the recent findings of Petitpierre *et al.* (2012) showing widespread invaders have similar climate niches in their native and invasive ranges. The clear importance of climate variables in controlling *Corbicula*'s distribution in the logistic regression makes our use of the climate-only MaxEnt model particularly appropriate. The switch to climate-only modelling with MaxEnt is also advantageous
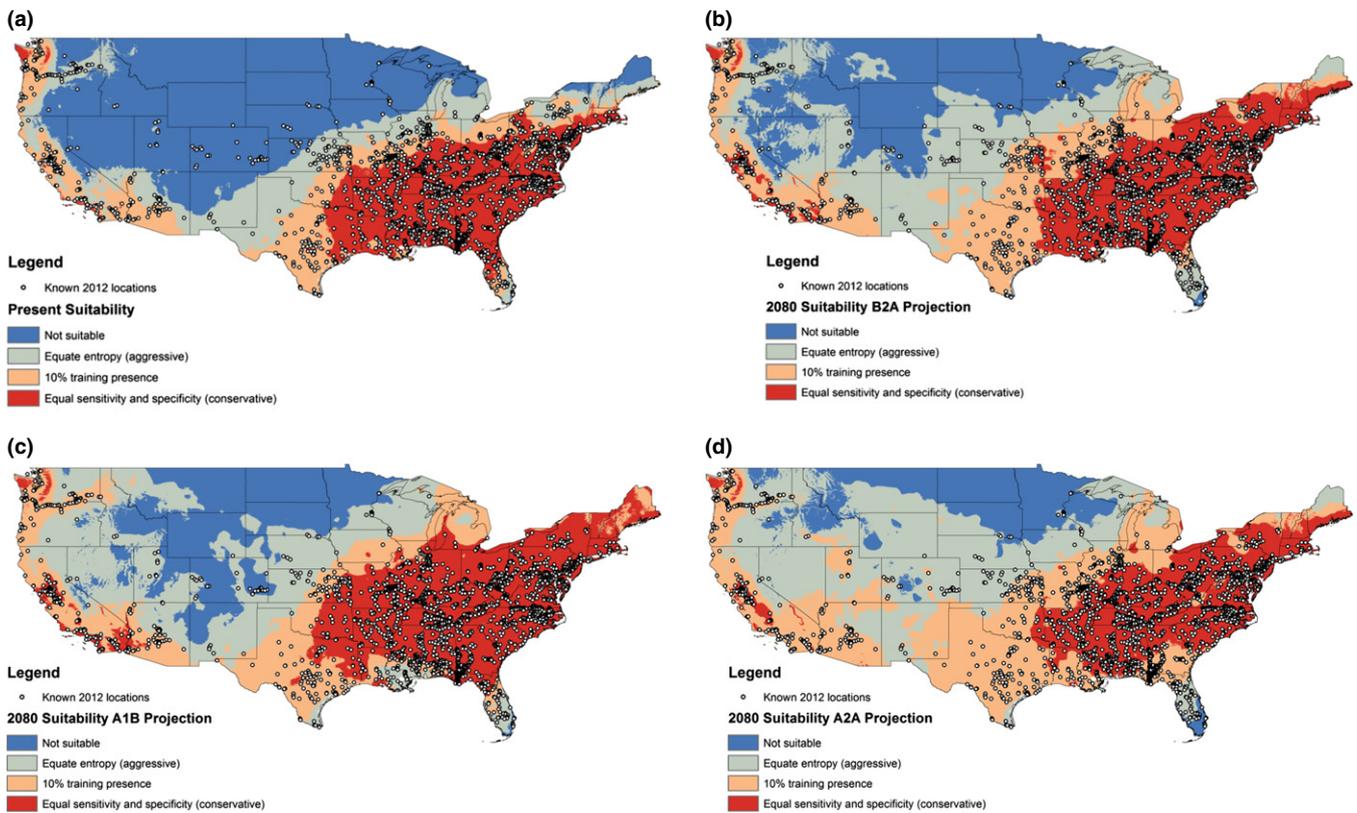
(a)

