

Representing calcification in distribution models for aquatic invasive species: surrogates perform as well as CaCO_3 saturation state

Alexander W. Latzka, John T. Crawford, Avery S. Koblings, Yuri Caldeira, Emily Hilts & M. Jake Vander Zanden

Hydrobiologia

The International Journal of Aquatic Sciences

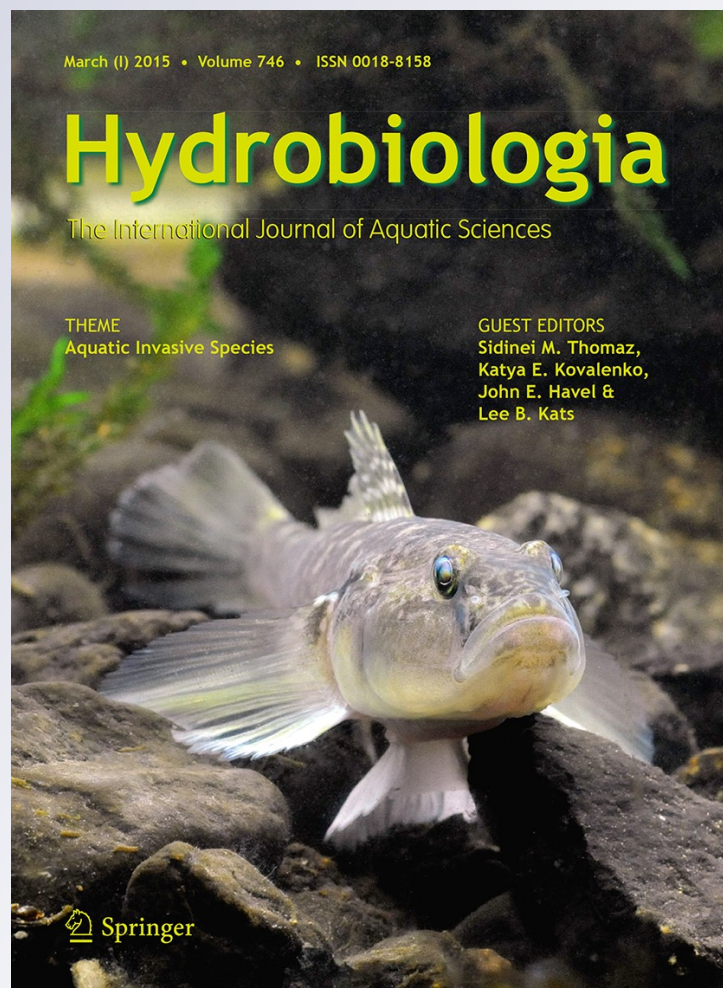
ISSN 0018-8158

Volume 746

Number 1

Hydrobiologia (2015) 746:197-208

DOI 10.1007/s10750-014-2001-7



Your article is protected by copyright and all rights are held exclusively by Springer International Publishing Switzerland. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Representing calcification in distribution models for aquatic invasive species: surrogates perform as well as CaCO_3 saturation state

Alexander W. Latzka · John T. Crawford ·
Avery S. Koblings · Yuri Caldeira · Emily Hilts ·
M. Jake Vander Zanden

Received: 14 January 2014 / Revised: 14 July 2014 / Accepted: 25 July 2014 / Published online: 8 August 2014
© Springer International Publishing Switzerland 2014

Abstract Modeling of invasive species' potential distributions is critical for preventing new invasions. Distribution models use environmental characteristics of observed distributions to determine the suite of conditions that allows invasion. Some variables are directly related to species' biological needs, while others are surrogates that do not directly influence distributions. For aquatic invasive species (AIS) with calcified shells and exoskeletons, one common limiting factor is the chemical restriction on forming calcium carbonate (CaCO_3), which is usually represented by conductivity or dissolved calcium. We predicted that conductivity and calcium would not serve as accurate surrogates for CaCO_3 saturation state, which is more directly linked to calcification potential. To test this prediction, we fit classification trees to field surveys of northern Wisconsin lakes for three calcifying AIS—rusty crayfish (*Orconectes rusticus*), banded mystery snail (*Viviparus georgianus*), and Chinese mystery snail (*Cipangopaludina chinensis*). Along with lake area, variables associated with calcification were the best predictors of species

distributions. Even though saturation state is the most biologically mechanistic variable, it did not improve model predictions or explain more variability in species presence than its surrogates. Overall, surrogates are sufficient for species distribution modeling in these lakes that exhibit a wide range of chemical states.

Keywords Aquatic invasive species · Distribution models · Calcification · Saturation state · *Orconectes rusticus* · *Cipangopaludina chinensis* · *Viviparus georgianus*

Introduction

Invasive species are a main driver of biodiversity loss (Sala et al., 2000) and can cause changes in ecosystem function (Vitousek, 1990). These changes can disrupt the delivery of ecosystem services (Pejchar & Mooney, 2009) and result in significant economic costs (Pimentel et al., 2005). Freshwaters seem to be especially vulnerable to the impacts of invasive species (Sala et al., 2000). Identifying vulnerable sites prior to their invasion enables efficient prevention efforts, which are significantly cheaper than control efforts after an invasion has occurred (Leung et al., 2002). These predictions often involve estimating the habitat suitability of sites using data on their environmental characteristics (Vander Zanden & Olden,

Guest editors: Sidinei M. Thomaz, Katya E. Kovalenko, John E. Havel & Lee B. Kats / Aquatic Invasive Species

A. W. Latzka (✉) · J. T. Crawford · A. S. Koblings ·
Y. Caldeira · E. Hilts · M. J. Vander Zanden
Center for Limnology, University of Wisconsin-Madison,
680 N. Park St, Madison, WI 53706, USA
e-mail: latzka@wisc.edu

2008), which may be directly or indirectly related to a species' biological requirements (Austin, 2007).

One factor that often limits the distributions of many aquatic taxa is the amount of available dissolved calcium and carbonate that are needed to form and maintain calcium carbonate (CaCO_3) exoskeletons (e.g., Capelli & Magnuson, 1983; Ramcharan et al., 1992; Olden et al., 2006; Oliveira et al., 2010). Both calcium and bicarbonate are at least partially supplied from ambient water, although some organisms may internally store CaCO_3 to be used when needed (Horne et al., 2009; Akiva-Tal et al., 2011; Marin et al., 2012). Low concentrations of dissolved calcium can impose limits on CaCO_3 mineralization, so dissolved calcium and correlated parameters are used in biological models. However, these models ignore carbonate, which also controls rates of calcification (Langdon et al., 2000). In contrast, the CaCO_3 saturation state (often denoted by Ω , unitless) describes the potential for precipitation or dissolution (Doney et al., 2009). Saturation state is defined as the product of the concentrations of the reactants divided by a solubility product which depends on temperature, salinity, pressure, and the particular mineral phase of the compound ($\Omega = [\text{Ca}^{2+}][\text{CO}_3^{2-}]/K'_{\text{sp}}$), where values above 1.0 indicate the potential for precipitation, and values below 1.0 indicate the potential for dissolution (Doney et al., 2009). Saturation state is thus a biologically mechanistic variable that could be used in distribution models and forecasting. In ocean acidification research, CaCO_3 saturation states have been used to explain and predict declines of calcifying corals and other marine taxa (Hoegh-Guldberg et al., 2007; Doney et al., 2009). Although saturation states are equally applicable in freshwater, they have been rarely measured or used in distribution models.

