Landscape-scale patterns in aquatic invasions:
Prevalence, colonization, establishment, and impacts

By

Alexander W. Latzka

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Freshwater and Marine Sciences)

at the
UNIVERSITY OF WISCONSIN-MADISON
2015

Date of Final Oral Examination: 5/13/2015

Committee members:

- M. Jake Vander Zanden, Professor, Zoology
- Stephen R. Carpenter, Professor, Zoology
- Robert W. Provencher, Professor, Applied Economics and Management
- Monica G. Turner, Professor, Zoology
- Ellen I. Damschen, Assistant Professor, Zoology

Table of Contents

ACKNOWLEDGMENTS

First and foremost, I am delighted to thank my advisor, Dr. Jake Vander Zanden, for his guidance and training throughout this long, crazy process. Looking back, it's clear I was frequently tantalized by some new question or opportunity, and how frustrating that must have been for you at times. But you granted me enough freedom, and more importantly, put up the bumpers when you could tell I was about to roll a gutter-ball, to get me to this point. And, most importantly, you made these years enjoyable. Without that, it would have been impossible to get here. So, thank you for everything you put into this. I'll be grateful my entire career.

There have been many others inside the Hasler Lab that have also played huge roles in the development of this thesis and of myself over the last few years. Dan Oele, John Crawford, and Ryan Batt: it all got started on Drake St. In between basement sessions, dance parties, and avoiding the music emanating from Ryan's room, we were able to figure out what it meant to be grad students at the CFL. Ryan, you've become an awesome scientist, and I'm always excited to hear your thoughts on my work. John, you're a force. I'm never going to keep up with you, but it sets the bar high. Thanks for getting us going on the CaCO3 paper – it didn't work out exactly as we planned, but at least we got something out of it. Dan, thanks for being the voice of sanity in the 118 and bringing it at Sapelo. Only many sandwiches could fuel that fieldwork. Next, the many residents of the 118: Steve Powers, Jereme Gaeta, Dan Oele, Evan Childress, Luke Loken, Etienne Fluet-Chouinard. So many good times, from hip-hop Fridays at the beginning all the way to month-long pull-up challenges in the approach to my defense. The 118 was always a great place for a quick science (or darts) session, and I'm sure without you guys, I'd still be banging my head on my keyboard over some R code. JVZ lab, early years: Gretchen Hansen and Matt

Kornis – you two were my big sister and brother in AIS work at the CFL for awhile. I look up to you both, and am excited that we’ve continued to work together up to this point. I’ve learned a lot from you both already, and can’t wait to see where both of you end up. JVZ lab, mid and late years: Jake, Ali, Chelsey, Colin, and many many postdocs (but, I’ll give a quick shout out to Mona, Mireia, Jereme, Tyler, and Eric) – thanks for all the great science discussions, feedback, and direction.

I don’t think I would have survived through this without some semblance of balance outside of work, so I’d like to thank everyone that enabled such habits. First, Dan, Jereme, Zach: whenever someone needed a reset (or alternatively, on a cloudy Saturday), the riv or the lakes were there (but not necessarily always muskies or monster trout) for that much needed reboot. Thanks for bringing me out, showing me spots, and most importantly, netting my fish (or, hand-landing it and tossing it into the snowbank, when sans net). I would say I’m going to miss those times, but I know I’ll be out on the water with all of you guys again before too long. Next, the Bottomfeeders (too many to name, but core members being Dan & Jess, Sam & John): I didn’t realize it until a few years in, but getting back into sports was huge for me. I had such a blast dominating recreational and intermediate level flag football (and scraping by in competitive too, I guess), and it was always good to get out of the office and break a sweat. For those of you who stay in Madison, I trust that the multi-sport dynasty of the Bottomfeeders will not end anytime soon. Be sure to pass it down. Lastly, Jami: thank you for putting up with me through the end of this. Dealing with my high stress and long hours probably wasn’t easy for you, but I know you made it easier for me, and I’ll be eternally grateful for your spirit and your contributions. Looking forward to what’s next ☺

Now, I recognize I never would have made it to the CFL, let alone to finishing my thesis, without those of you who made me who I am today. Mom & Brian, Dad & K: thank you for raising me to believe in hard work and to value intelligence. Thank you for making it okay to pursue a passion instead of a paycheck. Thank you for supporting me. Erek, Mariah, Jill, Brian: thanks for the constant competitiveness. I know it’s helped us all succeed in our own ways. One point of pride I hold for our family is that we not only all made it out of Hudson, but seemingly spread equidistantly around the continent. Erek especially, as my older brother – I look up to you, and you continue to set the bar high. I can’t imagine what it would have been like being the oldest, or having an older brother who didn’t make me want to succeed like you do. I’m thankful I don’t have to know what that’s like. I love you all, even though I doubt you’ll read much beyond this page.

So, thank you everyone for all of your knowledge, guidance, encouragement, prodding, interest, patience, and support. You’ve made this thesis possible, and the process immensely enjoyable. On Wisconsin.

ABSTRACT

Species invasions are a leading driver of biological change. They cause drastic ecological and economic impacts across the globe. Freshwaters are especially vulnerable to these negative effects. Due to the irreversibility and high cost associated with controlling established invaders, prevention is the key to management. The goal of prevention of invasive species, on landscapes where the species has already arrived, is to disrupt the species' secondary spread. Secondary spread consists of colonization in new habitats, establishment, and impact. In this thesis, I evaluate aquatic invasive species with respect to these three stages. In Chapter 1, I investigate the extent to which aquatic invasions are linked to human activity variables including lake distance to road and shoreline residential development. Using a database of over 400 lakes across Wisconsin that spans gradients in both anthropogenic variables, we found high correlations between human activity and the number of species present, suggesting that prevention efforts targeting human pathways are worthwhile. In Chapter 2, I evaluate predictors of establishment potential for invasive snails and crayfish, all of which need sufficient chemical conditions to support calcium carbonate exoskeletons. We found that a surrogate predictor – conductivity, which is more easily measurable and widely available – was a sufficient replacement for biologically mechanistic variables in distribution models. In Chapter 3, I estimated the prevalence of invasive species across Wisconsin, and found that existing species occurrence records drastically underrepresent invasive species distributions. In Chapter 4, I use simulations and empirical data to estimate how the impacts of invasive species may be spread across landscapes. Estimates of impact are rarely taken into account in management decisions, and the variety of distributions that is possible suggests that such information may be valuable. Each of

these chapters seeks to answer a question or test a hypothesis about how invasions play out at landscape scales, and provides guidance for future invasive species research and management.

INTRODUCTION

Recent rapid and pervasive ecological change, driven by agricultural, industrial, technological and population growth by humans, has ushered our planet into a new geological epoch – the Anthropocene, in which the effects of humans on the planet have become the dominant force shaping Earth's geological and biological characteristics (Crutzen 2006, Steffen et al. 2007). The scope and degree of changes induced by humans have created ecological conditions that may have no prior analog for comparison (Williams and Jackson 2007, Radeloff et al. *in review – coauthored paper*). Such novel ecosystems stretch across the globe, and may hamper ecologists' abilities to predict changes, while also forcing policymakers and managers to abandon past approaches, such as using historical baselines as goals for ecological management, in favor of radically different ones better suited to the new status quo (Radeloff et al. *in review - coauthored paper*). A primary driver of this novelty is the human-mediated introduction of species to environments outside of their native ranges (Lugo et al. 2012, Traveset et al. 2013).

Although species have always dispersed to new environments, current rates of species introductions between continents are several orders of magnitude higher than those in previous geological eras (Ricciardi 2007). This rapid increase in the rate of biotic exchange was driven by a global transportation network that has linked countries and continents that would otherwise be biologically distinct (Hulme 2009). As a result, at global and regional scales, biota are becoming more and more homogenized (Olden and Poff 2003, Olden 2006), which threatens the sustained conservation of genetic, taxonomic, and functional biodiversity (Olden et al. 2004). Of the species that are introduced, only a fraction become invasive (Williamson and Fitter 1996, Jeschke and Strayer 2005). When they do become invasive, there may be negative consequences, and they may be substantial (Simberloff 2011). Invasive species can compete with (Callaway and

Aschehoug 2000, Callaway and Ridenour 2004), prey upon (Cox and Lima 2006), parasitize (Hatcher et al. 2012), and otherwise negatively affect native species, sometimes putting them at risk of extirpation, especially in insular habitats like oceanic islands (Sax and Gaines 2008). Thus, invasive species are one of the leading drivers of global biodiversity loss (Vitousek et al. 1996, Sala et al. 2000). Invasive species can function as ecosystem engineers that alter ecosystem structure and shift ecosystem processes (Vitousek and Walker 1989, Crooks 2002). Such changes can disrupt the delivery of ecosystem services to humans (Charles and Dukes 2007, Pejchar and Mooney 2009), leading to losses and damages costing up to $120 billion per year in the United States alone (Pimentel et al. 2005). Moreover, species introductions and the ecological changes they induce are exceedingly difficult to reverse (Mack and Lonsdale 2002). Thus, the prevention of new invasions is the most cost-effective method of invasive species management (Leung et al. 2002).

Freshwaters may be particularly vulnerable to the effects of species invasions (Sala et al. 2000, Moorhouse and Macdonald 2015), and so preventing invasions of freshwaters may be particularly important. Additionally, freshwaters provide provisioning, regulating, and cultural ecosystem services (Postel and Carpenter 1997). In areas where freshwaters make up a dominant feature of the landscape, they may be a crucial component of regional economies and cultural identities. Such areas, like the Northern Highlands Lake District of Wisconsin, where some of this work was carried out, may be vulnerable to aquatic invasive species (Peterson et al. 2003). For instance, invasive rainbow smelt can reduce recruitment of walleye, a popular gamefish in the region (Mercado-Silva et al. 2007), and the presence of an invasive macrophyte on lakes in the region is associated with decreases in home values of up to 15% (Horsch and Lewis 2009, Provencher et al. 2012). Prevention of new invasions in these regions requires stopping the secondary spread of aquatic invasive species that have already established in the region (Vander Zanden and Olden 2008). Many aquatic invasive species in Wisconsin were first introduced to the region in the Great Lakes via canal openings (sea lamprey) and ballast water (round goby, zebra mussel, quagga mussel), or via the aquarium (several invasive macrophytes), aquaculture (common carp, Asian carp), and culinary (Chinese mystery snail) trades. Once these species have arrived in the region, they may spread to new lakes, often via hitchhiking on boats (Johnson et al. 2001, Rothlisberger et al. 2010) or by transporting angling bait to new locations (Litvak and Mandrak 1993, DiStefano et al. 2009), although natural dispersal via surface water connections (Havel and Shurin 2004) or dispersal via waterfowl are also possible (Green and Figuerola 2005), the latter of which has been observed at least since Darwin first documented this occurrence in The Origin of Species. The goal of prevention is to interrupt this process of secondary spread.

Secondary spread involves three stages, or filters: colonization, establishment, and impact (Vander Zanden and Olden 2008). For an invasive species to colonize a new lake, it must disperse to or be brought to a new environment by a vector, which may or may not be a human pathway. In **Chapter 1** of this dissertation, we evaluate the extent to which invasions of Wisconsin lakes are associated with human activities, and whether lakes that receive very few human visitors are uninvaded. We use lakes' distance to their nearest roads and the degree of lakeshore development to represent human accessibility and activity on lakes. We show that lakes that are far from roads serve as a refuge from invasive species, lakes that are close to roads may become invaded, and lakes that are close to roads with high levels of development are likely to be invaded by multiple species. Even though waterfowl may act as an occasional disperser of aquatic invasive species, these results suggest that humans are the primary and most important vector, and that prevention efforts ought to continue targeting human pathways.

Once an invasive species arrives at a new location, it may or may not establish a population. Establishment is a function of the propagule size (are enough individuals introduced at the same time to overcome Allee effects?) (Drake and Lodge 2006) and environmental matching (does the environment match the species' niche?) (Thuiller et al. 2005). In order to effectively prevent new invasions on a landscape with thousands of lakes, we need to identify and target prevention efforts on the subset of lakes that are environmentally suitable. Thus, we require predictions of environmental suitability based on empirical measurements of abiotic conditions. However, we often use surrogate variables to make these predictions, rather than variables that are closely associated with the biological mechanisms that determine whether a population can establish or not. In **Chapter 2**, we test whether mechanistic variables can improve model predictions from those based on surrogates. Many aquatic invasive species in Wisconsin, such as rusty crayfish, Chinese mystery snails, and banded mystery snails, need water conditions that favor the construction and maintenance of calcium carbonate exoskeletons. The ability to form these exoskeletons is dictated by the water's calcium carbonate saturation state, a variable used widely in ocean acidification research to forecast declines of corals and marine bivalves. However, calcium carbonate saturation state is difficult to measure, so data exist for relatively few lakes. In lieu of saturation state, freshwater ecologists use correlated surrogates like calcium and conductivity to predict habitat suitability. We built classification trees predicting lake suitability for calcifying invasive species, and substitute surrogate variables for saturation state to test whether model accuracy suffers. Our results show that surrogates do an adequate job of representing saturation state in models.

The outcome of colonizing and establishing in new sites is a landscape that is dotted with invasive species in some proportion of sites. We can define this proportion of sites that is invaded as the invasive species prevalence. Despite prevalence being a rudimentary and fundamental measure of the scale and impact of invasions, few studies have explicitly measured it for invasive species. Moreover, existing datasets of species occurrence may underrepresent certain habitats and species (Gardiner et al. 2012). Indeed, many invasive species datasets have poor taxonomic and spatial coverage (Crall et al. 2006). In **Chapter 3**, we test whether the prevalence of invasive species in lakes in Wisconsin is adequately represented by existing databases of invasive species occurrences. Although the Wisconsin Department of Natural Resources occurrence dataset contained more entries for most species than publicly accessible datasets, it was still under-representative of overall prevalence, which we calculated from a set of stratified surveys. Occurrence records indicate a statewide prevalence of 11.9%, but stratified surveys lead to a much higher estimate of prevalence of 39.0%. We saw similar differences for species-specific prevalences, although the disagreement was smallest for commonly-known species like zebra mussels and largest for less-known species like Chinese mystery snail. These results suggest that the prevalence of invasive species on our landscapes may be much higher than might be imagined, which stresses that the overall impacts of invasive species are quite substantial.

Predicting colonization and establishment is common in invasive species research, and can help management identify sites that should be considered in a management plan. However, not all invasions result in negative consequences, and the lakes that are vulnerable to impacts should be those that are the highest management priority. Unfortunately, defining and measuring impacts is complex (Jeschke et al. 2014), and models predicting impacts are rare and often unsuccessful (Kulhanek et al. 2011). Thus, quantitative estimates of impact are typically not considered in management decision-making. In **Chapter 4**, we explore the range of potential patterns in how impacts may distributed across sites that are invaded. To explore potential patterns, we simulate species' abundance across a set of sites, and allow for different relationships between abundance and impact. The resulting distributions of impact are highly variable – some most frequently exhibit high impacts, some commonly exhibit low impacts, while others are bimodal. Such different distributions may have very different implications for management, yet the lack of measurements and models of impact prevent these details from being incorporated into decisions. This variability, created by changing simple assumptions about how often a species achieves high abundance and how abundance determines impact, highlights the need for further emphasis of impact and its determinants in future research and management of invasive species.

Literature Cited

Callaway, R. M., and E. T. Aschehoug. 2000. Invasive plants versus their new and old neighbors: A mechanism for exotic invasion. Science 290:521–523.

Callaway, R. M., and W. M. Ridenour. 2004. Novel weapons: invasive success and the evolution of increased competitive ability. Frontiers in Ecology and the Environment 2:436–443.

Charles, H., and J. S. Dukes. 2007. Impacts of invasive species on ecosystem services. Pages 217–237 in W. Nentwig, M. M. Caldwell, G. Heldmaier, R. B. Jackson, O. L. Lange, H. A. Mooney, E. D. Schulze, and U. Sommer, editors. Biological Invasions. Springer Berlin Heidelberg.

Cox, J. G., and S. L. Lima. 2006. Naiveté and an aquatic–terrestrial dichotomy in the effects of introduced predators. Trends in Ecology & Evolution 21:674–680.

Crall, A. W., L. A. Meyerson, T. J. Stohlgren, C. S. Jarnevich, G. J. Newman, and J. Graham. 2006. Show me the numbers: what data currently exist for non-native species in the USA? Frontiers in Ecology and the Environment 4:414–418.

Crooks, J. A. 2002. Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. Oikos 97:153–166.

Crutzen, P. J. 2006. The “Anthropocene.” Pages 13–18 in P. D. E. Ehlers and D. T. Krafft, editors. Earth System Science in the Anthropocene. Springer Berlin Heidelberg.

DiStefano, R. J., M. E. Litvan, and P. T. Horner. 2009. The bait industry as a potential vector for alien crayfish introductions: Problem recognition by fisheries agencies and a Missouri evaluation. Fisheries 34:586–597.

Drake, J., and D. Lodge. 2006. Allee effects, propagule pressure and the probability of establishment: Risk analysis for biological invasions. Biological Invasions 8:365–375.

Gardiner, M. M., L. L. Allee, P. M. Brown, J. E. Losey, H. E. Roy, and R. R. Smyth. 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment 10:471–476.

Green, A. J., and J. Figuerola. 2005. Recent advances in the study of long-distance dispersal of aquatic invertebrates via birds. Diversity and Distributions 11:149–156.

Hatcher, M. J., J. T. A. Dick, and A. M. Dunn. 2012. Disease emergence and invasions. Functional Ecology 26:1275–1287.

Havel, J. E., and J. B. Shurin. 2004. Mechanisms, effects, and scales of dispersal in freshwater zooplankton. Limnology and Oceanography:1229–1238.

Horsch, E. J., and D. J. Lewis. 2009. The effects of aquatic invasive species on property values: evidence from a quasi-experiment. Land Economics 85:391–409.

Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology 46:10–18.

Jeschke, J. M., S. Bacher, T. M. Blackburn, J. T. A. Dick, F. Essl, T. Evans, M. Gaertner, P. E. Hulme, I. Kühn, A. Mrugała, J. Pergl, P. Pyšek, W. Rabitsch, A. Ricciardi, D. M. Richardson, A. Sendek, M. Vilà, M. Winter, and S. Kumschick. 2014. Defining the impact of non-native species. Conservation Biology 28:1188–1194.

Jeschke, J. M., and D. L. Strayer. 2005. Invasion success of vertebrates in Europe and North America. Proceedings of the National Academy of Sciences of the United States of America 102:7198.

Johnson, L. E., A. Ricciardi, and J. T. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications 11:1789–1799.

Kulhanek, S. A., A. Ricciardi, and B. Leung. 2011. Is invasion history a useful tool for predicting the impacts of the world’s worst aquatic invasive species? Ecological Applications 21:189–202.

Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis, and 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.

Litvak, M. K., and N. E. Mandrak. 1993. Ecology of freshwater baitfish use in Canada and the United States. Fisheries 18:6–13.

Lugo, A. E., T. A. Carlo, and J. M. Wunderle. 2012. Natural mixing of species: novel plant–animal communities on Caribbean Islands. Animal Conservation 15:233–241.

Mack, R. N., and W. M. Lonsdale. 2002. Eradicating invasive plants: hard-won lessons for islands. Pages 164–172 in C. R. Veitch and M. N. Clout, editors. Turning the tide: the eradication of invasive species. IUCN SSC Invasive Species Specialist Group, IUCN, Gland, Switzerland and Cambridge, UK.

Mercado-Silva, N., G. G. Sass, B. M. Roth, S. Gilbert, and M. J. V. Zanden. 2007. Impact of rainbow smelt (Osmerus mordax) invasion on walleye (Sander vitreus) recruitment in Wisconsin lakes. Canadian Journal of Fisheries and Aquatic Sciences 64:1543–1550.

Moorhouse, T. P., and D. W. Macdonald. 2015. Are invasives worse in freshwater than terrestrial ecosystems? Wiley Interdisciplinary Reviews: Water 2:1–8.

Olden, J. D. 2006. Biotic homogenization: a new research agenda for conservation biogeography. Journal of Biogeography 33:2027–2039.

Olden, J. D., N. LeRoy Poff, M. R. Douglas, M. E. Douglas, and K. D. Fausch. 2004. Ecological and evolutionary consequences of biotic homogenization. Trends in Ecology & Evolution 19:18–24.

Olden, J. D., and N. L. Poff. 2003. Toward a mechanistic understanding and prediction of biotic homogenization. The American Naturalist 162:442–460.

Pejchar, L., and H. A. Mooney. 2009. Invasive species, ecosystem services and human well-being. Trends in Ecology & Evolution 24:497–504.

Peterson, G. D., T. D. Beard Jr, B. E. Beisner, E. M. Bennett, S. R. Carpenter, G. S. Cumming, C. L. Dent, and T. D. Havlicek. 2003. Assessing future ecosystem services: a case study of the Northern Highlands Lake District, Wisconsin. Conservation Ecology 7:1.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52:273–288.

Postel, S., and S. Carpenter. 1997. Freshwater ecosystem services. Pages 195–214 in G. Daily, editor. Nature’s services: societal dependence on natural ecosystems. Island Press.

Provencher, B., D. J. Lewis, and K. Anderson. 2012. Disentangling preferences and expectations in stated preference analysis with respondent uncertainty: The case of invasive species prevention. Journal of Environmental Economics and Management 64:169–182.

Radeloff, V. J. W. Williams, B. L. Bateman, K. D. Burke, S. K. Carter, E. S. Childress, K. J. Cromwell, C. Gratton, A. O. Hasley, B. M. Kraemer, A. W. Latzka, E. Marin-Spiotta, C. D. Meine, S. E. Munoz, T. M. Neeson, A. M. Pidgeon, A. R. Rissman, R. J. Rivera, L. M. Szymanski, and J. Usinowicz. *In review*. The rise of novelty in ecosystems. Ecological Applications.

Ricciardi, A. 2007. Are modern biological invasions an unprecedented form of global change? Conservation Biology 21:329–336.

Rothlisberger, J. D., W. L. Chadderton, J. McNulty, and 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, and D. H. Wall. 2000. Global biodiversity scenarios for the year 2100. Science 287:1770–1774.

Sax, D. F., and S. D. Gaines. 2008. Species invasions and extinction: The future of native biodiversity on islands. Proceedings of the National Academy of Sciences 105:11490–11497.

Simberloff, D. 2011. How common are invasion-induced ecosystem impacts? Biological Invasions 13:1255–1268.

Steffen, W., P. J. Crutzen, and J. R. McNeill. 2007. The Anthropocene: Are humans now overwhelming the great forces of nature. AMBIO: A Journal of the Human Environment 36:614–621.