(b)

(c)

(d)

Fig. 3 A map of modelled likelihood of *Corbicula* presence, from a maximum entropy model based on current climate conditions of three BioClim variables: minimum temperature coldest month, precipitation in the driest quarter, precipitation in the warmest quarter (Bio 6, 17, 18) for: (a) current climate conditions, and projected conditions in 2080 under (b) B2A emissions scenario, (c) A1B emissions scenario and (d) A2A emissions scenario (listed in increasing warmness). The map is presented using various thresholds of suitability, as calculated by MaxEnt. Equate entropy is a more liberal threshold, which will determine a larger area to be suitable than the 10% training presence threshold or the most conservative threshold, equal sensitivity and specificity.

because it allows us to use the much larger set of known *Corbicula* locations where habitat factors were unmeasured. Climate envelope models are often criticised for failing to include other important variables or failing to compare the importance of climate variables relative to other biotic and abiotic factors (Pearson & Dawson, 2003). By explicitly making this comparison, we have demonstrated that a climate-only MaxEnt model can capture a major portion of the variability in a species distribution at the continental scale.

Our results show that a significant expansion in *Corbicula*'s range throughout the northern and central United States is probable with future temperature increases. This may already be occurring, as a population was identified in the Merrimack River, New Hampshire, while this analysis was being conducted (New Hampshire Department of Environmental Services (NH DES), 2012). This range expansion could be rapid and occur soon, because climate changes appear to be occurring at the higher range of predicted temperature changes (Fasullo &

Trenberth, 2012). The models also identified waterbodies with higher pH and fine gravel substrate as likely to be suitable for *Corbicula*, which supports previous studies that indicated *Corbicula*'s preference for sand and fine gravel substrates (McMahon, 1999; McMahon & Bogan, 2001). These results can help catchment managers to target and monitor vulnerable habitats. Because of this, catchment managers in New England and the northern central United States should be aware that rivers with neutral to basic pH and sandy substrate are particularly at risk of invasion. Monitoring for early detection coupled with an eradication or control plan may be able to prevent widespread establishment, especially in lentic environments, where populations of *Corbicula* have been reduced using gas impermeable barriers (Wittman *et al.*, 2012). Given *Corbicula*'s high per biomass filtration rates (McMahon & Bogan, 2001; Atkinson *et al.*, 2010), high densities and influential impacts, newly invaded rivers could experience significant changes in energy flow and community structure, as observed in previously invaded rivers

(Cohen *et al.*, 1984; Lauritsen, 1986; Phelps, 1994; Hakenkamp & Palmer, 1998; Sousa, Antunes & Guilhermino, 2008a).

Under the A2A climate scenario, the one with the highest emissions, the area deemed climatically suitable in the southeastern United States is predicted to shrink in southern Florida. The A2A projections predict a decline in precipitation in the driest quarter and an increase in overall temperature. The interaction between increasing temperatures and declining precipitation is probably responsible for the classification of this area as unsuitable according to the more conservative thresholds. There is no reason why increasing minimum temperature would lead to areas being unsuitable for *Corbicula*; however, maximum temperature and minimum temperature covary. Increased maximum temperatures and decreased precipitation could lead to mass mortality events such as those observed during periods of low water level and high temperature (Cohen *et al.*, 1984; Ilarri *et al.*, 2011; Bodis, Toth & Sousa, in press; W.G. McDowell pers. obs.), due to the inability of *Corbicula* to regulate its oxygen consumption under stressful conditions (McMahon, 2002). In all these observed cases, however, *Corbicula* populations quickly rebounded from the mass mortality events (and were probably sped in such recovery from fast growth in warmer water) and were not removed from a river entirely. Regarding the future suitability of southern Florida, it is important to note, however, that under the A2A climate scenario, a combination of precipitation and temperature conditions in Florida arises that does not exist in the training data, so the model results are extrapolated beyond training data and should be interpreted cautiously (Elith *et al.*, 2006).