Instead, distribution models for freshwater species have relied on more easily measurable and more widely available surrogates for CaCO_3 saturation state. However, the effects of surrogate usage in general are often not evaluated (Austin, 2007), so effects on model accuracy are unknown. For CaCO_3 saturation state, surrogate variables include dissolved calcium, conductivity, hardness, and alkalinity. These variables have been shown in cross-site studies and experiments to correlate with species distributions (Herbst et al., 2008; Boets et al., 2013), mortality

(Hammond et al., 2006), shell or carapace strength (Czarnołęski et al., 2006; Glass & Darby, 2009), and predation success (Lewis & Magnuson, 1999; Czarnołęski et al., 2006) for various calcifying taxa, so they may adequately substitute for saturation state. However, their correlation with saturation state is not perfect (Fig. 1a, b; data from sampled lakes), and thus some of the variance in these responses may be better described by saturation state. One study that did explicitly use saturation state found that it was an important predictor of habitat suitability for the invasive golden mussel (*Limnoperna fortunei*) in Brazilian rivers, although saturation state was not important in large North American rivers where calcium and pH tended to be well above the golden mussel's physiological threshold (Oliveira et al., 2010), but comparisons were not made to other calcification variables.

Two taxa that form CaCO_3 structures are aquatic crustaceans and gastropods. In North American temperate lakes, three common invasive species in these groups are the rusty crayfish (RC; *Orconectes rusticus*, Girard 1852), the Chinese mystery snail (CMS; *Cipangopaludina chinensis*, Gray 1863, also called *Bellamya chinensis*, Smith, 2000), and the banded mystery snail (BMS; *Viviparus georgianus*, Lea 1834). Together, these are three of the most common aquatic invasive species in our study region in terms of number of lakes invaded (Latzka, unpublished data). The impacts of RC are broad and well studied, and include competition with and extirpations of native crayfish (Olden et al., 2006), predation on benthic invertebrates, and ecosystem-level shifts caused by the consumption and destruction of macrophytes (Lodge et al., 1994; Hansen et al., 2013). The snails (CMS and BMS) are potentially less harmful (Johnson et al., 2008; Solomon et al., 2009), but their impacts are not well studied, and they can sometimes reach very high abundances (i.e., >280 BMS/m²; Latzka, unpublished data).

We hypothesized that the calcification of these species is linked closely to saturation states, causing distribution models based on saturation state to outperform those based on surrogates. Specifically, we expected predictions for lakes with near-threshold calcification conditions to be improved by saturation state (Fig. 1c, d). We tested this hypothesis using a

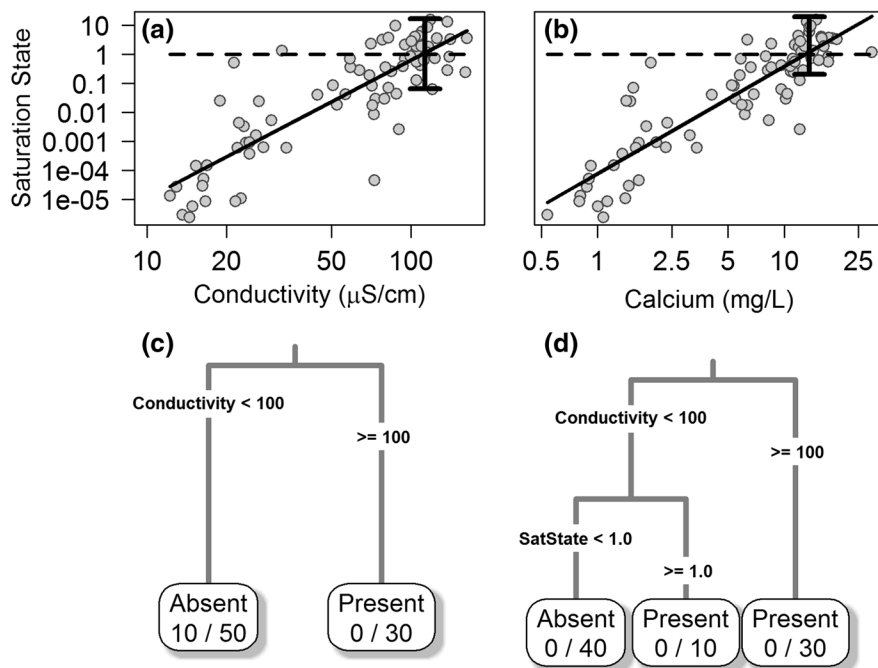


Fig. 1 **a, b** Scatterplots of surrogates (conductivity (**a**) and calcium (**b**)) and saturation state, with values plotted on the log scale and raw values shown on the axes. *Solid lines* indicate the regression lines (conductivity: $P < 0.001$, $R^2 = 0.72$; calcium: $P < 0.001$, $R^2 = 0.78$), *dashed lines* indicate a hypothetical threshold at a saturation state of 1.0 (log of sat. state = 0), and the *brackets* indicate the range in values surrogates may take at that threshold. **c, d** Hypothetical classification trees demonstrate a potential consequence of this variation in the surrogate around

the biological threshold. In **c** the tree based on conductivity alone incorrectly predicts 10 of 80 lakes, whereas in **d** saturation state allows these lakes to be predicted correctly, even if the surrogate or another variable forms the first node of the tree. Each split in the tree lists the variable and its associated threshold, which determines which branch to follow for a given lake. The resulting nodes, their presence/absence predictions, and the number of lakes misclassified out of the total in that node are noted in the *boxes*

survey of lakes that span a calcification gradient for the presence of these three invasive species. We constructed distribution models for each and analyzed the predictive abilities of three calcification variables: conductivity, dissolved calcium concentration, and CaCO_3 saturation state.

Materials and methods

Site description

The Northern Highlands Lake District (NHL; 45°54'N, 89°38'W) in Wisconsin contains approximately 1,500 lakes, to which AIS are perceived as a dominant threat (Beardmore et al., unpublished data). AIS are spread primarily by recreational boaters (Rothlisberger et al., 2010). These lakes vary in landscape position, with many small bog lakes, seepage lakes, and

large drainage lakes that vary in their relative groundwater input (Kratz et al., 1997). Thus, significant gradients of dissolved calcium and pH (and thus calcification potential) exist, potentially constraining the distributions of many invasive species in the region.

Data collection

During the summers of 2010 and 2011, we sampled 81 NHL lakes for BMS, CMS, and RC and for several water chemistry variables related to calcification. We selected sites that varied along a calcification gradient while controlling for other variables that influence invasibility, using a stratified random sample of lakes at least 1 ha in surface area. Lakes were first stratified by the presence or absence of a boat launch, then by high or low conductivity (since conductivity data were available prior to sampling; using a 3rd quartile cutoff), then by large or small (3rd quartile cutoff)

surface area, resulting in eight unique strata. At least 10 lakes from each stratum were chosen for sampling.