Thuiller, W., D. M. Richardson, P. Pyšek, G. F. Midgley, G. O. Hughes, and M. Rouget. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11:2234–2250.

Traveset, A., R. Heleno, S. Chamorro, P. Vargas, C. K. McMullen, R. Castro-Urgal, M. Nogales, H. W. Herrera, and J. M. Olesen. 2013. Invaders of pollination networks in the Galápagos Islands: emergence of novel communities. Proceedings of the Royal Society of London B: Biological Sciences 280:20123040.

Vander Zanden, M. J., and 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., C. M. D’Antonio, L. L. Loope, and R. Westbrooks. 1996. Biological invasions as global environmental change. American Scientist 84:468–478.

Vitousek, P. M., and L. R. Walker. 1989. Biological invasion by Myrica faya in Hawai’i: Plant demography, nitrogen fixation, ecosystem effects. Ecological Monographs 59:247–265.

Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment 5:475–482.

Williamson, M., and A. Fitter. 1996. The varying success of invaders. Ecology 77:1661–1666.

CHAPTER 1

ROADS AND RESIDENTIAL DEVELOPMENT AS DRIVERS OF AQUATIC INVASIVE SPECIES IN LAKES

To be submitted for publication* with coauthors: Scott Van Egeren, Maureen Ferry, and M. Jake Vander Zanden
*Target journal: Conservation Biology

Abstract:

Invasive species are a major driver of adverse ecological change. The presence and number of invasive species has been found to be strongly associated with measures of human activity in many different types of habitats. For lake ecosystems, no studies have directly evaluated the association between aquatic invasive species and measures of human activity such as the degree of lakeshore residential development and distance to roads. Using field survey data for six invasive species in Wisconsin lakes, we tested if the presence and number of invasive species in lakes was related to lakeshore residential development and distance from roads. The number of invasive species present in environmentally suitable lakes, modeled using zero-inflated Poisson regressions, and the probability of occurrence of most individual invasive species, modeled using logistic regressions, were closely associated with the proximity to roads and residential development. In species-specific models, all species distributions were significantly associated with proximity to roads, whereas the two invasive aquatic plant species were also significantly related to residential development. No lakes classified as ‘remote’ contained even a single aquatic invasive species. Our findings reveal a strong association between aquatic invasive species and human activity, which has two key implications. First, it indicates that humans are the primary vector for aquatic invasive species dispersal, thus highlighting the continued need for

efforts to stop human-mediated spread of aquatic invasive species. Second, that protecting lakes from road and housing development offers the additional benefit of protection against aquatic invasive species.

Introduction

Invasive species are a key driver of ecological change (Pyšek & Richardson 2010) and global biodiversity loss (Sala et al. 2000). Their negative effects add up to nearly $120 billion in damages and losses per year in the United States (Pimentel et al. 2005). These negative effects can best be mitigated by preventing new invasions, which relies on understanding how invasive species spread and establish. Invasive species spread and establishment can be strongly influenced by both anthropogenic and natural factors. There are two key processes that underlie and ultimately determine invasive species distributions. First, the habitat must be environmentally suitable in order to support a species (Peterson 2003). Suitability is determined by the habitat's underlying natural conditions, but may also be influenced by human activities such as disturbance, which can promote invasions by some species (Hobbs & Huenneke 1996; Bunn & Arthington 2002; Marvier et al. 2004; Lake & Leishman 2004; Hansen & Clevenger 2005; Buckley et al. 2007; Chytrý et al. 2008). Second, an invasive species must be able to disperse to a given location (Lockwood et al. 2005). Humans are a dominant dispersal agent (Johnson et al. 2001; Hulme 2009), though once established in a region, invasive species can also disperse via natural processes, such as wind dispersal (Kowarik & Lippe 2011) or via waterways (Panov et al. 2009) to new sites. An understanding of the role of anthropogenic factors in determining the distribution of invasive species is vital to ongoing management efforts aimed at minimizing their spread and impacts (Gallardo & Aldridge 2013; Gallardo et al. 2015b).

For terrestrial systems, a large body of literature has evaluated links between invasive species and measures of human activity. For example, the presence, cover, and species richness of invasive plants are negatively related to distance from road (Harrison et al. 2002; Gelbard & Belnap 2003; Flory & Clay 2005; Kalwij et al. 2008; Christen & Matlack 2008). Roads can disturb natural habitats, or act as corridors for vehicle-dispersed invasive species, especially plants (Lippe & Kowarik 2007). Consequently, roadless habitats sometimes provide a refuge from invasive plants (Gelbard & Harrison 2003). Similarly, there are positive relationships between the number of terrestrial invasive species and metrics of human development such as urbanization (Blair 1996; Riley et al. 2005; Lussier et al. 2006).

In contrast with terrestrial systems, we have a remarkably weak understanding of the role of anthropogenic factors in determining the distributions of aquatic invasive species (AIS). Many AIS, including macrophytes, mollusks, crustaceans, macroinvertebrates, and plankton may undergo human-mediated dispersal as “hitchhikers” on recreational boats (Rothlisberger et al. 2010). For instance, spread of zebra mussels (*Dreissena polymorpha*) among lakes is widely known to be caused by boaters, and is often modeled through boater pathways (Johnson & Carlton 1996; Bossenbroek et al. 2001, 2007; Leung et al. 2006). Additionally, some AIS may benefit from human-induced disturbance. However, our understanding of how these factors then influence AIS distributions at landscape scales is limited. Capelli & Magnuson (1983) note a positive relationship between invasive rusty crayfish (*Orconectes rusticus*) abundance and lakeshore residential development, though the study only included a single invasive species and did not include any undeveloped lakes. Johnson et al. (2008) report positive associations between the presence of several AIS and the number of boat launches for lakes, but did not differentiate the magnitude or types of human activities occurring on lakes that did not have public boat launches. Such lakes may experience shoreline development and/or boater presence via private access or surface water connections to other lakes, or may be more remote and experience very little human activity. Our understanding of the link between the occurrence and number of AIS and anthropogenic variables is weak. This is surprising considering the fact that freshwater systems, and lakes specifically, are particularly vulnerable to the negative ecological effects of invasive species (Sala et al. 2000; Moorhouse & Macdonald 2015). When invaded, lakes tend to exhibit decreased abundance and diversity of native communities and altered ecosystem properties and processes (Gallardo et al. 2015a). In addition, some AIS including Eurasian watemilfoil (*Myriophyllum* spicatum) and zebra mussels cause drastic economic costs on the lakes they invade (Connelly et al. 2007; Horsch & Lewis 2009; Provencher et al. 2012).

The present study was designed specifically to evaluate the links between aquatic invasive species and two key variables reflective of human activity - distance to the nearest road and degree of lakeshore residential development. Our study includes a wide range of lake types, spanning from remote wilderness lakes to highly developed lakes. It also includes a suite of six different invasive species spanning a broad range of taxonomic groups. Understanding the links between human activity and the occurrence of aquatic invasive species has important implications for ongoing management efforts aimed at minimizing their spread and adverse impacts – preventing human-vectored spread to new lakes requires that we understand the natural and anthropogenic factors that put lakes at greatest risk of invasion.

Methods

Study region

We evaluated the relationship between invasive species presence and human activity on inland lakes in Wisconsin, USA. Wisconsin is a lake-rich landscape with nearly 15,000 lakes, ponds, or reservoirs greater than 1 ha in area (2007 USGS National Hydrography Dataset). The northern portion of the state is largely forested, with relatively small year-round human populations and many lakes (Carpenter et al. 2007). The southern half of the state has fewer lakes and is human-dominated, with much of the land use being agricultural, urban, or suburban (Carpenter et al. 2007).

Data collection

Our dataset is comprised of field surveys conducted for this study along with two other recent field surveys aimed at mapping AIS presence/absence in Wisconsin lakes. Our surveys included a stratified random sample (stratified across boat launch presence or absence, high or low conductivity, and large or small lake area) of lakes in the Northern Highlands Lake District, a random sample of lakes with boat launches across Wisconsin, and a survey of randomly selected lakes that were both far from roads and had no residential development. The remaining two field surveys, which were conducted prior to this study, were both regional efforts to document all invasive species present in lakes within smaller regions of the state (1-4 counties). We included their results with the results of our own surveys to maximize coverage of recent AIS survey data across the state. Additional information on all surveys is provided in the Supporting Information.

All lakes in our dataset were surveyed for the presence or absence of the following AIS: Chinese mystery snail (CMS; *Cipangopaludina chinensis*), banded mystery snail (BMS; *Viviparus georgianus*), Eurasian watermilfoil (EWM; *Myriophyllum spicatum*), curly-leaf pondweed (CLP; *Potamogeton crispus*), rusty crayfish (RC; *Orconectes rusticus*), and zebra mussel (ZM; *Dreissena polymorpha*). These six species comprise 2,312 of 2,564 known AIS occurrences in Wisconsin (Wisconsin DNR, 2012). Attempts to detect RC were not carried out on many lakes, and several surveys at a small number of additional lakes excluded certain species, so these lakes were left out of the species-specific analyses. Because all six species are visually conspicuous and reside in shallow water during the summer when surveys were conducted, littoral zones were surveyed by visual assessment via snorkeling or canoeing, with some surveys incorporating more intensive sampling at specific sites. At boat launches, we surveyed via snorkeling for 30 minutes. At five sites selected to represent the littoral habitat variability (e.g., substrate type, shoreline morphometry, bathymetry) present in the lake, we conducted 10-min timed snorkels and deployed 6 crayfish traps (Gee-style minnow traps with widened openings) baited with beef liver (Capelli & Magnuson 1983) that were left overnight and inspected for RC the next morning. When low water clarity interfered with visual assessment of snails and mussels, we took 6 D-net samples each covering 0.5x0.5 m of substrate to inspect for AIS.

We evaluated the presence of these species with respect to three independent variables. We used lake conductivity as a measure of habitat suitability because it was measured for the majority of lakes in our dataset and because it correlates positively with other metrics of suitability for aquatic macrophytes (Hutchinson 1970), and mollusks, crustaceans and other species with calcium carbonate exoskeletons (Latzka et al. 2015). We used geographic information systems (GIS) to calculate 2 metrics associated with human activity on each lake in our survey – distance to the nearest road and the degree of lakeshore development. We chose these variables because they could be calculated using GIS, and are closely related to human activity. For each lake, we calculated the linear distance from the nearest road, using lake polygons in the USGS National Hydrography dataset (2007) and US Census TIGER shapefiles for roads (2010). To represent the degree of lakeshore development, we calculated the proportion of land surrounding each lake that was developed. We created a 100 m buffer around each lake, summed areas for all types of developed land use (including High, Medium, and Low Intensity, and Developed, Open Space categories) in the 2011 National Land Cover Dataset (30 m resolution) within the buffer, and divided by the buffer's non-water area. We chose 100 m as a suitable distance for the buffer to characterize shoreline residential development, as wider buffers tend to include substantial amounts of non-lakefront residential development, while narrower buffers fail to include lakefront properties that are set back from the lake (Schnaiberg et al. 2002).

Data analysis

Predicting the number of invasive species

To model the numbers of AIS, we ran a zero-inflated Poisson (ZIP) regression. ZIP models are appropriate for response data represented by counts that include many zeroes (Zeileis et al. 2008). In such data, there is an elevated probability of the response having a value of zero – these observations are first fit by the zero-inflation model – a binomial model with a logit link. The second component models a count using a Poisson process with a log link, and predicts a response of 0, 1, 2, and so on. For aquatic invasive species in lakes, certain conditions – such as being far from roads or having no nearby development – may eliminate the potential for that lake to become invaded. If these conditions are not met, a lake still may or may not become invaded, and it may be invaded by more than one species. Independent variables in our models included distance to road, percent development, and conductivity. We also explored how our model changed when including and substituting the following covariates: lake area, lake order (i.e., landscape position, which incorporates lake's connectivity to other lakes and streams), and the presence of a public boat launch. We explored model predictions in two complementary ways. First, we made predictions across each anthropogenic variable independently, while holding the other constant at its minimum, median and 99th percentile values (chosen instead of maximum to allow for predictions across a larger, observed range of the other variable), using bootstrapped 95% confidence intervals. Second, we plotted the prediction for each possible combination of distance to road and development, which allowed for closer inspection of different combinations of variables. All predictions were made for a hypothetical lake with a median level of conductivity (97.4 µS/cm), which is predicted to have suitable environmental conditions for all species (Latzka et al. 2015 and unpublished data). Analyses were performed using the pscl R package (Zeileis et al. 2008; Jackman 2015).

Characterizing invasions among lake types

We compared rates of invasions among lakes classified as "remote," "marginal human activity," and "high human activity," and tested whether the number of AIS differed among groups. Remote lakes had 0% development within the 100 m buffer, and were at least 100 m from the nearest road – thus, their shorelines consisted entirely of natural land covers. High human activity lakes were less than 100 m from the nearest road and at least 20% developed. Marginal human activity had one but not both of these characteristics – they were either less than 100 m from a road with less than 20% development, or greater than 100 m from a road with any positive level of development. We verified these classifications with the degree of development and accessibility from field observations and aerial imagery – lakes classified as high activity always had high levels of development, marginal activity lakes had indications of human activity, but to a much lesser degree, and remote lakes showed no signs of human activity.

Because the sample size of remote lakes was small and the distribution of the number of AIS present was non-normal, we compared the number of AIS observed in lakes of each category using the non-parametric Kruskal-Wallis test and post-hoc pairwise comparisons using a Nemenyi test (Demšar 2006) (PMCMR R package - Pohlert 2015).

Predicting probability of presence by species

We tested if individual aquatic invasive species were associated with lake attributes using logistic regression. These models predict the probability of a lake being invaded by a particular species based on distance to road, percent development, and conductivity. For each species, we fit single-predictor models to test relationships between species presence and each variable, independent of other attributes. In addition, we fit models using all possible subsets of the three predictors and selected the best model based on AIC. For single-predictor models and for each variable in the best multiple-predictor model, we report odds ratios, which are more easily interpretable than regression coefficients. We scaled the predictors so that these odds ratios can be interpreted as the change in the odds of a lake being invaded associated with either a 1% increase in development or a 10 m increase in a lake's distance from a road.

All analyses were performed in R version 3.1.2 (R Core Team 2014).

Results

Our dataset included a total of 448 lakes, but we removed 11 lakes due to lack of conductivity data. Of the remaining 437 lakes, 116 had 0 AIS, 131 had 1 AIS, and 190 had multiple AIS (Fig. 1). The frequency distribution of values of distance from road was right-skewed, whereby most lakes were within a few meters of the nearest road (Fig. 2). The distribution of development values was also right-skewed, whereby most lakes had moderate-to-low levels of development (Fig. 2).

Predicting the number of invasive species

Distance from road and proportion developed were each significant predictors of the number of AIS in our two-component ZIP model. In the first component, which predicts the probability of having zero AIS – the zero-inflated model – conductivity and distance from road were significant predictors of AIS presence ($p<0.001$), where distance from road was positively associated with the probability of a having zero AIS, and conductivity was negatively associated. In the second component, the Poisson count model, all variables were significant. The degree of shoreline development and conductivity were each significantly and positively related to the number of AIS ($p<0.001$), while distance from road was negatively related ($p<0.01$). Our ZIP model outperformed a standard Poisson regression (Vuong test, $p<0.001$), and the model including all three predictors had the lowest AIC (1155.8) compared to models including all possible subsets of variables (1167.0-1302.0). Of the additional covariates tested, only lake area improved this model when substituted for distance from road (AIC=1148.7) or development (1153.2). Further, when all covariates were included in the same model (all variance inflation factors were less than 2), distance from road and development remain as significant predictors, indicating their importance in determining the presence and number of AIS in lakes. Thus, we found that our original model, where the presence of AIS is a function of conductivity and distance from road, and the number of AIS is a function of conductivity, distance from road, and development, was sufficient for evaluating the effects of distance from road and development.

We used our model to predict the number of AIS at minimum, median, and near-maximum (99th percentile) values of each predictor (e.g., distance from road or development) across the gradient of observed values for the other predictor (Fig. 3), while holding conductivity at its median level. At high values of distance from road (99^{th} percentile – 384 m), in which lakes were all also relatively undeveloped, lakes were unlikely to be invaded by any species (Fig. 3a). At minimum (0 m) and median (12.5 m) values of distance from road, the predicted number of AIS increased from about 1 to 3 as the proportion of development increases from 0 to 1 (Fig. 3b,c). At high levels of development (99^{th} percentile – 74%), lakes were relatively near to roads, and were predicted to have 2-3 AIS (Fig. 3d). At median (15%) and minimum levels of development (0%), the predicted number of AIS decreased from about 1.5 at 0 m from a road to 0 around 200 m from a road (Fig. 3e,f). The heatmap of predictions illustrates these results for all possible combinations of distance to road and development, demonstrating that lakes near roads with high development were predicted to have 2-3 AIS, lakes far from roads (> about 200 m) were predicted to have 0 AIS, and lakes near roads with low-to-moderate levels of development were predicted to have 1-2 AIS (Fig. 4).

Characterizing invasions among lake types

AIS were not detected in any remote lake. In contrast, moderate and high human activity lakes had significantly higher numbers of AIS (Kruskal-Wallis test, $p<0.001$), with medians of 1 and 2 species, respectively (Fig. 5). Nemenyi pairwise comparisons indicated significant differences among all groups ($p<0.01$).

Predicting probability of presence by species

In species-specific single predictor models, where the presence of each species was a function of each variable independently, significant relationships were found for each variable for all species, except development for CMS and distance from road for BMS and ZM (Table 1).

Odds ratios associated with a 1% increase in development ranged from 0.90 for CMS (although non-significant) to 1.06 for CLP, which was most strongly associated with development. Odds ratios associated with a 10 m increase in distance from road ranged from 0.57 for RC, which was most strongly associated with proximity to roads, to 0.97 for BMS. The best multiple-predictor models chosen by AIC included significant relationships with distance from road for all species, development for both macrophyte species (EWM and CLP), and conductivity for all species except RC (Table 1).

Discussion

Our results demonstrate a clear statistical relationship between measures of human activity and residential development and the magnitude of species invasions in lakes. We found that remote lakes that are far from roads serve as refugia from invasive species. For lakes that were near roads, lakes that were more developed had higher numbers of invasive species. The strong connection between human activity on lakes and invasive species provides insights into their vectors of spread, and supports prevention efforts that target human pathways of invasion. Such prevention efforts are important given the substantial ecological and economic impacts of invasions.

Explanation of correlations between AIS and human activity

These results can be explained primarily by AIS dispersal vectors. The lack of invasive species in remote lakes fits assumptions that the primary vectors for these species are human pathways such as recreational boating (Johnson et al. 2001). In addition, developed lakes tend to be larger and attract higher numbers of boaters (Reed-Andersen et al. 2000), and thus may receive higher propagule pressure, enabling the establishment of multiple species. Higher numbers of AIS may also be driven by higher levels of habitat disturbance on developed lakes, Some invasive species tend to have traits, such as high growth rates and wide physiological tolerances, that allow them to cope well with anthropogenic disturbance (McMahon 2002; Marvier et al. 2004; Marchetti et al. 2004), so they may be particularly well-suited to inhabit developed lakes that undergo more anthropogenic disturbance than undeveloped lakes.

The species-specific results lend insights into what types of lakes different species are able to invade. First, both human activity metrics were significantly related with the presence of most AIS species, verifying that AIS introductions are human-mediated. CMS, however, had a non-significant relationship with development, indicating that they may be better able to take advantage of infrequent human vectors, or nonhuman vectors altogether, which could be explained by their extreme resistance to desiccation (Havel 2011). BMS and ZM were not related to distance from road in univariate models, but these relationships were significant when conductivity was included as a predictor. Overall, the best logistic models for each species emphasize the importance of proximity to roads in determining AIS distributions, as it was significant in each of the best models. Invasive macrophytes may be additionally facilitated by particularly high levels of human activity associated with lakeshore development, as they were the only species with significant relationships with development after conductivity and distance from road were accounted for. Both EWM and CLP have relatively short desiccation times (Bruckerhoff et al. 2015), so they may be limited to lakes that receive more frequent boater traffic, which will likely be the more developed lakes. Alternatively, these species may be better adapted to establishing in lakes that experience littoral habitat disturbance associated with development (Christensen et al. 1996; Jennings et al. 2003).

Alternative explanations

While previous studies have noted correlations of some species presences with human factors, none have investigated the relationship between multiple metrics of human activity and the number of AIS present. This study used a large presence/absence dataset spanning gradients in both the lakes' degree of residential development and distance to the nearest roads. To our knowledge, this represents the first study of aquatic invasive species to sample relatively inaccessible lakes that were far from roads.

Despite this unique dataset and improvement upon previous studies in terms of the range of sites included, there are two noteworthy limitations. First, because remote lakes were difficult to find and access, we obtained a sample size of only 11 lakes, although nonparametric tests and our ZIP model both indicated that the number of AIS present in these lakes was significantly lower than that in other lakes. Second, distance from roads and degree of development may correlate with other potentially important environmental variables. Indeed, environmental variables were slightly more important than anthropogenic variables (which were still significant) in predicting the distributions of many invasive species in northwest Europe (Gallardo et al. 2015b). Similarly, in Wisconsin, native species distributions in lakes tend to be more affected by environmental variables than dispersal-related variables (Magnuson et al. 1998). Thus, it is possible that environmental variables that correlate with distance from road and development could lead to the general results we found. In general, Wisconsin lakes that have no development and are far from roads tend to be small. Small lakes tend to have a higher landscape position, and thus have fewer surface-water connections with rivers and other lakes and lower levels of dissolved calcium, and thus lower conductivity (Kratz et al. 1997; Riera et al. 2000). These covariates may also reduce the likelihood that AIS will be able to colonize and establish in remote, undeveloped lakes. To account for this, we sampled small, low calcium lakes across both gradients of human activity. We also included conductivity, landscape position, and lake area in our models. While conductivity was a key variable in explaining habitat suitability, both anthropogenic variables remained significant after including it in models. In contrast, landscape position did not improve models. Models did improve slightly when area was substituted for distance from road and development. However, lake area may correlate with invasion likelihood in two separate ways. First, area correlates with environmental conditions like dissolved calcium and conductivity, and thus may be an indirect indicator of habitat suitability. Second, area is a main driver of boat traffic to lakes (Reed-Andersen et al. 2000), and thus may correlate with propagule pressure. For these reasons, area can be a strong predictor of invasions, but its role in shaping species distributions is difficult to interpret. Because covariates other than area did not perform as well as distance from road and development in our models, we are confident in our finding that AIS distributions in environmentally suitable Wisconsin lakes are largely determined by human activity.

