INVASIVE POTENTIAL OF INVASIVE PLANTS IN THE FOREST OF THE SOUTHERN REGION, UNITED STATES

Dawn Lemke, John W. Coulston, Philip Hulme, Clint Paterson, and Jennifer A. Brown[1]

Abstract.—Alien plants introduced for commercial or landscaping use have caused substantial problems as invaders of natural and managed ecosystems. The magnitude of the problem has dramatically increased over the past few decades with accelerated land disturbance, land use changes, and global and internal transportation. In the southern region of the United States, invasive plants are one of the threats to the long-term sustainability of our forest ecosystems along with climate change and land use change. We assessed the potential distribution of invasive plants in forests of the southern region using data from the invasive species component of the U.S.Forest Service Forest Inventory and Analysis (FIA) Program and freely available digital data including elevation, climate, and land use. Using an ensemble modeling approach, we integrated maximum entropy algorithms, logistic regression, random forest, boosted regression trees, and support vector machine. Areas of agreement between models were considered areas of high probability. This suggests the importance of adaptive management and long-term monitoring programs and the need for further development of methods for assessing probable future climate conditions. We have used this approach to evaluate the relative importance of dependent variables and the application and selection of modeling techniques.

INTRODUCTION

Invasive species pose a major threat to the sustainability of natural ecosystems through biotic homogenization and loss of biodiversity, with negative consequences for both social and economic systems (Miller et al. 2012). Invasive species are considered a major component of global environmental change (Vitousek et al. 1997). Identifying areas of potential invasion is an important part of ecosystem management, and one tool that can be applied to this is species distribution models (SDMs) (Gallien et al. 2010). SDMs can be used to predict spatial patterns of potential biological invasions and prioritize locations for early detection and control of invasion outbreaks. SDMs combine concepts from ecology and natural history with more recent developments in statistics and geospatial information systems (Franklin 2009). In this paper we focused on two questions specific to the application of SDMs: 1) Which modeling technique(s) is most appropriate for this study?; and 2) Do environmental determinates remain consistent among models? To address these questions we developed SDMs for 22 plants invasive to the forests of the southern region of the United States using five SDM methods.

METHODS

Invasive plants considered for this study included all species with more than 100 plot occurrences in the invasive plant component of the FIA database (USDA

[1] Ph.D. Candidate and Research Associate (DL), Canterbury University, New Zealand and Alabama A&M University, Normal, AL; Supervisory Research Forester (JWC), U.S. Forest Service, Southern Research Station; Professor of Plant Biosecurity (PH), Lincoln University, New Zealand; Research Associate (CP), Alabama A&M University; Professor of Statistics (JAB), Canterbury University, New Zealand. DL is corresponding author: to contact, call 256-372-4562, or email at dawn.lemke@aamu.edu.

FS 2007) (Table 1). Twenty-two environmental variables derived from the national land cover database, digital elevation models, and Bioclim data were used (full details given in Lemke et al. 2011) (Table 2). Environmental variables were checked for intercorrelation and the statistical package R was used to develop the following five models for each species: maximum entropy algorithms, logistic regression, random forest, boosted regression trees, and support vector machine (R Core Team 2012). Data were down sampled to give a 1:4 ratio (for every occurrence location, four random absence records were selected) for logistic regression, random forest, boosted regression trees, and support vector machine models to balance the data. Models were derived using a manual backward selection method where variables that had little or no impact on the model were removed based on the results of 10 model runs (Lemke et al. 2011). The key variables in determining the occurrence of each species were identified by their percent contribution to the final model and with a jack-knife test on gain and influence on the area under the curve (AUC). This approach assisted in reducing a model that over fits. Three techniques were used to assess model reliability: the performance of test and training data, the omission rate, and AUC. Data were randomly split with 30 percent in test and 70 percent in training datasets for the regional models and were run 10 times with random selections. The omission rate was calculated using a threshold value defined by the maximized sum of sensitivity and specificity. Models with an omission rate less than 0.25 and an AUC of greater than 0.75 were considered acceptable. Ensemble models were built for each species that had more than one acceptable model using only the acceptable models. When more than 75 percent of the models agreed in occurrence, these areas were considered highly likely to be invaded, when less than 25 percent of the models agreed in occurrence, these areas were considered highly unlikely to be invaded,

Table 1.—Comparison of five species distribution modeling techniques (boosted regression trees [BRT], logistic regression [LR], maximum entropy algorithms [ME], random forest [RF], and support vector machine [SVM]) for 22 species invasive to the forest of the Southern region. An omission rate (OR) less than 0.25 and an area under the curve (AUC) of greater than 0.75 were considered acceptable models, shown in bold.