At each site, the presence or absence of each species was determined using snorkel surveys and crayfish traps. We conducted timed snorkel surveys, with 30-min snorkels at the primary boat launch or access point on each lake, and 10-min snorkels at each of five sites that were selected to represent the littoral habitat variability (e.g., substrate type, bathymetry, shoreline morphometry) present in the lake. We searched for both crayfish and snails when snorkeling. Crayfish were collected in Gee-style minnow traps with widened openings baited with beef liver (Capelli & Magnuson, 1983). At each of the five sites, we placed two four-meter transects perpendicular to shore beginning between 0.5- and 1-m depth, with three minnow traps placed evenly along each transect. Traps were collected after one night, and all crayfish were identified.

We also measured several water chemistry variables related to calcification potential at each lake. We measured conductivity, temperature, and pH using a Thermo Scientific Orion 5* Plus handheld meter between 1- and 2-m depth in the center of each lake. To minimize the error associated with field measurements of pH, we stored the pH electrode in electrode storage solution between lakes, and conducted a three-point calibration at values of 4, 7, and 10 at least once every 3 days. Secchi depth was also measured at this location. For measures of dissolved inorganic and organic carbon (DIC/DOC) and dissolved calcium, we collected filtered water samples (through 0.70 μm polycarbonate filter) from 0.5 m below the surface. Thirty mL samples were stored in sealed glass containers with zero head space to estimate DIC/DOC concentrations and a second container with 50 ml of filtered sample was acidified with 0.5-ml ultrapure concentrated hydrochloric acid to estimate calcium concentrations. All samples were stored on ice while in the field, and subsequently refrigerated until laboratory analysis. DIC/DOC samples were analyzed with a Shimadzu TOC-V-csh Total Organic Carbon Analyzer and calcium samples with an inductively coupled plasma optical emission spectrophotometer (North Temperate Lakes LTER Research Protocol: <http://lter.limnology.wisc.edu/research/protocols>).

For each lake, we calculated the carbonate concentration according to the equation

$$[\text{CO}_3] = \frac{C_t * K_{a1} * K_{a2}}{K_{a1} * K_{a2} + K_{a1} * [H^+] + [H^+]^2}$$

from Butler (1991), based on our measurements of DIC (converted to a molar concentration C_t) and pH (converted to $[H^+]$). We assumed temperatures of 25°C for values of K_{a1} (4.47×10^{-7}) and K_{a2} (4.68×10^{-11}). Calculated carbonate concentrations, measured calcium concentrations, and K_{sp} constants were then used to calculate saturation state ($\Omega = [\text{Ca}^{2+}][\text{CO}_3^{2-}]/K'_{sp}$). We chose the K_{sp} constant by assuming a temperature of 25°C and the calcite mineral phase of CaCO_3 : $K_{sp} = 3.36 \times 10^{-9}$. We used the calcite form because calcite is present in the crayfish cuticle (Roer & Dillaman, 1984) (unknown for CMS and BMS). However, using aragonite instead would not affect our conclusions.

Data analysis

We used classification and regression tree (CART) analyses to model the presence/absence of these species among 79 sampled lakes (two omitted due to contamination of water samples). We investigated how each calcification variable affected the structure and performance of the best trees. CART analysis repeatedly splits a dataset into increasingly homogeneous groups with respect to some response variable (De'ath & Fabricius, 2000). CART produces a sequence of binary questions (i.e., is the lake above or below a certain concentration of calcium?) about the data that can be visualized as a tree (De'ath & Fabricius, 2000). Data inputs are flexible, as CART works well with both continuous and categorical predictors, even if some are non-normal or collinear, as well as with linear and nonlinear relationships (Prasad et al., 2006). The optimal tree is produced by finding the sequence of splits that explains the most variation in the response variable while limiting the number of splits by how much variation each explains (De'ath & Fabricius, 2000). For this analysis, classification trees have the added advantage of being able to use a single variable at multiple levels in a tree. Since we sampled a broad range of lake types, we expected that conductivity and calcium would as reliable as saturation state at the tops of trees, where the broad range of calcification potential could be split into smaller groups. However, we also expected that

groups of lakes with calcification conditions near biological thresholds, which would occur further down the tree, may be better predicted by saturation state (Fig. 1c, d). In line with this idea, we expected that calcium and conductivity would perform similarly to saturation state in predicting presence/absence at the tops of trees across broad ranges of the variables, but that saturation state would enable more accurate predictions further down the trees (Fig. 1c, d). CART depicts these complex relationships in simple trees, making detecting and understanding them easier than in analogous methods like logistic regression.

We ran classification (presence/absence response) trees for each species. Independent variables were related to calcification (conductivity, calcium, and saturation state), boater access (public access presence, distance to road, distance to highway), and other lake characteristics (area, maximum depth, Secchi depth). We chose these variables because they account for the likelihood of initial introductions (boater access) and establishment (suitability associated with calcification and lake characteristics). Other than the highly correlated calcification variables, these variables are not collinear, with variance inflation factors all below 2.3. In each tree, we required a minimum of 5 lakes in each node to reduce spurious splits caused by low sample sizes. We created the initial trees to minimize misclassification rates and then pruned to prevent overfitting. We pruned to the final trees by running a ten-fold cross-validation, where 9/10 of the dataset were used to build a tree that made predictions for the remaining 1/10, and repeated for each fold. This cross-validation process allows estimation of the error associated with each potential split in the tree. Splits that have high error are then likely to be pruned back. We pruned trees to minimize the overall relative risk (defined as the sum of the relative error in the tree and a cost associated with having too many variables) (Therneau et al., 2014). However, two trees remained overfit after this process, and two were over-pruned so that they had no splits and just one node, due to low species prevalence. In these cases, we re-pruned the full trees by requiring any additional split in the model to account for a decreased misclassification rate of at least 4 (5%; for 3 trees), and 9 lakes (11%; 1 tree). For each species, we constructed trees with two different sets of independent variables. First, we constructed trees with one calcification variable each, and evaluated changes in

tree structure and performance. Second, we constructed trees using all variables in order to determine which calcification variables were selected in the best overall trees. To evaluate tree performance, we calculated Cohen's Kappa for each tree to measure the accuracy of the trees' predictions (Landis & Koch, 1977; Fielding & Bell, 1997; Mouton et al., 2010). However, since tree complexity was not consistent among the variables used in each tree, Kappa's could not be compared among trees to evaluate whether surrogates improved trees overall.

Instead, we calculated each variable's overall importance by including all variables into a single model for each species and evaluating their isolated effects on misclassification rates. At any split, we allowed the primary variable to be substituted out for another variable, and monitored the increase in misclassification rate. This substitution was done for each split in the tree with each variable. In this way, we quantified a variable's importance as the sum of its effects on the tree's total goodness of fit (number classified correctly) at all of the splits in which the variable was either used or substituted in for another variable (Breiman et al., 1984; Therneau et al., 2014). Importance values for a split are the increase in the number of lakes correctly predicted compared to predictions based on majority rule, where all lakes are given the majority response value. We use these importance values to evaluate how well each calcification variable performed in the trees, using both pruned and unpruned trees. All analyses were performed in R version 2.15.2 (R Development Core Team, 2012) using the rpart package (Therneau et al., 2014; R package version 4.1–4.8), which was shown to perform well for ecological predictions (Prasad et al., 2006).