An alternative dispersal mechanism for AIS may be transportation by waterbirds. Observations of bird-mediated dispersal date back at least to Darwin's Origin of Species, with many more recent observations for aquatic invertebrates, snails and plants (Green & Figuerola 2005; van Leeuwen et al. 2012b, 2012a; García-Álvarez et al. 2015). However, evidence for such bird-mediated dispersal of AIS is largely anecdotal, and no studies have compared the relative magnitude of spread of AIS by waterfowl versus humans (Reynolds et al. 2015). Our results highlight a strong association between invasions and human activity on lakes, indicating an important role for human transport vectors. Waterbirds may still be a possible vector, as their dispersal patterns may be influenced by lake area and shoreline development (Traut & Hostetler 2004). However, if bird-mediated invasions were as common as human-mediated invasions, we would expect some invasions in remote lakes. Indeed, small, shallow lakes may have relatively high bird richness (Scheffer et al. 2006). Such lakes were entirely absent of AIS in our surveys. Even though bird-mediated spread is important in species dispersal across geologic time-scales and in shaping species' evolutionary histories (Wesselingh et al. 1999), we believe it is unlikely that birds would transport AIS frequently enough to account for the rapid rates of spread associated with new invasions (Olden et al. 2006).

Implications

Our general finding that the number of AIS increases with development and decreases with distance from road is likely applicable to other regions. Indeed, the distributions of selected terrestrial, freshwater, and marine invasive species in northwest Europe all correlate with anthropogenic variables including human population density and road proximity (Gallardo et al. 2015b), However, the relative importance of human-aided dispersal vs. natural dispersal vs. underlying environmental conditions may vary depending on the landscape. Wisconsin lakes exhibit large gradients in distance from road and development, making it a useful region for studying these relationships. Regions without such variability – for instance, regions in which lakes are rare and uniformly developed, or wilderness regions where all lakes are relatively remote – may be better predicted by natural dispersal variables or environmental conditions. In addition, invasive species that are more closely associated with non-human pathways of dispersal would not have distributions that would be well-predicted by our models.

The ecological and economic impacts of aquatic invasive species can be substantial. Ecological impacts include strong and consistent negative effects on native communities and changes in ecosystem functions (Gallardo et al. 2015a). For instance, many macrophytes, including Eurasian watermilfoil, can outcompete native plants by forming mats that shade out other species (Madsen et al. 1991; Boylen et al. 1999; Stiers et al. 2011). Rusty crayfish tend to outcompete native crayfish (Olden et al. 2006) and reduce macrophyte communities, leading to food web shifts (Lodge et al. 1994; Mccarthy et al. 2006; Hansen et al. 2013). The impacts of invasive filter-feeders like zebra and quagga mussels can cause shifts in ecosystem properties and processes, including shifting energy flows from pelagic to benthic pathways with concomitant changes in phytoplankton, zooplankton, and macrophyte biomass (Strayer et al. 1999; Zhu et al. 2006; Higgins & Vander Zanden 2010). Additionally, many AIS have substantial economic costs. For instance, Eurasian watermilfoil tends to decrease housing values by approximately 15% (Horsch & Lewis 2009; Provencher et al. 2012). Control costs for zebra mussels at power plants throughout North America were estimated at $267 million (Connelly et al. 2007).

All of these risks, along with the relative irreversibility of invasions (Mack & Lonsdale 2002), have made prevention and containment the cornerstone of most invasive species management plans (Leung et al. 2002; Lodge et al. 2006). Prevention efforts largely focus on reducing introductions by interrupting boater pathways via boat cleaning, draining, drying, and disinfecting (Rothlisberger et al. 2010). This management approach relies on the idea that AIS introductions are primarily caused by human activities. If other dispersal vectors, including natural dispersal via surface waters or transportation by birds were major vectors for AIS, these prevention efforts may be undermined and not worthwhile. Our results point to the strong role played by human activity in invasions of lakes, and validate the belief that continued prevention of human-mediated introductions is worthwhile.

Our results clearly show an absence of invasive species in remote lakes, low levels of invasion in lakes that are near roads but have little development, and high numbers of invasive species in highly developed lakes. These results suggest that protection of lakes from development may not only conserve terrestrial habitat in protected areas, but also protect lakes from the effects of invasions.

Supporting information

The description of field surveys that we compiled to construct the dataset used in this analysis (Appendix S1) is available online.

Acknowledgements

This material is based upon work supported by the National Science Foundation under #CNH-0909281 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. Additional support from the Wisconsin Department of Natural Resources funded portions of the data collection for this study. We would like to thank the staff of UW Trout Lake Station for their assistance with fieldwork and logistics.

Literature Cited

Blair, R. B. 1996. Land use and avian species diversity along an urban gradient. Ecological Applications **6**:506–519.

Bossenbroek, J. M., L. E. Johnson, B. Peters, and D. M. Lodge. 2007. Forecasting the Expansion of Zebra Mussels in the United States. Conservation Biology **21**:800–810.

Bossenbroek, J. M., C. E. Kraft, and J. C. Nekola. 2001. Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. Ecological Applications **11**:1778–1788.

Boylen, C. W., L. W. Eichler, and J. D. Madsen. 1999. Loss of native aquatic plant species in a community dominated by Eurasian watermilfoil. Pages 207–211 in J. Caffrey, P. R. F. Barrett, M. T. Ferreira, I. S. Moreira, K. J. Murphy, and P. M. Wade, editors. Biology, Ecology and Management of Aquatic Plants. Springer Netherlands.

Bruckerhoff, L., J. Havel, and S. Knight. 2015. Survival of invasive aquatic plants after air exposure and implications for dispersal by recreational boats. Hydrobiologia **746**:113–121.

Buckley, Y. M., B. M. Bolker, and M. Rees. 2007. Disturbance, invasion and re-invasion: managing the weed-shaped hole in disturbed ecosystems. Ecology Letters **10**:809–817.

Bunn, S. E., and A. H. Arthington. 2002. Basic Principles and Ecological Consequences of Altered Flow Regimes for Aquatic Biodiversity. Environmental Management **30**:492–507.

Capelli, G. M., and J. J. Magnuson. 1983. Morphoedaphic and biogeographic analysis of crayfish distribution in northern Wisconsin. Journal of Crustacean Biology **3**:548–564.

Carpenter, S. R. et al. 2007. Understanding regional change: a comparison of two lake districts. BioScience **57**:323–335.

Christen, D. C., and G. R. Matlack. 2008. The habitat and conduit functions of roads in the spread of three invasive plant species. Biological Invasions **11**:453–465.

Christensen, D. L., B. R. Herwig, D. E. Schindler, and S. R. Carpenter. 1996. Impacts of lakeshore residential development on coarse woody debris in north temperate lakes. Ecological Applications **6**:1143–1149.

Chytrý, M., V. Jarošík, P. Pyšek, O. Hájek, I. Knollová, L. Tichý, and J. Danihelka. 2008. Separating habitat invasibility by alien plants from the actual level of invasion. Ecology **89**:1541–1553.

Connelly, N. A., C. R. O. Jr, B. A. Knuth, and T. L. Brown. 2007. Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. Environmental Management **40**:105–112.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. **7**:1–30.

Flory, S. L., and K. Clay. 2005. Invasive shrub distribution varies with distance to roads and stand age in eastern deciduous forests in Indiana, USA. Plant Ecology **184**:131–141.

Gallardo, B., and D. C. Aldridge. 2013. The “dirty dozen”: socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland. Journal of Applied Ecology **50**:757–766.

Gallardo, B., M. Clavero, M. I. Sánchez, and M. Vilà. 2015a. Global ecological impacts of invasive species in aquatic ecosystems. Global Change Biology. *In press*.

Gallardo, B., A. Zieritz, and D. C. Aldridge. 2015b. The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine Invaders. PLoS ONE **10**:e0125801.

García-Álvarez, A., C. H. A. van Leeuwen, C. J. Luque, A. Hussner, A. Vélez-Martín, A. Pérez-Vázquez, A. J. Green, and E. M. Castellanos. 2015. Internal transport of alien and native plants by geese and ducks: an experimental study. Freshwater Biology **60**:1316–1329.

Gelbard, J. L., and J. Belnap. 2003. Roads as conduits for exotic plant invasions in a semiarid landscape. Conservation Biology **17**:420–432.

Gelbard, J. L., and S. Harrison. 2003. Roadless habitats as refuges for native grasslands: interactions with soil, aspect, and grazing. Ecological Applications **13**:404–415.

Green, A. J., and J. Figuerola. 2005. Recent advances in the study of long-distance dispersal of aquatic invertebrates via birds. Diversity and Distributions **11**:149–156.

Hansen, G. J. A., C. L. Hein, B. M. Roth, M. J. Vander Zanden, J. W. Gaeta, A. W. Latzka, and S. R. Carpenter. 2013. Food web consequences of long-term invasive crayfish control. Canadian Journal of Fisheries and Aquatic Sciences **70**:1109–1122.

Hansen, M. J., and A. P. Clevenger. 2005. The influence of disturbance and habitat on the presence of non-native plant species along transport corridors. Biological Conservation **125**:249–259.

Harrison, S., C. Hohn, and S. Ratay. 2002. Distribution of exotic plants along roads in a peninsular nature reserve. Biological Invasions **4**:425–430.

Havel, J. E. 2011. Survival of the exotic Chinese mystery snail (Cipangopaludina chinensis malleata) during air exposure and implications for overland dispersal by boats. Hydrobiologia **668**:195–202.

Higgins, S. N., and M. J. Vander Zanden. 2010. What a difference a species makes: a meta-analysis of dreissenid mussel impacts on freshwater ecosystems. Ecological monographs **80**:179–196.

Hobbs, R. J., and L. F. Huenneke. 1996. Disturbance, Diversity, and Invasion: Implications for Conservation. Pages 164–180 Ecosystem Management. Springer New York. Available from http://link.springer.com/chapter/10.1007/978-1-4612-4018-1_16 (accessed August 12, 2015).

Horsch, E. J., and D. J. Lewis. 2009. The effects of aquatic invasive species on property values: evidence from a quasi-experiment. Land Economics **85**:391–409.

Hulme, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. Journal of Applied Ecology **46**:10–18.

Hutchinson, G. E. 1970. THE CHEMICAL ECOLOGY OF THREE SPECIES OF MYRIOPHYLLUM (ANGIOSPERMAE, HALORAGACEAE) l*”. OCEANOGRAPHY **15**. Available from http://www.avto.aslo.info/lo/toc/vol_15/issue_1/0001.pdf (accessed May 28, 2014).

Jackman, S. 2015. pscl: Classes and methods for R developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, California. Available from http://pscl.stanford.edu/.

Jennings, M. J., E. E. Emmons, G. R. Hatzenbeler, C. Edwards, and M. A. Bozek. 2003. Is Littoral Habitat Affected by Residential Development and Land Use in Watersheds of Wisconsin Lakes? Lake and Reservoir Management **19**:272–279.

Johnson, L. E., and J. T. Carlton. 1996. Post-Establishment Spread in Large-Scale Invasions: Dispersal Mechanisms of the Zebra Mussel Dreissena Polymorpha. Ecology **77**:1686–1690.

Johnson, L. E., A. Ricciardi, and J. T. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications **11**:1789–1799.

Johnson, P. T., J. D. Olden, and M. J. Vander Zanden. 2008. Dam invaders: impoundments facilitate biological invasions into freshwaters. Frontiers in Ecology and the Environment **6**:357–363.

Kalwij, J. M., S. J. Milton, and M. A. McGeoch. 2008. Road verges as invasion corridors? A spatial hierarchical test in an arid ecosystem. Landscape Ecology **23**:439–451.

Kowarik, I., and M. von der Lippe. 2011. Secondary wind dispersal enhances long-distance dispersal of an invasive species in urban road corridors. NeoBiota **9**:49–70.

Kratz, T. K., K. E. Webster, C. J. Bowser, J. J. Magnuson, and B. J. Benson. 1997. The influence of landscape position on lakes in northern Wisconsin. Freshwater Biology **37**:209–217.

Lake, J. C., and M. R. Leishman. 2004. Invasion success of exotic plants in natural ecosystems: the role of disturbance, plant attributes and freedom from herbivores. Biological Conservation **117**:215–226.

Latzka, A. W., J. T. Crawford, A. S. Koblings, Y. Caldeira, E. Hilts, and M. J. Vander Zanden. 2015. Representing calcification in distribution models for aquatic invasive species: surrogates perform as well as CaCO3 saturation state. Hydrobiologia **746**:197–208.

Leung, B., J. M. Bossenbroek, and D. M. Lodge. 2006. Boats, Pathways, and Aquatic Biological Invasions: Estimating Dispersal Potential with Gravity Models. Biological Invasions **8**:241–254.

Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis, and 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.

Lippe, M. Von Der, and I. Kowarik. 2007. Long-distance dispersal of plants by vehicles as a driver of plant invasions. Conservation Biology **21**:986–996.

Lockwood, J. L., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology & Evolution **20**:223–228.

Lodge, D. M., M. W. Kershner, J. E. Aloi, and A. P. Covich. 1994. Effects of an omnivorous crayfish (Orconectes rusticus) on a freshwater littoral food web. Ecology:1265–1281.

Lodge, D. M., S. Williams, H. J. MacIsaac, K. R. Hayes, B. Leung, S. Reichard, R. N. Mack, P. B. Moyle, M. Smith, and D. A. Andow. 2006. Biological invasions: recommendations for US policy and management. Ecological Applications **16**:2035–2054.

Lussier, S. M., R. W. Enser, S. N. Dasilva, and M. Charpentier. 2006. Effects of habitat disturbance from residential development on breeding bird communities in riparian corridors. Environmental Management **38**:504–521.

Mack, R. N., and W. M. Lonsdale. 2002. Eradicating invasive plants: hard-won lessons for islands. Pages 164–172 in C. R. Veitch and M. N. Clout, editors. Turning the tide: the eradication of invasive species. IUCN SSC Invasive Species Specialist Group, IUCN, Gland, Switzerland and Cambridge, UK. Available from https://portals.iucn.org/library/efiles/documents/SSC-OP-028.pdf#page=170 (accessed April 30, 2015).

Madsen, J. D., J. W. Sutherland, J. A. Bloomfield, L. W. Eichler, and C. W. Boylen. 1991. The decline of native vegetation under dense Eurasian watermilfoil canopies. Journal of Aquatic Plant Management **29**:94–99.

Magnuson, J. J., W. M. Tonn, A. Banerjee, J. Toivonen, O. Sanchez, and M. Rask. 1998. Isolation vs. extinction in the assembly of fishes in small northern lakes. Ecology **79**:2941–2956.

Marchetti, M. P., P. B. Moyle, and R. Levine. 2004. Invasive species profiling? Exploring the characteristics of non-native fishes across invasion stages in California. Freshwater Biology **49**:646–661.

Marvier, M., P. Kareiva, and M. G. Neubert. 2004. Habitat Destruction, Fragmentation, and Disturbance Promote Invasion by Habitat Generalists in a Multispecies Metapopulation. Risk Analysis **24**:869–878.

Mccarthy, J. M., C. L. Hein, J. D. Olden, and M. Jake Vander Zanden. 2006. Coupling long-term studies with meta-analysis to investigate impacts of non-native crayfish on zoobenthic communities. Freshwater Biology **51**:224–235.

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

Moorhouse, T. P., and D. W. Macdonald. 2015. Are invasives worse in freshwater than terrestrial ecosystems? Wiley Interdisciplinary Reviews: Water **2**:1–8.

Olden, J. D., J. M. McCarthy, J. T. Maxted, W. W. Fetzer, and M. J. Vander Zanden. 2006. The rapid spread of rusty crayfish (Orconectes rusticus) with observations on native crayfish declines in Wisconsin (U.S.A.) over the past 130 years. Biological Invasions **8**:1621–1628.

Panov, V. E. et al. 2009. Assessing the risks of aquatic species invasions via european inland waterways: from concepts to environmental indicators. Integrated Environmental Assessment and Management **5**:110–126.

Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. The quarterly review of biology **78**:419–433.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics **52**:273–288.

Pohlert, T. 2015. PMCMR: Calculate pairwise multiple comparisons of mean rank sums. Available from http://CRAN.R-project.org/package=PMCMR.

Provencher, B., D. J. Lewis, and K. Anderson. 2012. Disentangling preferences and expectations in stated preference analysis with respondent uncertainty: The case of invasive species prevention. Journal of Environmental Economics and Management **64**:169–182.

Pyšek, P., and D. M. Richardson. 2010. Invasive species, environmental change and management, and health. Annual Review of Environment and Resources **35**:25–55.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org.

Reed-Andersen, T., E. M. Bennett, B. S. Jorgensen, G. Lauster, D. B. Lewis, D. Nowacek, J. L. Riera, B. L. Sanderson, and R. Stedman. 2000. Distribution of recreational boating across lakes: do landscape variables affect recreational use? Freshwater Biology **43**:439–448.

Reynolds, C., N. A. F. Miranda, and G. S. Cumming. 2015. The role of waterbirds in the dispersal of aquatic alien and invasive species. Diversity and Distributions **21**:744–754.

Riera, J. L., J. J. Magnuson, T. K. Kratz, and K. E. Webster. 2000. A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. Freshwater Biology **43**:301–318.

Riley, S. P. D., G. T. Busteed, L. B. Kats, T. L. Vandergon, L. F. S. Lee, R. G. Dagit, J. L. Kerby, R. N. Fisher, and R. M. Sauvajot. 2005. Effects of urbanization on the distribution and abundance of amphibians and invasive species in southern California streams. Conservation Biology **19**:1894–1907.

Rothlisberger, J. D., W. L. Chadderton, J. McNulty, and 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. et al. 2000. Global biodiversity scenarios for the year 2100. Science **287**:1770–1774.

Scheffer, M., G. J. Van Geest, K. Zimmer, E. Jeppesen, M. Søndergaard, M. G. Butler, M. A. Hanson, S. Declerck, and L. De Meester. 2006. Small habitat size and isolation can promote species richness: second-order effects on biodiversity in shallow lakes and ponds. Oikos **112**:227–231.

Schnaiberg, J., J. Riera, M. G. Turner, and P. R. Voss. 2002. Explaining human settlement patterns in a recreational lake district: Vilas County, Wisconsin, USA. Environmental Management **30**:24–34.

Stiers, I., N. Crohain, G. Josens, and L. Triest. 2011. Impact of three aquatic invasive species on native plants and macroinvertebrates in temperate ponds. Biological Invasions **13**:2715–2726.

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

Traut, A. H., and M. E. Hostetler. 2004. Urban lakes and waterbirds: effects of shoreline development on avian distribution. Landscape and Urban Planning **69**:69–85.

van Leeuwen, C. H. A., G. van der Velde, J. M. van Groenendael, and M. Klaassen. 2012a. Gut travellers: internal dispersal of aquatic organisms by waterfowl. Journal of Biogeography **39**:2031–2040.

van Leeuwen, C. H. A., G. van der Velde, B. van Lith, and M. Klaassen. 2012b. Experimental quantification of long distance dispersal potential of aquatic snails in the gut of migratory birds. PLoS ONE **7**:e32292.

Wesselingh, F. P., G. C. Cadée, and W. Renema. 1999. Flying high: on the airborne dispersal of aquatic organisms as illustrated by the distribution histories of the gastropod genera Tryonia and Planorbarius. Geologie en Mijnbouw **78**:165–174.

Zeileis, A., C. Kleiber, and S. Jackman. 2008. Regression models for count data in R. Journal of Statistical Software **27**. Available from http://www.jstatsoft.org/v27/i08/.

Zhu, B., D. G. Fitzgerald, C. M. Mayer, L. G. Rudstam, and E. L. Mills. 2006. Alteration of ecosystem function by zebra mussels in Oneida Lake: Impacts on submerged macrophytes. Ecosystems **9**:1017–1028.

Table 1. Sampling results and odds ratio for each significant predictor in species-specific logistic regressions. Upper table corresponds to survey results and univariate logistic regressions and bottom table shows the best models selected by AIC using all variables

	Sampling results		*Odds Ratio (95% CI) for univariate* ***Development model*** *(for 1% increase)*	*Odds Ratio (95% CI) for univariate* ***Distance from Road model*** *(for 10 m increase)*
	Lakes present	*Lakes surveyed*		
Chinese mystery snail	160	437	0.99	0.80***
Banded mystery snail	139	436	1.03***	0.97
Rusty crayfish	52	295	1.02*	0.57***
Zebra mussel	22	429	1.04***	0.83
Eurasian watermilfoil	118	436	1.06***	0.90*
Curly-leaf pondweed	138	434	1.06***	0.85**

	Odds Ratio (95% CI) for best model out of all subsets		
	Conductivity (for 10 μs/cm increase)	*Development (for 1% increase)*	*Distance from Road (for 10 m increase)*
Chinese mystery snail	0.97 (0.95 – 0.99)		0.79 (0.69 – 0.89)
Banded mystery snail	1.06 (1.04 – 1.07)		0.96 (0.91 – 1.00)
Rusty crayfish			0.57 (0.42 – 0.75)
Zebra mussel	1.11 (1.08 – 1.16)		0.82 (0.59 – 0.99)
Eurasian watermilfoil	1.06 (1.04 – 1.08)	1.04 (1.02 – 1.05)	0.92 (0.82 – 0.99)
Curly-leaf pondweed	1.07 (1.05 – 1.09)	1.04 (1.03 – 1.06)	0.85 (0.75 – 0.95)

*significant at α=0.01, **significant at α=0.001,***p<0.001

Figure Captions

Fig. 1 Results of invasive species surveys across Wisconsin, USA. White dots indicate lakes that were not invaded by any species. The size of gray dots is proportional to the number of invasive species detected. Map inset indicates location of the study within the Laurentian Great Lakes region.

Fig. 2 The distance to road and proportion developed of sampled lakes. Curves along the axes are kernel density plots for the corresponding variable. Hotter/Lighter colors represent higher densities of points. Boxes distinguish remote and high human activity lakes from marginal human activity lakes.

Fig 3 Predicted number of aquatic invasive species based on zero-inflated Poisson regression. Panels show predictions based on percent development in a 100 m buffer (a-c) and minimum distance to road (d-f), while holding the other variable at 99% quantile, median, and minimum levels. Black lines show model prediction and gray polygons indicate a bootstrapped 95% CI for the prediction. Predictions are restricted to the space in which there are observations to fit the model.

Fig 4 Heatmap showing the predicted number of aquatic invasive species from a zero-inflated Poisson regression. Predictions are truncated to the space in which there are observations to fit the model.

Fig. 5 Violin plots showing the number of aquatic invasive species (AIS) at three levels of human activity. Categories match the cutoffs shown in Fig. 2. Groups are significantly different from each other ($p<0.001$). Black dots indicate the median, white rectangles indicate the interquartile range, and whiskers indicate minima and maxima. Gray polygons are mirrored kernel density plots. Numbers indicate the number of lakes surveyed.

