Predicting Potential Invasive Species Distribution: An Application to New Zealand Mudsnails in the Pacific Northwest

Youngah Lim
Agricultural and Resource Economics
Munisamy Gopinath
Agricultural and Resource Economics
Samuel Chan
Oregon Sea Grant
Michael Harte
Marine Resource Management

Oregon State University
213 Ballard Hall
Corvallis, OR 97331
limy@onid.orst.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

Predicting Potential Invasive Species Distribution: An Application to New Zealand Mudsnails in the Pacific Northwest

Introduction

Over the past few decades, invasive species (IS) with harmful effects on ecosystems, economy, and human health have become a major concern for researchers, private industries, and policymakers. Losses from such bioinvasion include declining productivity of ecosystems, pecuniary as well as aesthetic, increased production and management costs to businesses, and monitoring, enforcement and management costs to the public (Perrings et al. 2002). For example, Sala et al. (2000) predicted that biotic exchange, arising from introduction and establishment of non-indigenous species, would cause large biodiversity changes within each biome, while Pimentel et al. (2005) estimated that IS damages and losses add up to $120 billion per year in the United States. Moreover, Mann et al. (2010) predict that managing zebra and quagga mussels in the Columbia River Basin, if introduced, would annually cost up to $50 million.

Resource managers and policymakers often consider the task of IS management with highly limited budgets. Depending on IS risk and current/potential damage, management options include prevention, early-detection, eradication, containment, and control. However, decision makers face significant uncertainty in risk assessment and potential damage, which severely limits their evaluation of alternative management strategies. Furthermore, qualitative assessments of risk and damage constrain prioritizing IS and identifying optimal response to bioinvasions. While spatial and mapping techniques have helped provide some quantitative measures of IS risk, very few studies have proposed or examined quantitative approaches to evaluating alternative management strategies. The objective of this study is to develop, test and calibrate a quantitative and predictive approach to IS risk assessment, which would help IS managers choose cost-efficient strategies.

In the following, we will use the term *invasive species* to refer an introduced species that is likely to cause economic, environmental, or human health losses. Introduced species are often termed as invasive, alien, exotic, foreign, non-native, naturalized, immigrant, and non-indigenous species (Carlton, 2001). More precisely, Carlton (2001) defined *introduced species* as species that have been transported by human activities—intentionally or unintentionally—into a region in which they did not historically occur and now reproducing in the wild. Executive Order 13112 (1999) defined *invasive species* as an alien species whose introduction causes or would cause economic, environmental, or human health damages.

For geographic prediction of IS dispersal, this study will employ a gravity model and maximum entropy (Maxent) method to predict potential IS distribution. The gravity model will represent the anthropogenic introduction of the species from infected to uninfected sites, while Maxent will characterize environmental suitability of available habitats for the introduced species. Thus, we can quantify the role of anthropogenic and biological factors in the geographic prediction of IS dispersal and establishment risk.

Non-indigenous species are generally introduced into a region by three modes: natural spread and host range extension, accidental introduction, and intentional introduction (Maynard and Nowell, 2009). Geographical barriers usually limit the scope of natural dispersal, but human-mediated dispersal, intentional or accidental, can cover any location beyond the bio-geographical barriers. For instance, if someone wanted to bring a non-indigenous species into a country, e.g. ornamental plants or profitable aquaculture products, legitimate border processes or public agencies govern such introductions. Thus, the effect of the intentionally introduced species would be much easier to predict and estimate than those unintentionally introduced. In this context, Carlton and Ruiz (2005) suggested that a key IS vector is the unintentional species transportation by humans.

Whether or not introduced species will survive in a new habitat is the next step in the analysis of IS risk. Although ecological theories about species distribution, community structure, and biodiversity can explain and predict species distribution, ecological niche modeling has dramatically grown recently (e.g. Peterson, Papeş, and Eaton, 2007). The niche concept is based on joint environmental conditions which allow the birth rate of a local population to be equal to or greater than the death rate, and per capita effects of the species on these environmental conditions (Chase and Leibold, 2003). Based on the niche concept, species distribution models consider multiple environmental conditions, e.g. temperature, precipitation, water quality, topography, which are required for the species occurrence and establishment (Franklin and Miller, 2009).

The application in this study will focus on the current New Zealand mudsnails (NZMS) distribution in the Pacific Northwest by integrating two concepts: anthropogenic introduction and habitat suitability. A gravity model and a species distribution model (maximum entropy) will be used to characterize recreational boat movement, a key vector of anthropogenic introduction, and habitat suitability, respectively. The outcomes of these two models will be utilized to predict potential geographic invasive species (IS) distribution. Thus, the predictions include a representation of relative risk of IS introduction and its probability to establish successfully in the introduced regions. For this purpose, we will use the hydrologic unit as the basic spatial unit because it reflects the geological habitat information better than other segments such as administrative districts or counties.

Conceptual Framework

Spatial Interaction Model: Gravity Model

Previous research has defined spatial interaction as movement or communication between regions under a decision making process (Haynes and Fortheringham, 1984; Fortheringham and O'Kelly, 1989). Spatial interaction models are mathematical frameworks to explain and/or predict the spatial interactions based on attributes of origins and destinations, and spatial separation variables, e.g. distance. A gravity model for spatial interaction is a relatively simple framework consisting of the above three components (Fortheringham and O'Kelly, 1989). It is usually written as:

(1) $$T_{ij} = f\left(\alpha w_i;\ \beta w_j;\ \delta d_{ij}\right),$$

where T_{ij} represents a flow between origin i and destination j, w_i and w_j represent levels of attractiveness (or repulsiveness) of i and j, and d_{ij} is the distance between i and j. The Greek alphabets are parameters to be inferred or estimated. Despite its simple structure, the gravity model is one of the best known models with a high level of goodness-of-fit, i.e. explaining spatial interaction (Fortheringham and O'Kelly, 1989; Ortúzar and Willumsen, 1992).

