Bioeconomic forecasting of invasive species by ecological syndrome

JOHN PAUL SCHMIDT,† MICHAEL SPRINGBORN, AND JOHN M. DRAKE

1Odum School of Ecology, University of Georgia, Athens, Georgia, USA
2Department of Environmental Science and Policy, University of California, Davis, California, USA


Abstract. Invasive non-native species cause enormous economic damage. Although there is both regulative and legislative precedent for policies restricting introduction of potentially invasive species, lack of a unified theory of invasions—particularly with respect to plants—has impeded efforts to implement screening despite empirical patterns suggesting the existence of “invasion syndromes”. Motivated by recent advances in the comparative biology of invasive species, we sought to develop a cost-sensitive model that would associate groups of species according to biological traits and assign them to risk categories based on their invasion potential. Focusing on invasive plants in the US, which are estimated to generate costs of $US 34.7 billion/year, we then combined this scheme with estimates of the per species expected economic losses associated with forgoing trade and with benchmark values for the economic costs associated with plant pests to obtain a decision tool that would maximize economic benefits. If used for screening, this tool is estimated to yield expected net benefits of $80,000–$140,000 per species assessed under very conservative estimates of losses due to invasion.

Key words: bioeconomics; boosted regression trees; comparative biology; cost-sensitive learning; ecological forecasting; environmental economics; invasive plant species; prediction.

INTRODUCTION

Extensive environmental damage has resulted from the introduction of invasive non-indigenous species, with significant economic costs (Parker et al. 1999, Mack et al. 2000, Pimentel et al. 2005, Colautti et al. 2006, Vilà et al. 2010). Most species introduced into new geographic regions do not become established, and most naturalized non-native species do not become pests (Reichard and Hamilton 1997). Of those species which have become invasive, many were intentionally introduced (Reichard 1994, Mack et al. 2000, Mack 2005). While intentional introductions occur across taxonomic groups, we focus here on invasive plants for which the importation of nursery stock has been the major pathway to introduction (Mooney and Hobbs 2000, McNeely 2001, Ruiz and Carlton 2003, APHIS 2007). Indeed, approximately 50% of species now invading wildlands in the continental U.S. and Canada were imported for horticultural purposes (APHIS 2007) and 85% of the woody plant species that have naturalized in North America were introduced primarily via trade in landscape ornamentals while only 14% were introduced for agriculture or production forestry (Reichard and Hamilton 1997, Williamson and Fitter 1999). The trade in biological materials carries inherent environmental risks since the intent of plant...
importers is to propagate and spread new introductions and species are accordingly chosen for an environmental match. There is an urgent need to reduce the number of invasive species introduced and spread by human activities. Given the deliberate nature of most plant introductions, there is considerable scope for intervention.

Motivated by biosecurity concerns, many countries have incorporated restrictions on the importation of biological materials into national environmental policy, and several international agreements address the issue specifically. Australia, New Zealand, and South Africa, particularly, have implemented stringent policies. By contrast, U.S. and E.U. regulations are less restrictive. In the context of live animal imports, the U.S. relies on a reactive approach based on the Lacey Act of 1900. While other countries like Australia and New Zealand review proposed imports proactively, such policies are currently stalled in the U.S. (e.g., H.R. 6311 and H.R. 669) in large part due to concerns regarding the economic burden. However, importation of plants in the U.S. can be restricted under Quarantine 37 (1918). Yet, because Q37 regulations were drafted at a time when the risks associated with plant introductions were perceived to be minor relative to their potential economic value, importation of new species for horticultural purposes is presently allowed without formal risk analysis. Instead, the regulatory design of Q37 relies on a list of prohibited plant taxa (currently 95 vascular plant species and 5 parasitic vascular genera). Taxa not specifically listed as restricted may be imported without a permit (APHIS 2007). To improve vigilance, the Animal and Plant Health Inspection Service (APHIS) has proposed a new regulatory mechanism, NAPPRA (“Not Authorized Pending Pest Risk Analysis”), for screening plants to determine which species should be restricted pending more formal pest risk analysis (74 FR 140, 2009-7-23). Of course, to be environmentally and economically effective, both initial screening and subsequent risk assessment require reliable predictive models.

To date, three major obstacles have precluded development of economically efficient predictive models of invasiveness. (1) Databases of biological traits that are required for comparative studies are still largely incomplete and lack quantitative information in different ways for different groups of species. Missing data limits sample sizes for estimation, and, as a result, the strength of statistical inferences (De'ath and Fabricius 2000). (2) Information about a propensity for invasiveness can be obscured by the low frequency of pests relative to the numbers of non-pest naturalized and introduced species although that frequency appears to be much higher for vertebrates than for plants (Smith et al. 1999, Williamson and Fitter 1999, Jeschke and Strayer 2005). When invasive species form a small set of the possible examples, any signal which might be detected by pattern recognition algorithms is overwhelmed by the noise of the contrast class. (3) Costs of mistakes are asymmetric (Hall et al. 2005). For example, if most introduced species do not become invasive, on an individual basis the naïve prediction that no species would become a pest would have high predictive performance. Such a policy would not be economically efficient, however, since the average losses from an invasive species mistakenly introduced (a false negative) typically greatly exceed the cost of mistakenly denying entry to a non-invasive species (a false positive). In the case of plants, loss estimates (Pimentel et al. 2005) and estimates of the value of the ornamental plant trade (Hall et al. 2005) suggest that for the U.S. the costs of an invasive species exceed the opportunity costs of a benign species mistakenly rejected for trade by a factor ranging from 22 to 5700 (see Appendix A). If economic efficiency is even a minor consideration, it follows that this cost imbalance must be reflected in tools used for decision making (Elkan 2001, Maloof 2003, van der Heijden et al. 2004).