Figure 1

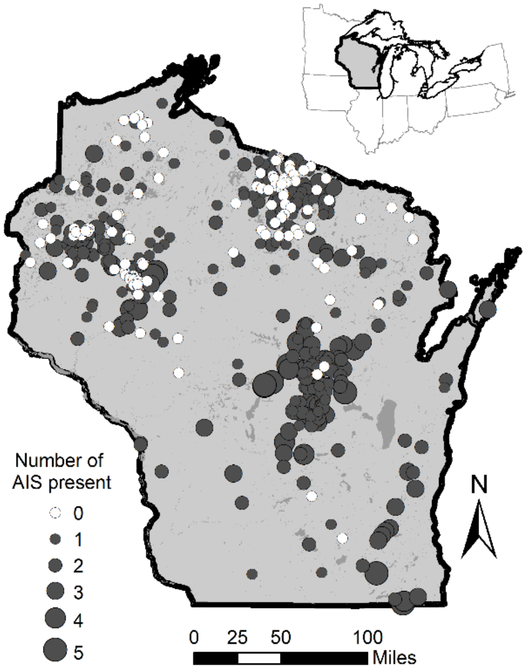

Fig. 1 Results of invasive species surveys across Wisconsin, USA. White dots indicate lakes that were not invaded by any species. The size of gray dots is proportional to the number of invasive species detected. Map inset indicates location of the study within the Laurentian Great Lakes region.

Figure 2

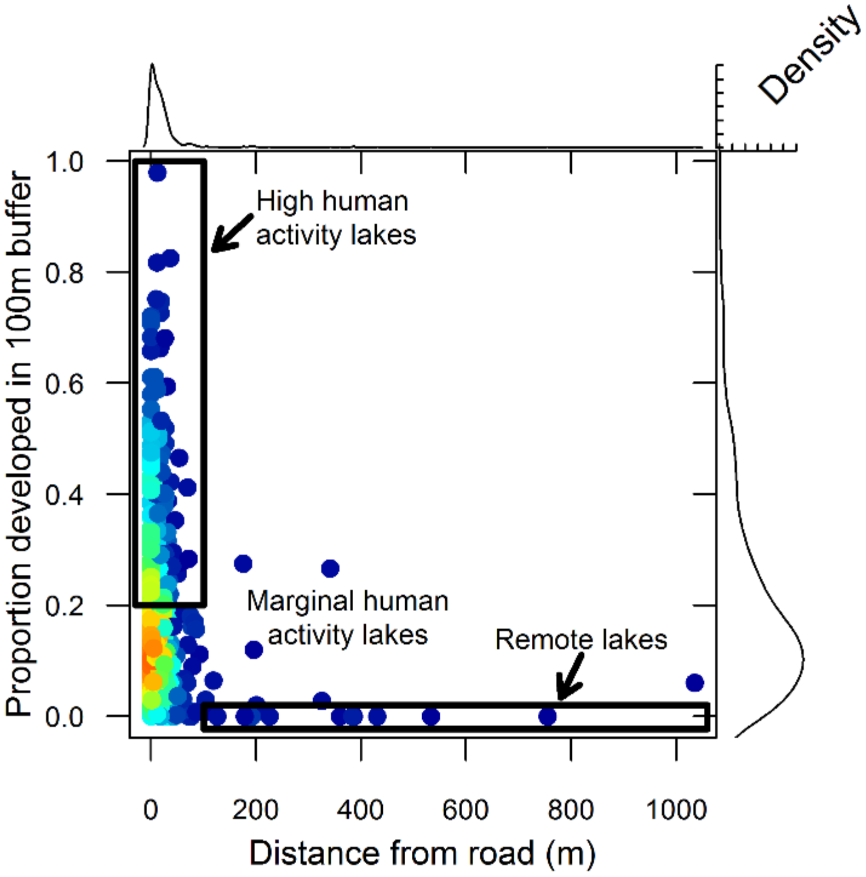

Fig. 2 The distance to road and proportion developed of sampled lakes. Curves along the axes are kernel density plots for the corresponding variable. Hotter/Lighter colors represent higher densities of points. Boxes distinguish remote and high human activity lakes from marginal human activity lakes.

Figure 3

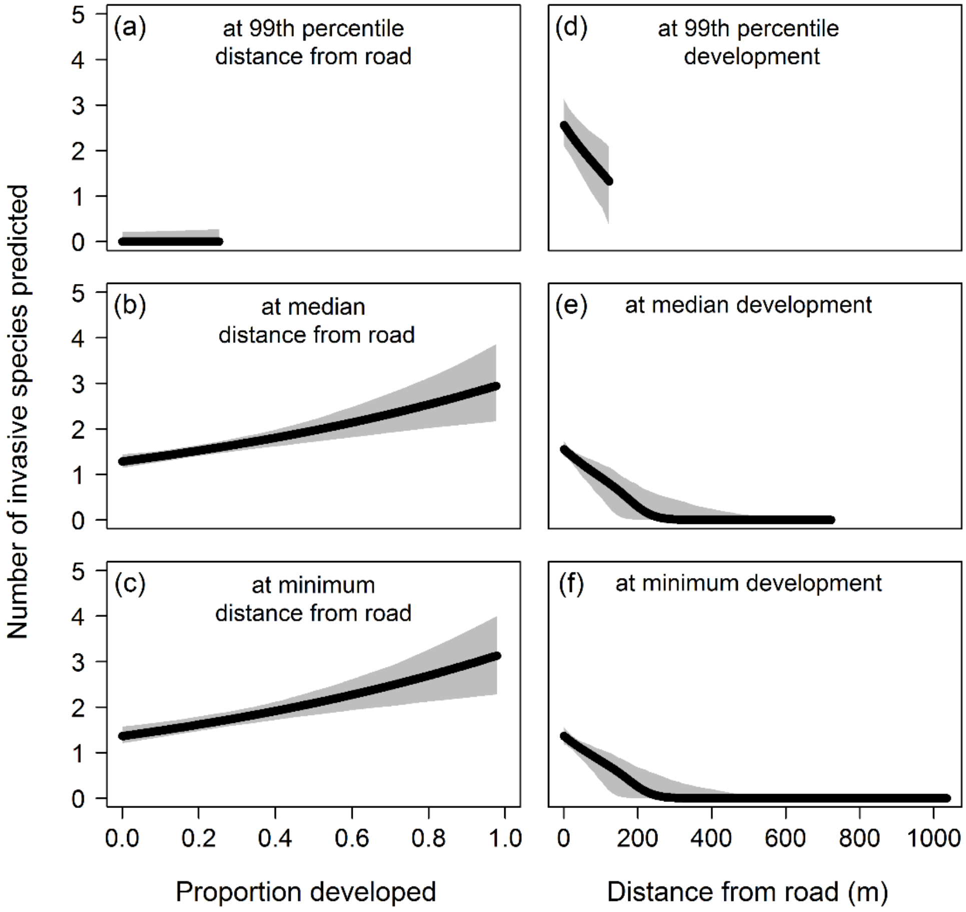

Fig 3 Predicted number of aquatic invasive species based on zero-inflated Poisson regression. Panels show predictions based on percent development in a 100 m buffer (a-c) and minimum distance to road (d-f), while holding the other variable at 99% quantile, median, and minimum levels. Black lines show model prediction and gray polygons indicate a bootstrapped 95% CI for the prediction. Predictions are restricted to the space in which there are observations to fit the model.

Figure 4

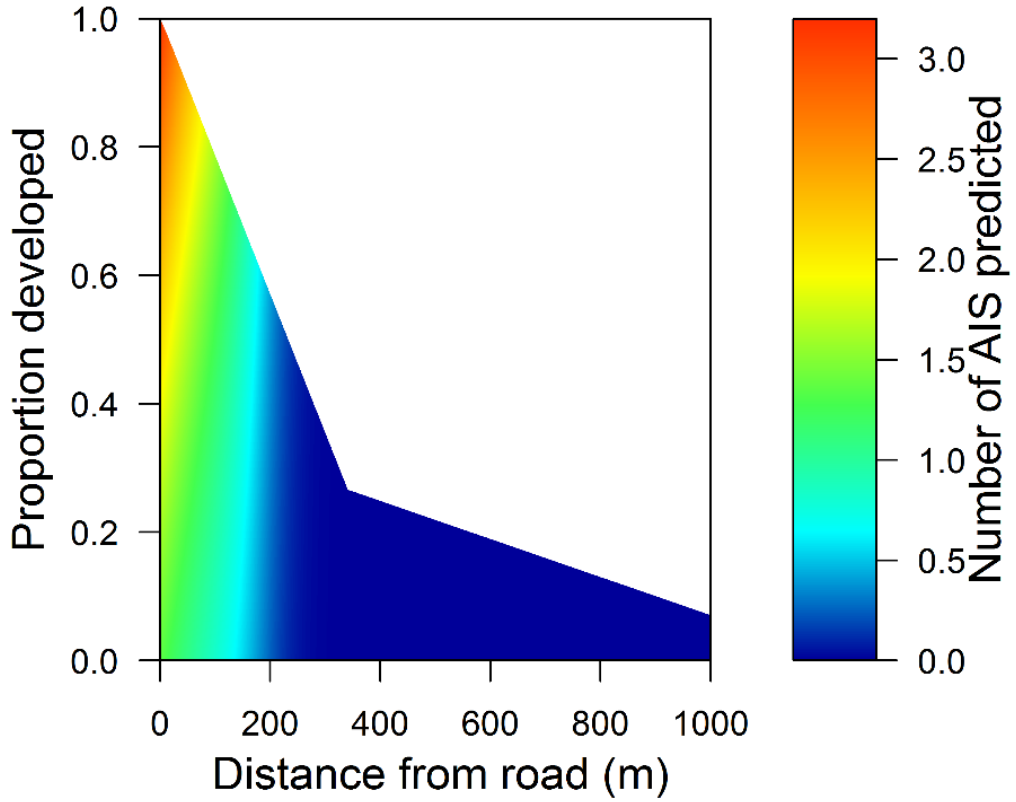

Fig 4 Heatmap showing the predicted number of aquatic invasive species from a zero-inflated Poisson regression. Predictions are truncated to the space in which there are observations to fit the model.

Figure 5

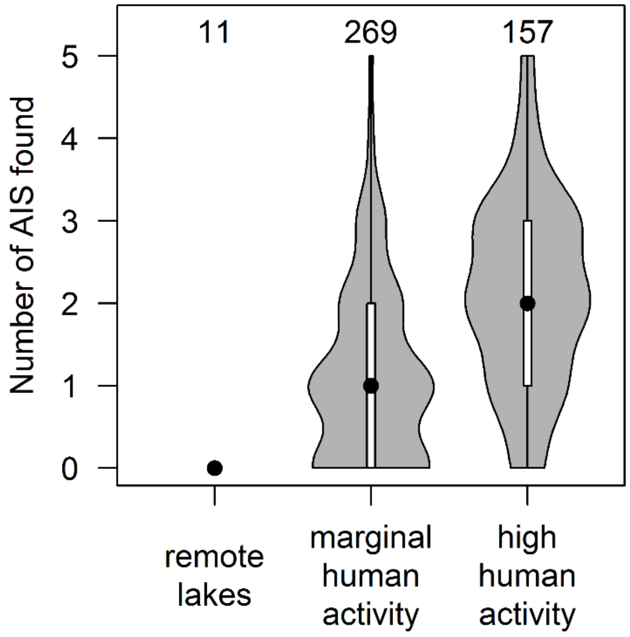

Fig. 5 Violin plots showing the number of aquatic invasive species (AIS) at three levels of human activity. Categories match the cutoffs shown in Fig. 2. Groups are significantly different from each other ($p<0.001$). Black dots indicate the median, white rectangles indicate the interquartile range, and whiskers indicate minima and maxima. Gray polygons are mirrored kernel density plots. Numbers indicate the number of lakes surveyed.

Appendix S1 – Description of invasive species surveys used in this analysis

We combined the results of two prior field surveys with surveys conducted for this study, with all sampling taking place between 2007 and 2012. The authors led two separate surveys that were incorporated into this dataset.

AWL and JVZ led a survey of 81 lakes in the Northern Highlands Lake District (Vilas and Oneida counties). Lake selection and survey methods are described in Latzka et al. 2015.

MF and SVE led a statewide Early Detection survey for the Wisconsin DNR, aimed mostly at lakes with public access. Lakes were selected randomly, but made to vary geographically in proportion to the relative occurrence of lakes in different regions across the state. Survey methods follow those in Latzka et al. 2015, although rusty crayfish were not sought in these surveys.

The two prior surveys were each attempts to census aquatic invasions within smaller regions of the state across all lakes that were reasonably accessible. We chose to combine these data with our own surveys in order to maximize coverage of difficult-to-access, and potentially less-invaded lakes. Data were supplied by original collectors to the authors.

The earlier of these studies was lead by Anna Mares, an employee of the Wisconsin DNR. The survey was conducted from 2007-09, and covered the counties of Barron, Chippewa, Dunn, Eau Claire, and Rusk counties, and was funded by WDNR Grant # AEPP-054-07, Xcel Energy Natural Resources Fund, and the Wisconsin. The survey covered 126 lakes, and covered all species listed in the Methods section. Most lakes had some form of public access (boat launch or pier). Each lake was surveyed 3 times during a single season, for one season out of the 3 years of the study, although data provided to the authors aggregated these results. Sampling methods included a rake-sampling method for aquatic macrophytes, crayfish trapping for rusty crayfish, net tows for zebra mussel veligers, and visual surveys for invasive snails.

The second provided dataset was from Paul Skawinski, from a consultant group contracted by the DNR. These surveys were conducted in 2009-2011 and covered lakes in Portage, Wood, Waushara, Waupaca, and Marathon counties. All species were sampled via visual detection during meandering kayak surveys through the littoral zone of each lake, although underwater cameras and rakes were sporadically used to supplement visual assessments. All six species covered in our analyses were sampled in these surveys.

Literature Cited

Latzka, A. W., J. T. Crawford, A. S. Koblings, Y. Caldeira, E. Hilts, and M. J. Vander Zanden. 2015. Representing calcification in distribution models for aquatic invasive species: surrogates perform as well as CaCO3 saturation state. Hydrobiologia **746**:197–208.

CHAPTER 2

REPRESENTING CALCIFICATION IN DISTRIBUTION MODELS FOR AQUATIC INVASIVE SPECIES: SURROGATES PERFORM AS WELL AS $CaCO_3$ SATURATION

Published as:
Latzka, A. W., J. T. Crawford, A. S. Koblings, Y. Caldeira, E. Hilts, and M. J. Vander Zanden. 2015. Representing calcification in distribution models for aquatic invasive species: surrogates perform as well as $CaCO_3$ saturation state. Hydrobiologia 746:197–208.

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.

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, 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 ($\Omega = [Ca^{2+}][CO_3^{2-}]/K'_{sp}$), 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; Free et al., 2009; Boets et al., 2013; Maceda-Veiga 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/m^2; 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 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 (NHLD; 45°54'N, 89°38'W) in Wisconsin contains approximately 1500 lakes, to which AIS are perceived as a dominant threat (Beardmore et al., 2015). 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 NHLD 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 minute snorkels at the primary boat launch or access point on each lake, and 10 minute 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 1m 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 2m 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 three 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

$$[CO_3] = \frac{C_t * K_{a1} * K_{a2}}{K_{a1} * K_{a2} + K_{a1} * [H^+] + [H^+]^2}$$

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 ($\Omega = [Ca^{2+}][CO_3^{2-}]/K'_{sp}$). We chose the K_{sp} constant by assuming a temperature of 25°C and the calcite mineral phase of $CaCO_3$: $K_{sp}=3.36*10^{-9}$. 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 homogenous 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 be 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 10-fold cross-validation, where 9/10 of the dataset was 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 4.1-8), 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 µS/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 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.

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 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 µS/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 µS/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 a 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 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 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 (Czarnołęski et al., 2006; Glass & Darby, 2009). High shell strength deters predation from fish and crayfish (Lewis & Magnuson, 1999; Czarnołęski 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 Jr., 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 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.

Acknowledgements

This material is based upon 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.

Literature Cited

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 CaCO3 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.

Beardmore, B., 2015. Boater perceptions of environmental issues affecting lakes in northern Wisconsin. JAWRA Journal of the American Water Resources Association 51: 537–549.

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, N.Y.

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 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.

Czarnołęski, 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.

Giraud, 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.

Guisan, A., & W. Thuiller, 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993–1009.

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, *Paranephrops 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. Hatziolos, 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 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 lamellibranchs. 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. Aloi, & A. P. Covich, 1994. Effects of an omnivorous crayfish (*Orconectes rusticus*) on a freshwater littoral food web. Ecology 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 (U.S.A.) 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. R Foundation for Statistical Computing, Vienna, Austria, 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.

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.

Figure Legends:

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

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

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 – conductivity: a, d, g; calcium: b, e, h; saturation state: c, f, h). Shaded boxes indicate trees that were selected as the best overall trees 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 figure 1 and units of measure are shown in figure 3.

Fig. 5 Calculated importance for each variable used in each (a) unpruned and (b) pruned classification trees, the latter of which are displayed in figure 4. A variable's importance is the sum of the decrease in error (number misclassified) by 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

Figure 1

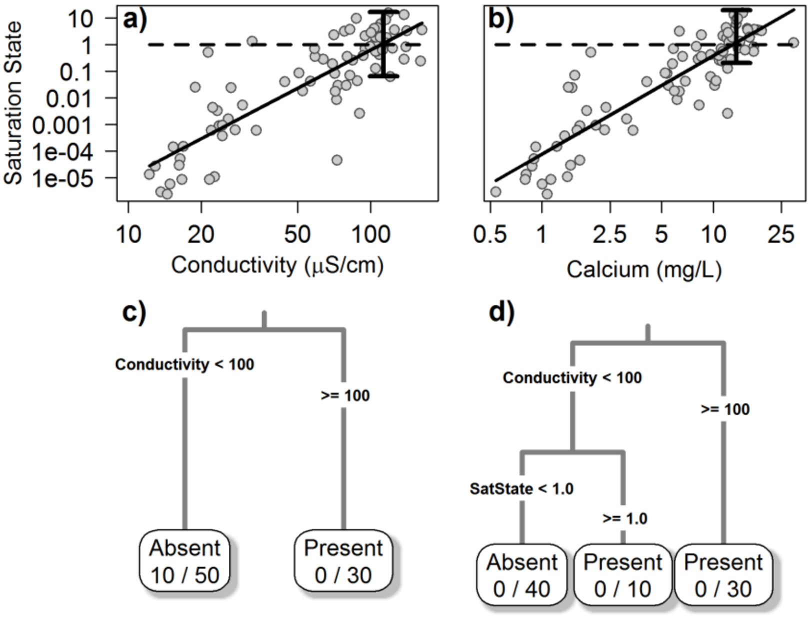

Fig. 1 a,b) Scatterplots of surrogates (conductivity (a) and calcium (b)) and saturation state, with values plotted on the log scale and raw values shown on the axes. Solid lines indicate the regression lines (conductivity: $p<0.001$, $R^2= 0.72$; calcium: $p<0.001$, $R^2=0.78$), dashed lines indicate 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

Figure 2

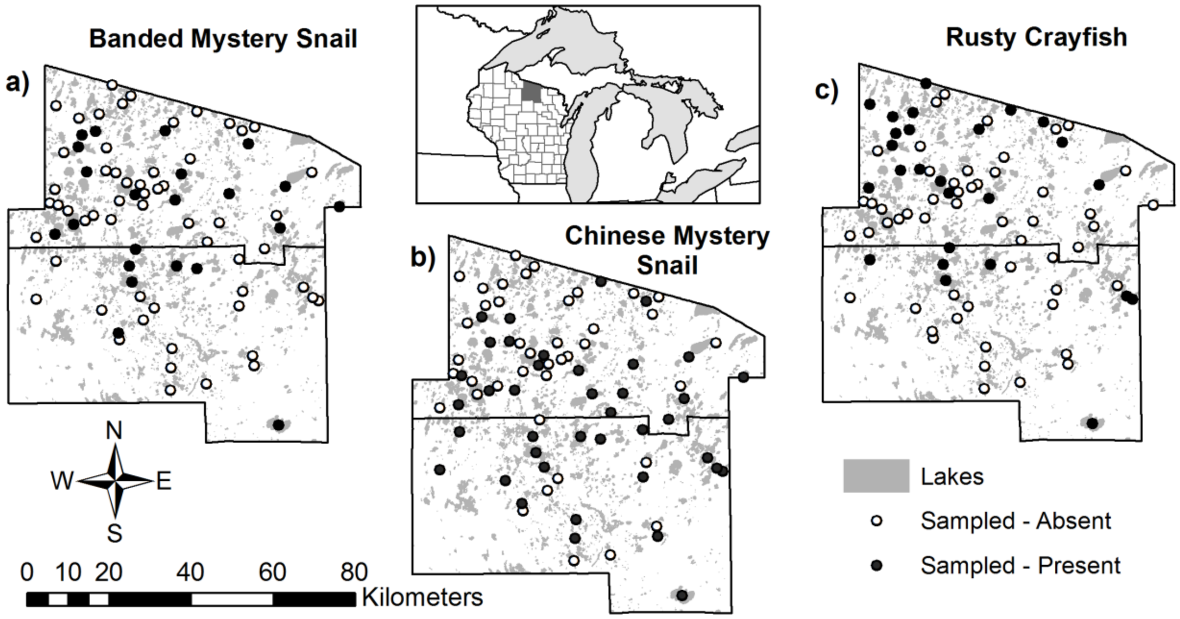

Fig. 2 Maps of surveyed lakes and the presence (black points) or absence (white points) of a) banded mystery snail, b) Chinese mystery snail, and c) rusty crayfish in Vilas and Oneida counties of Wisconsin, USA. Inset indicates the surveyed counties in dark gray within the Laurentian Great Lakes region, USA