Several empirical studies have used the gravity model to predict the distribution of aquatic invasive species, but most focused on zebra mussels (e.g. Schneider et al. 1998; Bossenbroek et al. 2001; Leung et al. 2004; Bossenbroek et al. 2007; Leung and Mandrak, 2007). The type of gravity model depends on the availability of attribute information for the origin and destination. An unconstrained, a production-constrained, and an attraction-constrained gravity model can be used when researchers have information on both origin and destination, on destination characteristics only, and on origin information only, respectively (Fortheringham and O'Kelly, 1989; Noronha and Goodchild, 1992). In this study, we will employ the unconstrained gravity model to model recreational boat flows, T_{ij}, by utilizing origin and destination characteristics and the distance between them. Two common functional forms used for the gravity model are exponential and power function, and the choice between the two forms relies on how fast the flow decreases with the distance (Fortheringham and O'Kelly, 1989; Potapov and Lewis, 2008). We will estimate both functional forms and compare the results to determine which model better explains observed flow patterns. Hence, the two models to estimate are:

(2) $T_{ij} = {w_i}^{\alpha} {w_j}^{\beta} \exp\left(-\delta d_{ij}\right)$ and

(3) $T_{ij} = {w_i}^{\alpha} {w_j}^{\beta} {d_{ij}}^{-\delta}$.

The number of boats moving from an infected region i to an uninfected region j can be interpreted as propagule pressure. That is, propagule pressure is a measure of the number of non-indigenous species introduced into a region, and such pressure is an important determinant of successful bioinvasion (Leung et al. 2004; Lockwood et al. 2005; Leung and Mandrak, 2007). Propagule pressure is measured as the absolute number of organisms involved in one release event (propagule size) and the number of release events (propagule number). Formally , the number of recreational boats, T_{ij}, is a proxy for propagule pressure in terms of the number of boats from infected regions (propagule number) and the level of infection of the origin (propagule size). In other words, the probability of IS introduction is high when more boats come from an infected region, or a boat comes from a highly infected region.

Species Distribution Models: Maximum Entropy Method

Species distribution models, often referred to as ecological niche models, bioclimatic envelopes, or habitat suitability models, can estimate habitat suitability for observed species and predict the probability of species occurrence when environmental information and predictors are available

(Segurado and Araújo, 2004; Elith and Graham, 2009). If presence and absence data are available, the outcome of niche models is usually more accurate than that of models based on presence-only data (Franklin and Miller, 2009). However, absence data is only obtainable under strict and consistent sampling, requiring caution in interpreting an unobservable occurrence as species absence or other possibilities. For example, species absence may be due to lack of survey of specific regions or locations or the previous absence data have not been updated to check recent infestations. Niche models using presence-only data include: ecological niche factor analysis, generic algorithms for rule production, and maximum entropy (Franklin and Miller, 2009). All three are non-parametric methods which identify rules of species distribution between species occurrence and environmental conditions without a pre-determined statistical distribution or parameter.

This study will employ maximum entropy (Maxent) method mainly because of presence-only data availability. Moreover, Maxent provides a probabilistic output, which can be easily combined with the outcome of the gravity model. Advantages of Maxent are: (1) it requires presence-only data, (2) it can employ both categorical and continuous data, (3) it converges well to the optimum (maximum) entropy, (4) the Maxent probability distribution has a mathematical definition, (5) over-fitting can be prevented, (6) it has potential to handle sampling bias, and (7) its output is continuous, so fine distinction is possible among different regions. The disadvantages are that Maxent remains in development and as a result, is sensitive to the number of regulations that can be placed. Additionally, conventional statistics software cannot estimate the model. Despite its disadvantages, Maxent appears to have outperformed other presence-only data methods such as a genetic algorithm for rule production and envelope method (e.g.

BIOCLIM) in several studies (e.g. Elith et al. 2006; Phiilips et al. 2006; Elith and Graham, 2009).

The following describes Maxent species distribution modeling (Phillips et al. 2006). Let π be an unknown distribution over a finite set X, which can be considered as grids of ecological domains. The distribution π assigns a probability $\pi(x)$ to each element of X. Let $\hat{\pi}$ denote an approximation of the unknown distribution. The entropy of $\hat{\pi}$,

(4) $$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x)\ ln\hat{\pi}(x).$$

Equation (4) is an application of Shannon (1948)'s a measure of entropy representing uncertainty of a set of events. Based on given incomplete information (here, species observation and environmental conditions), our aim is to maximize this entropy value to find the unknown distribution. Assume that $f_1, \ldots, f_n$ are known functions of features, e.g. environmental conditions, on X, and $f_j(x)$ is a realized value of the function. Accordingly, the expectation of features under π is defined as $\sum_{x \in X} \pi(x) f_j(x)$ and denoted by $\pi[f_j]$. We can approximate this expectation by using empirical average of features based on independently drawn m samples:

(5) $$\tilde{\pi}[f_j] = \frac{1}{m} \sum_{i=1}^{m} f_j(x_i).$$

Our objective is to find the probability distribution $\hat{\pi}$ of maximum entropy subject to $\hat{\pi}[f_j] = \tilde{\pi}[f_j]$, which makes the approximate expected features of unknown distribution equal to the approximate expected features of empirical samples. In fact, the derived means cannot be equal to the true means, so the constraint will be relaxed as $|\hat{\pi}[f_j] - \tilde{\pi}[f_j]| \leq \beta_j$. As a result, the raw output of Maxent is the exponential function that assigns a probability to each site. The results are not necessarily proportional to probabilities of the species presence since their sum must be equal to one and we usually handle a limited size of geographic area. That is, the raw result of Maxent is not an absolute probability of species occurrence, but it represents a ranking of risk among regions. We will compare relative risk of NZMS across regions using the Maxent outcome.

A few studies have employed other species distribution modeling to analyze NZMS distribution, the focus of my application. For example, Vinson et al. (2007) utilized ecological niche factor analysis to predict NZMS distribution in Idaho, Montana, Utah, and Wyoming, while Loo et al. (2007) used the generic algorithm for rule production to predict NZMS distribution in Australia and North America. Simple logistic regression was used by Schreiber et al. (2003) to study NZMS dispersal in southern Victorian streams, Australia.

Data and Estimation

Gravity Model

The presence data of New Zealand mudsnails (NZMS) in the Pacific Northwest (Idaho, Oregon and Washington) are obtained from the U.S. Geological Survey's Nonindigenous Aquatic Species (USGS NAS) and Montana State University (MSU) databases. Density information is nearly unavailable in both databases: USGS NAS does not collect or report density data, while MSU database provide less confidence on the sparsely available density from a few infected sites.