Results

CMS were the most common of the three species—present in 39 of 79 lakes (49.4%)—followed by RC in 26 (32.9%), and BMS in 22 lakes (27.8%) (Fig. 2). Conductivity ranged from 12.2 to 163.2 $\mu\text{S}/\text{cm}$ (mean = 71.9; median = 73.3; SD = 42.4) (Fig. 3a). Dissolved calcium ranged from 0.5 to 29.4 mg/l (mean = 8.2; median = 8.2; SD = 6.0) (Fig. 3b). Calculated saturation states ranged from 2×10^{-6} to 15.33 (mean = 1.33, median = 0.19, SD = 25) (Fig. 3c). Twenty-two of the 79 sampled lakes had saturation states above

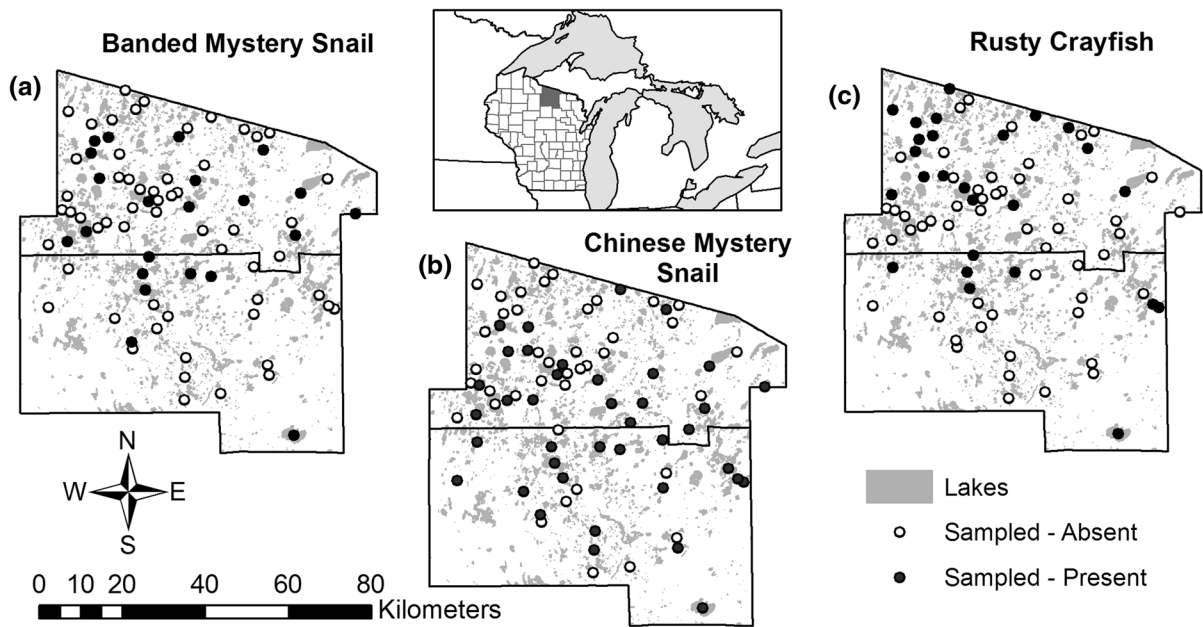


Fig. 2 Maps of surveyed lakes and the presence (*black points*) or absence (*white points*) of **a** banded mystery snail, **b** Chinese mystery snail, and **c** rusty crayfish in Vilas and Oneida counties

of Wisconsin, USA. *Inset* indicates the surveyed counties in dark gray within the Laurentian Great Lakes region, USA

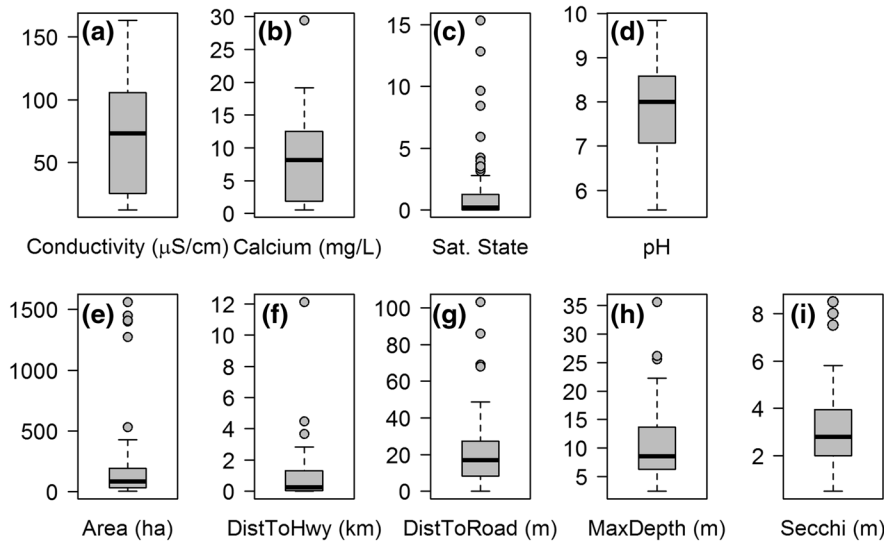


Fig. 3 Boxplots of calcification variables and others used in classification trees. The *first row* displays those variables that are related to calcification potential: **a** conductivity, **b** dissolved calcium, **c** saturation state, and **d** pH, which is kept out of the classification trees because it is used in calculating saturation

state. The *second row* shows other potentially important variables including: **e** lake area, **f** distance to highway, **g** distance to nearest road, **h** maximum depth, and **i** Secchi depth. Boxplots display minimum, 1st quartile, median, 3rd quartile, and maximum non-outlier observations

1.0 indicating potential for purely chemical precipitation of CaCO_3 . CMS were least limited by saturation state as they were present in lakes with

saturation states as low as 0.0002, while both RC and BMS were limited to lakes with saturation state above 0.04.