Figure 3

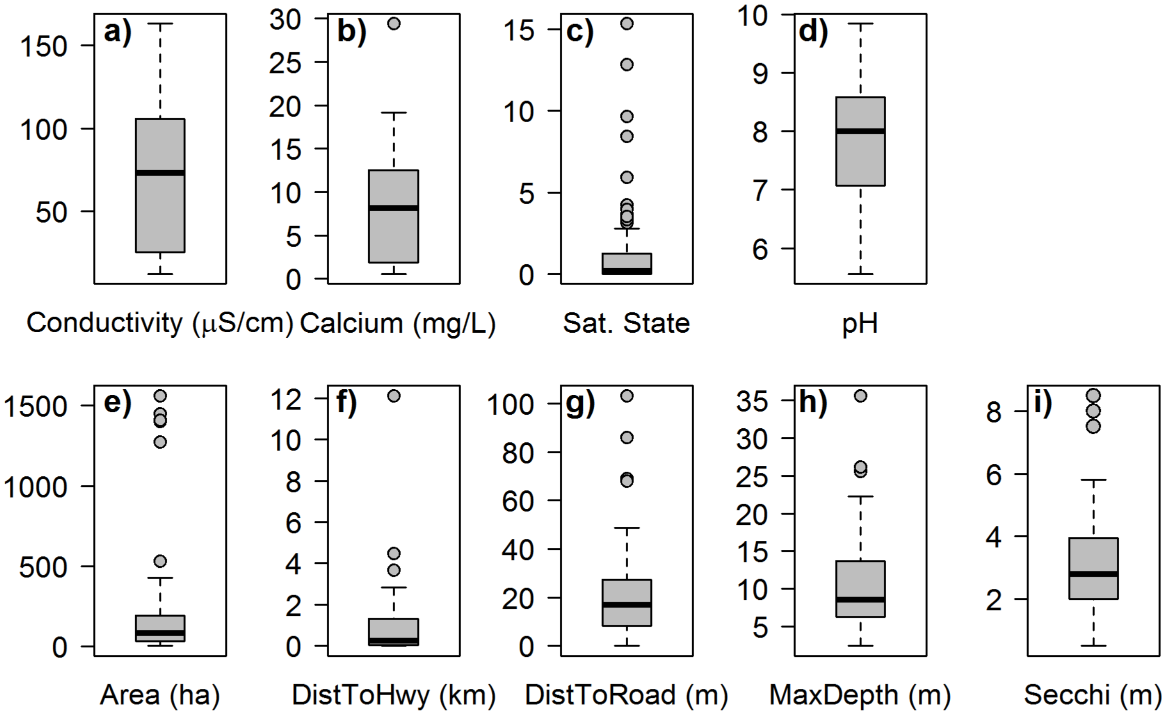

Fig. 3 Boxplots of calcification variables and others used in classification trees. The first row displays those variables that are related to calcification potential: a) conductivity, b) dissolved calcium, c) saturation state, and d) pH, which is kept out of the classification trees because it is used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

Figure 4

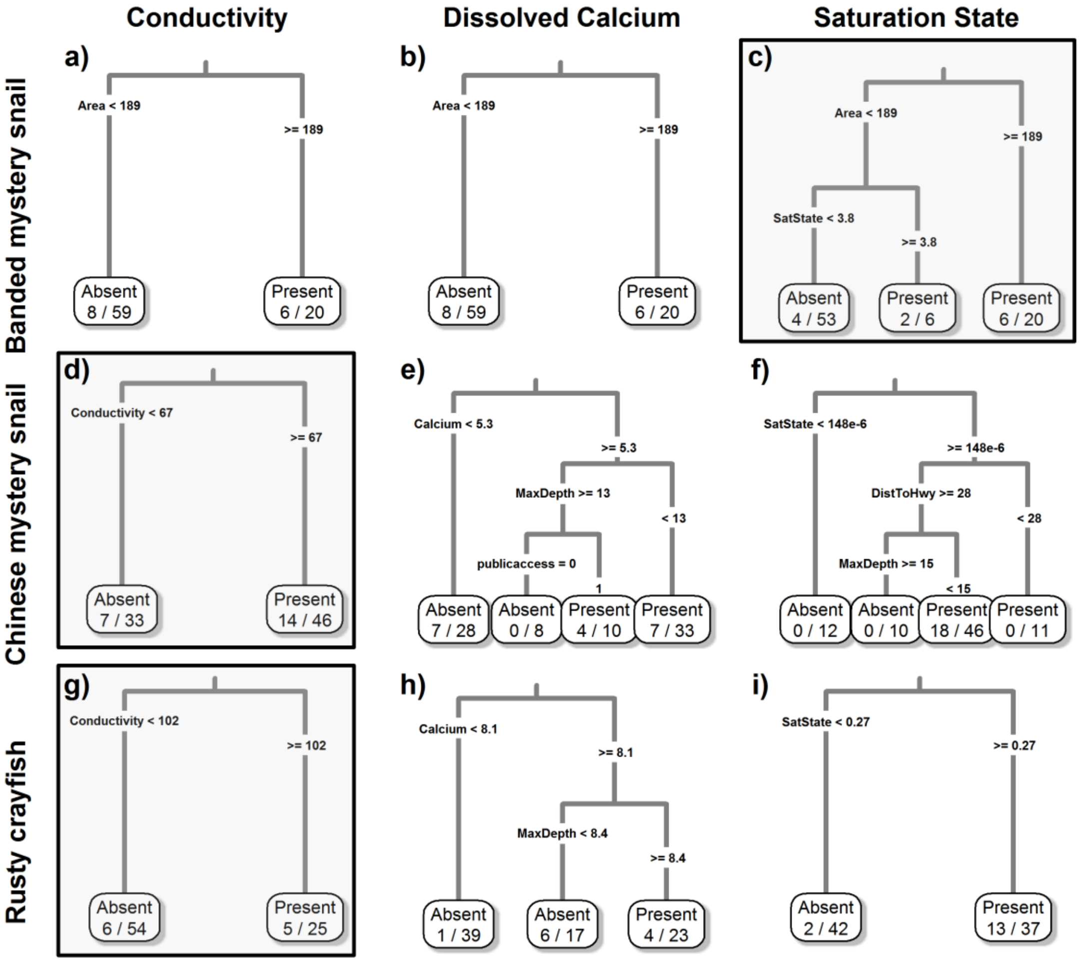

Fig. 4 Classification trees for each species (by row; a-c) banded mystery snail, d-f) Chinese mystery snail, g-i) rusty crayfish) and for each calcification variable that was used (by column – conductivity: a, d, g; calcium: b, e, h; saturation state: c, f, h). Shaded boxes indicate trees that were selected as the best overall trees 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 figure 1 and units of measure are shown in figure 3.

Figure 5

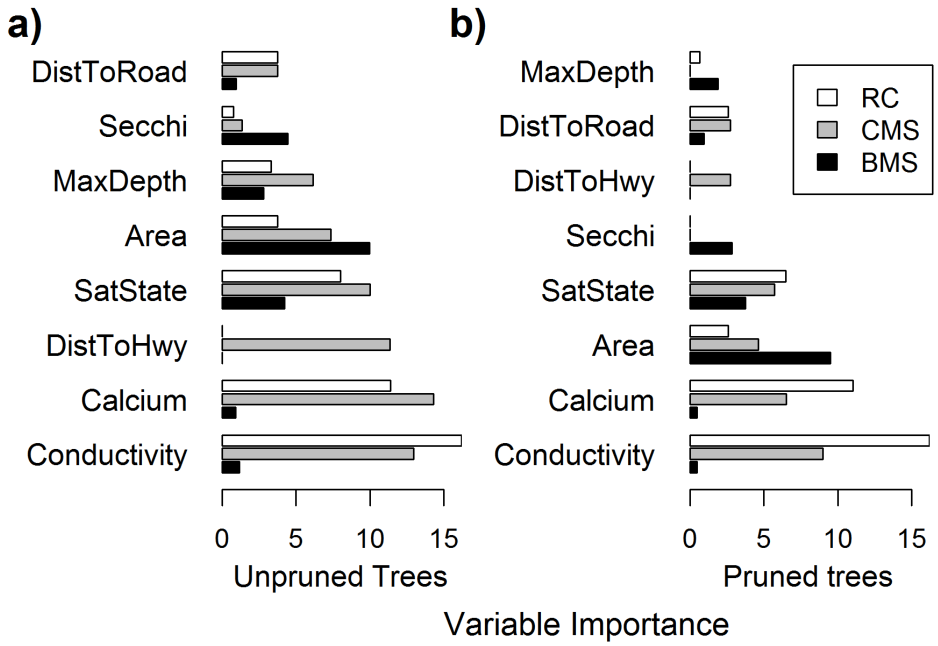

Fig. 5 Calculated importance for each variable used in each (a) unpruned and (b) pruned classification trees, the latter of which are displayed in figure 4. A variable's importance is the sum of the decrease in error (number misclassified) by using the variable either as a primary node, or as a substitute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

CHAPTER 4

HOW INVADED ARE WE? OCCURRENCE RECORDS UNDER-REPRESENT PREVALENCE OF INVASIONS

To be submitted for publication* with coauthors: Scott Van Egeran, Jake Vander Zanden
*Target journal: Frontiers in Ecology and the Environment

Abstract:

Invasive species are a leading driver of global change. Despite much research aimed at understanding their ecology and impacts, we have rarely assessed a fundamental feature of invasive species across landscapes: their prevalence. Invasive species prevalence may be important in making management decisions and assessing the scale of impacts. Moreover, we traditionally rely on invasive species occurrence datasets to inform management and predictive models, but these datasets may be vulnerable to data gaps, and are rarely compared to systematic surveys. In this study, we contrast existing species occurrence records with several estimates of AIS prevalence. We determine the prevalence of six invasive species (Chinese mystery snail [*Bellamya chinensis*], banded mystery snail [*Viviparus georgianus*], Eurasian watermilfoil [*Myriophyllum spicatum*], curly-leaf pondweed [*Potamogeton crispus*], rusty crayfish [*Orconectes rusticus*], and zebra mussel [*Dreissena polymorpha*]) from a stratified sample of 458 lakes in Wisconsin by scaling lake-specific rates of invasion up to the statewide scale. We found that prevalence, defined as the proportion of lakes invaded out of the total number of lakes in the state, is 39.0%, yet occurrence records document invasive species in 11.0%. This degree of underrepresentation was present across all species and lake types assessed, but strongest for small, inaccessible lakes and lesser-known species, suggesting that invasive species datasets may have data gaps similar to those in citizen science. Such data gaps may cause poor model

performance in distribution models, sub-optimal management decisions, and low estimates of overall invasive species impacts. Moreover, the difference between records and reality may be more extreme in states with less funding for invasive species management.

Introduction

Invasive species have frequent and sometimes high impacts on organisms, populations, communities, ecosystems, and human economies (Pyšek and Richardson 2010; Kumschick *et al.* 2015; Gallardo *et al.* 2015). Globally, invasions homogenize the global biota and contribute to the development of novel communities and ecosystems (Olden 2006; Qian and Ricklefs 2006; Ricciardi 2007). Because many of these effects are undesired, the cumulative economic costs exceed $120 billion annually in the US alone (Pimentel *et al.* 2005). While aggregate impacts are undoubtedly substantial, quantitative estimates are likely not well-constrained, because we have rarely tried to answer a key question: how common are invasive species on our landscapes? Fundamentally, how can we have a good understanding of a species' impacts, and thus make informed decisions on allocating management resources, without a good understanding of how common invasive species are across landscapes?

We can define prevalence simply as the proportion of sites invaded within a range or other geographic region of interest. In addition to being a fundamental aspect of our understanding of invasive species, an accurate picture of the scale of invasion may benefit management effectiveness, which relies on accurate predictions of spread (Vander Zanden and Olden 2008). If species distribution data are incomplete and prevalence is therefore underestimated, we do not have accurate information about source of spread, which may lead to poor model predictions and ineffective prevention efforts. Additionally, invasive species prevalence may affect management strategies. For instance, when a species is undergoing secondary spread, if it is present in a majority of sites, a defensive approach where sites without the invasive should be protected, but if the species is rare, an offensive approach where propagules from the few invaded sites are turned away at the source is optimal, all else being equal (Potapov *et al.* 2007; Drury and Rothlisberger 2008).

There are surprisingly few studies that have explicitly calculated invasive species prevalence, potentially due to insufficient data. Data on invasive species distributions and occurrences are compiled and housed in several publicly accessible databases, including MISIN (Midwest Invasive Species Information Network), GISIN (Global Invasive Species Information Network), USGS BISON (Biodiversity Information Serving Our Nation; not specific to invasive species), and EDDMapS (Early Detection and Distribution Mapping System). The goal guiding these databases is to include every known occurrence of each invasive species to inform rapid response efforts to new invasions. However, in order to accurately estimate prevalence, datasets must capture the variability present within a study area. Few invasive species occurrence datasets meet this standard. Of 252 datasets on invasive species presence or distributions, only 19% were deemed to have high quality spatial coverage (Crall et al. 2006). Even if a dataset does have full spatial coverage, the environmental heterogeneity within the study area may not be fully covered. Moreover, these data are often opportunistically collected or originate from a range of sources, including citizens, and are thus may be limited in their applicability, as other citizen science datasets sometimes demonstrate biased spatial and taxonomic coverage (Gardiner *et al.* 2012). Comparisons of such datasets to more complete data collected systematically are rare (but see (Szabo *et al.* 2012). Nonetheless, these datasets often form the empirical foundation on which many management decisions rest.

In contrast to potentially under-representative and/or biased species occurrence datasets, systematic sampling across various habitat or environmental strata can provide reliable prevalence estimates. For instance, information on the lionfish invasion traditionally comes from diver-based surveys, creating a depth-based bias, and data from stratified surveys revealed lionfish occurrence at depths beyond those reached by recreational divers, and thus, higher estimates of prevalence (Switzer *et al.* 2015). More generally, systematic or stratified sampling can be effective for informing species distribution models (Hirzel and Guisan 2002) and making inferences across an entire study region (Bunce *et al.* 1983; Rew *et al.* 2006).

Here, we use field surveys on a stratified sample of lakes to test whether existing records accurately represent the prevalence of several aquatic invasive species (AIS) in Wisconsin lakes. We compare prevalence estimates to the results of stratified surveys, using species that are thought to be relatively common or impactful in the state and which we predict to be well-documented in existing species occurrence records. Species include zebra mussels and Eurasian watermilfoil, which are often viewed as the AIS poster-children in the state, Chinese mystery snails and banded mystery snails, which are morphologically unique and visually conspicuous, and rusty crayfish and curly-leaf pondweed, which are moderately conspicuous and moderately well-known. Thus, we expect any differences in prevalence estimates to be conservative.

Methods

Data collection

Wisconsin, USA is a lake-rich landscape with 14,364 lakes, ponds, or reservoirs greater than 1 ha in area (WDNR/USGS Hydrography Layer 2005). To assess invasion of these lakes, we combined the results of four separate field surveys conducted between 2007 and 2012. Two regional (3-5 counties) field surveys sampled nearly all accessible lakes in the regions. Another regional survey in the Northern Highlands Lake District led by the authors used a stratified random sample of lakes from the region, which included isolated lakes (greater than 115 meters from the nearest road with less than 20% development in a 100m buffer). A survey led by the Wisconsin Department of Natural Resources selected lakes using a random selection or lakes with boat launches. After all surveys were completed, lakes were combined into one dataset and stratified according to the presence of a boat launch, high or low conductivity, and large or small surface area, plus isolated lakes. This stratification resulted in 9 lake types ranging from highly accessible and environmentally suitable to those with near-zero human activity during the open water season.

We surveyed all lakes for the presence or absence of several common AIS. Species surveyed included Chinese mystery snail (CMS; *Bellamya chinensis*), banded mystery snail (BMS; *Viviparus georgianus*), Eurasian watermilfoil (EWM; *Myriophyllum spicatum*), curly-leaf pondweed (CLP; *Potamogeton crispus*), rusty crayfish (RC; *Orconectes rusticus*), and zebra mussel (ZM; *Dreissena polymorpha*) (Figure 2 D-I, respectively). These species numerically dominate invasions in the state, making up 2,312 of 2,564 known AIS populations (http://dnr.wi.gov/lakes/invasives/, March, 2012). One survey did not target RC, and several sampling events at a small number of additional lakes excluded certain species, so these lakes were left out of the species-specific analyses. Since these species are visually conspicuous in shallow water, littoral zones were surveyed by snorkeling or canoeing. In the two regional surveys, the entire littoral zone was surveyed. In the statewide and NHLD surveys, boat launches and 5 sites among different shoreline habitats were snorkeled. When low water clarity interfered with visibility, 6 D-net samples of the benthos (~0.25m^2 each) were examined at each site.

Invasive species records

We compared the results of our field surveys to previously documented invasions. We retrieved datasets (as of November, 2013) from four different public invasive species databases, including MISIN (Midwest Invasive Species Information Network), GISIN (Global Invasive Species Information Network), USGS BISON (Biodiversity Information Serving Our Nation; not specific to invasive species), and EDDMapS (Early Detection and Distribution Mapping System). All data consisted of documented invasions by each species with associated geographic coordinates, which we spatially merge with a map of Wisconsin lakes. We considered any occurrence inside of or within 100 m of a lake to be an invasion in that lake. This process generated the list of lakes that were invaded by each species according to each dataset. In addition to these datasets, we retrieved data from the Wisconsin Department of Natural Resources, who aggregates invasive species occurrence records for informing management efforts. These records document all confirmed occurrences of invasive species in the state's waterbodies. Data can come from many sources, but encompass results of field surveys conducted by the DNR, local groups, and researchers, standardized sampling by a network of over 1000 volunteers, and voluntary citizen reports (http://dnr.wi.gov/lakes/invasives/).

For each database of invasive species records, we compared the number of lakes invaded by each species across all lakes in the state. We also compared the number of lakes invaded by each species in our field surveys to the number invaded for the same set of lakes in each occurrence dataset.

Invasive species prevalence

By scaling up the survey results from our stratified random sample of lakes, we estimated the proportions of lakes invaded across the state, for each lake type and for each species. We calculated the proportion of each lake type invaded and the associated 95% Wilson score confidence interval. To estimate the total proportion of Wisconsin lakes invaded, we summed the products of the proportion of each lake type invaded and the number of lakes in that type across the state (among 8004 lakes with data needed for classification). We generated a 95% confidence interval for the overall proportion invaded by propagating the standard error components of each Wilson score confidence interval (square root of sum of squared standard errors).

We then used these estimates of prevalence to make more detailed comparisons with occurrence records. We chose to focus these comparisons on the WDNR records, because they are used directly in management decisions. We compared our prevalence estimate among the 8,004 classifiable lakes in Wisconsin to the proportion of those lakes identified as being invaded in WDNR records. We made the same comparison for each specific species. Additionally, we calculated the proportion of each lake type known to be invaded, based on WDNR records, and multiplied by the number of lakes in that lake type across the state.

Results

Survey results

Of 458 lakes surveyed, 338 had at least one invasive species (Figure 1a). CMS were the most commonly found species (35.6% of lakes), followed by BMS (31.9%), CLP (31.2%), EWM (26.2%), and RC (12.4%), with ZM being the most rare (5.2%) (Table 1).

Occurrence records

Occurrence records demonstrated generally similar invasion statistics (Table 1). MISIN and GISIN, in particular, report very similar numbers, only differing by a maximum of four lakes across the state. BISON, on the other hand, contained very few occurrences of CMS, EWM, and CLP, and the highest number of occurrences of RC among non-DNR records. EDDMapS prioritizes invasive plants in its datasets, and resultingly had zero occurrences for CMS, BMS, RC, or ZM, but the numbers for EWM and CLP were comparable to those from MISIN and GISIN. The numbers of lakes invaded in WDNR records were higher than those from public databases for each species, except ZM. WDNR records indicate EWM as the most common AIS in the state, as it is present in 470 lakes, followed by CLP in 338 lakes, CMS in 291 lakes, RC in 244 lakes, BMS in 172 lakes, and ZM in 90 lakes.

Among sampled lakes, the patterns among the four public databases remained the same as above, and WDNR records had the highest number of occurrences for each species, including ZM. However, none of the records databases, including the WDNR dataset, contained nearly as many occurrences as documented in field surveys. Compared to the surveys, the WDNR dataset included only 68 of 163 occurrences of CMS, 78 of 146 occurrences of BMS, 83 of 143 occurrences of CLP, 46 of 57 occurrences of RC, 104 of 120 occurrences of EWM, and 20 of 24 occurrences of ZM.

Invasive species prevalence across Wisconsin and among lake types: Surveys vs. records

In our field surveys, lakes with public access were more invaded than those with only private access (Figure 1b). Isolated lakes were entirely uninvaded (Figure 1b). Within access classes, high conductivity and large surface area were associated with higher invasion rates

(Figure 1b). Similarly, when comparing only among lakes of similar size and conductivity, the presence of boat launches was also associated with higher invasion rates (Figure 1b). These patterns largely held true for single species patterns. However, relative to other species, especially EWM and ZM, CMS were notably present at higher rates in less accessible lakes (Figure 2).

Although 338 surveyed lakes had at least one non-native species, only 227 were indicated as invaded in WDNR records (Figure 1a). For all lake classes (except the isolated lakes) and species, invasion rates were underrepresented by WDNR records. The differences between prevalence estimates for surveys and records was largest for small, less accessible lakes and for CMS, BMS, and CLP than large lakes with boat launches and EWM, RC, and ZM (Figures 1b;2).

We can also compare estimates of statewide invasive species prevalence between WDNR records and surveys. In WDNR records, 955 of the 8,004 classifiable lakes in Wisconsin are listed as invaded, so we would estimate 11.9% of lakes in Wisconsin to be invaded (11.9%). From our stratified sample of lakes in Wisconsin, where we multiply the rate of invasion in each lake type by the number of lakes of that type across Wisconsin, and sum across lake types, we estimate actual invasive species prevalence to be 39.0% (95% CI bounds: 36.4-49.2%). The field surveys, therefore, increase our estimate of invasive species prevalence by a factor of 3.28. In the same comparison, the prevalence of each species increased by factors of 1.10 (ZM), 1.3 (EWM), 1.7 (RC), 3.2 (CLP), 7.6 (BMS), and 5.1 (CMS), for final prevalence estimates of 1.16%, 7.83%, 6.17%, 13.29%, 15.87%, and 18.47%, respectively.

Discussion

These results show that invasive species are much more prevalent than may be inferred from existing databases of occurrence records. Occurrence records managed by the Wisconsin DNR were slightly higher quality than publicly accessible databases (MISIN, GISIN, BISON, EDDMapS), but still failed to document many invasive species found in a stratified sample of Wisconsin lakes. We estimate prevalence of invasive species in Wisconsin lakes at 39%, whereas just 11.9% of the same lakes were previously listed as invaded. Although the records consistently under-represented the prevalence of invasions, the degree of under-representation varied among species and lake types. These results highlight the considerable scale at which invasions occur across our landscapes. Due to the substantial negative impacts of invasive species, our results also suggest that the landscape-wide impacts of invasions may be even more severe than we would infer from invasive species occurrence records.