In terms of recreational boat flow, we project potential boat flows within each state and across states because there is no direct data about the exact number of boats transporting between regions. For the projection, we use boat registration data from each state and its respective survey of boat owners, along with boat inspection data in Idaho and Washington. Boat registration data will represent the number of boats in origin regions (2009 Idaho, 2010 Oregon, and 2010 Washington), and survey (2009 Idaho, 2007 Oregon (published in 2008), and 2007-09

Washington) and inspection data (2009 Idaho and 2008-10 Washington) will provide boat movement/flow information. To be specific, boat registration data describes the total number of boats registered in each county (Oregon and Washington) and city (Idaho), and boater's survey and inspection data includes where the survey participant lived and visited. We will extrapolate boat movements based on the survey and inspection data to the entire population of boat owners in each state. Moreover, the observational unit will be adjusted to hydrologic units using geographic information system (GIS), so I will recalculate the number of boats based on the relative size and population of a given area. The hydrologic unit reflects information on drainage, hydrography, and hydrologic boundaries and codes of four hierarchical units (Seaber et al. 1987). The distance between regions will be calculated based on relevant layers such as hydrologic units and the administrative boundary of counties, and water bodies. The layers are to be obtained from the National Atlas of the United States and the U.S. Census Bureau.

As spatial attributes of origins and destinations, we consider water body size, water quality, and accessibility of water bodies. That is, people prefer larger water bodies, higher quality of water bodies, and more conveniently accessible areas. We calculate the size of water bodies directly from GIS, and the accessibility can be expressed by the density of major roads in each region, or distance between specific water bodies and major roads. The water quality will be characterized by the level of chemicals or turbidity in water. These data will be obtained from the U.S. Environmental Protection Agency (USEPA) Storage and Retrieval (STORET) database.

In the gravity model, the proposed dependent variable will be continuous: $\frac{\sum_i T_{ij}}{\sum_i \sum_j T_{ij}}$. However, the dependent variable can be censored at zero for some hydrological units. In order to reflect this data structure, we estimate a tobit (a censored regression) model by utilizing the maximum likelihood estimation method. The likelihood function of the tobit is known to achieve the global maximum regardless of its initial estimator or hypothesized values (Maddala, 1983).

Maximum Entropy Method

The Maxent analysis uses presence-only data of NZMS due to data availability and the lack of consistent sampling. In the USGS NAS and MSU databases, the occurrence data have been reported by different researchers at different time, so the sampling methods are likely not consistent across observations. Maximum entropy (Maxent) method will also employ possible environmental factors that would affect establishment of NZMS. Following the NZMS habitat characterization from the University of California, Santa Barbara, Riparian Invasion Research Laboratory (UCSB RIVRLab), we can consider the following environmental variables affecting NZMS establishment: seasonality, substrate, light, water temperature, velocity, salinity, conductivity, and locations near cities. In addition to these variables, stream order, pH of water, precipitation, and elevation can be additional environmental factors. Among these factors, velocity, elevation, and stream order are expected to represent natural dispersal of NZMS. Data on environmental characteristics and elevation data are taken from USEPA's STORET database and the PRISM Climate Group, and city information is downloaded from the U.S. Census Bureau.

As noted earlier, standard statistical software cannot be used for Maxent analysis. Thus, we use the Maxent program developed by Steven Phillips (available at Schapire's webpage, see reference). The above program utilizes raster geographic information systems (GIS) data and provides users a predicted map, simple statistic value, and response curves to each environmental factor. The GIS data involve two spatial features: vector and raster data. The vector data involves points, lines, and polygons, while the raster data is composed of grids and each grid has a corresponding value for a location. Therefore, the raster data is known to be more useful to model continuous spatial variation such as precipitation, elevation, and soil erosion (Chang, 2010). The default output of Maxent is n the logistic form, which ranges between 0 and 1 and can be interpreted as the probability of species presence (Phillips and Dudík, 2008).

Preliminary Results for Oregon

Gravity Model Data and Results

For the Oregon gravity model, boat flows from hydrologic unit i to j are estimated as a function of the distance between regions and relative attractiveness of each area. In the following, we provide the basics in the construction of relevant variables and then, provide preliminary estimation results for Oregon.[1] Boat flows, T_{ij}, are derived from 2010 Oregon boat registration data (county level) and 2008 Oregon triennial boat survey data. Using the survey, observed boat flows from one hydrologic unit to another are derived. Then, with the assumption that the survey results apply to the entire population of Oregon boat owners, we project potential boat flows between hydrological units i and j. For the projection, we use water body size, hydrologic unit size, 2009 population of cities, and 2010 population estimate of each county as weights. The distance between hydrologic units d_{ij} is measured in miles by the distance between centroids of hydrologic units.

To represent relative attractiveness w_i and w_j, i.e. geographic attributes of each area—the size of each hydrologic unit, water body dispersion inside the hydrologic unit, and adjacency to the Pacific Ocean, accessibility (road density), and water quality are considered. Water body

[1] We anticipate completing the Idaho and Washington models and risk assessment for the entire Pacific Northwest before presentation of the study at the 2011 AAEA Meeting in Pittsburgh.

dispersion is measured by the Herfindahl index which is commonly used to measure market concentration (Cabral, 2000). That is, the water body dispersion is derived as the sum of the squares of water body size proportions (shares) of the relevant hydrologic unit. Adjacency to the Pacific Ocean is included as a dummy variable depending on whether or not the destination hydrologic unit is adjacent to the Pacific Ocean. The road density is calculated by the total length of road (meters) divided by the hydrologic unit size (squared kilometers).

In order to measure water quality, we employ the National Sanitation Foundation's water quality index (Oram 2010) by utilizing data from the U.S. Environmental Protection Agency (USEPA) Storage and Retrieval (STORET) database and Oregon Department of Environmental Quality (DEQ) Laboratory Analytical Storage and Retrieval (LASAR). Because of data limitations, water quality index is calculated from five water characteristics: pH, turbidity, biochemical oxygen demand (BOD), nitrates, and total suspended solids (TSS), although calculating the original index requires additional characteristics. We derived monthly water characteristics in each water body by utilizing monitoring sites' geographic coordinates and transformed the values into hydrologic unit water characteristics by averaging them with size of water body as the weight. Then, an average of monthly water characteristics for each quality over the calendar year is obtained. Finally, an annual water quality index was calculated as a weighted sum of year-based individual water characteristics, following the approach of National Sanitation Foundation. A lower value for the above index implies a higher water quality in the given hydrologic unit. Summary statistics on the variables used in the gravity model estimation are presented in Table 1.