Focusing on the U.S. and Canada, where detailed data are publicly available, we conducted a cost-sensitive study of the relationship between pest class and candidate biological predictors. This extends work by Koop et al. (2011), who developed a weed risk assessment model for the U.S., but made the simplifying assumption that the costs of misclassification were symmetric. Further, our work complements two recent studies that developed cost-sensitive models for efficient classification: Lieli and Springborn (in press) consider Australian weeds while Springborn et al. (2011) examine U.S. herpetofaunal imports. Cost-sensitive classifica-
tion of plant imports in North America has yet to be treated. While all three of the above analyses use standard logistic regression models, here we numerically evaluated boosted regression tree models. The per species expected net benefit (ENB) of implementing such a risk assessment model was calculated for a range of estimated economic parameters. Our study shows that invasive species can be predicted by both “full” and “reduced” models with the predictive performance needed to be cost-effective at the spatial scale relevant to national policy. The models we fit have been visualized in a series of graphical decision boundaries and are therefore easy for decision analysts or regulators to use without special training or expert knowledge. There are therefore now no technical limitations to the application of these or similar technologies to screen new groups of non-indigenous plants to identify species likely to become pests. Furthermore, by demonstrating the importance of incorporating cost, these results have implications for the modeling of invasion risk globally across a range of taxa.

MATERIALS AND METHODS

Scope

From the Plants National Database (U.S. Department of Agriculture, Natural Resources Conservation Service 2012) we compiled a list of 4,744 non-native vascular plant species which have become naturalized in the United States and Canada. Within this geographic area, Plants National Database lists species as native or introduced to the lower 48 states, Canada, Alaska, Hawaii, Puerto Rico, and the Virgin Islands, but does not indicate whether species have been introduced between U.S. regions and become noxious. Therefore, we also included 210 vascular plant species identified by Plants National Database as pests within, but also native to the lower 48 states. We classified these species according to the following criteria. “Weeds,” the broadest category, were defined as species which are considered invasive or potentially invasive by state and federal resource managers, state Exotic Plant Pest Councils, or academic weed specialists. “Noxious species” are species legally designated as such by at least one state. Of the 4,953 species in our database 1,110 (22.4%) are weeds, and 435 (8.7%) are noxious. Importantly, these categories are strict subsets (noxious $\subseteq$ weed). The goal of our study was to develop a decision procedure that would reliably predict membership in these nested categories using readily available data on biotic and ecological traits.

Predictor data

Biotic traits.—As candidate predictors of invasiveness, and in keeping with our goal to provide a turnkey tool for pest classification, we chose traits which were likely to be predictive based on previous studies and for which data were freely available. From Plants National Database we compiled data on life history, growth form, wetland habitat affinity, and hybrid status. Data were complete for all of these variables. Further variables were available for a subset of species. Maximum height data for 1,074 species were gathered from Plants National Database, the Ecological Flora of the British Isles (Fitter and Peat 1994), the LEDA Traitbase (Kleyer et al. 2008), and the Flora of China, Flora of North America, and Flora of Pakistan (eFloras 2008). Based on Plants National Database and the sources above, 948 species were identified as evergreen or deciduous. Values for average seed mass were obtained from the Royal Botanic Gardens Kew Seed Information Database (2008) for 2,800 species. Values for highest reported chromosome number for each species were obtained from the Index to Plant Chromosome Numbers (Goldblatt and Johnson 1979) and the Kew Gardens Plant DNA C-values Database (Bennett and Leitch 2001) for 3,230 species. For seed mass, maximum height, and chromosome number, we compiled values for individual species, and average values for genera when data were available for two or more species within a genus. For these three traits, we calculated a standardized value $\tilde{x}_s = (x_s - \bar{x}_g) / \sigma_g$ where $x_s$ is the measured trait value for species $s$, $\bar{x}_g$ is the mean value for the genus of $s$, and $\sigma_g$ is the genus standard deviation for each species belonging to a genus where data existed in either online database for more than one species. We included standardized values for multiple reasons: (1) the value of species relative to congeners controls in part for relatedness, (2) relative chromosome number can serve as a surrogate for ploidy, and (3) can reveal whether,
for a continuous trait, invasive species diverge from closely related species. From Germplasm Resources Information Network (GRIN) (U.S. Department of Agriculture, Agricultural Research Service 2011), we obtained data on native range size (whether present in one or more of 8 regions: Africa, Temperate Asia, Tropical Asia, Australasia, Europe, North America, South America) and developed summary value ranging from 1 to 8 to quantify the extent for 3,367 species. Finally, to control for phylogenetic non-independence, we derived an identity matrix (each species has a label of 1 or 0 for each taxonomic label) from the Angiosperm Phylogeny Group (Stevens 2001) and included APG II genus, family, and order designations for each species in the set of candidate predictors. Above the order level of taxonomy, we employed a speciational (rather than gradual or continuous) model of character evolution by including all nodes in the APG II supertree while setting all branch lengths to be equal. Since we focus on the repeatability of predictions generated, we were most interested in determining whether trait relationships are highly specific to taxonomic groups, and therefore limit the generality of predictions.