Differences among species and lake types in the degree of mismatch between records and surveys are intuitive given standard data sources. Invasive species occurrence records, like much biodiversity data, accumulates occurrences from various data sources, including from the public and citizen scientists. Such datasets may be biased towards more easily accessible areas and toward charismatic or targeted rare species. For instance, citizen scientists studying lady beetles tended to misidentify common species as rare species (Gardiner *et al.* 2012). Additionally, plant observations are more frequently near roads than in a random distribution (Kadmon *et al.* 2004), and occurrences of passerine birds in sub-Saharan Africa are clustered around cities, roads, and rivers (Reddy and Dávalos 2003). Although these examples are not for invasive species datasets, the same patterns apply. Because the data are not collected systematically, they may be clustered in more easily accessible areas. We found that invasive species records were most reliable on large, easily accessible lakes. The difference in prevalence estimates between records and surveys increased as lakes got smaller and less accessible. There are fewer boaters on such lakes (Reed-Andersen *et al.* 2000), which reduces the probability that any invasive species will be detected and reported. Additionally, we found that occurrence records were most reliable for Eurasian watermilfoil and zebra mussels, which are the poster-children for AIS awareness in Wisconsin. Thus, they may be more readily identified and reported than lesser-known species like Chinese and banded mystery snails, which despite being the most common species in our surveys, demonstrated the largest differences between records and surveys, with prevalence estimates 5.1 and 7.6 times those that would be made from the records. These results suggest that public awareness of invasive species issues may in fact improve data quality. In addition to being the poster-children in invasive species outreach, zebra mussels and Eurasian watermilfoil likely have much higher impacts than these snails (Horsch and Lewis 2009; Solomon *et al.* 2009; Higgins and Vander Zanden 2010). Thus, records may be more reliable for highly impactful species.

Invasive species occurrence records, despite the limitations shown here, are becoming an increasingly common source of knowledge for invasion biology and management. There are many different databases in addition to those analyzed in this study, including Invasive Plant Atlases, the Weed Information Management System, the National Biological Information Infrastructure Invasive Species Information node, the Global Organism Detection and Mapping system, and the National Institute for Invasive Species Science database. These databases have been set up by agencies and universities, with several groups cooperating and sharing data through the Inter-American Biodiversity Information Network's Invasive Information Network (Simpson *et al.* 2009). As is the case in Wisconsin, environmental agencies may also keep their own dataset. GPS tools, websites, and cell phone apps have been data submission for some of these databases very simple for citizen contributions (Rice 2007; Graham *et al.* 2007). Thus, it is likely that these datasets will continue to grow and be widely utilized by researchers and managers. However, our results highlight that despite the high numbers of occurrences reached in some of these databases, they still may vastly under-represent local invasive species distributions and be treated with caution.

Perhaps the most common use of these data in research is to fit and validate invasive species distribution models. Such models predict the spread of invasive species to new areas, and thus inform prioritization efforts for prevention (Peterson and Vieglas 2001; Guisan and Thuiller 2005). However, datasets with incomplete environmental coverage – as we have demonstrated is the case for Wisconsin invasive species data with respect to small, inaccessible lakes – can strongly influence predicted responses and reduces the applicability of models to other areas and into the future (Thuiller *et al.* 2004). Thus, the poor quality of these datasets may worsen model predictions when used inappropriately. However, invasive species occurrence data are still a valuable contribution and should not be ignored. Our results show that simple, stratified surveys can reveal data gaps. Such surveys can also be used to fill data gaps and be used in conjunction with invasive species records to produce more reliable distribution models (Jarnevich *et al.* 2006; Szabo *et al.* 2012). Thus, we encourage cautious use of widely available species occurrence data in invasive species distribution modeling and, in particular, suggest carefully evaluating data for any data gaps that may arise with opportunistic sampling.

Estimating invasive species prevalence is rare, but it is a basic and crucial component in understanding invasions across landscapes. For instance, it can affect whether management should prioritize offensive or defensive management (i.e., containment vs. prevention) (Potapov *et al.* 2007; Drury and Rothlisberger 2008). Our results show that prevalence estimates can vary by factors ranging from 1.1 to 7.6, with the overall prevalence of AIS being 3.3 times higher than indicated in species occurrence records. These degrees of under-representation by occurrence records may cause managers to opt for offensive management, thinking that the species is/are only present at a small number of sites, but, depending on the costs of each strategy, it may actually be optimal to use the opposite strategy.

Likewise, prevalence can be used as an indicator of the scale of invasions, and can thus be used to evaluate costs and inform risk assessments. In order to accurately assess the costs of invasions, whether ecological impacts or an economic valuation of costs, we must have an idea of how many or what proportion of sites are invaded. In formalizations of a species' impact, prevalence is usually not directly accounted for. Parker *et al.* (1999) presented a useful framework for calculating species impact, wherein impact is the product of the species' invaded range area, its per capita impacts and its average abundance. Although prevalence is not include in this model or comparable models (Thomsen *et al.* 2011) directly, it is embedded within the abundance term, where a species' average abundance across its range is a function of the proportion of sites it has invaded within the range (i.e., prevalence). In impact assessments that involve summing across habitats or sites (Thiele *et al.* 2010; Barney *et al.* 2013; Latzka et al. *in revision*), an accurate picture of the prevalence of the invasive species is crucial. Underrepresenting prevalence by a factor of 3.3 in such models would lead to estimates of impact that would be 3.3 times less than the species' actual impact. Thus, we emphasize the importance of estimating and reporting invasive species prevalence to better inform our general understanding of invasive species impacts.

In addition to coordinating a network of over 1,000 volunteers to collect invasive species data on lakes, Wisconsin spent over US$12 million on aquatic invasive species management and research in 2009 and 2010 (Rosaen *et al.* 2012). In contrast, the other Great Lakes states spent between $92,732 (Ohio) and $7.8 million (Minnesota) (Rosaen *et al.* 2012). Such expenditures and data collection efforts could indicate that invasive species records for Wisconsin would be reliable. However, even with these expenditures, invasive species prevalence in records is underrepresented by a factor of 3.3. Given the lesser spending in other states, data gaps are likely to be much larger. Thus, we expect that our finding that invasive species records underrepresent the prevalence of invasive species is not unique, and compared to other regions, may actually be the best-case scenario.

In this study, we surveyed a stratified sample of lakes and found invasive species prevalence to be approximately 39%, which was 3.3 times higher than indicated in the best of several invasive species records databases. This surprising mismatch stresses that are knowledge of even well-studied invasive species is incomplete, and that our estimates of their impacts are poorly constrained, because we do not have reliable data to infer the basic prevalence of invasive species. With poor distribution data and under-representative picture of prevalence, management efforts may be undermined by poor distribution model performance. In areas with less spending and poorer data on invasive species, mismatches between occurrences records and reality may be even greater.

Acknowledgements

This material is based upon 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 Chelsey Blanke for assistance in compiling invasive species records databases, as well as staff of UW Trout Lake Station for their assistance with fieldwork and logistics.

References

Barney JN, Tekiela DR, Dollete ES, and Tomasek BJ. 2013. What is the “real” impact of invasive plant species? *Front Ecol Environ* **11**: 322–9.

Bunce RGH, Barr CJ, and Whittaker HA. 1983. A stratification system for ecological sampling. In: Fuller RM (Ed). Ecological mapping from ground, air and space. Cambridge: Institute of Terrestrial Ecology.

Drury KLS and Rothlisberger JD. 2008. Offense and defense in landscape-level invasion control. *Oikos* **117**: 182–90.

Gallardo B, Clavero M, Sánchez MI, and Vilà M. 2015. Global ecological impacts of invasive species in aquatic ecosystems. *Glob Change Biol*: n/a – n/a.

Gardiner MM, Allee LL, Brown PM, *et al.* 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Front Ecol Environ* **10**: 471–6.

Graham J, Newman G, Jarnevich C, *et al.* 2007. A global organism detection and monitoring system for non-native species. *Ecol Inform* **2**: 177–83.

Guisan A and Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol Lett* **8**: 993–1009.

Higgins SN and Zanden MJ Vander. 2010. What a difference a species makes: a meta-analysis of dreissenid mussel impacts on freshwater ecosystems. *Ecol Monogr* **80**: 179–96.

Hirzel A and Guisan A. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecol Model* **157**: 331–41.

Horsch EJ and Lewis DJ. 2009. The effects of aquatic invasive species on property values: evidence from a quasi-experiment. *Land Econ* **85**: 391–409.

Jarnevich CS, Stohlgren TJ, Barnett D, and Kartesz J. 2006. Filling in the gaps: modelling native species richness and invasions using spatially incomplete data. *Divers Distrib* **12**: 511–20.

Kadmon R, Farber O, and Danin A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol Appl* **14**: 401–13.

Kumschick S, Gaertner M, Vilà M, *et al.* 2015. Ecological Impacts of Alien Species: Quantification, Scope, Caveats, and Recommendations. *BioScience* **65**: 55–63.

Latzka, AW, Hansen GJA, Kornis M, and Vander Zanden, MJ. Spatial heterogeneity in invasive species impacts at the landscape scale. *In revision.*

Olden JD. 2006. Biotic homogenization: a new research agenda for conservation biogeography. *J Biogeogr* **33**: 2027–39.

Parker IM, Simberloff D, Lonsdale WM, *et al.* 1999. Impact: toward a framework for understanding the ecological effects of invaders. *Biol Invasions* **1**: 3–19.

Peterson AT and Vieglas DA. 2001. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem. *BioScience* **51**: 363–71.

Pimentel D, Zuniga R, and Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ* **52**: 273–88.

Potapov AB, Lewis MA, and Finnoff DC. 2007. OPTIMAL CONTROL OF BIOLOGICAL INVASIONS IN LAKE NETWORKS. *Nat Resour Model* **20**: 351–79.

Pyšek P and Richardson DM. 2010. Invasive species, environmental change and management, and health. *Annu Rev Environ Resour* **35**: 25–55.

Qian H and Ricklefs RE. 2006. The role of exotic species in homogenizing the North American flora. *Ecol Lett* **9**: 1293–8.

Reddy S and Dávalos LM. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *J Biogeogr* **30**: 1719–27.

Reed-Andersen T, Bennett EM, Jorgensen BS, *et al.* 2000. Distribution of recreational boating across lakes: do landscape variables affect recreational use? *Freshw Biol* **43**: 439–48.

Rew LJ, Maxwell BD, Dougher FL, and Aspinall R. 2006. Searching for a needle in a haystack: evaluating survey methods for non-indigenous plant species. *Biol Invasions* **8**: 523–39.

Ricciardi A. 2007. Are modern biological invasions an unprecedented form of global change? *Conserv Biol* **21**: 329–36.

Rice B. 2007. Invasive species data applications and data sharing across the Americas. *Front Ecol Environ* **5**: w15–6.

Rosaen AL, Grover EA, and Spencer CW. 2012. The Costs of Aquatic Invasive Species to Great Lakes States. East Lansing, MI: Anderson Economic Group.

Simpson A, Jarnevich C, Madsen J, *et al.* 2009. Invasive species information networks: collaboration at multiple scales for prevention, early detection, and rapid response to invasive alien species. *Biodiversity* **10**: 5–13.

Solomon CT, Olden JD, Johnson PTJ, *et al.* 2009. Distribution and community-level effects of the Chinese mystery snail (Bellamya chinensis) in northern Wisconsin lakes. *Biol Invasions* **12**: 1591–605.

Switzer TS, Tremain DM, Keenan SF, *et al.* 2015. Temporal and Spatial Dynamics of the Lionfish Invasion in the Eastern Gulf of Mexico: Perspectives from a Broadscale Trawl Survey. *Mar Coast Fish* **7**: 10–7.

Szabo JK, Fuller RA, and Possingham HP. 2012. A comparison of estimates of relative abundance from a weakly structured mass-participation bird atlas survey and a robustly designed monitoring scheme. *Ibis* **154**: 468–79.

Thiele J, Kollmann J, Markussen B, and Otte A. 2010. Impact assessment revisited: improving the theoretical basis for management of invasive alien species. *Biol Invasions* **12**: 2025–35.

Thomsen MS, Olden JD, Wernberg T, *et al.* 2011. A broad framework to organize and compare ecological invasion impacts. *Environ Res* **111**: 899–908.

Thuiller W, Brotons L, Araújo MB, and Lavorel S. 2004. Effects of restricting environmental range of data to project current and future species distributions. *Ecography* **27**: 165–72.

Zanden MJ Vander and Olden JD. 2008. A management framework for preventing the secondary spread of aquatic invasive species. *Can J Fish Aquat Sci* **65**: 1512–22.

Table 1. Number of lakes invaded by each species according to various invasive species occurrence databases, Wisconsin DNR records, and survey results for all lakes in Wisconsin and surveyed lakes

	Number of documented invasions for all lakes in Wisconsin (n=14,364):					
	Database:					
Species	BISON	GISIN	MISIN	EddMAPS	WDNR (verified and vouchered)	
BMS	164	171	171		172	
CMS	0	229	229		291	
EWM	6	517	513	500	470	
CLP	5	365	362	358	338	
RC	164	129	129		244	
ZM	110	111	112		90	
	Number of documented invasions for surveyed lakes (n=458):					
	Database:					
Species	BISON	GISIN	MISIN	EddMAPS	WDNR (verified and vouchered)	Surveys
BMS	61	63	63		78	146
CMS	0	55	55		68	163
EWM	0	106	106	103	104	120
CLP	0	102	102	102	83	143
RC	61	27	27		46	57
ZM	17	18	18		20	24

Figure Legends

Fig. 1 Statewide comparison of Wisconsin Department of Natural Resources invasive species occurrence records and our survey results. Blue dots and bars are lakes that were not invaded according to surveys or records, yellow dots and bars are lakes that were already known to be invaded based on records, and red dots and bars are lakes where new invasions were found. (a) A map of sampled lakes. (b) The proportion of lakes invaded in each lake type category.

Fig. 2 Survey and occurrence records comparisons broken down by species and lake type. Colors match those in Fig. 1.

Figure 1.

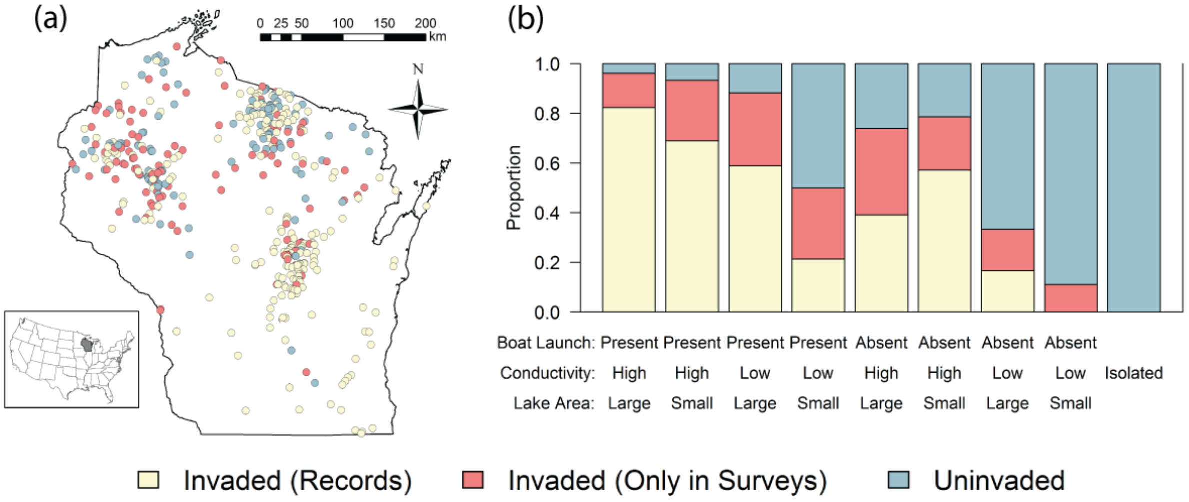

Fig. 1 Statewide comparison of Wisconsin Department of Natural Resources invasive species occurrence records and our survey results. Blue dots and bars are lakes that were not invaded according to surveys or records, yellow dots and bars are lakes that were already known to be invaded based on records, and red dots and bars are lakes where new invasions were found. (a) A map of sampled lakes. (b) The proportion of lakes invaded in each lake type category.

Figure 2.

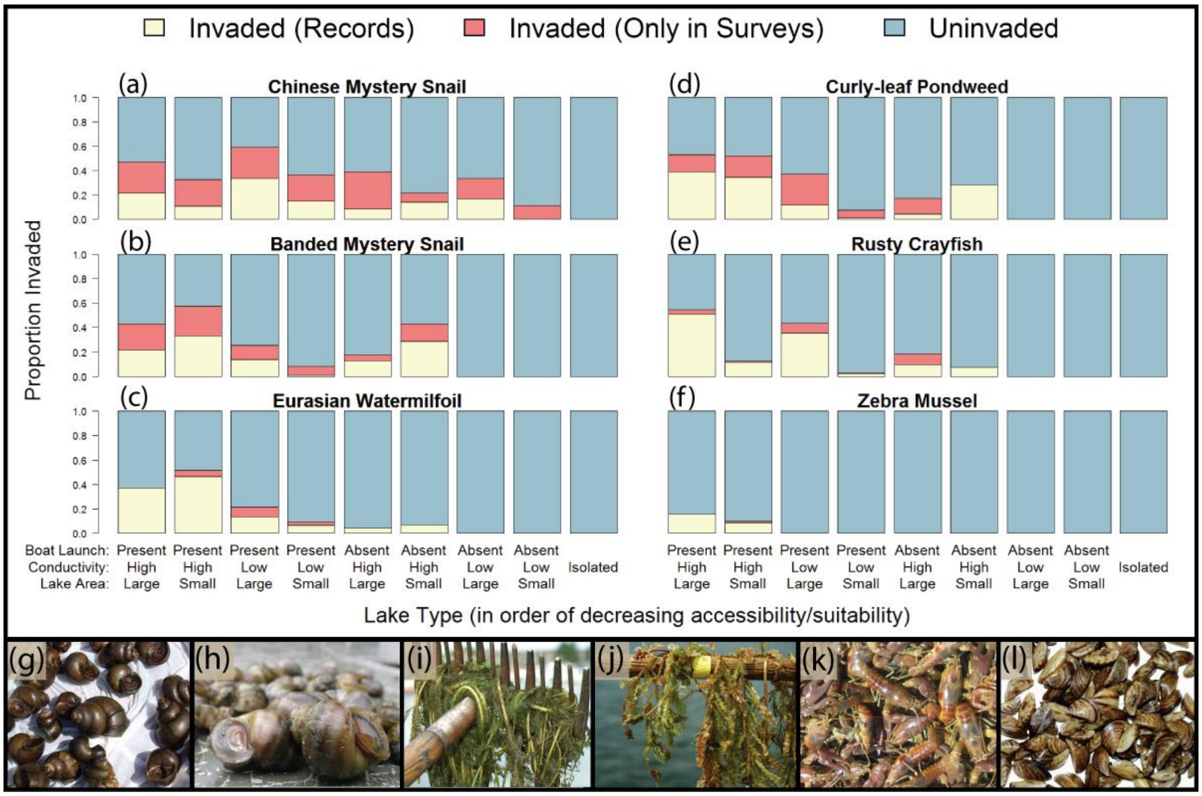

Fig. 2 Survey and occurrence records comparisons broken down by species and lake type. Colors match those in Fig. 1.

CHAPTER 4

SPATIAL HETEROGENEITY IN INVASIVE SPECIES IMPACTS AT THE LANDSCAPE SCALE

This chapter has been submitted for publication, and is currently *in revision* at Ecosphere, with coauthors: Gretchen J. A. Hansen, Matthew Kornis, and M. Jake Vander Zanden.

Abstract:

Invasive species have substantial impacts across the globe. While management efforts should aim to minimize undesirable impacts, we have a poor understanding of how impacts of a given invasive species vary spatially. Here, we develop a framework for considering heterogeneity of invasive species impacts that allows us to explore the range of possible spatial patterns of impact. This framework incorporates two factors – how invasive species abundance varies among sites (i.e., abundance distributions) and how invasive species impact varies as a function of abundance (i.e., abundance-impact curves). Combining these two factors allows for the creation of probability distributions that represent how invasive species impacts may vary spatially among sites. We simulate how different combinations of invasive species abundance distributions and abundance-impact curves produce different patterns of invasive species impact. These simulations illustrate a remarkably wide range of invasive species spatial impact patterns – probability distributions of impact were left-skewed, right-skewed, bimodal, and normal. Total landscape-level impacts, estimated by summing site-level impacts, were similarly variable depending on the distribution of site-level impacts. To demonstrate the real-world application of this framework, we generated impact distributions for two problematic invasive species – zebra mussel and Eurasian watermilfoil – across lakes in Wisconsin, USA, where invaded lakes are documented and abundance distributions and abundance-impact curves can be inferred from the literature. Impact distributions of these species tended to be right-skewed (i.e., the majority of sites had low impacts). Total impacts depended strongly on the type of abundance-impact curve. Our results indicate that invasive species abundance and abundance-impact curves ultimately affect how invasive species impacts are distributed across the landscape, which has important implications for invasive species management.

Introduction

Invasive species are among the leading drivers of global change (Sala et al. 2000), causing substantial ecological (Ricciardi et al. 2013) and economic impacts (Pimentel et al. 2005). Minimizing the undesirable impacts of invasions is a major goal of environmental agencies (Byers et al. 2002), but achieving this goal is non-trivial given the problem's scope and continued growth. The large number of invasive species to consider, vast landscapes with many invaded or vulnerable sites, heterogeneity of impacts among sites, and the suite of management options (prevention, containment, control, eradication) all contribute to the complexity of invasive species management (Byers et al. 2002).

Funding for invasive species management pales in comparison to the scale and complexity of the problem (Simberloff et al. 2005) and thus, understanding patterns of impact is necessary to use resources cost-effectively. Resource management is often confronted with this inequality between a problem's magnitude and available management resources, necessitating action plans that achieve the largest benefit per unit cost. Global biodiversity conservation, for instance, maps biodiversity hotspots to help maximize species protection given limited funding for land acquisition (Myers et al. 2000, Brooks et al. 2006, Game et al. 2013). An analogous approach for invasive species management may be to identify and target sites on the landscape where impact is likely to be highest in order to maximize the impact reduction per dollar spent (Epanchin-Niell and Hastings 2010). Unfortunately, current understanding of the spatial patterns of invasive species impacts is inadequate to inform this sort of prioritization. Although several impact assessment frameworks have been developed to estimate and compare species' total impacts (Parker et al. 1999, Thomsen et al. 2011 and citations therein), few have incorporated spatial heterogeneity of impacts among sites (but see Thiele et al. 2010, Barney et al. 2013).