The combined climate and habitat model explained a substantial amount of the variability in *Corbicula* presence and absence. Some of the remaining variability may be explained by biotic factors such as competition and predation, which are not included in these models. Within dense beds of native mussels, there are indications that competition may prevent establishment of *Corbicula* (Vaughn & Spooner, 2006). When examining presence and absence on a reach or river scale, however, this is unlikely to prevent establishment of *Corbicula*. Although predation has been shown to control densities (Robinson & Wellborn, 1988), there is no evidence that predators can remove *Corbicula* from a system entirely. Few barriers to *Corbicula* dispersal into suitable habitat seem to exist, given its rapid expansion in the southeastern United States following its introduction (McMahon & Bogan, 2001). Because *Corbicula* can disperse widely and has been present throughout the western United States for 80 years and the eastern United States for 30–40 years, *Corbicula* has probably reached most of the locations that are climatically suitable for it to survive under current climate conditions. For species with a shorter invasion history or larger barriers to dispersal, it is necessary to recognise that the current distribution of the species may not reflect climatic tolerances. Because of this, species distribution models for such species are less reliable, and it is critical to acknowledge the limitations of modelling in this regard.

There are two caveats to consider when interpreting our results. First, as is typical for almost all large-scale climate models for aquatic species, air temperature was used in place of water temperature in both the logistic regression and the MaxEnt models. Air temperatures typically track water temperatures well, and similar approaches have been successfully used to model aquatic species including freshwater diatoms, snails, salamanders and trout (Kumar *et al.*, 2009; Milanovich *et al.*, 2010; Wenger *et al.*, 2011; Blank & Blaustein, 2012; Byers *et al.*, 2013). There will be some discrepancies between air temperature and water temperature, and these differences will probably be greater for large bodies of water, systems with large amounts of groundwater inputs and hot springs. In fact, these site-specific discrepancies may explain some of the outlier points, especially those in large river systems where local water temperature may be buffered against extreme fluctuations in air temperature. Given that more than 2800 known *Corbicula* locations went into the model, it is not surprising that some of these points known to contain *Corbicula* fall outside what is considered to be a suitable climate. Nevertheless, only 3% fell outside the most liberal threshold presented. Also, it is important to recognise that significant departures from natural conditions may allow *Corbicula* to persist in environments that would otherwise be unsuitable. For example, several points in the USGS NAS Database include brief notes regarding outflow of water used in cooling operations. This warmer microclimate may allow *Corbicula* to persist in an area not predicted to be hospitable based on air temperature.

Second, our model does not make any predictions regarding densities, nor does it address how the effects of *Corbicula* on a previously colonised or newly invaded aquatic ecosystem might be altered by climate change. Density is one of the main factors controlling the impact of a non-indigenous species (Parker *et al.*, 1999), and areas that are deemed suitable habitat may support widely varying densities of *Corbicula*. Modelling densities go beyond the abilities of the approaches used in this paper and present an interesting challenge for

future work. The interactions between *Corbicula* behaviour and climate change could lead to alterations in their impacts. For example, increasing temperatures could lead to higher metabolic rates and increased filtration and nutrient uptake by *Corbicula* (Spooner & Vaughn, 2008). However, at very high temperatures, mass mortalities of *Corbicula* can occur and may negatively affect water quality and native mussels by increasing toxic concentrations of ammonia in porewater and the water column (Cherry *et al.*, 2005; Cooper, Bidwell & Cherry, 2005). Climate change could make these events more common.

In summary, it is critical to assess the importance of climate in controlling the distribution of a species, as this dramatically affects the weight that should be given to the impacts of climate change and more local drivers. By empirically testing the relative importance of climate and carefully considering the biology of the organism being studied, we have demonstrated that *Corbicula* is likely to expand well beyond its current distribution. Already widely distributed, *Corbicula* could become nearly ubiquitous throughout the conterminous United States. MaxEnt and the freely available BioClim data make constructing a species distribution model for current and future distributions deceptively easy, and their misuse opens them up to criticism (Woodward & Beerling, 1997; Haegeman & Loreau, 2008). However, the combination of multiple approaches allows us to recognise what the models are capturing well and can lead to a more robust result than any single model alone.