Table 1. Summary Statistics of Oregon Data

Variable	Observation	Mean	Standard Deviation	Minimum	Maximum
Number of boats	6084	34.55	200.61	0.00	4893.08
Hydrologic Area Size (km^2)	79	3064.59	1822.71	338.27	10727.08
Herfindahl[a]	79	12.63	63.21	6.25E-08	518.46
Dummy (Ocean)	79	0.18	0.38	0.00	1.00
Road Density (%)	79	14.93	13.82	0.74	75.51
Distance (m)	3081[b]	291271.90	140343.10	26468.27	674259.60
Water Quality Index	79	2.68	3.06	3.35E-05	10.64

[a] Herfindahl = 10,000 × sum of (water body/hydrologic unit size)^2.

[b] (the number of hydrologic units^2 – the number of hydrologic units) / 2 = (79^2 – 79) / 2 = 3081.

The empirical estimation of the gravity model employs a linear-log model, where the dependent variable is in linear form and the independent variables are in log form. The coefficient in the linear-log specification is interpreted as the absolute change of the dependent variable with respect to a relative change in the independent variable. The exponential and power functional form of gravity models in equations (2) and (3) are estimated using the linear-log specification. Result from the exponential form are quantitatively and qualitatively similar to those of the power function. We discuss results of the exponential form in the following since its log likelihood value is slightly higher than that of the power function in our estimation. The estimated equation is given by:

$$(6) \quad T_{ij} = \alpha_0 + \alpha_1 \ln(HUC\ size)_i + \alpha_2 \ln(Herfinhahl)_i + \alpha_3 \ln(Road\ density)_i$$
$$+\beta_1 \ln(HUC\ siz)_j + \beta_2 \ln(Herfindahl)_j + \beta_3 \ln(Road\ density)_j$$
$$+\gamma_1 Dummy(Ocean)_j + +\gamma_2 Water\ Quality_i + \gamma_3 Water\ Quality_j$$
$$-\delta d_{ij} + \varepsilon_{ij}.$$

The results of estimating equation (6) are presented in Table 2 in five alternative specifications of the exponential form of the gravity model. A censored normal regression (Tobit) model is estimated for each specification because 3589 of 6084 observations in the empirical data are censored at zero. The ordinary least square estimation is known to yield biased estimates when analyzing incompletely observed data (Maddala 1983; Cameron and Trivedi 2005).

The simplest version of our model only considers destination's characteristics with and without water quality indices (Table 1, Specification 1 and 3). Specification 2 includes origin's and destination's characteristics except water qualities, while Specification 4 includes both regions' attributes with water qualities. Note that the log-likelihood value of specifications with origin's characteristics is relatively lower than that with destination characteristics only. Based on specification tests, we choose Specification 5 for the discussion of results in the following.

The signs of coefficient estimates are robust across the 5 specifications. They can be interpreted as follows: a hydrologic unit attracts more boaters when (1) the boater lives in a smaller hydrologic unit, (2) the size of the destination unit is bigger, (3) the water bodies in living area (origin) are more scattered, (4) the water bodies in the destination are less dispersed, (5) the origin and the destination are accessible (high road-density), (6) the destination is closer from the origin, (7) the destination unit is adjacent to the Pacific Ocean, (8) the water quality of origin is low, and (9) the water quality of destination is high.

In order to compare the relative impact of independent variables on boat flows, we employ standardized coefficients (Table 1, Specification 5, bStdX). The coefficient bStdX measures the change in boat flows (dependent variable) when an independent variable is increased by one standard deviation, holding all else constant. Results show that water body dispersion and the road density of a destination increase boat flows more than that of the other variables on the right hand side of equation (6). That is, the presence of a large water body in a hydrological unit makes it very attractive to boaters, which increases the likelihood of unintended invasions. If such large water bodies coexist with high accessibility in the form roads, the attraction is larger further increasing the risk of IS introductions. Water quality does not appear to be a strong pulling force for boaters, but may be more important in habitat suitability analysis that follows. Interestingly, the sign of the water quality coefficient in the origin is the opposite of that of the destination area. Based on the estimates of Specification 5, we compute predicted boat flows for each hydrological unit j, i.e. $\sum_i \hat{T}_{ij}$, for use in relative risk of invasions.

Table 2. Gravity Model Results – Oregon (dependent variable: boat flows, T_{ij})

Independent Variables	Specification 1	Specification 2	Specification 3	Specification 4	Specification 5	
					Coefficient	bStdX
Constant	-1312.53*** (92.39)	-1367.64*** (126.27)	-1355.69*** (93.17)	-1382.23*** (127.40)	-1489.93*** (94.92)	.
ln (HUC size)$_i$	.	-16.42* (9.37)	.	-18.41* (9.53)	.	.
ln (HUC size)$_j$	129.61*** (10.30)	129.48*** (10.23)	130.63*** (10,33)	131.17*** (10.29)	131.94*** (10.33)	84.61
ln (Herfindahl)$_i$	.	-10.11*** (1.34)	.	-10.55*** (28.88)	.	.
ln (Herfindahl)$_j$	30.07*** (1.64)	30.09*** (1.63)	28.88*** (1.68)	28.88*** (1.67)	23.93*** (1.68)	123.64
ln (Road Density)$_i$	.	66.17*** (6.87)	.	68.41*** (7.15)	52.91*** (6.06)	48.41
ln (Road Density)$_j$	110.14*** (7.16)	110.55*** (7.13)	117.69*** (7.62)	119.00*** (7.59)	118.77*** (7.61)	108.00
Distance$_{ij}$	-3.52E-04*** (3.80E-05)	-3.47E-04*** (3.84E-05)	-3.59E-04*** (3.80E-05)	-3.48E-04*** (3.84E-05)	-3.27E-04*** (3.80E-05)	-45.77
Dummy (Ocean)$_j$	112.95*** (13.36)	112.36*** (13.24)	115.36*** (13.38)	115.24*** (13.29)	114.23*** (13.34)	43.62
Water Quality$_i$	.	.	7.95*** (1.50)	1.29 (1.61)	5.22*** (1.53)	20.77
Water Quality$_j$	.	.	-6.12*** (1.97)	-6.58*** (1.96)	-6.76*** (1.97)	-22.20
Log likelihood	-19384.68	-19311.01	-19370.54	-19304.96	-19331.52	.