Heterogeneity of invasive species impacts among sites encompasses two critical components: spatial heterogeneity in invasive species abundance, which is a universal driver of impact (Thomsen et al. 2011), and the relationship between a species' abundance and its impact (Fig. 1). Neither of these components is well-described empirically for the majority of invasive species. Variation in invasive species abundance across the landscape may take many possible forms (Fig. 1A). Native species generally occur at low abundance at most sites where they occur (Brown 1984), and a number of aquatic invasive species (AIS) follow similar patterns, producing right-skewed frequency distributions of abundance (hereafter "abundance distribution") (Hansen et al. 2013). In contrast, invasive species are sometimes characterized by their tendency to dominate a community, as demonstrated empirically for some monotype-forming wetland plants (Zedler and Kercher 2004, Frieswyk et al. 2007) and anecdotally for some terrestrial plants such as kudzu, producing left-skewed abundance distributions. Furthermore, understanding spatial variability in impacts requires knowledge of the relationship between abundance and impact (hereafter "abundance-impact curve). Although this relationship is likely to be positive in most cases (but see Kornis et al. 2014), the shape of this curve may take different forms for different species (Fig. 1B; Yokomizo et al. 2009). However, abundance-impact curves have not been quantified for most invasive species (Kulhanek et al. 2011), hindering estimates of impact from abundance data.

Here, we combine a range of possible abundance distributions with various abundance-impact curves to generate species' probability distributions of site-level impacts (hereafter "impact distributions" (Fig. 1). Our objective is to gauge the extent of possible invasive species impact distributions, and how those patterns are influenced by abundance-impact curves and patterns in abundance. We also explore how these impact distributions may influence an invasive species' cumulative impacts. We then apply our framework for generating impact distributions to two aquatic invasive species in Wisconsin in an empirically-informed management scenario that highlights which possibilities may be realistic, and how a manager could use currently available information to generate impact distributions.

Methods

Invasive species impacts on a heterogeneous landscape – a conceptual framework

In order to characterize heterogeneity in impacts across invaded landscapes, we first need to break the landscape into components of interest – lakes, habitat patches, pixels, etc. – which we refer to generally as sites ($s=1 \ldots n$). A subset of these sites ($\leq s_n$), determined by environmental characteristics matching species' ecological requirements and dispersal abilities, are inhabited by an invasive species. Across these inhabited sites, the species' abundance (a_s) will vary, and how it varies will differ among species. We can use probability distributions of abundance to characterize this variation ($P(a_s)$) (Fig. 1A), which may be theoretical or empirical if average or snapshot estimates of site-level abundance are available. Next, we need to describe the relationship between impact and abundance. Impacts at a site are likely to be positively related to species abundance, but the shape of this relationship will vary among species and among types of impact (Fig. 1B). We can thus use various abundance-impact curves to calculate a site-level impact (I_s) for any value of abundance drawn from an abundance distribution: $I_s = f_{AI}(a_s)$ (Fig. 1). For an invaded landscape, we can then sum all site-level impacts to determine the species' landscape-level impact: $I_T = \sum_{s=1}^{n}(I_s) = \sum_{s=1}^{n}(f_{AI}(a_s))$, where I_T is analogous to impact estimates in previous formalizations (e.g., Parker et al 1999).

Simulating site-level and landscape-level impacts

Using the above conceptual framework, we used abundance distributions and abundance-impact curves to calculate a range of potential impact distributions (Fig. 1). We used four hypothetical probability distributions to represent a broad range of species abundance patterns (Fig. 1A). In all distributions, abundance was expressed on a standardized scale ranging from 0 to 1. We used three Beta distributions with varying shape parameters α and β to represent centered (α=3, β=3), right-skewed (α=0.8, β=3), and left-skewed (α=3, β=0.8) distributions. We also used a uniform distribution (ranging from 0-1) to represent equal probability of invasive species occurring at any abundance.

We modeled the possible relationships between species' abundance and impact using the four curves described by Yokomizo et al. (2009). These curves include a low abundance threshold where high impacts occur at all but the lowest abundances ("Low threshold", Type 1), a sigmoidal curve where impacts start low then increase rapidly at moderate abundance and asymptote at high abundance ("Sigmoidal", Type 2), a linear response ("Linear", Type 3), and a high abundance threshold in which high impacts only occur at high abundances ("High threshold", Type 4) (Fig. 1B). We standardized abundance values to range from 0 to 1 and scaled impact on a percentage scale (i.e., 0=no impact, 100=maximum possible impact). Since the asymptotic curves never reach their maximum potential impact, we assumed that abundances greater than 0.9 produced the maximum impact (results were not sensitive to variations in this threshold).

To quantify the distributions of site-level impacts, we randomly drew 1,000 values from each of the four abundance distributions (Fig. 1A) to represent site-level abundances across an invaded range. For each value, we determined its associated impact using each of the four abundance-impact curves (Fig. 1B; described in Yokomizo et al. 2009 and Appendix 1), resulting in 16 unique combinations of abundance distributions and abundance-impact curves. We recognize that a single species may have multiple impacts, and that each impact may follow a different abundance-impact curve. For the purposes of this simulation, impact can be considered any measurable impact or an index that integrates multiple types of impact.

We calculated landscape-level total impacts by summing all site-level impacts. We recognize that assuming additivity may not always be appropriate – in some cases, managers may be particularly concerned with highly impacted sites, so that one site with maximum impact is judged to be worse than 10 sites with 10% of maximum impact. In other cases, managers may be concerned with the number of invasions, so that the 10 invaded sites with low impact are worse than one site with high impact. In this analysis, we kept the number of sites constant at 1,000 in order to isolate the effects of abundance distributions and abundance-impact curves on total impacts, but changes in the number of sites invaded would cause proportional changes in total impacts.

Results

Simulated impacts were highly variable among sites (Fig. 2). For most scenarios, the frequency distributions of site-level impacts were highly skewed: six with left skew and three with right, although some of these also display bimodality. When site-level abundances were left-skewed (row 2 in Fig. 2), three of the four scenarios produced left-skewed impact distributions. The exception, which paired a left-skewed abundance distribution with a high-abundance impact threshold, produced a slightly left-skewed bimodal impact distribution. When the abundance distribution was right-skewed (row 3 in Fig. 2), three of the four scenarios produced right-skewed impact distributions. The exception, which paired a right-skewed abundance distribution with a low-abundance impact threshold, produced a slightly left-skewed bimodal impact distribution. Impact distributions created by low- and high-abundance impact thresholds (columns 1 and 4 in Fig. 2) were similar to those created by left-skewed and right-skewed abundance distributions, respectively.

Landscape-level impacts were sensitive to both the abundance distribution and the abundance-impact curve (Fig. 3). Values ranged from 1,884 to 99,500, where a value of 100,000 would indicate that a species achieved its maximum possible impact of 100 at each of the 1,000 sites in the landscape. When abundance distributions were left-skewed, landscape-level impacts were always high, with three of four species exhibiting impacts greater than 85% of the maximum (i.e., >85,000) and all with impacts above 60% of the maximum. In contrast, when abundance distributions were right-skewed, three of the four scenarios had landscape-level impacts below 25% of the maximum observed total impact, and all had total impacts below 65% of the maximum. Similarly, abundance-impact curves that produced high impacts at low abundances tended to produce high landscape-level impacts, while those that only exhibit impacts at high abundances were less impactful.

Discussion

Impact distributions

Our results highlight how abundance distributions and abundance-impact relationships of invasive species affect patterns of heterogeneity in impacts across sites. The shapes of impact distributions were variable, where some scenarios produced a high relative probability of high site-level impacts, others a high probability of low impacts, and others normal, bimodal, or uniform distributions of impacts across sites (Fig. 2). In some cases, knowledge of just one of the attributes – the abundance distribution or abundance-impact curve – was enough to approximate the shape of potential impact distributions, making predicting the pattern of heterogeneity possible even with limited information. Invasive species with left-skewed abundance distributions or low-abundance impact thresholds will almost always exhibit left-skewed impact distributions. In these cases, targeted management that prioritizes some sites over others may be inefficient, because all sites have similarly high impacts. In contrast, species with right-skewed abundance distributions or high-abundance impact thresholds almost always exhibit right-skewed impact distributions, where high impact sites are relatively rare. In these cases, prioritized approaches to management could be promising, because reducing or eliminating impacts at a small number of highly impacted sites may significantly reduce total impacts.

Landscape-level impacts

Landscape-level impacts were sensitive to both the abundance distribution and the abundance-impact curve of the species, where species with left-skewed abundance distributions and low threshold abundance-impact curves exhibited the highest total impacts (Fig. 3). Species with these attributes could have total impacts much higher than those with right-skewed abundance distributions or high threshold abundance impact-curves, respectively (assuming equal number of occupied sites). Note that such a comparison assumes that the maximum site-level impacts of species are equal. As with the impact distributions, knowledge of just one attribute – the abundance distribution or abundance-impact curve – can greatly improve predictions of total impacts. For example, invasive species with left-skewed abundance distributions or low-abundance impact thresholds will almost always achieve relatively high total impacts, all else being equal.

Abundance distributions and abundance-impact curves

Despite their importance in determining landscape-level impacts and the heterogeneity therein, abundance distributions have rarely been quantified for invasive species. Hansen et al. (2013) reported right-skewed abundance distributions for several high-impact aquatic invasive species from a wide range of taxonomic groups and geographic locations. However, invasive species may not always follow right-skewed abundance distributions. Invasive terrestrial plants seem to have great potential for centered, uniform or left-skewed distributions, which may align with anecdotal evidence of invasive species “dominance”. For example, several invasive plants are known for forming monotypes where the invader has high abundance with few co-occurring species (Frieswyk et al. 2007). Zedler and Kercher (2004) examined 21 commonly occurring invasive wetland plants, and note that 15 of these species (71%) form monotypes in invaded regions. Monotype-forming invaders may be best described by left-skewed distributions - when present, these species are usually found at high abundance. While these examples indicate that left-skewed abundance distributions may be possible for invasive species, future work should focus on quantifying abundance distributions for a broad range of taxa. Existing site-level abundance data may be available in public databases, enabling quantification of abundance-impact curves for other taxa.

Similar to abundance distributions, abundance-impact curves are not yet described for even the most harmful invasive species (Kulhanek et al. 2011), despite their substantial effect on landscape-level impacts and their heterogeneity. Nonetheless, there are empirical examples supporting the abundance-impact relationships used here (initially proposed by Yokomizo et al. 2009). Linear responses frequently occur in nature for processes such as litter decomposition (Elgersma and Ehrenfeld 2011) and filter feeding or grazing (Holling 1965). For example, filter-feeding by zebra mussels (*Dreissena polymorpha*) reduces phytoplankton biomass, which in turn has a linear effect on zooplankton biomass (Higgins and Vander Zanden 2010). Elgersma and Ehrenfeld (2011) documented a high-abundance threshold response of the ratio of soil bacteria to fungi across abundances of invasive Japanese barberry (*Berberis thunbergii*). High-abundance thresholds may be common for impacts that are met with a certain amount of resistance at low abundance, such as the decline/extirpation of native species that would otherwise be impacted by predation or competition at high invasive species abundance. Low-abundance impact thresholds may occur for invasive species that function as ecosystem engineers (Crooks 2002) or keystone species, which have effects disproportionate to their density (Paine 1969, Power et al. 1996). Sigmoidal response curves are the least supported by examples from the literature. However, invasive species often serve as prey in invaded ecosystems (Kitchell et al. 2000, King et al. 2006). Predators of invasive species may switch from native prey following invasion if the invasive species' density is sufficient for detection and learning by predators. Such predator/prey relationships are best modeled by a sigmoidal curve (or Type-III functional response) (Holling 1965). It should be noted that these four abundance-impact curves are not exhaustive of all possible relationships – thresholds may occur anywhere along the abundance axis, for instance. Additionally, fundamentally different curves may occur for some species. For example, experimental evidence suggests that the negative effects of an invasive fish, round goby (*Neogobius melanostomus*), on the growth rates of native competitors peaks at moderate densities (Kornis et al. 2014). At high densities, increased intra-specific interactions among round gobies appeared to result in diminished impacts on native species (Kornis et al. 2014). Such alternative abundance-impact curves would potentially lead to different impact distributions, but could easily be incorporated into our conceptual framework.

Empirically informed management scenarios

Our framework has direct management applications for evaluating the spatial heterogeneity of invasive species impacts, especially if abundance distributions and abundance-impact relationships are known or can be reasonably estimated. To demonstrate, we used our framework to compare site- and species-specific impacts of two problematic invasive species in Wisconsin lakes: Eurasian watermilfoil (*Myriophyllum spicatum*; hereafter "milfoil") and zebra mussels (*Dreissena polymorpha*; hereafter "mussels"). We used invasive species occurrence records from the Wisconsin Department of Natural Resources (July 2012), estimated invasive species abundance from a large number of lakes (data in Hansen et al. 2013), and approximated invasive species abundance-impact relationships based on the literature and our general knowledge of how these species affect lakes. For each invaded lake in the occurrence records, we randomly drew a value from the abundance distribution (sampled with replacement) and applied the appropriate abundance-impact curve to estimate impact. This process generate a frequency distribution of impacts for invaded lakes, which were summed to estimate total landscape-level impact. We repeated this randomization process 1,000 times to simulate variability in the impact distribution and total impact estimates.

Zebra mussels – Occurrence records list zebra mussels in 177 Wisconsin lakes (WDNR 2012). Globally, zebra mussels have right-skewed abundance distributions in invaded sites (Hansen et al. 2013) (Fig. 4A). How abundance affects impact is likely to depend on the response variable of interest. Zebra mussel impacts on native mussels asymptote at relatively low zebra mussel abundance (Ricciardi 2003), which aligns with their status as a keystone species (Karatayev et al. 2002) (Fig. 4B). In contrast, Ricciardi (2003) also found native gammarid density to be linearly related with zebra mussel abundance (Fig. 4C). The modeled frequency distributions for each of these impacts were right skewed (Fig. 4B,C), although skewness was greater for impacts on gammarids. Summed landscape-level impacts on native mussels were nearly 4 times higher than those associated with gammarids (Fig. 4D).

Eurasian watermilfoil - Occurrence records list milfoil in 527 lakes (WDNR 2012). Although the distribution of milfoil abundance is also right-skewed, it exhibited a thicker tail than zebra mussels (Fig. 5A). Furthermore, milfoil abundance is likely to have different relationships with impact than mussels, but may still have different forms for different impacts. Milfoil's impacts on boaters may be linear (Fig. 5B), since rates of propeller entanglement are likely to be directly related to abundance, while milfoil's impacts on adjacent property values are more severe at high abundance (Zhang and Boyle 2010) (Fig. 5C). The modeled frequency distributions of impacts were right-skewed (Fig. 5B,C), with greater skewness when the linear abundance-impact curve associated with boat propeller entanglement was used. Summed landscape-level impacts were 6 times higher for linear abundance-impact curves than high-threshold curves, which may incur only small impacts at low to moderate abundance levels (Fig. 5D).

Comparing milfoil and mussels - Even though there is high variability in the types of impact distributions that are theoretically possible, these results indicate that right-skewed impacts may be most realistic. Each species had one impact distribution that was thick-tailed and another that was thin-tailed. For zebra mussels, the thick tail was can be attributed to the low threshold in the abundance-impact curve. For milfoil, the thick tail occurred with a linear abundance-impact curve, so the relatively high frequency of high impacts was caused by the thick tail in milfoil's abundance distribution. Without an objective weighting scheme to compare impacts of milfoil and mussels, we cannot say which generates higher landscape-level impacts. However, practitioners and stakeholders tasked with prioritizing invasive species management objectives may be reasonably expected to make such valuations. For example, if we determined that the maximum impacts of each species are the same, milfoil's landscape-level impacts would be higher, given their presence in nearly five times as many lakes and by the thick tail in their abundance distribution. On the other hand, if we determined that the impacts of mussels increase rapidly at low abundance and those of milfoil only increase at high abundance, then landscape-level impacts of mussels would actually be greater, despite their presence in many fewer lakes. Finally, if we judged the maximum impacts of mussels to be higher than those of milfoil, thus requiring a differential weighting scheme, then mussel impacts may easily outweigh milfoil impacts.

Management implications – single species on a landscape

Our results demonstrate the possibility for different patterns in an invasive species' impacts among sites on a landscape. Recognizing the potential for spatial heterogeneity in impacts allows for management approaches that extend beyond simple consideration of invasive species presence/absence. Prioritization of sites for protection from invasion or eradication, for instance, may provide an improved design only when the impact distributions are right-skewed, as was the case for milfoil and mussels. More generally, the number of sites that ought to be managed intensively to have the greatest reduction in an invasive species' overall impact is inversely related to the degree of right-skewedness in the impact distribution. If highly impacted sites are common (left-skewed) or if almost all sites are moderately impacted (unimodal or uniform), focusing on a small number of sites may only cause a small reduction in the total impact of an invasive species. In these cases, it may be relatively more beneficial to invest in more pervasive management techniques, such as education and outreach to organize a widespread volunteer network to participate in control and reduce further spread.

Management implications – multiple species

Assessments of species' total impacts exist, and were emphasized in inaugural issue of Biological Invasions, in which Parker et al. (1999) postulated that a species' total impact is the product of its range size, average abundance, and per capita effect. Several additional models have been adapted to incorporate various characteristics such as species and habitat attributes (for a review of these studies, see Thomsen et al. 2011). However, these approaches to assessing landscape-level impacts have rarely also considered the site-level impacts, hindering prioritization efforts.

Characterizing heterogeneity in impacts allowed us to estimate landscape-level impacts by summing across sites. Previous comparisons of species' impacts have either been qualitative or binary in nature or have not directly addressed site-to-site heterogeneity for a given species. For example, Kolar and Lodge (2002) identified five potential problematic invaders from a list of 66 fish species that could invade the US Great Lakes from the Ponto-Caspian region. Our results highlight the fact that impacts of these high-risk species are likely to be spatially heterogeneous. The pattern of heterogeneity, which we illustrate with impact distributions, depends on the shapes of the species' abundance distributions and abundance-impact curves, and can affect estimates of their total, landscape-wide impacts. This spatial texture underlying invasive species impacts is not widely acknowledged or understood – only recently have impact assessments begun to incorporate spatial heterogeneity (Thiele et al. 2010, Barney et al. 2013).

Our framework allows for simultaneous characterization of impact heterogeneity and estimation of landscape-level impacts. This information can then inform resource allocation to reduce impacts of a single species and enable rankings of species' total impacts. With presently available data and literature, managers may be able to make more informed decisions. However, our results show that the potential patterns of heterogeneity in impacts are diverse. Without quantifying abundance distributions and abundance-impact curves for more invasive species, our knowledge of the heterogeneity in invasive species impacts and ability to inform management is limited.

Acknowledgments

We would like to thank the many people who provided thoughts on this manuscript, including Steve Carpenter, Zach Lawson, Evan Childress, and Sam Christel. We would also like to thank Jennifer Filbert and the Wisconsin Department of Natural Resources for providing occurrence records. This material is based upon the work supported by the National Science Foundation under #CNH-0909281 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.

Literature Cited

Barney, J. N., D. R. Tekiela, E. S. Dollete, and B. J. Tomasek. 2013. What is the “real” impact of invasive plant species? Frontiers in Ecology and the Environment 11:322–329.

Brooks, T. M., R. A. Mittermeier, G. A. B. da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim, and A. S. L. Rodrigues. 2006. Global biodiversity conservation priorities. Science 313:58–61.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American Naturalist 124:255–279.

Byers, J. E., S. Reichard, J. M. Randall, I. M. Parker, C. S. Smith, W. M. Lonsdale, I. a. E. Atkinson, T. R. Seastedt, M. Williamson, E. Chornesky, and D. Hayes. 2002. Directing research to reduce the impacts of nonindigenous species. Conservation Biology 16:630–640.

Crooks, J. A. 2002. Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. Oikos 97:153–166.

Elgersma, K. J., and J. G. Ehrenfeld. 2011. Linear and non-linear impacts of a non-native plant invasion on soil microbial community structure and function. Biological Invasions 13:757–768.

Epanchin-Niell, R. S., and A. Hastings. 2010. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. Ecology Letters 13:528–541.

Frieswyk, C. B., C. A. Johnston, and J. B. Zedler. 2007. Identifying and characterizing dominant plants as an indicator of community condition. Journal of Great Lakes Research 33:125–135.

Game, E. T., P. Kareiva, and H. P. Possingham. 2013. Six common mistakes in conservation priority setting. Conservation Biology 27:480-485.

Hansen, G. J. A., M. J. Vander Zanden, M. J. Blum, M. K. Clayton, E. F. Hain, J. Hauxwell, M. Izzo, M. S. Kornis, P. B. McIntyre, A. Mikulyuk, E. Nilsson, J. D. Olden, M. Papeş, and S. Sharma. 2013. Commonly rare and rarely common: Comparing population abundance of invasive and native aquatic species. PLoS ONE 8:e77415.

Higgins, S. N., and M. J. Vander Zanden. 2010. What a difference a species makes: a meta-analysis of dreissenid mussel impacts on freshwater ecosystems. Ecological monographs 80:179–196.

Holling, C. S. 1965. The functional response of predators to prey density and its role in mimicry and population regulation. Memoirs of the Entomological Society of Canada 97:5–60.

Karatayev, A. Y., L. E. Burlakova, and D. K. Padilla. 2002. Impacts of zebra mussels on aquatic communities and their role as ecosystem engineers. Pages 433–446 *in* E. Leppäkoski, S. Gollasch, and S. Olenin, editors. Invasive Aquatic Species of Europe. Distribution, Impacts and Management. Springer Netherlands.

King, R. B., J. M. Ray, and K. M. Stanford. 2006. Gorging on gobies: beneficial effects of alien prey on a threatened vertebrate. Canadian Journal of Zoology 84:108–115.

Kitchell, J. F., S. P. Cox, C. J. Harvey, T. B. Johnson, D. M. Mason, K. K. Schoen, K. Aydin, C. Bronte, M. Ebener, M. Hansen, M. Hoff, S. Schram, D. Schreiner, and C. J. Walters. 2000. Sustainability of the Lake Superior fish community: Interactions in a food web context. Ecosystems 3:545–560.

Kolar, C. S., and D. M. Lodge. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science 298:1233–1236.

Kornis, M. S., J. Carlson, G. Lehrer-Brey, and M. J. V. Zanden. 2014. Experimental evidence that ecological effects of an invasive fish are reduced at high densities. Oecologia 175:325–334.

Kulhanek, S. A., A. Ricciardi, and B. Leung. 2011. Is invasion history a useful tool for predicting the impacts of the world's worst aquatic invasive species? Ecological Applications 21:189–202.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

Paine, R. T. 1969. A note on trophic complexity and community stability. The American Naturalist 103:91–93.