Acknowledgments

Author's contributions

WGM and JEB conceived of the study, AJB maintained and provided database for MaxEnt modelling, WGM and JEB conducted analyses; and AJB, WGM and JEB wrote and edited the article.

References

Akaike H. (1973) Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*. (Eds B.N. Petrov & F. Csaki), Akademiai Kiado, Budapest, Hungary, pp. 267–281.

Atkinson C.L., Opsahl S.P., Covich A.P., Golladay S.W. & Connor L.M. (2010) Stable isotopic signatures, tissue stoichiometry, and nutrient cycling (C and N) of native and invasive freshwater bivalves. *Journal of North American Benthological Society*, **29**, 496–505.

Blank L. & Blaustein L. (2012) Using ecological niche modeling to predict the distributions of two endangered amphibian species in aquatic breeding sites. *Hydrobiologia*, **685**, 121–134.

Bodis E., Toth B. & Sousa R. (In press) Massive mortality of invasive bivalves as a potential resource subsidy for the adjacent terrestrial food web. *Hydrobiologia*.

Boltovskoy D., Karateyev A., Burlakova L., Cataldo D., Karatayev V., Sylvester F. *et al.* (2009) Significant ecosystem-wide effects of the swiftly spreading invasive freshwater bivalve *Limnoperna fortunei*. *Hydrobiologia*, **636**, 271–284.

Burnham K.P. & Anderson D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, NY.

Byers J.E., McDowell W.G., Dodd S.R., Haynie R.S., Pintor L.M. & Wilde S.B. (2013) Climate and pH predict the potential range of the invasive apple snail (*Pomacea insularum*) in the Southeastern United States. *PLoS ONE*, **8**, e56812. doi:10.1371/journal.pone.0056812.

Byers J.E., Reichard S., Randall J.M., Parker I.M., Smith C.S., Lonsdale W.M. *et al.* (2002) Directing research to reduce the impacts of nonindigenous species. *Conservation Biology*, **16**, 630–640.

Caraco N.F., Cole J.J., Raymond P.A., Strayer D.L., Pace M.L., Findlay S.E.G. *et al.* (1997) Zebra mussel invasion in a large, turbid river: phytoplankton response to increased grazing. *Ecology*, **78**, 588–602.

Cataldo D. & Boltovsky D. (1999) Population dynamics of *Corbicula fluminea* (Bivalvia) in the Paraná River Delta (Argentina). *Hydrobiologia*, **380**, 153–163.

Cherry D.S., Scheller J.L., Cooper N.L. & Bidwell J.R. (2005) Potential effects of Asian clam (*Corbicula fluminea*) die-offs on native freshwater mussels (Unionidae) I: water-column ammonia levels and ammonia toxicity. *Journal of the North American Benthological Society*, **24**, 369–380.

Clark J.S., Carpenter S.R., Barber M., Collins S., Dobson A., Foley J.A. *et al.* (2001) Ecological forecasts: an emerging imperative. *Science*, **293**, 657–660.

Cohen A.N. & Carlton J.T. (1998) Accelerating invasion rate in a highly invaded estuary. *Science*, **279**, 555–558.

Cohen R.R.H., Dresler P.V., Phillips E.J.P. & Cory R.L. (1984) The effect of the Asiatic clam, *Corbicula fluminea*,

on phytoplankton of the Potomac River, Maryland. *Limnology and Oceanography*, **29**, 170–180.

Cooper N.L., Bidwell J.R. & Cherry D.S. (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**, 381–394.

Elith J., Graham C.H., Anderson R.P., Dudík M., Ferrier S., Guisan A. *et al.* (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography*, **29**, 129–151.

Fasullo J.T. & Trenberth K.E. (2012) A less cloudy future: the role of subtropical subsidence in climate sensitivity. *Science*, **338**, 792–794.

Graham M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology*, **84**, 2809–2815.

Haegeman B. & Loreau M. (2008) Limitations of entropy maximization in ecology. *Oikos*, **117**, 1700–1710.