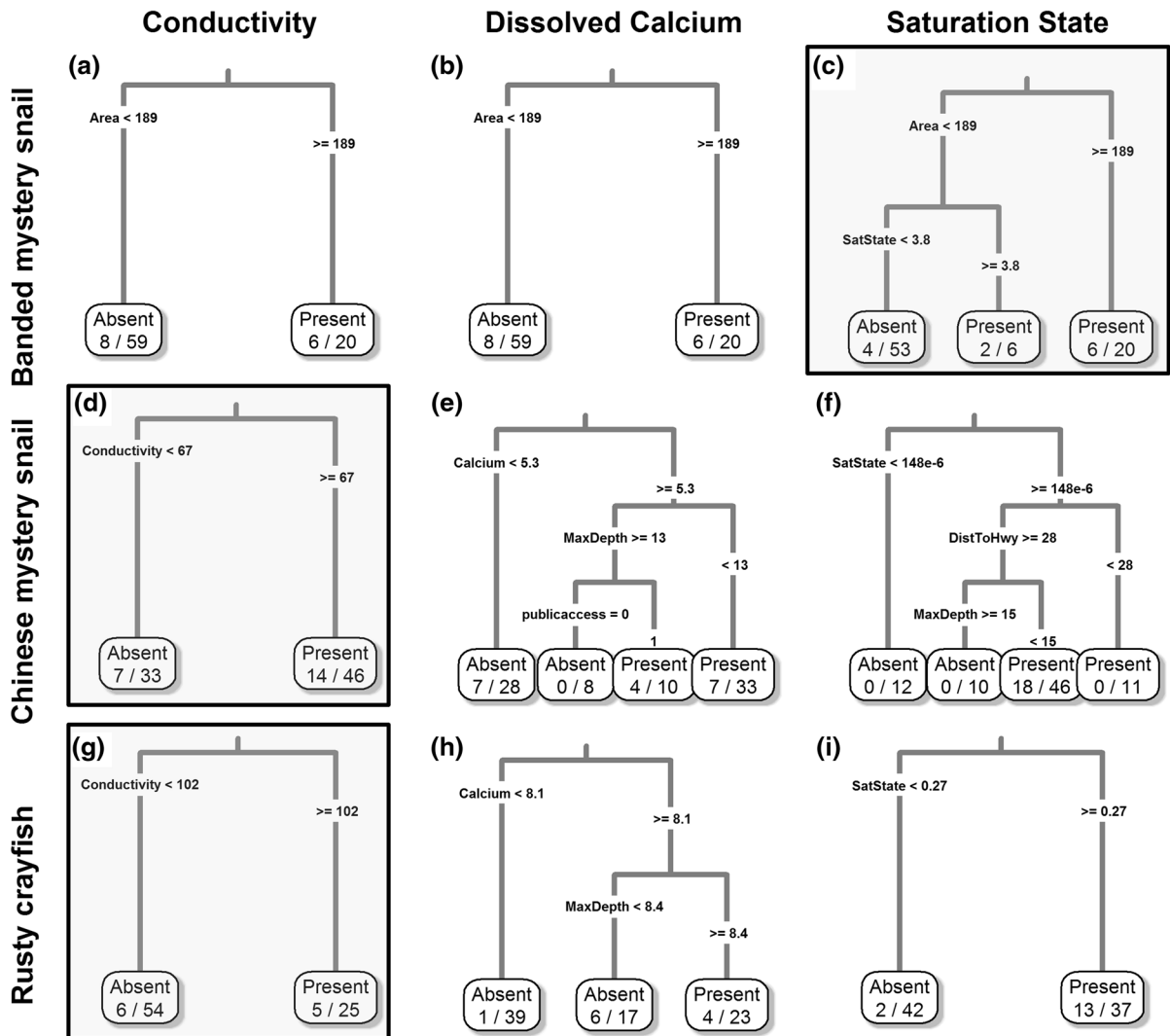


Fig. 4 Classification trees for each species (by row: **a–c** banded mystery snail, **d–f** Chinese mystery snail, **g–i** rusty crayfish) and for each calcification variable that was used (by column: **a, d, g**; conductivity; **b, e, h**; calcium; **c, f, h** saturation state). *Shaded boxes* indicate trees that were selected as the best overall trees

The classification trees were able to describe much of the variation in each species' distribution (Fig. 4). Minimum misclassification rates of the best pruned trees were 12/79 (15.2%) for BMS, 18/79 (22.8%) for CMS, and 11/79 (13.9%) for RC. Pruning resulted in trees that were fairly simple, as no tree had more than four nodes, and five had just two nodes (single split). However, additional gains in predictive accuracy in more complex trees were not large enough to justify additional splits. Two trees—those for CMS based on

when all calcification variables were included in model fitting, as well as for the corresponding calcification variable on its own. Tree structure follows that shown and described in Fig. 1 and units of measure are shown in Fig. 3

calcium and saturation state—did have more accurate, complex structures using our initial pruning technique. However, the additional splits were likely unreliable. In their lower branches, the CMS tree based on calcium predicted lakes with moderate calcium levels (1.2–1.6 mg/l) to be more likely to contain CMS than those with higher levels (1.6–5.3 mg/l). Similarly, the tree based on saturation state predicted lakes nearer to highways to be less likely to contain CMS. Since these results were likely cases of overfitting, we used

additional pruning to produce more reliable trees, each of which has four nodes. Unlike trees for CMS, those for BMS tended to be over-pruned: two of the trees from the initial pruning process had just a single node (no splits). In these cases, the secondary pruning method selected trees with just one split to generate predictions. The trees selected as best overall when all calcification variables were included were the saturation state model for BMS (Fig. 4c) and conductivity models for CMS (Fig. 4d) and RC (Fig. 4g). None of these models contained more than one calcification predictor.

The structures of the trees reveal how each variable affects the species' distribution. In each of the trees predicting BMS distributions (Fig. 4a–c), the first split was based on area, where lakes larger than 189 ha were predicted to contain BMS. For conductivity and calcium, area formed the only split ($Kappa = 0.55$ for each). When saturation state was included in the tree ($Kappa = 0.64$), it became an important variable to distinguish between BMS presence and absence in large lakes, where large lakes with saturation states below 3.8 were still predicted to be absent of BMS. However, saturation state's inclusion only led to two additional accurate predictions (Fig. 4a–c). For CMS (Fig. 4d–f), the first node in the trees was always the calcification variable. When conductivity was used, ($Kappa = 0.47$), it formed the only split with a threshold of 67 $\mu S/cm$ determining presence/absence. In those based on calcium ($Kappa = 0.54$) and saturation state ($Kappa = 0.55$), additional splits were made by maximum depth, the presence of public access, and distance to highway, where CMS presences tended to be associated with shallower, accessible lakes near highways, assuming they had sufficient calcification conditions. Like those of CMS, the trees for RC (Fig. 4g–i) also included calcification variables at their first nodes. Again, conductivity ($Kappa = 0.68$) formed the only split in its tree, where RC were predicted to be present in lakes above 102 $\mu S/cm$. Similarly, the saturation state tree for RC ($Kappa = 0.61$) had only one split, where RC were predicted to be present in lakes with saturation state values above 0.27. In the tree based on calcium ($Kappa = 0.68$), the calcium threshold was 8.1 mg/l, but predictions for high-calcium lakes also depended on maximum depth, where RC were associated with lakes deeper than 8.4 m.

The predictive abilities of each variable are made apparent in our analysis of variable importance.

Along with lake area, the calcification variables had the most predictive power, although less so for BMS (Fig. 5). Conductivity was the most important predictor for CMS and RC in pruned trees (Fig. 5b), although calcium was slightly better for CMS in unpruned trees (Fig. 5a). Area was the best predictor for BMS in both pruned and unpruned trees (Fig. 5). Saturation state was consistently important, but was never the best predictor (Fig. 5). This general result was consistent in both unpruned (Fig. 5a) and pruned trees (Fig. 5b).