**Statistical analyses**

**Boosted regression trees.**—We used boosted regression trees as the foundation for cost-sensitive classification models that we developed for each of the three classes of invasive plants. While traditional regression trees (e.g., Reichard and Hamilton 1997) produce a single, complicated “best” model or tree, boosted regression trees improve on this method by building and merging results from multiple models using a forward, stage-wise, numerical strategy to minimize the loss in predictive performance (Ridgeway 1999, Elith et al. 2008). Results reported here were obtained using the gbm package in R (Ridgeway 2007).

**Model tuning and selection.**—Because we wished to estimate the expected performance of the model for species that were not used for fitting and to avoid the unrealistically low error rates returned by in-sample comparisons, we randomly divided the data, stratifying by class, into training (75%), and test (25%) sets. Since the response variable in each model was binary, models were fit using a Bernoulli distribution. To maximize performance, samples from the positive (minority) class of noxious or weed species were inverse-weighted to compensate for the data imbalance, and model parameters were tuned in ten-fold cross-validation such that the minimum number of trees exceeded 1,000 (Elith et al. 2008). During fitting, relative importance of predictor variables was calculated as the number of times each variable was selected for splitting, weighted by the squared improvement to the model as a result of each split, averaged over all trees, and rescaled to sum to 100 (Friedman and Meulman 2003). Variables with low importance (<2%) were sequentially eliminated if model performance was not reduced. The final model was then used to predict the final 25% holdout test set, providing the performance estimates we report.

From the final fitted models we extracted the estimated probability of membership in each target class. To construct a decision rule from this result, a cutoff was chosen to serve as a decision boundary as follows. Following model estimation, we calculated the receiver-operator curve (ROC) on the holdout test set. ROC curves represent the tradeoff between the “true positive rate” (the proportion of invasive species correctly identified as “positives”) and the “false positive rate” (the proportion of non-invasive species incorrectly identified as “positives”) across the complete range of possible cutoffs. For ease of manipulation and presentation, we represented the cutoffs as deciles of the output probabilities generated by the fit model, i.e., ranging from 0 to 1. We compared models according to the area under the ROC curve (AUC), a value ranging between 0 and 1 which summarizes the probability that a randomly chosen positive case (invasive) has a higher predicted probability than a randomly chosen negative case (non-invasive). The closer AUC is to 1 the better a model is at discriminating pest from non-pest species. The prediction of withheld test data for final model evaluation was performed only after final selection of predictors and model parameters. To further our goal of supplying a user-friendly tool for decision analysis, after identifying the most predictive model, we fit a set of reduced models using only the three most important variables. Performance of the reduced
models was similarly evaluated for comparison.

**Improving decisions by including costs and benefits**

Because we anticipate the use of our results as a screening tool, our goal was not to minimize statistical error but rather maximize expected net benefits. Therefore the classification step—mapping the risk statistic (generated by the model) to a class label (e.g., “weed” or “non-weed”)—is made expressly dependent on the expected benefits and damages of the resulting imports. The chosen cutoff within the risk statistic domain is the one which generates the greatest expected net benefits—there is no requirement, of course, that this cutoff be associated with the point of equal probability (probability 0.5, odds ratio of 1.0).

To characterize the expected net benefits per species we let \( V_T > 0 \) represent the per species expected benefit of traded imports (equivalently, the forgone benefits of excluded imports) and \( V_L > 0 \) represent the per species expected losses *conditional* on the species truly belonging to the invasive class (equivalently, the forgone damages and control costs of excluding an invasive). We note that if a species is invasive the expected losses typically outweigh the expected benefits: \( V_L > V_T \). The base rate probability that a randomly selected species belongs to the invasive class is given by \( \pi \). Finally, let FPR represent the false positive rate and TPR represent the true positive rate.

If the pre-existing status quo is an “open door” policy for new imports (e.g., as is largely the case in the U.S.), the per species expected net benefits of screening are given by the *benefits* generated when an invasive species is excluded (a “true positive”), net of the *cost* generated when a benign species is mistakenly excluded (a “false positive”):

\[
\text{ENB} = \pi \times \text{TPR} \times (V_L - V_T) - (1 - \pi) \times \text{FPR} \times V_T
\]

\[
\approx \pi \times \text{TPR} \times \left[ \frac{V_L}{V_T} \right] - (1 - \pi) \times \text{FPR}
\]

where \( \pi \) is the probability that a species is invasive; TPR is the probability that an invasive species is so classified; \( V_L - V_T \) is the benefit of excluding an invasive species; \( 1 - \pi \) is the probability that a species is benign; FPR is the probability that a benign species is mistakenly excluded (a “false positive”), net of the cost generated when a false positive is generated when a benign species is mistakenly excluded (a “false positive”).