Hakenkamp C.C. & Palmer M.A. (1998) Introduced bivalves in freshwater ecosystems: the impact of *Corbicula* on organic matter dynamics in a sandy stream. *Oecologia*, **119**, 445–451.

Hall R.O., Tank J.L. & Dybdahl M.F. (2003) Exotic snails dominate nitrogen and carbon cycling in a highly productive stream. *Frontiers in Ecology and the Environment*, **1**, 407–411.

Hecky R.E., Smith R.E.H., Barton D.R., Guildford S.J., Taylor W.D., Charlton M.N. *et al.* (2004) The nearshore phosphorus shunt: a consequence of ecosystem engineering by dreissenids in the Laurentian Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 1285–1293.

Hellmann J.J., Byers J.E., Bierwagen B.G. & Dukes J.S. (2008) Five potential consequences of climate change for invasive species. *Conservation Biology*, **22**, 534–543.

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

Hinks S.S. & Mackie G.L. (1991) Effects of pH, calcium, alkalinity, hardness, and chlorophyll on the survival, growth, and reproductive success of zebra mussel (*Dreissena polymorpha*) in Ontario lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**, 2049–2057.

Ilarri M.I., Antunes C., Guilhermino L. & Sousa R. (2011) Massive mortality of the Asian clam *Corbicula fluminea* in a highly invaded area. *Biological Invasions*, **13**, 277–280.

Intergovernmental Panel on Climate Change (IPCC) (2007) *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, U.K.

Jarnevich C.S., Stohlgren T.J., Barnett D. & Kartesz J. (2006) Filling in the gaps: modeling native species richness and invasions using spatially incomplete data. *Diversity and Distributions*, **12**, 511–520.

Jeschke J.M. & Strayer D.L. (2008) Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences*, **1134**, 1–24.

Kumar S., Spaulding S.A., Stohlgren T.J., Hermann K.A., Schmidt T.S. & Bahls L.L. (2009) Potential habitat distribution for the freshwater diatom *Didymosphenia geminata* in the continental US. *Frontiers in Ecology and the Environment*, **7**, 415–420.

Lauritsen D.D. (1986) Filter-feeding in *Corbicula fluminea* and its effect on seston removal. *Journal of the North American Benthological Society*, **5**, 165–172.

Mack R.N., Simberloff D., Lonsdale M.W., Evans H., Clout M. & Bazzaz F. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.

McMahon R.F. (1999). Invasive characteristics of the freshwater bivalves, *Corbicula fluminea*. In: *Nonindigenous Freshwater Organisms: Vectors, Biology and Impacts* (Eds R. Claudi & J.H. Leach), pp. 315–343. Lewis Publishers, Boca Raton, FL.

McMahon R.F. (2002) Evolutionary and physiological adaptations of aquatic invasive animals; r selection versus resistance. *Canadian Journal Fisheries and Aquatic Sciences*, **59**, 1235–1244.

McMahon R.F. & Bogan A.E. (2001) Bivalves. In: *Ecology and Classification of North American Freshwater Invertebrates*, 2nd edn. (Eds J.H. Thorp & A.P. Covich), pp. 331–428. Academic Press, New York, NY.

Milanovich J.R., Peterman W.E., Nibbelink N.P. & Maerz J.C. (2010) Projected loss of a salamander diversity hotspot as a consequence of projected global climate change. *PLoS ONE*, **5**, e12189.

Moore D.S. & McCabe G.P. (1993) *Introduction to the Practice of Statistics*. Freeman, New York, NY.

New Hampshire Department of Environmental Services (NH DES) (2012) *Asian Clams in New Hampshire*. Available at: http://des.nh.gov/organization/commissioner/pip/factsheets/bb/documents/bb-62.pdf. (Last accessed May 14, 2013).

Parker I.M., Simberloff D., Lonsdale W.M., Goodell K., Wonham M., Kareiva P.M. *et al.* (1999) Impact: toward a framework for understanding the ecological effects of invaders. *Biological Invasions*, **1**, 3–19.