Discussion

Classification trees for BMS, CMS, and RC all performed well in predicting species occurrence, with all Kappa's falling between 0.47 and 0.68, where models with Kappa's below 0.4 are considered poor, between 0.4 and 0.75 are good, and above 0.75 are excellent (Landis & Koch, 1977). Although each species exhibited distributions that were described in part by calcification variables, our expectation that saturation state would improve classification trees was not supported. Instead, the best overall trees identified by misclassification rate and Kappa used saturation state only for BMS and instead used conductivity for CMS and RC (Fig. 3). In trees where all calcification variables were included, our results analyzing variable importance demonstrated that conductivity and calcium were at least as good as saturation state at predicting species occurrence (Fig. 5). Due to the heterogeneity in landscape position in our study region that drives heterogeneity in calcification conditions (Kratz et al., 1997), the importance of calcification variables in the models is not surprising. In a different region where habitat suitability is more limited by other variables, such as dissolved oxygen or temperature, these would likely be more important in classification trees. We also found the same general outcome using multiple logistic regressions (results not shown), but CART allowed for easy visual evaluation of our expectation that predictions for lakes with near-threshold conditions would improve when saturation state was used as a variable (Fig. 1).

We predicted that saturation state would improve the accuracy of distribution models for lakes with near-threshold conditions, since extremely high or low values of calcification would be adequately

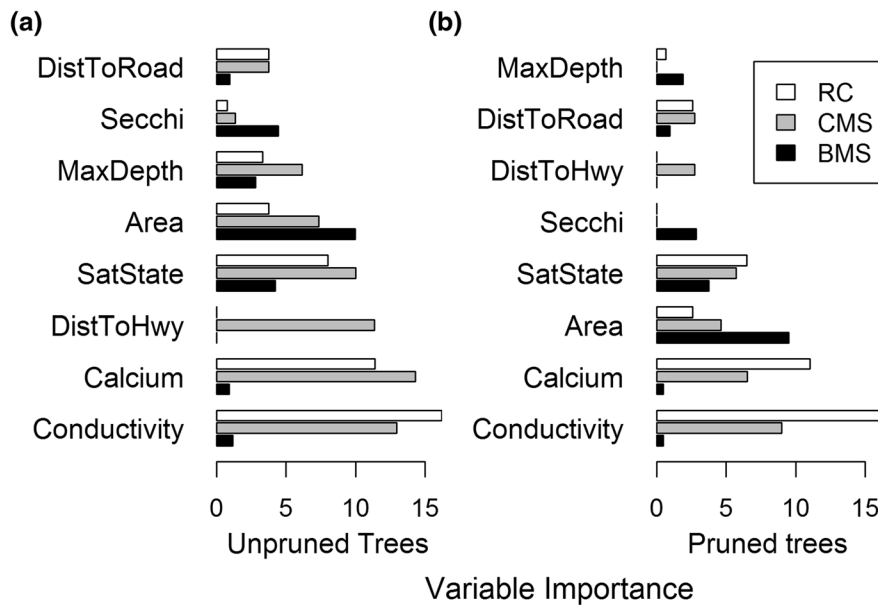


Fig. 5 Calculated importance for each variable used in each **a** unpruned and **b** pruned classification trees, the latter of which are displayed in Fig. 4. A variable's importance is the sum of the decrease in error (number misclassified) using the variable either as a primary node, or as a substitute for the primary

splitting variable. Values of variable importance can be thought of as the number of lakes correctly predicted by the model due specifically to that variable, where the improvement at a split is measured against assigning all lakes the majority response value

represented by surrogates (Fig. 1c, d). Calcification values formed the first node of classification trees for CMS and RC but not for BMS, which was better predicted by lake area. However, nodes further down the trees were rarely used to describe near-threshold calcification conditions (Fig. 4). The one exception was for BMS, whose classification tree had area as the initial node with a threshold of 189 ha. Lakes below that threshold were then predicted by saturation state, where lakes below 3.8 were typically absent of BMS. However, this tree only predicted two more lakes correctly than those based on area alone. For CMS and RC, calcification variables were only used at the first nodes, indicating that their distributions were best described by the wide calcification gradient alone, and that using saturation state to make predictions for lakes with near-threshold conditions did not improve the models.

Our results may have been influenced by our snapshots of lake conditions, which may have disregarded the much more variable environment that the organisms actually experience. We made all of our measurements during midday in the summer. However, seasonal changes in temperature and

metabolism-driven diel changes in pH likely cause changes in saturation state that are independent of changes in conductivity and dissolved calcium. Additionally, changes in precipitation and relative groundwater contribution to these lakes could cause all variables to vary by changing the delivery of weathering-derived bicarbonate, calcium and other solutes (Webster et al., 2000). These dynamics cause carbonate and pH to vary at multiple scales—diel, seasonal, annual, and decadal—simultaneously (Hanson et al., 2006). Thus, saturation state likely follows similar complex dynamics, and may not be adequately represented by a one-time measurement. Populations may be linked more closely to seasonal minima/maxima or average conditions than to snapshots. Furthermore, this one-time measurement is unlikely to be perfectly accurate if there was error in our estimates of pH or DIC, both of which are included in the calculation of saturation state. These sources of variability and uncertainty in our estimate of saturation state may partially explain why species occurrences were often better predicted by conductivity or calcium, but nonetheless do not change our conclusion that these surrogates adequately replace

our estimates of saturation state in distribution models.

In addition to influencing species distributions, the calcification conditions of a lake influence organismal and population-level processes. Crawling speed and aerial respiration of snails decrease when the potential for calcification is low, while cutaneous respiration increases (Dalesman & Lukowiak, 2010). Shell strength of snails and mussels is compromised at low levels (Czarneński et al., 2006; Glass & Darby, 2009). High shell strength deters predation from fish and crayfish (Lewis & Magnuson, 1999; Czarneński et al., 2006), so poor calcification conditions may also enable higher predation rates or altered antipredator behaviors, which was shown for RC at low calcium conditions (Edwards et al., 2013). These organismal and population-level responses to low calcification potential are likely to be more sensitive to minute differences in saturation state than species distributions, which are controlled by the integration of variable abiotic and biotic conditions and, for introduced species, the propagule pressure or colonization potential. We did not test for organismal or population-level effects, but may have been more likely to see different responses to our calcification variables than we did using species distributions.

Although saturation state was not the best predictor, it was selected in classification trees that achieved low misclassification rates and high Kappa's. While thresholds of conductivity and dissolved calcium are often determined experimentally (i.e., what is the lowest calcium level in which a species can survive?) or through distribution models (i.e., at what level of calcium is there a significant change in the probability a species will be present?), thresholds for saturation state are unknown for most freshwater species. Our results allow us to infer distribution-based thresholds, but not the true biological thresholds. Since other collinear variables may limit a species' presence at any one location, the true suitability threshold may not be revealed in a species distribution. In classification trees, saturation state formed thresholds at values of 3.8 for BMS (after accounting for lake area), 0.00015 for CMS, and 0.27 for RC (Fig. 4). These values are likely higher than their physiological thresholds because lakes with low saturation state that may be suitable are also smaller and less likely to receive propagules. Similarly, invasion of the golden mussel in Brazilian rivers is considered low-risk only where

saturation state is below 0.0001 (Oliveira et al., 2010). To calculate saturation states, we used K_{sp} 's for the calcite mineral phase, whereas saturation states for aragonite under the same conditions are lower. These results may indicate that the actual suitability thresholds for each species may be below 1.0, indicating that they may maintain CaCO_3 exoskeletons even when chemical conditions may favor dissolution. Indeed, all three species did occur in lakes with saturation states below 1.0 (minima of 0.00015 for CMS, and 0.04 for BMS and RC).