An important consideration is that discrimination was between weeds and non-weeds or state-listed noxious and non-noxious species—in either case our model was estimated using information on naturalized species. Non-naturalized species have not been considered because few records are available in the U.S. for the majority of species that have been introduced but have not naturalized. However, since we are concerned with imports of naturalizing and non-naturalizing species, we must characterize a base rate of invasion as the invasion propensity of a randomly chosen species for import, not a conditional draw from the pool of species known to naturalize. Thus in our calculations of ENB, we assumed the base rate of invasion (\( \pi \)) to be the current (2010) weed (4.4%) or noxious (1.74%) percentage of all introduced species (estimated at 25,000, Pimentel et al. 2005) rather than the percentage of naturalized species (our data set).

**Results**

**Prediction**

The best predictive models of weeds and noxious species included variables reflecting biological traits, taxonomy, and native range size (Appendix B: Table B1). Predictive performance of full models (models retaining all predictors with >2% importance) was similar for weeds (AUC = 0.77, where AUC = 1 represents perfect classification, and AUC = 0.5 represents random
classification) and noxious species (AUC = 0.75) (Appendix B: Table B2). However, predictive performance was only marginally lower (AUC = 0.73, 0.72) for a reduced model requiring only the three most important predictors for either weeds or noxious species. For the reduced model for weeds, the most important predictors were facultative wetland habitat association, seed mass, and maximum height. For noxious species, the three most important predictors were native range size, seed mass, and highest chromosome number (Appendix B: Table B3). While the two variables shared by both models suggest common causality among these two classes of pest plants, they are nevertheless discriminable. Phylogeny contributed little to prediction of either weeds or noxious species. A comparison of the relative importance of predictors in full models

![Image](image.jpg)

Fig. 1. Decision boundaries at the true-positive-benefit to false-positive-cost ratio ($\psi$) = 10, 20, and 500 for weeds (top), and ($\psi$) = 20, 30, 60, 150, and 500 for noxious (bottom) species. Using this figure, for a given cost ratio $\psi$, a decision-maker optimally rejects a given proposal if the attributes of the proposed species (given on the horizontal and vertical axes) map to a position that lies within the contour (towards lighter shading) for a given value of $\psi$. For instance, using the decision graph for noxious species and assuming the conservative value of $\psi$ = 20, a species with a large native range for which $\log(\text{seed mass}) = 1$ and $\log(\text{highest chromosome number}) = 1.5$ would be denied entry, because it falls within the $\psi = 10$ decision boundary. Predictive models were developed using the $gbm$ package of R (Ridgeway 2007). Contours were obtained from the raw model output by smoothing with the $image.smooth$ function in the $fields$ package of R (Furrer et al. 2009).
and graphs of the marginal plots and histograms for each predictor are available for inspection (Appendix B: Fig. B1). Seed size appears to interact with chromosome number (Fig. 1) such that species with high chromosome counts and seed mass less than ~250 but >0.1 mg are much more likely to be serious (noxious) pests. However, very small-seeded (less than ~0.1 mg) plants in the range of 1–3 m in height, if also facultatively associated with wetland habitats, are likely to be weeds. A similar pattern holds for noxious species: very small-seeded plants are unlikely to be noxious unless the log of highest chromosome number is in the range 1.5–2.0 and native range size is large. These results suggest that there exists a generalized invasive syndrome that can be related to seed mass, plant height, chromosome number, wetland affinity, and native range size.

Estimates of benefit/cost ratio

To use this classification model for admit/deny decisions requires imposing a decision boundary in the space of biotic traits to separate regions of this space where species are too risky to introduce from regions where the risk of invasiveness is acceptably low. The optimal location of the admit/deny decision boundary depends on both model predictive performance and the true-positive-benefit to false-positive-cost ratio $\varphi = (V_N - V_T)/V_N$, where $V_N$ is the level of expected forgone losses (damages and control costs) from excluding an undesirable species and $V_T$ is the expected forgone value of trade when a species is excluded. The true value of $\varphi$ is unknown. To obtain empirically derived bounds on $\varphi$, we estimated $V_T$ from estimates of the willingness-to-pay for imports and expenditures on plants for planting (P4P) imports to the United States during 2005–2009 of $281,200–$410,100 (Appendix A). Although the topic of numerous analyses (Mack et al. 2000, McNeely 2001, Pimentel et al. 2005), the value of $V_N$ has remained elusive and controversial. For our purposes, it suffices to consider a range of $V_N$ that we believe brackets the set of possible values. At the low end, we obtain $V_{N_L}$ by setting $V_T = V_{N_L} \pi$, where $\pi$ is the base rate of invasion (the proportion of naturalized species which have become invasive so far), in which case a decision maker is indifferent to accepting or rejecting a randomly chosen species (in the absence of any risk analysis system) because the expected import value is equal to the expected losses weighted by the likelihood that the species is weedy. Since this level of $V_N$ is far below a commonly cited published estimate discussed below, the cutoff based on this value of $V_N$ is conservative in the sense that it is less stringent (returns the decision to admit for a greater number of species) than would result from any higher estimates of $V_N$. At these levels of $V_N$ and $V_T$, we estimate the benefit-cost ratios, $\varphi_{weed} = 22$ for weeds and $\varphi_{noxious} = 56$ for noxious species (Appendix A), yielding per species ENB values of $80,000–$140,000. As a higher estimate we take the published estimate of the total losses from weeds and noxious species with the broadest scope of impacts considered (Pimentel et al. 2005) and divide by the number of species in these categories. These estimates for $V_N$ result in a range of $\varphi_{weed} = 1500–2200$ for weeds and $\varphi_{noxious} = 3900–5700$ for noxious species (Appendix A). While these bounds are not strict, we believe they bracket the range of opinion on this quantity.