***, ** and * indicate 1%, 5% and 10% level of significance, respectively.
bStdX represents the change in the dependent variable resulting from one standard deviation increase of the independent variable.

Maxent Results

In characterizing NZMS habitat suitability, we considered the following environmental characteristics: elevation, 2010 monthly precipitation, 2010 monthly maximum temperature, 2010 monthly minimum temperature, 2010 county population density, and closeness to primary and secondary roads. All data are transformed to raster data with the same spatial reference (WGS72) and cell sizes. As noted earlier, the NZMS occurrence data come from the U.S. Geological Survey Nonindigenous Aquatic Species database and Montana State University NZMS website. In the following, we report Maxent results based on 25% test samples.

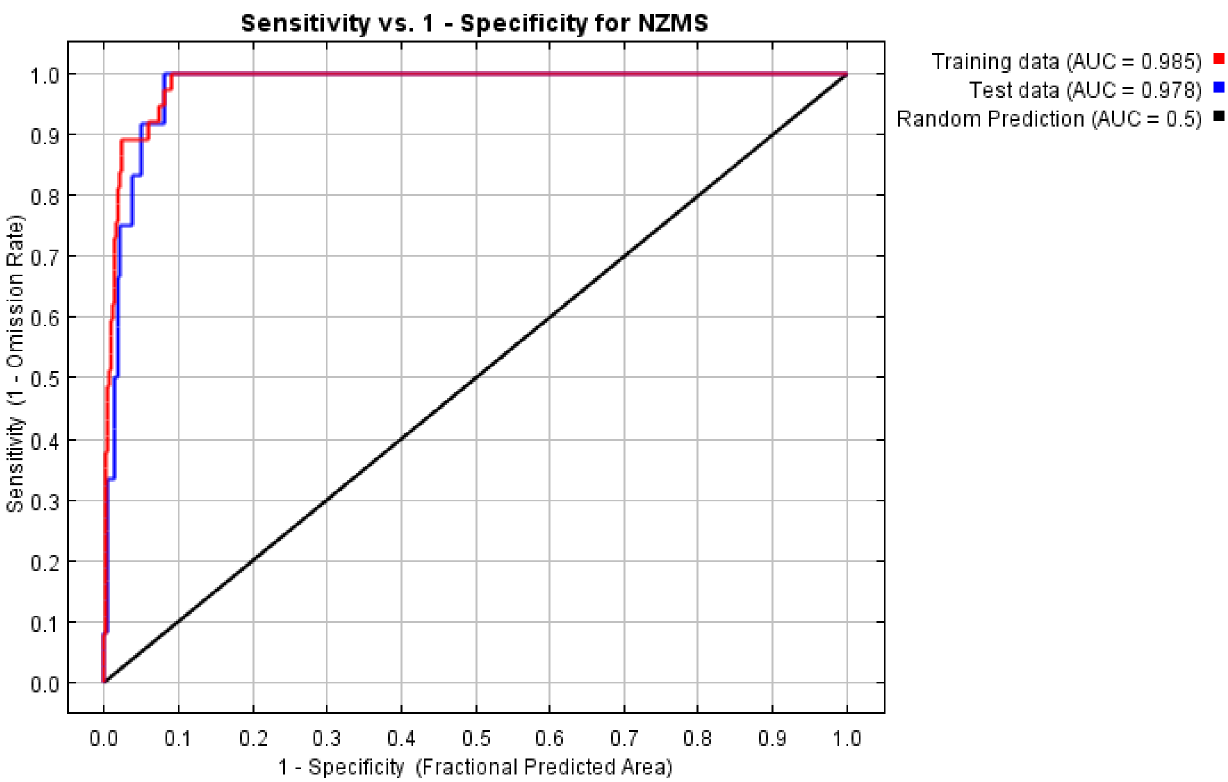

Figure 1. The Receiver-Operating Characteristic (ROC) Curve

Figure 1 presents the receiver-operation characteristic (ROC) and area under the curve (AUC) thereof. The x-axis of the figure is the false-positive rate which measures the prediction of species presence when absence is observed, while the y-axis is the true positive rate which identifies prediction of presence when presence is observed. The higher the AUC (area under the ROC ranging between 0.5 and 1.0), the better is the predictive performance of the Maxent model (Franklin and Miller, 2009). Based on AUC, our application of the Maxent model has predicted NZMS occurrence with high accuracy.

To identify the relative importance of various environmental characteristics for NZMS habitat suitability, we employ two outcomes of the Maxent model: percent contribution and permutation importance of each characteristic. In each algorithmic step, the program heuristically calculates a model gain (maximum entrophy), which becomes the basis for computing a characteristic’s percent contribution. However, the percent contribution heavily relies on the chosen algorithm, and does not reflect correlation among characteristics. Hence, permutation importance appears to be a better measure of a characteristic's explanatory power since it is path (algorithm) independent (Phillips, 2011).

Based on permutation importance, Table 3 shows that elevation ranked first among environmental characteristics suitable to NZMS establishment. Maximum temperature in March, county population density, and precipitation in December are also important factors determining NZMS habitat suitability. Among the characteristics considered, relative importance of county population density and closeness to major roads appears similar to that in the Tobit model (Table 2).

Table 3. Permutation Importance of Environmental Characteristics

Variable	Elevation	Max Temp March	County Pop Density	Precipitation December	Precipitation November	Precipitation May	Precipitation July	Max Temp October	Closeness to Road
Permutation Importance	**71.8**	**6.3**	**4.6**	**4.5**	**3.5**	**1.7**	**1.4**	**1.4**	**1.2**

Figure 2 shows the permutation importance of each monthly variable: precipitation, and maximum and minimum temperatures. The maximum temperature in March appears more important than other variables because temperature at the start of spring season is likely to critically determine initial NZMS population (Allee effect) and its ability to naturally reproduce. The higher importance of the precipitation in November and December shows NZMS habitat suitability's reliance on freshwater. In general, minimum temperatures do not play a critical role in determining NZMS habitat suitability.