Suitability thresholds of saturation state below 1.0 for calcifying organisms are not unprecedented. Corals are able to up-regulate pH, which effectively increases saturation state at the site of calcification (McCulloch et al., 2012). Crayfish may be able to construct hard carapaces by forming gastroliths—extracellular calcium carbonate storage deposits—that store amorphous calcium carbonate in which the calcium is readily available prior to molting (Akiva-Tal et al., 2011). Additionally, the enzyme carbonic anhydrase has been demonstrated to be important in CaCO_3 formation in many taxa including mollusks, crustaceans, birds, and others (Common, 1941; Costlow, 1959; Freeman, 1960; Giraud, 1981; Henry & Kormanik, 1985). Carbonic anhydrase increases the supply of bicarbonate and carbonate ions (Horne et al., 2002; Maren, 2006) and mobilizes CaCO_3 reserves during calcification (Istin & Girard, 1970; Henry & Kormanik, 1985), allowing for mineralization even when ambient carbonate supplies are low (Horne et al., 2002). However, maintenance and construction of CaCO_3 in sub-optimal conditions may be energetically costly, so populations may persist but be impacted in other ways, as discussed above. By focusing on species presence and absence, our study ignored other potential impacts of low saturation states on individuals and populations. Future research could test whether populations are affected by measuring exoskeleton strength or structure, mortality due to predation, and behavioral traits like crawling speed across a range of saturation states.

A core objective of this work was to determine whether biologically mechanistic variables can improve species distribution models that traditionally use surrogate variables. Surrogates may often be used in models because they are more easily measurable and thus available for a higher number of sites than mechanistic variables. Data on many environmental

variables at broad spatial scales are becoming increasingly easy to access (via remote sensing for instance). Data availability for a higher number of sites and over longer time periods enables models with better spatial and temporal coverage, but these may come at the cost of accuracy at specific sites or times if surrogate variables are used. In our study, surrogate measures of the potential for calcification were at least as powerful in distribution models as the mechanistic variable they replaced. In fact, conductivity may be the best variable related to calcification to use in distribution models across lakes with a broad range of calcification potential because it is the most easily measured. However, comparing across narrower ranges of calcification potential or using alternative responses such as species abundance may lead to different conclusions about the effectiveness of surrogates. Similarly, not all mechanistic variables may be as reliably replaced by surrogates, so their use in distribution models ought to be continually evaluated.

Acknowledgments This material is based upon the work supported by the National Science Foundation under #CNH-0909281 with additional support from the North Temperate Lakes Long-Term Ecological Research program #DEB-0822700 and a traineeship provided to A. Latzka from NSF-IGERT award DGE-1144752: Novel ecosystems, rapid change, and no-analog conditions: the future of biodiversity conservation in human-dominated landscapes. We thank A. Krause, S. Neymark, J. Trcka, and J. Skaggs, as well as staff of UW Trout Lake Station for their assistance with fieldwork and logistics. We also thank E. Stanley for logistical support and helpful thoughts in the study design. We are grateful for helpful feedback from J. Havel and 2 anonymous reviewers.

References

- Akiva-Tal, A., S. Kababya, Y. S. Balazs, L. Glazer, A. Berman, A. Sagi & A. Schmidt, 2011. In situ molecular NMR picture of bioavailable calcium stabilized as amorphous CaCO_3 biomineral in crayfish gastroliths. *Proceedings of the National Academy of Sciences* 108: 14763–14768.
- Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling* 200: 1–19.
- Boets, P., K. Lock & P. L. M. Goethals, 2013. Modelling habitat preference, abundance and species richness of alien macrocrustaceans in surface waters in Flanders (Belgium) using decision trees. *Ecological Informatics* 17: 73–81.
- Breiman, L., J. Friedman, C. J. Stone & R. A. Olshen, 1984. *Classification and Regression Trees*. Chapman and Hall/CRC, New York, NY.
- Butler, J. N., 1991. *Carbon Dioxide Equilibria and Their Applications*. CRC Press, Reading, MA.
- Capelli, G. M. & J. J. Magnuson, 1983. Morphoedaphic and biogeographic analysis of crayfish distribution in northern Wisconsin. *Journal of Crustacean Biology* 3: 548–564.
- Common, R. H., 1941. The carbonic anhydrase activity of the hen's oviduct. *The Journal of Agricultural Science* 31: 412–414.
- Costlow, Jr., J. D., 1959. Effect of carbonic anhydrase inhibitors on shell development and growth of *Balanus improvisus* Darwin. *Physiological Zoology* 32: 177–184.
- Czarneński, M., J. Kozłowski, P. Kubajak, K. Lewandowski, T. Müller, A. Stańczykowska & K. Surówka, 2006. Cross-habitat differences in crush resistance and growth pattern of zebra mussels (*Dreissena polymorpha*): effects of calcium availability and predator pressure. *Archiv für Hydrobiologie* 165: 191–208.
- Dalesman, S. & K. Lukowiak, 2010. Effect of acute exposure to low environmental calcium on respiration and locomotion in *Lymnaea stagnalis* (L.). *Journal of Experimental Biology* 213: 1471–1476.
- De'ath, G. & K. E. Fabricius, 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.
- Doney, S. C., V. J. Fabry, R. A. Feely & J. A. Kleypas, 2009. Ocean acidification: the other CO_2 problem. *Annual Review of Marine Science* 1: 169–192.
- Edwards, B. A., V. R. E. Lewis, F. H. Rodd & D. A. Jackson, 2013. Interactive effects of calcium decline and predation risk on the potential for a continuing northward range expansion of the rusty crayfish (*Orconectes rusticus*). *Canadian Journal of Zoology* 91: 328–337.
- Fielding, A. H. & J. F. Bell, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38–49.
- Freeman, J. A., 1960. Influence of carbonic anhydrase inhibitors on shell growth of a fresh-water snail, *Physa heterostropha*. *Biological Bulletin* 118: 412–418.
- Girard, M.-M., 1981. Carbonic anhydrase activity in the integument of the crab *Carcinus maenas* during the intermolt cycle. *Comparative Biochemistry and Physiology* 69A: 381–387.
- Glass, N. & P. Darby, 2009. The effect of calcium and pH on Florida apple snail, *Pomacea paludosa* (Gastropoda: Ampullariidae), shell growth and crush weight. *Aquatic Ecology* 43: 1085–1093.
- Hammond, K. S., J. W. Hollows, C. R. Townsend & P. M. Lokman, 2006. Effects of temperature and water calcium concentration on growth, survival and moulting of freshwater crayfish, *Paraneohrops zealandicus*. *Aquaculture* 251: 271–279.
- Hansen, G. J. A., C. L. Hein, B. M. Roth, M. J. Vander Zanden, J. W. Gaeta, A. W. Latzka & S. R. Carpenter, 2013. Food web consequences of long-term invasive crayfish control. *Canadian Journal of Fisheries and Aquatic Sciences* 70: 1109–1122.
- Hanson, P. C., S. R. Carpenter, D. E. Armstrong, E. H. Stanley & T. K. Kratz, 2006. Lake dissolved inorganic carbon and dissolved oxygen: changing drivers from days to decades. *Ecological Monographs* 76: 343–363.
- Henry, R. P. & G. A. Kormanik, 1985. Carbonic anhydrase activity and calcium deposition during the molt cycle of the