Cost-sensitive decision criteria

Model outputs were expressed in terms of probabilities of weed or noxious class membership which we refer to as the “invasion risk”. Using models of weeds and noxious species restricted to the three most important predictors, we impose a decision boundary in the space of biotic traits to separate regions containing “costly/risky” species from regions where the costs are acceptably low, subject to the tradeoff criterion, $\varphi$. To visualize these results, we graphed optimal decision boundaries for a range of benefit-cost ratios $\varphi$ (Fig. 1). These boundaries discriminate species that exceed the optimal cutoff (high risk) from other species (low risk) for a given value of the benefit-cost ratio and thereby can be used as a “turnkey” method for risk assessment of plants proposed for introduction. Table 1 lists values for key traits for a selection of noxious species falling within the most restrictive ($\varphi = 20$) decision boundary. The models visualized here were designed to be directly applicable to regulation. Where more data and technical expertise are available, cost-sensitive boosted regression tree models with more input variables could also be used.
DISCUSSION

Invasive species policy

We have shown that prediction of nuisance plant species can be achieved using a small set of biotic traits as predictors. Data on the traits we found to be important for prediction are easily compiled from large and growing databases on the Internet, which should enable rapid screening of potentially invasive plants. By evaluating performance in terms of economic outcomes rather than classification error, we provide a decision-making framework for predicting pest status of introduced plant species that is optimized with respect to a tradeoff criterion, $\phi$. Thus, our approach decomposes the problem into its statistical and economic elements. While data to inform the ratio $\phi$ remain imprecise, published recommendations and decisions in practice are nonetheless being made while essentially ignoring available information on $V_L$ and $V_T$ (thus embedding an implicit assumption on the level of $\phi$ which will be ideal only by chance). Therefore, the decomposition presented here will, we hope, promote transparency over elements of the model thereby facilitating an open debate as information improves.

The net present value of a risk assessment policy increases with the base-rate of invasion and the time horizon considered, and decreases with the discount rate (Keller et al. 2007). The true base rate currently may be higher than the historic base rate for several reasons. A larger fraction of species introduced today than historically may be capable of naturalizing because horticultural knowledge and practice has become more sophisticated leading to better climate matching and an increased scale of production. Moreover, potentially suitable climate regions for many introduced exotic species are likely to expand under global change scenarios. Therefore, expected net benefits of the risk assessment method we report on may be much higher especially if the costs of invasive species are reckoned to be higher.

Potential improvements to weed screening methods

Our aim was to show that a small set of biotic traits can be used to identify future plant pests with sufficient predictive performance to allow cost-sensitive screening at large scales. At small

<table>
<thead>
<tr>
<th>Species</th>
<th>Common name</th>
<th>log(HCN)</th>
<th>log(seed mass)</th>
<th>Native range size</th>
<th>FWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abutilon theophrasti</td>
<td>velvetleaf</td>
<td>1.62</td>
<td>0.95</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Aegilops cylindrica</td>
<td>jointed goatgrass</td>
<td>1.62</td>
<td>1.37</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Allaria petiolata</td>
<td>garlic mustard</td>
<td>1.62</td>
<td>0.42</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Cirsium arvense</td>
<td>Canada thistle</td>
<td>1.83</td>
<td>0.11</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Cytisus scoparius</td>
<td>Scotch broom</td>
<td>1.68</td>
<td>0.95</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Euphorbia esula</td>
<td>leafy spurge</td>
<td>1.81</td>
<td>0.54</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Hieracium aurantiacum</td>
<td>orange hawkweed</td>
<td>1.80</td>
<td>−0.80</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Hypericum perforatum</td>
<td>St. Johnswort</td>
<td>1.68</td>
<td>−0.70</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Ligustrum vulgare</td>
<td>common privet</td>
<td>1.66</td>
<td>1.34</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Lythrum salicaria</td>
<td>purple loosestrife</td>
<td>1.78</td>
<td>−0.17</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Phragmites australis</td>
<td>common reed</td>
<td>2.08</td>
<td>−0.70</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Rorippa sylvestris</td>
<td>creeping yellow cress</td>
<td>1.68</td>
<td>−0.96</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Salvia kali</td>
<td>Russian thistle</td>
<td>1.73</td>
<td>0.76</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sorghum halepense</td>
<td>Johnson grass</td>
<td>1.78</td>
<td>0.58</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Zygophyllum fabago</td>
<td>Syrian beancaper</td>
<td>1.45</td>
<td>0.59</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
scales, other screening tools—such as those that have been developed for Australia (Pheloung et al. 1999), modified for Hawaii (Daehler et al. 2004), and tested successfully in a variety of regions (Gordon et al. 2007)—have been shown to have even higher accuracies. For example, high predictive performance was achieved for Hawaii and Pacific Islands by incorporating native range size and/or a binary weediness score (Daehler et al. 2004). It follows that models specific to states or regions—which are likely to differ both in invasibility (particularly as related to climate) and vigilance with respect to invasive species—might perform better than our inclusive model for all the US and Canada. It is conceivable, therefore, that several regional models could be combined to give an overall prediction of invasiveness for the US. However, these earlier weed risk assessments are localized and require data on a list of traits or risk factors, some of which are difficult to acquire. We emphasize the utility of a method that can be applied to large and climatically varied geographic areas while relying on a small set of traits. Finally, given the evidence for time lags between naturalization and multiplication to pest densities (Kowarik 1995), many false positives in our analysis may eventually come to be classed as weeds or noxious species, and model performance is therefore underestimated.