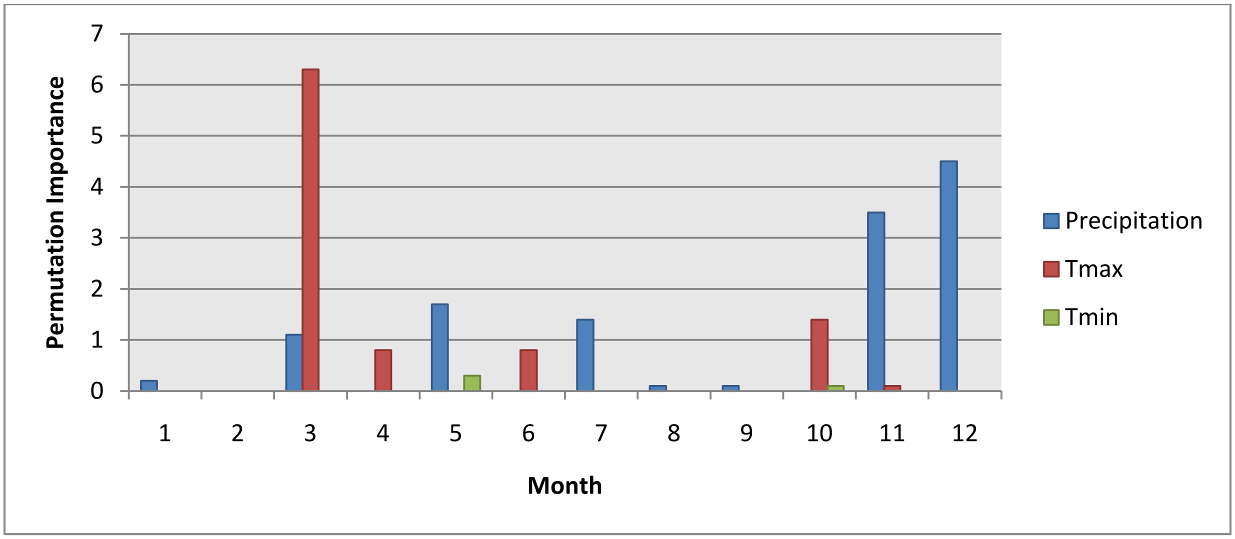

Figure 2. Permutation Importance of Monthly Precipitation and Temperature

Relative Risk Assessment

In this section, we combine the results of the Tobit model and Maxent method for relative risk assessment. First, from the Tobit model we predict boat flows from i to j, i.e. unconditional expected value of the dependent variable (y):

(7) $$E(y|\boldsymbol{X}) = P(y > 0|\boldsymbol{X}) \cdot E(y|\boldsymbol{X}, y > 0) = \Phi(\frac{\boldsymbol{X\beta}}{\sigma})[\boldsymbol{X\beta} + \sigma\lambda\left(\frac{\boldsymbol{X\beta}}{\sigma}\right)]$$

where $\boldsymbol{X}$ is the set of independent variables in equation (6), $\Phi(\cdot)$ is the standard normal cumulative density function, σ is a standard deviation of the normal distribution, and $\lambda(\cdot)$ is the inverse Mills ratio (Wooldridge, 2002). Denoting the expected value as $\hat{T}_{ij}$, we compute total boat flows in each hydrologic unit j as $\sum_i \hat{T}_{ij}$. Since the relative risk of NZMS introduction is proportional to total boat inflow, we normalize it as follows: $\tau_j = \sum_i \hat{T}_{ij} / \sum_j \sum_i \hat{T}_{ij}$. In Figure 3, MAP 1 shows τ_j values across hydrologic units in Oregon.

MAP 2 of Figure 3 charts the relative probability of NZMS establishment from the Maxent estimation, while MAP 3 scales up the values of MAP 2 by those from MAP 1, i.e. integrated relative risk. All three panels paint a similar picture. The maximum relative risk of introduction through boat flows is 0.0342, while that of establishment is 0.9631. Each point in MAP 3 of Figure 3 is the product of relative risk of introduction and establishment probability.

The relative risk assessment in Figure 3 shows, not surprisingly, that areas already infected face the largest threat. However, two hydrologic units (near Portland and Salem), although not infected, face serious threat of NZMS introduction and establishment. The introduction risk comes from higher boat flows to these units, which also appear ecologically suitable for NZMS establishment.