- blue crab *Callinectes sapidus*. Journal of Crustacean Biology 5: 234–241.
- Herbst, D. B., M. T. Bogan & R. A. Lusardi, 2008. Low specific conductivity limits growth and survival of the New Zealand mud snail from the Upper Owens River, California. Western North American Naturalist 68: 324–333.
- Hoegh-Guldberg, O., P. J. Mumby, A. J. Hooten, R. S. Steneck, P. Greenfield, E. Gomez, C. D. Harvell, P. F. Sale, A. J. Edwards, K. Caldeira, N. Knowlton, C. M. Eakin, R. Iglesias-Prieto, N. Muthiga, R. H. Bradbury, A. Dubi & M. E. Hatzios, 2007. Coral reefs under rapid climate change and ocean acidification. Science 318: 1737–1742.
- Horne, F., S. Tarsitano & K. L. Lavalli, 2002. Carbonic anhydrase in mineralization of the crayfish cuticle. Crustaceana 75: 1067–1081.
- Horne, F., S. Tarsitano & K. L. Lavalli, 2009. Aspects of mineralization of the cuticle of the crayfish, *Procambarus clarkii* (Decapoda, Cambaridae). Crustaceana 82: 1057–1065.
- Istin, M. & J. P. Girard, 1970. Carbonic anhydrase and mobilisation of calcium reserves in the mantle of lamelli-branchs. Calcified Tissue Research 5: 247–260.
- Johnson, P. T. J., J. D. Olden, C. T. Solomon & M. J. Vander Zanden, 2008. Interactions among invaders: community and ecosystem effects of multiple invasive species in an experimental aquatic system. Oecologia 159: 161–170.
- Kratz, T. K., K. E. Webster, C. J. Bowser, J. J. Magnuson & B. J. Benson, 1997. The influence of landscape position on lakes in northern Wisconsin. Freshwater Biology 37: 209–217.
- Landis, J. R. & G. G. Koch, 1977. The measurement of observer agreement for categorical data. Biometrics 33: 159–174.
- Langdon, C., T. Takahashi, C. Sweeney, D. Chipman, J. Goddard, F. Marubini, H. Aceves, H. Barnett & M. J. Atkinson, 2000. Effect of calcium carbonate saturation state on the calcification rate of an experimental coral reef. Global Biogeochemical Cycles 14: 639–654.
- Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis & G. Lamberti, 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society B: Biological Sciences 269: 2407–2413.
- Lewis, D. B. & J. J. Magnuson, 1999. Intraspecific gastropod shell strength variation among north temperate lakes. Canadian Journal of Fisheries and Aquatic Sciences 56: 1687–1695.
- Lodge, D. M., M. W. Kershner, J. E. Aloï & A. P. Covich, 1994. Effects of an omnivorous crayfish (*Orconectes rusticus*) on a freshwater littoral food web. Ecology 75: 1265–1281.
- Maren, T. H., 2006. The general physiology of reactions catalyzed by carbonic anhydrase and their inhibition by sulfonamides. Annals of the New York Academy of Sciences 429: 568–579.
- Marin, F., N. Le Roy & B. Marie, 2012. The formation and mineralization of mollusc shell. Frontiers in Bioscience 4: 1099–1125.
- McCulloch, M., J. Falter, J. Trotter & P. Montagna, 2012. Coral resilience to ocean acidification and global warming through pH up-regulation. Nature Climate Change 2: 623–627.
- Mouton, A. M., B. De Baets & P. L. M. Goethals, 2010. Ecological relevance of performance criteria for species distribution models. Ecological Modelling 221: 1995–2002.
- Olden, J. D., J. M. McCarthy, J. T. Maxted, W. W. Fetzer & M. J. Vander Zanden, 2006. The rapid spread of rusty crayfish (*Orconectes rusticus*) with observations on native crayfish declines in Wisconsin (USA) over the past 130 years. Biological Invasions 8: 1621–1628.
- Oliveira, M., S. Hamilton & C. Jacobi, 2010. Forecasting the expansion of the invasive golden mussel *Limnoperna fortunei* in Brazilian and North American rivers based on its occurrence in the Paraguay River and Pantanal wetland of Brazil. Aquatic Invasions 5: 59–73.
- Pejchar, L. & H. A. Mooney, 2009. Invasive species, ecosystem services and human well-being. Trends in Ecology & Evolution 24: 497–504.
- Pimentel, D., R. Zuniga & D. Morrison, 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52: 273–288.
- Prasad, A. M., L. R. Iverson & A. Liaw, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9: 181–199.
- R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna [http://www.R-project.org].
- Ramcharan, C. W., D. K. Padilla & S. I. Dodson, 1992. Models to predict potential occurrence and density of the zebra mussel, *Dreissena polymorpha*. Canadian Journal of Fisheries and Aquatic Sciences 49: 2611–2620.
- Roer, R. & R. Dillaman, 1984. The structure and calcification of the crustacean cuticle. American Zoologist 24: 893–909.
- Rothlisberger, J. D., W. L. Chadderton, J. McNulty & D. M. Lodge, 2010. Aquatic invasive species transport via trailered boats: what is being moved, who is moving it, and what can be done. Fisheries 35: 121–132.
- Sala, O. E., F. S. Chapin, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker & D. H. Wall, 2000. Global biodiversity scenarios for the year 2100. Science 287: 1770–1774.
- Smith, D., 2000. Notes on the taxonomy of introduced *Bellamya* (Gastropoda: Viviparidae) species in northeastern North America. Nautilus-Sanibel 114(2): 31–37.
- Solomon, C. T., J. D. Olden, P. T. J. Johnson, R. T. Dillon & M. J. Zanden, 2009. Distribution and community-level effects of the Chinese mystery snail (*Bellamya chinensis*) in northern Wisconsin lakes. Biological Invasions 12: 1591–1605.
- Therneau, T., B. Atkinson, & R. Ripley, 2014. rpart: Recursive Partitioning and Regression Trees [http://cran.r-project.org/web/packages/rpart/index.html] Accessed date: 1 July 2014.
- Vander Zanden, M. J. & J. D. Olden, 2008. A management framework for preventing the secondary spread of aquatic invasive species. Canadian Journal of Fisheries and Aquatic Sciences 65: 1512–1522.
- Vitousek, P. M., 1990. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. Oikos 57: 7–13.
- Webster, K. E., P. A. Soranno, S. B. Baines, T. K. Kratz, C. J. Bowser, P. J. Dillon, P. Campbell, E. J. Fee & R. E. Hecky, 2000. Structuring features of lake districts: landscape controls on lake chemical responses to drought. Freshwater Biology 43: 499–515.