Biotic traits

We found that maximum height, seed mass, facultative wetland habitat association, and highest chromosome number were important predictors of weedy and noxious species. While the value of seed mass and height as predictors has been somewhat equivocal in the past (Pyšek and Richardson 2007), we emphasize that the size of our data set, and the ability of boosted regression trees and similar machine learning approaches to accommodate non-linear relationships and interactions between predictors allows for greater predictive power. Critically, seed size appears to interact with chromosome number, and increasing evidence suggests that relative and absolute chromosome number may serve as surrogates for ploidy, and may be related to invasiveness (Pandit et al. 2011, Schmidt and Drake 2011). In addition, seed mass and standardized seed mass may relate to the ability of invasive species to spread once they have been introduced as surrogates for dispersal, fecundity (seed production), and allocation to seed number versus size. This is consistent with the finding (Moles et al. 2004) that seed mass was positively related to plant height, time to first reproduction, plant life span, and plant reproductive life span. We further suspect that the interaction of seed mass and maximum height is related to performance traits such as leaf-area allocation, growth rate, and fitness, traits for which invasive plants were found to have higher values than non-invasive plants in a meta-analysis investigating differences between the two groups (van Kleunen et al. 2010). While forming small minorities within our data set, species which are facultatively associated with wetlands (16% of naturalized species in North America) and species with large native range sizes (<1% of the 1291 species for which we were able to gather range size data) were disproportionately likely to be classified as either weeds or noxious (Appendix B: Fig. B1). Facultative wetland species are species tolerant of a range of environmental conditions. For such species, wetlands may offer an entry point for invasion since they are highly productive habitats, typically exhibiting a high degree of natural and anthropogenic disturbance and which are connected by water bodies that serve as dispersal corridors. Thus, although a clear and general set of mechanisms by which non-indigenous plants establish, spread, and multiply to pest densities is not well-established (Pyšek and Richardson 2007, van Kleunen et al. 2010), our results (Appendix B: Tables B3 and B4) suggest that there may nevertheless be a set of broad “invasiveness syndromes” encompassing both agricultural weeds and more aggressive natural area invaders characterized by interactions among just five traits: seed mass, plant height, chromosome number/ploidy, wetland affinity, and native range size.

Conclusions

Given the lengthy effect of invasions once initiated and potential changes to species distributions as a result of climate change, a risk assessment with the level of performance we have achieved can provide substantial economic benefits. These results are of immediate relevance to environmental policy. As a means to guiding
policy through the process that already exists under Quarantine 37, we have demonstrated that risk assessment methods can achieve sufficiently high levels of prediction with only a few easily acquired predictors, and we provide graphical decision trees that support application of this model to new species not considered by us. Critically, we evaluated performance in terms of economic outcomes rather than classification error. The result is a decision-making framework for predicting pest status of introduced plant species that is, therefore, optimized with respect to overall social welfare. Furthermore, this framework may be easily updated as understanding of key parameters ($\pi, \varphi$) and economic valuation of potential outcomes ($V_D, V_I$) is improved.

ACKNOWLEDGMENTS

Research was funded by USDA agreement no. 10-21-RR551 to John M. Drake. We thank Reuben Keller, and David Lodge for useful comments, Kevin Samples, Krisztian Magori, and Drew Kramer for technical assistance and Lynn Garrett and Cathy Sue at USDA’s Animal and Plant Health Inspection Service for assistance with the PPQ 264 plant import database. J.P. Schmidt compiled data, and performed boosted regression tree analysis. M. Springborn guided the economic analysis. All three authors contributed to writing, logically organizing, and editing this article.

LITERATURE CITED


efloras. 2008. Missouri Botanical Garden, St. Louis, Missouri and Harvard University Herbaria, Cambridge, Massachusetts, USA. http://www.efloras.org


Goldblatt, P., and D. E. Johnson, editors. 1979. Index to plant chromosome numbers. Missouri Botanical Garden, St. Louis, Missouri, USA.


Malloof, M. 2003. Learning when data sets are imbalanced and when costs are unequal and unknown. ICML Workshop on Imbalanced Datasets.


SUPPLEMENTAL MATERIAL
APPENDIX A

BENEFIT–COST ESTIMATION METHODS

Here we describe our approach for estimating the expected present value of losses (damage and control costs) from a new invasive plant species, $V_L$, and welfare lost from forgone imports of a new plant species, $V_T$, to establish a range for the benefit-cost ratio $\varphi = (V_L - V_T)/V_T$.