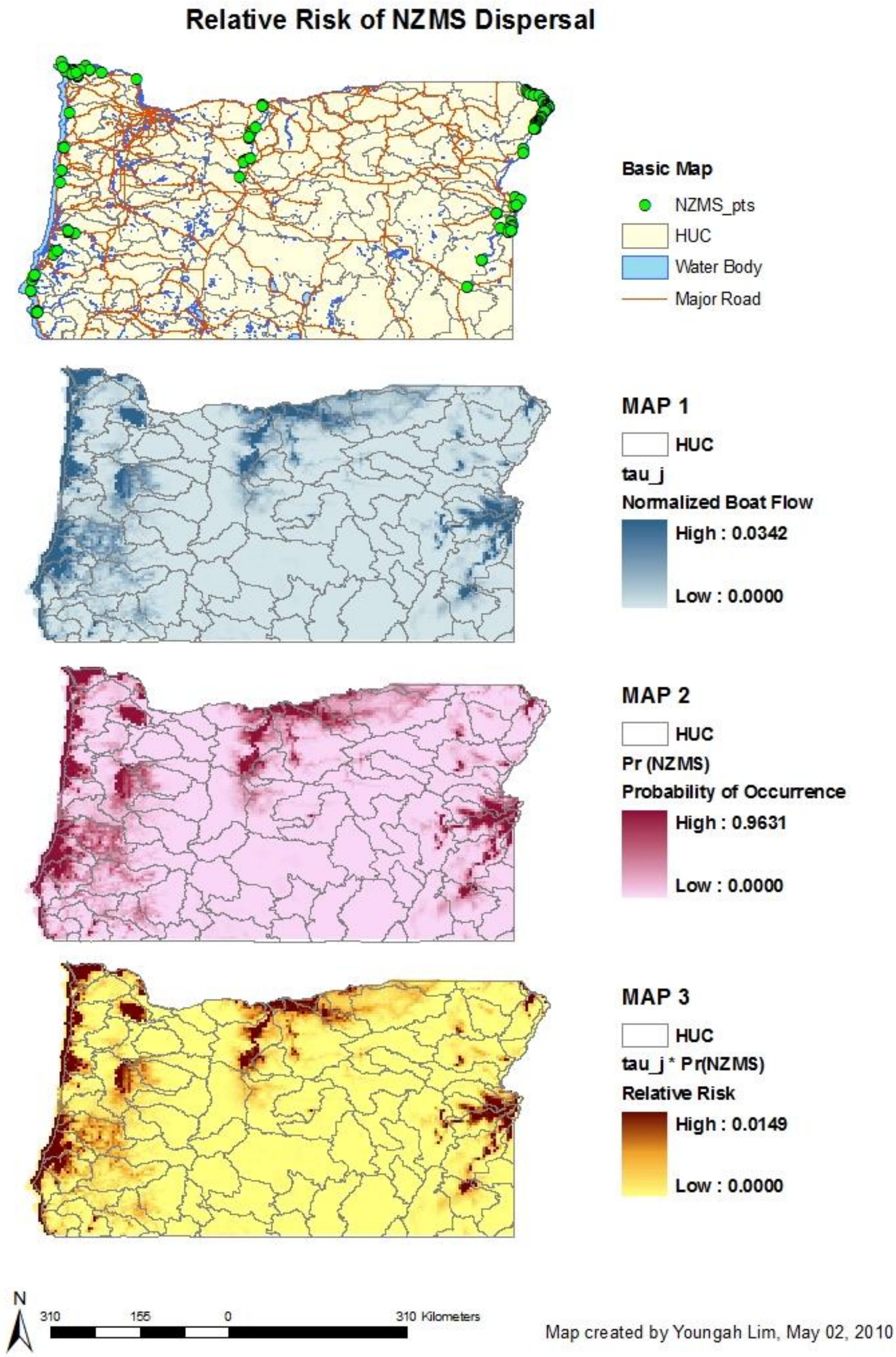

Figure 3. NZMS Relative Risk Map
Normalized Boat Flows, Habitat Suitability and Integrated Risk

Summary and Conclusions

When dealing with invasive species threats, the first step is to assess risk of introduction and establishment. Since anthropogenic and biological factors determine successful bioinvasion, it is important to identify pathways and their relative importance for introduction along with environmental suitability for survival and establishment. In this study, we have proposed a quantitative approach to IS risk assessment by employing a gravity model for species introduction and an ecological niche model for establishment. Our application focuses on New Zealand mudsnails in the Pacific Northwest.

The gravity model relates recreational boat movements, i.e. key pathway of unintentional IS introduction, to distance between water bodies (hydrological units) and the attractiveness of an origin relative to a destination. Results for Oregon suggest that the destination would be more attractive to boaters if it is larger; the boater lives in a smaller area; the water bodies in origin are scattered, but less so in destination; distance between origin and destination; accessibility of destination; the destination is adjacent to the ocean, and the water quality of destination is better than the origin. Results from the Maxent method show that NZMS can establish more successfully in regions with high elevation or high population and closer to major roadways. Also, the maximum temperature of spring and precipitation during winter are important characteristics of an environment suitable for NZMS establishment.

Finally, spatial relative risk of NZMS is derived as the product of outcomes from the gravity and ecological niche models. We have uncovered non-infested areas of Oregon facing high risk of NZMS introduction and current infested areas facing serious establishment. These quantitative risk measures help resource mangers prioritize management and can be replicated for other invasive species posing a threat to the Pacific Northwest.

References

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(6): 1778-1788.

Bossenbroek, J.M. Ladd E. Johnson, Brett Peters, and David M. Lodge. 2007. Forecasting the Expansion of Zebra Mussels in the United States. *Conservation Biology* **21**(3): 800-810.

Cabral, Luis M.B. 2000. *Introduction to Industrial Organization*. Cambridge and London: The MIT Press.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, and Sao Paulo: Cambridge University Press.

Carlton, J.T. 2001. *Introduced species in U.S. coastal waters: Environmental impacts and management priorities.* Pew report. Pew Oceans Commission.

Carlton, J.T. and Gregory M. Ruiz. 2005. Vector Science and Integrated Vector Management in Bioinvasion Ecology: Conceptual Frameworks, pp. 36-58 in Mooney, Harold A., Richard N. Mack, Jeffrey A. McNeely, Laurie E. Neville, Peter Johan Schei, and Jeffrey K. Waage (eds.). *Invasive Alien Species: A New Synthesis.* Washington, Covelo, and London: Island Press.

Carson, Richard T., Nicholas E. Flores, Kerry M. Martin, and Jennifer L. Wright. 1996. *Land Economics*, **72**(1): 80-99.

Chang, Kang-Tsung. 2010. *Introduction to Geographic Information Systems.* 5th ed. New York: The MacGraw-Hill Companies.

Chase, Jonathan M. and Mathew A. Leibold. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. Chicago and London: The University of Chicago Press.

Elith, Jane and Catherine H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32**(1): 66-77.

Elith, Jane et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**(2): 129-151.

Fortheringham, A.S. and M.E. O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications*. Dordrecht, Boston, and London: Kluwer Academic Publishers.

Franklin, Janet, and Jennifer A. Miller. 2009. *Mapping Species Distributions*. New York: Cambridge University Press.

Haynes, Kingsley E. and A. Stewart Fortheringham. 1984. *Gravity and Spatial Interaction Models*. Beverly Hills, London, and New Delhi: SAGE Publications.

Joint Nature Conservation Committee. *Potamopyrgus antipodarum*. N.d. [online] URL: http://www.jncc.gov.uk/page-1713. Accessed Mar 15, 2011.

Leung, Brian, John M. Drake, and David M. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. *Ecology* **85**(6): 1651-1660.

Leung, Brian and Nicholas E. Mandrak. 2007. The risk of establishment of aquatic invasive species: joining invasibility and propagule pressure. *Proceedings of the Royal Society Biological Sciences* **274**(1625): 2603-2609.

Lockwood, J.L., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution* **20**(5): 223-228.

Loo, Sarina E., Ralph Mac Nally, and P.S. Lake. 2007. Forecasting New Zealand mudsnail invasion range: Model comparisons using native and invaded ranges. *Ecological Applications* **17**(1): 181-189.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Mann, Roger, Hans Radtke, Susan S. Hanna, John Duffield, Joel R. Hamilton, and Noelwah R. Netusil. 2010. *Economic Risk Associated with the Potential Establishment of Zebra and Quagga Mussels in the Columbia River Basin*. Independent Economic Analysis Board Task Number 159.