Parker, I. M., D. Simberloff, W. M. Lonsdale, K. Goodell, M. Wonham, P. M. Kareiva, M. H. Williamson, B. Von Holle, P. B. Moyle, J. E. Byers, and others. 1999. Impact: toward a framework for understanding the ecological effects of invaders. Biological invasions 1:3–19.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecological Economics 52:273–288.

Power, M. E., D. Tilman, J. A. Estes, B. A. Menge, W. J. Bond, L. S. Mills, G. Daily, J. C. Castilla, J. Lubchenco, and R. T. Paine. 1996. Challenges in the quest for keystones. BioScience 46:609–620.

Ricciardi, A. 2003. Predicting the impacts of an introduced species from its invasion history: an empirical approach applied to zebra mussel invasions. Freshwater Biology 48:972–981.

Ricciardi, A., M. F. Hoopes, M. P. Marchetti, and J. L. Lockwood. 2013. Progress toward understanding the ecological impacts of nonnative species. Ecological Monographs 83:263–282.

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, and D. H. Wall. 2000. Global biodiversity scenarios for the year 2100. Science 287:1770–1774.

Simberloff, D., I. M. Parker, and P. N. Windle. 2005. Introduced species policy, management, and future research needs. Frontiers in Ecology and the Environment 3:12–20.

Thiele, J., J. Kollmann, B. Markussen, and A. Otte. 2010. Impact assessment revisited: improving the theoretical basis for management of invasive alien species. Biological Invasions 12:2025–2035.

Thomsen, M. S., J. D. Olden, T. Wernberg, J. N. Griffin, and B. R. Silliman. 2011. A broad framework to organize and compare ecological invasion impacts. Environmental Research 111:899–908.

Yokomizo, H., H. P. Possingham, M. B. Thomas, and Y. M. Buckley. 2009. Managing the impact of invasive species: the value of knowing the density-impact curve. Ecological Applications 19:376–386.

Zedler, J. B., and S. Kercher. 2004. Causes and consequences of invasive plants in wetlands: Opportunities, opportunists, and outcomes. Critical Reviews in Plant Sciences 23:431–452.

Zhang, C., and K. J. Boyle. 2010. The effect of an aquatic invasive species (Eurasian watermilfoil) on lakefront property values. Ecological Economics 70:394–404.

Figure Legends

Fig. 1. The process and inputs we used to simulate probability distributions of the site-level invasive species impacts across a landscape. All abundance values are standardized on a 0-1 scale. A) Potential frequency distributions of a species' site-level abundance, where each site is characterized by a single value for the species' abundance. Centered and skewed distributions are beta distributions, and the uniform distribution represents an equal probability of occurrence for any value of abundance. B) The four types of relationships between abundance and impact (i.e. abundance-impact curves) used in our simulations (modified from Yokomizo et al. 2009). Impact is standardized on a 0-100 scale. To generate distributions of impact, we randomly draw an abundance value (α) from an abundance distribution and plug it into an abundance-impact curve to calculate impact for that site (I_1 or I_2 depending on curve used, although any curve could be used), as shown in gray. For each unique combination of these abundance distributions and abundance-impact curves, we repeated the process 1,000 times to calculate impact for 1,000 sites and sum all values of I to calculate total landscape-level impacts.

Fig. 2. Probability distributions of impacts for each combination of abundance distribution (rows) and abundance-impact curve (columns), which correspond to those shown in Fig. 1, simulated by randomly drawing 1,000 values from each abundance distribution and calculating impact using each abundance-impact curve. Impacts are standardized so that 0=no impact and 100=maximum possible impact.

Fig. 3. Landscape-level (summed) impacts for each combination of abundance distribution (by shape and line type) and abundance-impact curve (on x-axis), which correspond to those shown in Fig. 1. Units for impacts are the same as Figs. 1 and 2, such that maximum impacts at a site

equal 100, and thus, maximum possible landscape-level impacts across 1,000 sites sum to 100,000.

Fig. 4. A) Histogram of site-level densities for zebra mussels (from Hansen et al. 2013). B, C) Abundance values are sampled with replacement for all invaded lakes in Wisconsin (n=177) and combined with a low threshold abundance-impact curve (B inset) and a linear abundance-impact curve (C inset) to calculate 2 potential distributions of zebra mussel impacts. D) Boxplots of landscape-level summed impacts across all lakes, where each box displays the quartiles of 1,000 impact estimates from repeated resampling of abundance data.

Fig 5 A) Histogram of site-level frequency of occurrence for Eurasian watermilfoil (from Hansen et al. 2013). B, C) Abundance values are sampled with replacement for all invaded lakes in Wisconsin (n=527) and combined with a linear abundance-impact curve (B inset) and a high threshold abundance-impact curve (C inset) to calculate 2 potential distributions of Eurasian watermilfoil impacts. D) Boxplots of landscape-level summed impacts across all lakes, where each box displays the quartiles of 1,000 impact estimates from repeated resampling of abundance data.

Figure 1

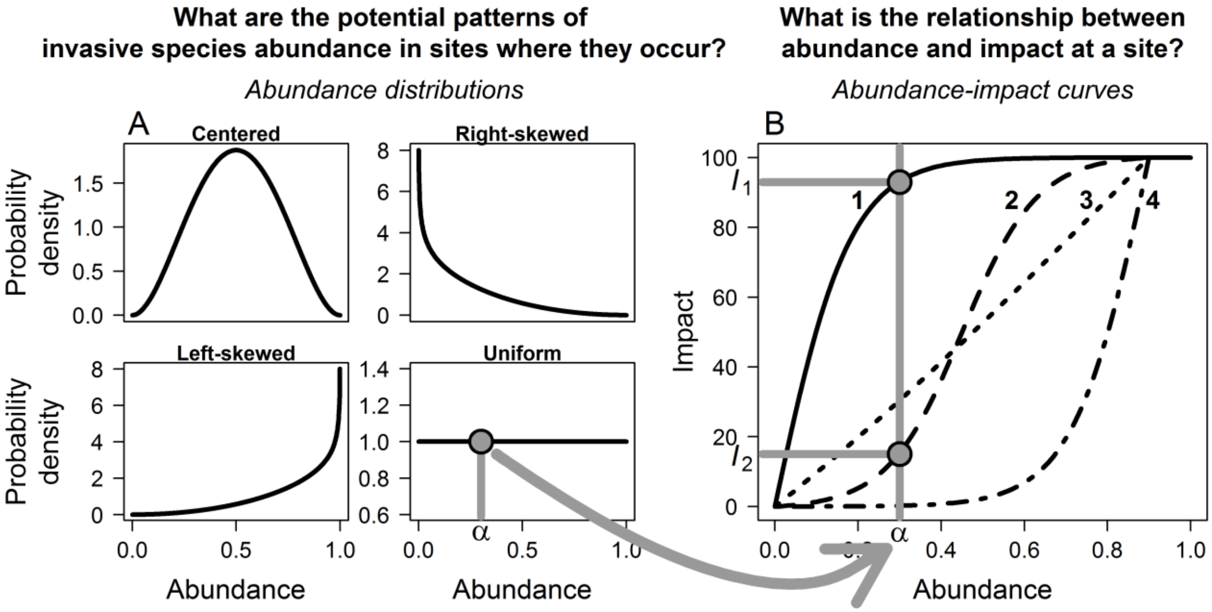

Fig. 1. The process and inputs we used to simulate probability distributions of the site-level invasive species impacts across a landscape. All abundance values are standardized on a 0-1 scale. A) Potential frequency distributions of a species' site-level abundance, where each site is characterized by a single value for the species' abundance. Centered and skewed distributions are beta distributions, and the uniform distribution represents an equal probability of occurrence for any value of abundance. B) The four types of relationships between abundance and impact (i.e. abundance-impact curves) used in our simulations (modified from Yokomizo et al. 2009). Impact is standardized on a 0-100 scale. To generate distributions of impact, we randomly draw an abundance value (α) from an abundance distribution and plug it into an abundance-impact curve to calculate impact for that site (I_1 or I_2 depending on curve used, although any curve could be used), as shown in gray. For each unique combination of these abundance distributions and abundance-impact curves, we repeated the process 1,000 times to calculate impact for 1,000 sites and sum all values of *I* to calculate total landscape-level impacts.

Figure 2

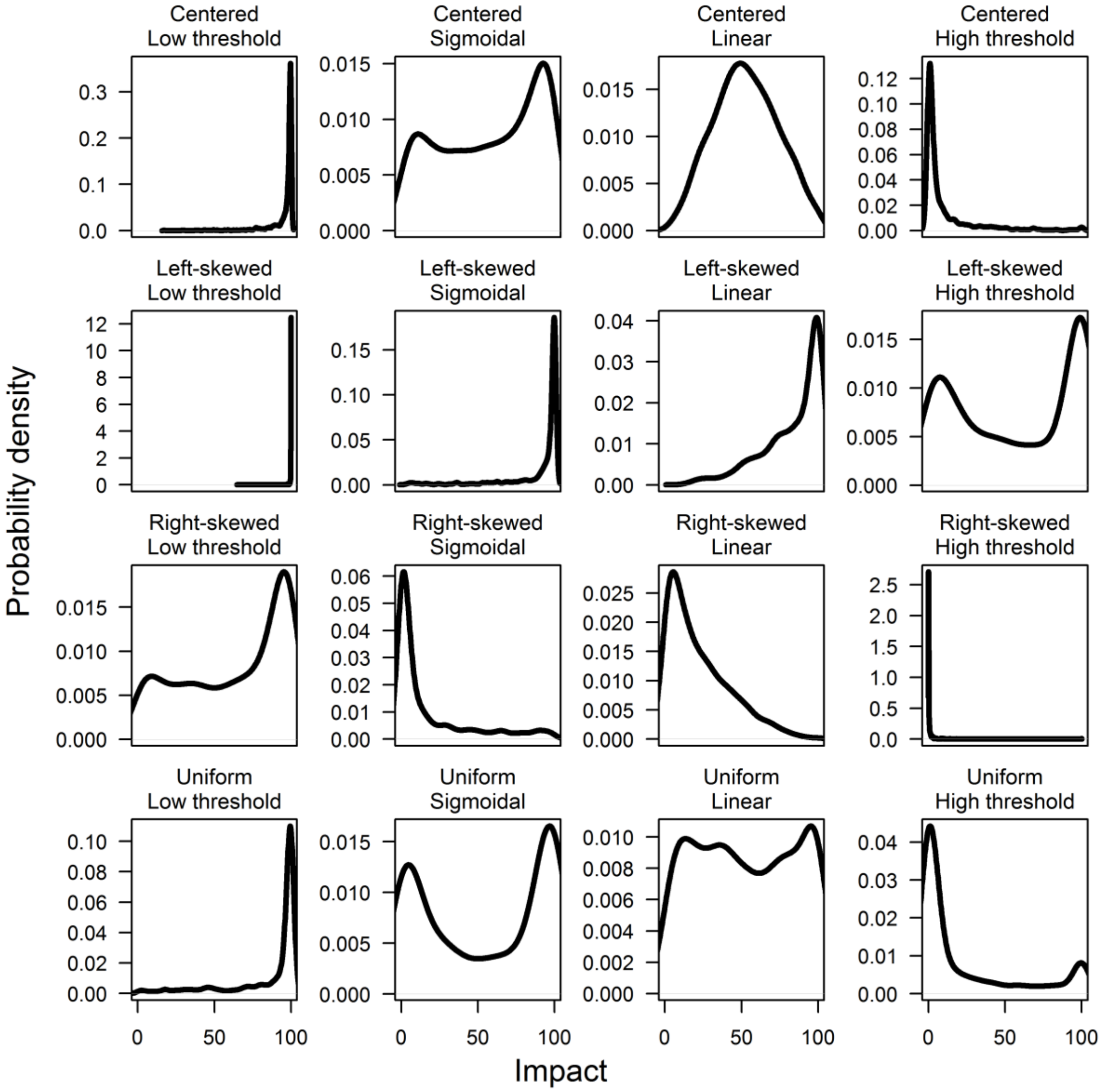

Fig. 2. Probability distributions of impacts for each combination of abundance distribution (rows) and abundance-impact curve (columns), which correspond to those shown in Fig. 1, simulated by randomly drawing 1,000 values from each abundance distribution and calculating impact using each abundance-impact curve. Impacts are standardized so that 0=no impact and 100=maximum possible impact.

Figure 3

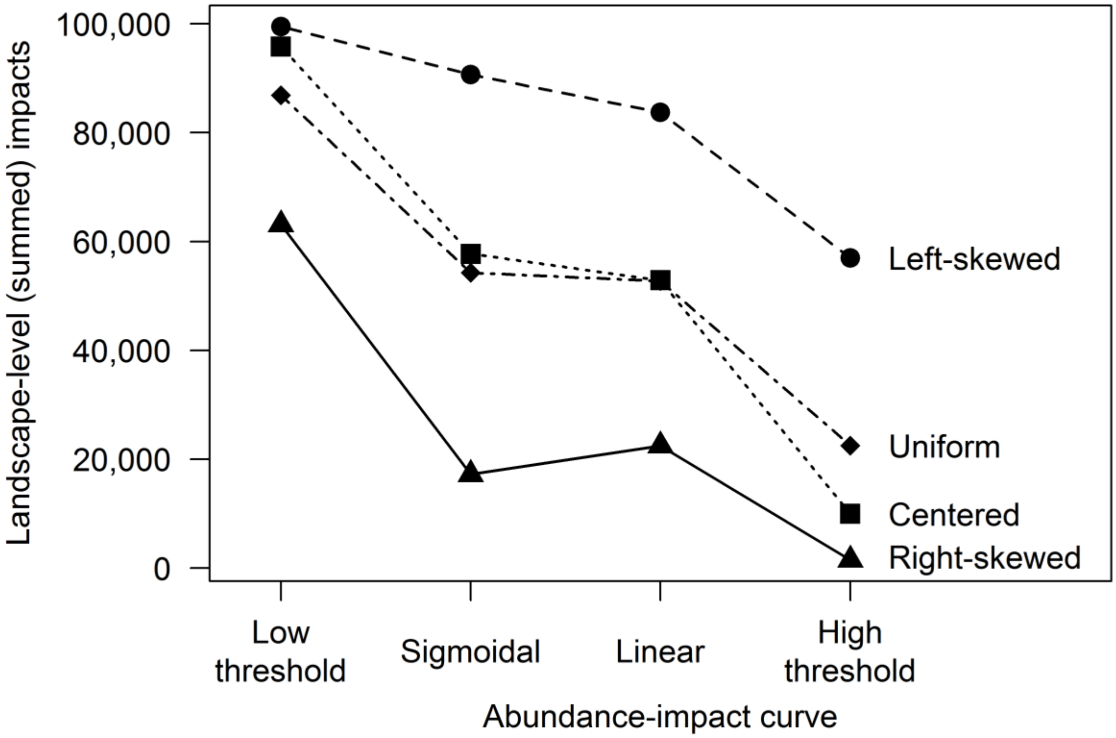

Fig. 3. Landscape-level (summed) impacts for each combination of abundance distribution (by shape and line type) and abundance-impact curve (on x-axis), which correspond to those shown in Fig. 1. Units for impacts are the same as Figs. 1 and 2, such that maximum impacts at a site equal 100, and thus, maximum possible landscape-level impacts across 1,000 sites sum to 100,000.

Figure 4

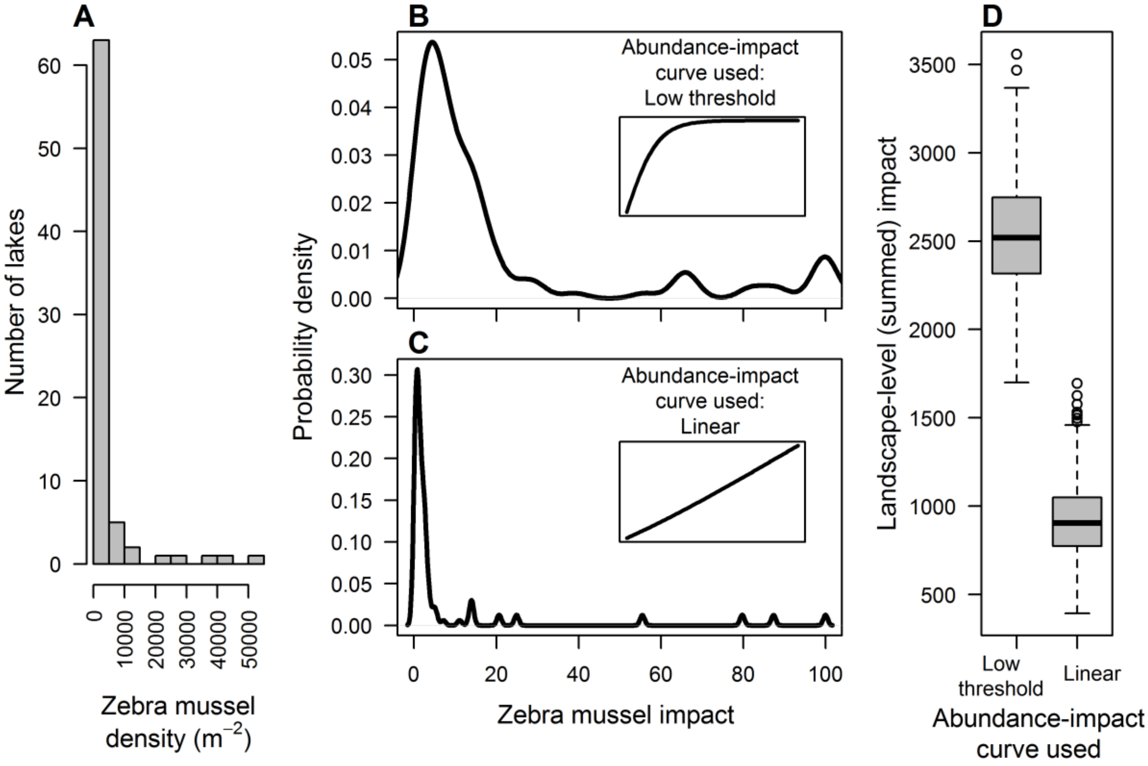

Fig. 4. A) Histogram of site-level densities for zebra mussels (from Hansen et al. 2013). B, C) Abundance values are sampled with replacement for all invaded lakes in Wisconsin (n=177) and combined with a low threshold abundance-impact curve (B inset) and a linear abundance-impact curve (C inset) to calculate 2 potential distributions of zebra mussel impacts. D) Boxplots of landscape-level summed impacts across all lakes, where each box displays the quartiles of 1,000 impact estimates from repeated resampling of abundance data.

Figure 5

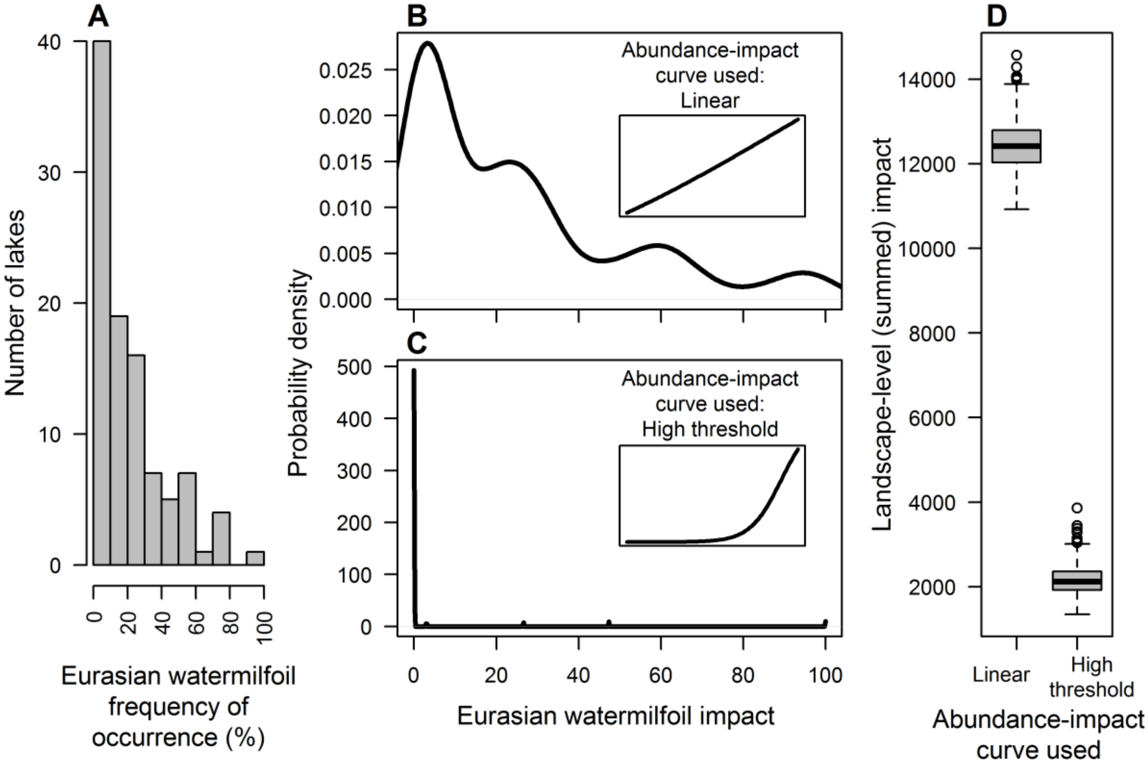

Fig 5 A) Histogram of site-level frequency of occurrence for Eurasian watermilfoil (from Hansen et al. 2013). B, C) Abundance values are sampled with replacement for all invaded lakes in Wisconsin (n=527) and combined with a linear abundance-impact curve (B inset) and a high threshold abundance-impact curve (C inset) to calculate 2 potential distributions of Eurasian watermilfoil impacts. D) Boxplots of landscape-level summed impacts across all lakes, where each box displays the quartiles of 1,000 impact estimates from repeated resampling of abundance data.

Appendix 1

The abundance-impact curves used in this analysis follow those from Yokomizo et al. 2009 take the form:

$$I_s = f_{AI}(\alpha_s)$$

where,

$$f_{AI} \ = \begin{cases} M * C * 1/(1 + \ e\hat{}((-\alpha_s/(Y-u))/b)\ ) - B \\ \qquad\qquad\qquad\qquad M \end{cases}$$

$$B \ = 1/(1 + e\hat{}(u/b))$$

$$C \ = (1 + e\hat{}((1-u)/b))/(1 - B * (1 + e\hat{}((1-u)/b))\ .$$

We used the following parameters to create the four abundance-impact curves:

Parameter	Low-abundance threshold	Sigmoidal	Linear	High-abundance threshold
u	0	0.5	1	1
b	0.1	0.1	1	0.1
Y	0.9	0.9	0.9	0.9
M	100	100	100	100