We took a two-step approach to estimating $V_T$—the present expected value of lost gains in trade from forgoing trade of a randomly chosen species—as follows. First, we estimated the aggregate welfare generated by all plants for planting (P4P) imports to the U.S. Then, following Keller et al. (2007), we divided this figure by the estimated number of unique plant species imported to the U.S. to approximate the mean species value. To obtain the aggregate P4P import welfare component, we calculated the lost domestic surplus from forgoing imports using Feenstra’s (2004) import demand model. Feenstra (2004) demonstrated how lost welfare from constraining imports can be approximated from information on import demand (willingness to pay for imported goods) and import expenditures. The price elasticity of import demand (hereafter, “demand elasticity”) is represented by $\varepsilon < 0$ and specifies the percentage change in the quantity of imports demanded given a 1 percent change in price of imports. Letting $R$ represent expenditures on P4P imports, lost surplus from forgoing trade was estimated from $W = R/(−2\varepsilon)$ (Feenstra 2004), a form of the equation that is convenient given the availability of both trade data to characterize $R$ and estimates of import demand elasticities from the empirical economic literature to set $\varepsilon$.

We obtained data (2005–2009) on import expenditures (U.S. Department of Commerce 2010) via the Global Agricultural Trade System (GATS) online database. Nominal import values were converted to real 2009 dollars using the Consumer Price Index (Bureau of Labor and Statistics 2010). Demand elasticities for U.S. imports at a highly disaggregated scale (six-digit Harmonized System code level) were recently estimated by Kee et al. (2008). These elasticity parameters enabled us to estimate $W$ individually for eight sub-groups within the P4P collection of goods. Over the past five years, annual expenditures on P4P imports have ranged from a minimum of $206,200,000 in 2009 to a maximum of $262,100,000 in 2005, with an average of $242,400,000.

To estimate the number of unique plant species available for import to the U.S. we examined records from the PPQ 264 database maintained by the USDA’s Animal and Plant Health Inspection Service identifying U.S. plant imports from 2005 to the present. These records only identify plants to the genus level. Therefore, we also estimated the average number of species traded per genus, as described below. From the PPQ 264 data, our desired statistic was the number of unique genera available for import in a given year. We used the number of unique genera actually imported as a lower bound on this number since the number available for import can be no less than the number actually imported. We also constructed an alternative estimate which assumes that by the final, sixth year in the available record the vast majority of all genera available for import will have appeared in the record. Using the increase in unique genera between the fifth and sixth years as an estimate of year-on-year increase in unique genera available, we estimated the number of unique genera available for import over the past six years. To map number of unique genera to number of unique species, we required an estimate of the average number of species in trade per genus. The New Zealand Ministry of Agriculture and Forestry (NZMAF 2010) maintains an extensive list of plant species which is used as a reference in the border assessment of imported plants. This Biosecurity Index includes...

www.esajournals.org 12 May 2012  Volume 3(5)  Article 46
approximately 3,500 unique genera with an average of 6.86 species per genus (with a standard deviation of 18.74).

Combining our estimates of (1) the overall welfare value of imports, (2) the unique number of genera imported, and (3) the average number of species per genus, we estimated the per species average annual benefit of imports. Using the "actually traded" unique genera we calculated an annual value of $20,500 per species and using the "estimated available" unique genera we calculate an annual value of $14,100 per species. Our final objective was to identify the present value of the annual stream of expected benefits from imports. Assuming a discount rate of $q = 0.05$, our high and low estimates of $V_T$ are given by $\frac{20,500}{0.05} = \frac{14,100}{0.05} = 410,100$ or $281,200$. (Note, however, that $\varphi$ is invariant to the choice of discount rate since it can be factored out of the numerator and denominator and canceled. Thus, the optimal decision of whether to reject or accept a proposed species is not sensitive to the choice of discount rate, although estimates of the expected net benefits from risk assessment are a function of $\rho$.)

The most difficult parameter to estimate is $V_L$, the present value of expected long-run losses from a weedy or noxious species. While the best available estimate for weed losses is arguably from Pimentel et al. (2005), the methodology has been criticized (Lovell et al. 2006). Pimentel (2009) notes that while the estimate is perhaps upward biased due to an implicit assumption that native weeds will not substitute for avoided exotic weeds, the loss estimates are also downward biased since they do not include many ecosystem impacts, for example from herbicide and pesticides. Instead of identifying a preferred estimate for $V_L$, we specify a reference or benchmark value which is characterized in relation to import benefits ($V_T$) which are better understood. For the benchmark case we set $V_T = V_L \pi$. In this case, a decision maker is indifferent to accepting or rejecting a randomly chosen species (in the absence of any risk analysis system) because the expected import value ($V_T$) is equal to the expected weed loss weighted by the likelihood that the species is weedy ($V_L \pi$).

### Table A1. Expected net benefits per species assessed of weed risk assessment under two approaches to characterizing $V_L$, a conservative “minimal” estimate and values based on Pimentel et al. (2005).