Maynard, G. and D. Nowell. 2009. Biosecurity and quarantine for preventing invasive species, pp. 1-18 in Clout, M.N. and P.A. Williams (eds.). *Invasive Species Management*. New York: Oxford University Press.

Montana State University. New Zealand Mudsnails in the Western USA. Nov 05, 2007. [online] URL: http://www.esg.montana.edu/aim/mollusca/nzms/index.html. Accessed Feb 28, 2011.

National Atlas of the United States. Mar 01, 2011. [online] URL: http://www.nationalatlas.gov/index.html. Accessed Mar 07, 2011.

National Invasive Species Council. 2003. *General Guidelines for the Establishment and Evaluation of Invasive Species Early Detection and Rapid Response Systems*. Version 1. National Invasive Species Council.

National Park Service. New Zealand Mud Snail, Baseline Distribution and Monitoring Study. N.d. [online] URL: http://www.nps.gov/yell/planyourvisit/mudsnail.htm. Accessed Mar 04, 2011.

Noronha, Valerian T. and Michael F. Goodchild. 1992. Modeling Interregional Interaction: Implications for Defining Functional Regions. Annals of the Association of American Geographers **82**(1): 86-102.

Oram, Brian. 2010. The Water Quality Index: Monitoring the Quality of Surfacewaters. November 3. [online] URL: http://www.water-research.net/watrqualindex/index.htm. Accessed Apr 27, 2011.

Oregon Department of Environmental Quality. Laboratory Analytical Storage and Retrieval. n.d. [online] URL: http://deq12.deq.state.or.us/lasar2. Accessed Mar 15, 2011.

Oregon Sea Grant. 2010. *New Zealand Mudsnails: How to prevent the spread of New Zealand mudsnails through field gear*. Oregon Sea Grant.

Ortúzar, Juan De Dios and Luis G. Willumsen. 1994. *Modelling Transport*. Chichester, New York, Brisbane, Toronto, and Singapore: John Wiley & Sons.

Perrings et al. 2002. Biological Invasion Risks and the Public Good: an Economic Perspective. *Conservation Ecology* **6**(1): 1. [online] URL: http://www.consecol.org/vol6/iss1/art1.

Peterson, A. Townsend, Monica Papeş, and Muir Eaton. 2007. Transferability and model evaluating in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**(4): 550-560.

Phillips, Steven J. and Miroslav Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**(2): 161-175.

Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**(3-4): 231-259.

Phillips, Steven. 2011. A Brief Tutorial on Maxent. [online] URL: http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc. Accessed March 5, 2011.

Pimentel, David, Rodolfo Zuniga, and Doug Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* **52**(3): 273-288.

Potapov, A.B. and M.A. Lewis. 2008. Allee Effect and Control of Lake System Invasion. *Bulletin of Mathematical Biology* **70**(5): 1371-1397.

Presidential Document. 1999. *Executive Order 13112*. Federal Register **64**(25): 6183-6186.

PRISM Climate Group. N.d. [online] URL: http://www.prism.oregonstate.edu/index.phtml. Accessed Mar 15, 2011.

Sala, Osvaldo E. et al. 2000. Global Biodiversity Scenarios for the Year 2100. *Science* **287**(5459): 1770-1774.

Schapire, Robert. Maxent software for species habitat modelingand datasets. N.d. [online] URL: http://www.cs.princeton.edu/~schapire/maxent. Accessed Mar 05, 2011.

Schneider, Daniel W., Christopher D. Ellis, and Kevin S. Cummings. 1998. A Transportation Model Assessment of the Risk to Native Mussel Communities from Zebra Mussel Spread. *Conservation Biology* **12**(4): 788-800.

Schreiber, E.S.G., G.P. Quinn, and P.S. Lake. 2003. Distribution of an alien aquatic snail in relation to flow variability, human activities and water quality. *Freshwater Biology* **48**(6): 951-961.

Seaber, Paul, F. Paul Kapinos, and George L. Knapp. 1987. *Hydrologic Unit Maps*. United States Geological Survey Water-Supply Paper 2294. U.G. Geological Survey.

Segurado, Pedro and Miguel B. Araújo. 2004. An evaluation of methods for modeling species distributions. *Journal of Biogeography* **31**(10): 1555-1568.

Shannon, C.E. 1948. A Mathematical Theory of Communication. *The Bell System Technical Journal* **27**(3): 379-423.

UCSB RIVRLab (University of California, Santa Barbara, Riparian Invasion Research Laboratory). New Zealand Mud Snail. N.d. [online] URL: http://rivrlab.msi.ucsb.edu/NZMS/index.php. Accessed Jan 21, 2011.

U.S. Census Bureau. Geography Division. Feb 24, 2011. [online] URL: http://www.census.gov/geo/www. Accessed Mar 07. 2011.

U.S. Environmental Protection Agency. Storage and Retrieval Data. Mar 10, 2011. [online] URL: http://www.epa.gov/storet. Accessed Mar 15, 2011.

U.S. Fish & Wildlife Service. Hagerman National Fish Hatchery. Feb 16, 2011. [online] URL: http://www.fws.gov/hagerman. Accessed Mar 15, 2011.

U.S. Geological Survey. USGS Geographic Data Download. Mar 04, 2010. [online] URL: http://edc2.usgs.gov/geodata/index.php. Accessed Mar 15, 2011

U.S. Geological Survey. Nonindigenous Aquatic Species. *Potamopyrgus antipodarum*. Mar 14, 2011. [online] URL: http://nas.er.usgs.gov/queries/factsheet.aspx?SpeciesID=1008. Accessed Mar 15, 2011.

Vinson, Mark, Tarita Harju, and Eric Dinger. 2007. *Status of New Zealand Mud Snails (Potamopyrgus antipodarum) in the Green River downstream from Flaming Gorge Dam: Current Distribution; Habitat Preference and Invertebrate Changes; Food Web and Fish Effects; and Predicted Distributions*. Final Report for Project Agreements: USFWS—601815G405, NPS—J1242050058, BLM—JSA041003.

Wooldridge, Jeffrey. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge and London: The MIT Press.