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

Paulsen S.G., Hawkins C.P., Van Sickle J., Yuan L.L. & Holdsworth S.M. (2008) An invitation to apply national survey data to ecological research. *Journal of the North American Benthological Society*, **27**, 1017–1018.

Pearson R.G. & Dawson T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.

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

Petitpierre B., Kueffer C., Broennimann O., Randin C., Daehler C. & Guisan A. (2012) Climatic niche shifts are rare among terrestrial plant invaders. *Science*, **335**, 1344–1348.

Phelps H.L. (1994) The Asiatic clam (*Corbicula fluminea*) invasion and system-level ecological change in the Potomac River Estuary near Washington, DC. *Estuaries*, **17**, 614–621.

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

Phillips S.J., Anderson R.P. & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modeling*, **190**, 231–259.

R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org.

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

Robinson J.V. & Wellborn G.A. (1988) Ecological resistance to the invasion of a freshwater clam, *Corbicula fluminea*: fish predation effects. *Oecologia*, **77**, 445–452.

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

Sousa R., Rufino M., Gaspar M., Antunes C. & Guilhermino L. (2008b) Abiotic impacts on spatial and temporal distribution of *Corbicula fluminea* (Muller, 1774) in the River Minho Estuary, Portugal. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 98–110.

Spooner D.E. & Vaughn C.C. (2008) A trait based approach to species' roles in stream ecosystems: climate change, community structure, and material cycling. *Oecologia*, **158**, 307–317.

Strayer D.L. (2010) Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, **55**, 152–174.

Strayer D.L., Caraco N.F., Cole J.J., Findlay S. & Pace M.L. (1999) Transformation of freshwater ecosystems by bivalves. *BioScience*, **49**, 19–27.

Vaughn C.C. & Hakenkamp C.C. (2001) The functional role of burrowing bivalves in freshwater ecosystems. *Freshwater Biology*, **46**, 1431–1446.

Vaughn C.C. & Spooner D.E. (2006) Scale-dependent associations between native freshwater mussels and invasive *Corbicula*. *Hydrobiologia*, **568**, 331–339.

Warton D.I. & Hui F.K.C. (2011) The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, **92**, 3–10.

Wenger S.J., Isaak D.J., Luce C.H., Neville H.M., Fausch K.D., Dunham J.B. *et al.* (2011) Flow regime, temperature, and biotic interactions drive differential declines of trout species under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 14175–14180.

Werner S. & Rothhaupt K.O. (2008) Mass mortality of the invasive bivalve *Corbicula fluminea* induced by a severe low-water event and associated low water temperatures. *Hydrobiologia*, **613**, 143–150.

Whittier T.R., Ringold P.L., Herlihy A.T. & Pierson S.M. (2008) A calcium based invasion risk assessment for zebra and quagga mussels (*Dreissena* spp). *Frontiers in Ecology and the Environment*, **6**, 180–184.

Wittman M.E., Chandra S., Reuter J.E., Schladow S.G., Allen B.C. & Webb K.J. (2012) The control of an invasive bivalve, *Corbicula fluminea*, using gas impermeable benthic barriers in a large natural lake. *Environmental Management*, **49**, 1163–1173.

Woodward F.I. & Beerling D.J. (1997) The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. *Global Ecology and Biogeography Letters*, **6**, 413–418.

Wrona F.J., Prowse T.D., Reist J.D., Hobbie J.E., Levesque L.M.J. & Vincent W.F. (2006) Climate change effects on aquatic biota, ecosystem structure and function. *Ambio*, **35**, 359–369.

Zarnetske P.L., Skelly D.K. & Urban M.C. (2012) Biotic multipliers of climate change. *Science*, **336**, 1516–1518.

Supporting Information

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

Table S1. Full list of EPA EMAP Habitat Variables Considered. Variables that were selected for inclusion in the final logistic regression model are marked with a †.

Table S2. List of environmental variables from the BIOCLIM dataset used in the MaxEnt model.

(Manuscript accepted 5 December 2013)