<table>
<thead>
<tr>
<th>Category</th>
<th>Risk</th>
<th>$V_T$</th>
<th>$V_L$</th>
<th>$\pi$</th>
<th>$\varphi$</th>
<th>TPR</th>
<th>FPR</th>
<th>ENB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>Noxious</td>
<td>High</td>
<td>$410,100$</td>
<td>$23,568,966$</td>
<td>0.0174</td>
<td>56</td>
<td>0.94</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Noxious</td>
<td>Low</td>
<td>$281,200$</td>
<td>$16,160,920$</td>
<td>0.0174</td>
<td>56</td>
<td>0.94</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Weed</td>
<td>High</td>
<td>$410,100$</td>
<td>$9,320,435$</td>
<td>0.044</td>
<td>22</td>
<td>0.59</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Weed</td>
<td>Low</td>
<td>$281,200$</td>
<td>$6,390,909$</td>
<td>0.044</td>
<td>22</td>
<td>0.59</td>
<td>0.23</td>
</tr>
<tr>
<td>Pimentel 2005</td>
<td>Noxious</td>
<td>High</td>
<td>$410,100$</td>
<td>$1,593,563,218$</td>
<td>0.0174</td>
<td>5666</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Noxious</td>
<td>Low</td>
<td>$281,200$</td>
<td>$1,593,563,218$</td>
<td>0.0174</td>
<td>3885</td>
<td>1</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Weed</td>
<td>High</td>
<td>$410,100$</td>
<td>$630,181,818$</td>
<td>0.044</td>
<td>2240</td>
<td>0.99</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Weed</td>
<td>Low</td>
<td>$281,200$</td>
<td>$630,181,818$</td>
<td>0.044</td>
<td>1536</td>
<td>0.99</td>
<td>0.58</td>
</tr>
</tbody>
</table>
## Appendix B

### Importance Values of Predictors and Marginal Plots of Boosted Regression Tree Analysis

Table B1. Predictive performance of models measured by area under the ROC curve (AUC). Results under training/training are for the training sets predicted on the same. Results under training/test are for training sets predicted on the holdout test set.

<table>
<thead>
<tr>
<th>Model</th>
<th>Weed</th>
<th>Noxious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training/training</td>
<td>Training/test</td>
</tr>
<tr>
<td>Full model with taxonomy</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>Full model</td>
<td>0.8</td>
<td>0.77</td>
</tr>
<tr>
<td>Reduced model</td>
<td>0.75</td>
<td>0.73</td>
</tr>
<tr>
<td>Smoothed reduced model</td>
<td>0.73</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table B2. Full models (without native range size) of weed and noxious classes which include all predictors with importance values (IV) > 2 during training. Values for importance below are from the model testing stage when training sets are used to predict a 25% holdout test set.

<table>
<thead>
<tr>
<th>Weed</th>
<th>IV</th>
<th>Noxious</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facultative wetland association</td>
<td>34.5</td>
<td>log10(seed mass)</td>
<td>25.3</td>
</tr>
<tr>
<td>log10(seed mass)</td>
<td>20.2</td>
<td>Native range size</td>
<td>15.2</td>
</tr>
<tr>
<td>log10(maximum height)</td>
<td>12.7</td>
<td>log10(highest chromosome number)</td>
<td>14.7</td>
</tr>
<tr>
<td>log10(highest chromosome number)</td>
<td>7.3</td>
<td>Seed mass standardized by genus</td>
<td>9.4</td>
</tr>
<tr>
<td>Obligate wetland association</td>
<td>5.8</td>
<td>log10(maximum height)</td>
<td>9.0</td>
</tr>
<tr>
<td>Native range size</td>
<td>5.3</td>
<td>Facultative wetland association</td>
<td>6.5</td>
</tr>
<tr>
<td>Berry</td>
<td>3.4</td>
<td>Obligate wetland association</td>
<td>6.4</td>
</tr>
<tr>
<td>Parasitic</td>
<td>3.3</td>
<td>Highest chromosome num.</td>
<td>6.3</td>
</tr>
<tr>
<td>Highest chromosome num.</td>
<td>3.3</td>
<td>Num. standardized by genus</td>
<td>6.2</td>
</tr>
<tr>
<td>Standardized by genus</td>
<td>3.3</td>
<td>Evergreen</td>
<td>4.9</td>
</tr>
<tr>
<td>Seed mass standardized by genus</td>
<td>2.8</td>
<td>Vine</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table B3. Top predictors ranked by importance value (IV) for reduced models which include only the top three predictors based on importance. Values for importance below are from the model testing stage when training sets are used to predict a 25% holdout test set.

<table>
<thead>
<tr>
<th>Weed</th>
<th>IV</th>
<th>Noxious</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facultative wetland association</td>
<td>47.6</td>
<td>log10(seed mass)</td>
<td>44.0</td>
</tr>
<tr>
<td>log10(seed mass)</td>
<td>32.7</td>
<td>log10(highest chromosome number)</td>
<td>28.3</td>
</tr>
<tr>
<td>log10(maximum height)</td>
<td>19.7</td>
<td>Native range size</td>
<td>27.6</td>
</tr>
</tbody>
</table>
Fig. B1. Weed model. Partial dependence plots (improvements of the final GBM model as a function of a single predictor using the methods of Friedman (2001)) overlaid on a frequency histogram (summarizing over the entire dataset of 5953 species) of each predictor. The y-axes are on the logit scale and are centered to have zero mean over the data distribution. Graphs are arranged by order of predictor importance.
Fig. B1 (continued).
Fig. B2. Noxious model. Partial dependence plots (improvements of the final GBM model as a function of a single predictor using the methods of Friedman (2001)) overlaid on a frequency histogram (summarizing over the entire dataset of 5953 species) of each predictor. The y-axes are on the logit scale and are centered to have zero mean over the data distribution. Graphs are arranged by order of predictor importance.
Fig. B2 (continued).