		BRT		LR		ME		RF		SVM	
Species	**n**	**AUC**	**OR**	**AUC**	**OR**	**AUC**	**OR**	**AUC**	**OR**	**AUC**	**OR**
Tree of heaven (*Ailanthus altissima*)	854	**0.88**	**0.14**	**0.81**	**0.18**	**0.88**	**0.13**	**0.89**	0.43	**0.80**	**0.17**
Silktree (*Albizia julibrissin*)	677	**0.76**	0.32	0.69	0.42	**0.76**	**0.24**	0.72	0.80	0.59	0.49
Princesstree (*Paulownia tomentosa*)	231	**0.81**	**0.22**	0.73	**0.23**	**0.80**	0.26	**0.79**	0.67	0.64	0.33
Chinaberry (*Melia azedarach*)	468	**0.87**	**0.20**	**0.79**	**0.16**	**0.86**	**0.22**	**0.86**	0.52	**0.82**	0.29
Tallowtree (*Triadica sebifera*)	930	**0.93**	**0.14**	**0.88**	**0.22**	**0.93**	**0.13**	**0.94**	0.30	**0.89**	**0.22**
Autumn olive (*Elaeagnus umbellate*)	327	**0.88**	**0.16**	**0.75**	**0.20**	**0.98**	**0.19**	**0.89**	0.44	**0.82**	**0.17**
Privets (*Ligustrum* L.)	7580	**0.81**	**0.23**	0.72	0.31	**0.77**	**0.21**	N/A		0.62	0.29
Bush honeysuckles (*Diervilla* spp.)	499	**0.89**	**0.22**	0.74	0.38	**0.89**	**0.18**	**0.90**	0.40	**0.84**	**0.18**
Nandina (*Nandina* Thumb.)	143	**0.81**	**0.21**	**0.77**	0.36	**0.80**	0.38	**0.81**	0.67	0.73	0.35
Nonnative roses (*Rosa* spp.)	3031	**0.87**	**0.18**	**0.76**	**0.21**	**0.85**	**0.16**	**0.88**	0.46	0.74	0.35
Climbing yams (*Dioscorea* L.)	120	**0.78**	0.35	0.62	0.31	**0.83**	**0.22**	**0.79**	0.52	0.61	0.31
English ivy (*Hedera helix*)	104	**0.85**	0.29	**0.79**	0.33	**0.84**	**0.21**	**0.88**	0.58	**0.81**	0.36
Japanese honeysuckle (*Lonicera japonica*)	15931	**0.82**	**0.21**	0.69	**0.23**	0.72	**0.20**	N/A		0.71	0.30
Kudzu (*Pueraria* spp.)	280	**0.79**	0.33	0.71	**0.24**	**0.82**	0.30	**0.82**	0.64	0.62	0.28
Periwinkles (*Vinca* spp.)	115	0.74	0.42	0.67	0.47	0.74	0.42	0.77	**0.77**	0.61	**0.20**
Nonnative wisterias (*Wisteria* spp.)	113	**0.78**	0.35	**0.75**	**0.25**	**0.80**	0.30	0.74	0.80	0.65	**0.23**
Tall fescue (*Schedonorus phoenix* (Scop.) Holub)	810	**0.85**	**0.22**	0.72	0.28	**0.82**	**0.25**	**0.85**	0.52	0.62	0.32
Nepalese browntop (*Microstegium vimineum*)	1740	**0.86**	**0.15**	0.73	**0.22**	**0.83**	**0.12**	**0.86**	0.51	0.56	**0.21**
Japanese climbing fern (*Lygodium japonicum*)	1299	**0.93**	**0.11**	**0.89**	**0.15**	**0.92**	**0.08**	**0.97**	0.27	**0.90**	**0.08**
Garlic mustard (*Alliaria petiolata*)	105	**0.95**	**0.17**	**0.82**	**0.11**	**0.97**	**0.19**	**0.97**	**0.19**	**0.94**	**0.13**
Shrubby lespedeza (*Lespedeza frutescens*)	964	**0.82**	0.35	0.67	0.39	**0.79**	0.30	**0.79**	0.68	0.54	**0.19**
Chinese lespedeza (*Lespedeza cuneata*)	1909	**0.77**	0.26	0.62	0.39	**0.76**	0.27	**0.78**	0.66	0.53	**0.10**
Percent of acceptable models		86%	59%	45%	55%	91%	68%	90%	5%	36%	50%

Table 2.—Number of models using each of 22 environmental variables across five species distribution modeling techniques (boosted regression trees [BRT], logistic regression [LR], maximum entropy algorithms [ME], random forest [RF], and support vector machine [SVM]) for 22 species invasive to the forests of the Southern region

	Elevation	Slope	Dist River	Min Temp	Rainfall	Rain in the Wettest Month	Rain in the Driest Month	Rain in the Warmest Quarter	Temp Range	Temp in the Wettest Quarter	Temp in the Driest Quarter	Pop Density	Dist City	Dist Main Road	Dist Road	% Farming	% Forest	Change in Forest	% Grass	% Pine	Change in Pine	Residential
Tree of heaven	3	1	0	4	2	4	2	2	0	2	3	2	1	0	0	1	1	0	1	0	1	1
Silktree	4	0	0	4	2	3	2	2	1	2	4	3	1	0	2	3	1	0	0	0	0	5
Princesstree	3	4	1	4	2	3	2	3	0	1	2	2	2	0	0	1	1	0	0	1	0	1
Chinaberry	3	0	0	5	2	3	3	3	2	3	2	0	0	0	0	5	0	1	0	0	0	0
Tallowtree	1	0	0	5	2	2	5	3	1	1	3	2	0	0	0	0	3	0	0	0	0	0
Autumn olive	4	0	0	3	3	4	1	1	0	4	2	4	0	0	0	1	2	0	1	0	0	1
Privets	3	0	1	3	1	1	1	2	1	2	1	3	0	0	0	4	0	0	0	0	0	1
Bush honeysuckles	4	0	0	4	1	4	2	2	1	2	2	3	0	0	0	3	2	0	0	0	0	1
Nandina	2	0	0	5	3	3	1	2	2	3	1	5	2	0	1	2	1	0	0	0	0	4
Nonnative roses	4	0	0	4	1	1	2	3	0	1	2	0	0	0	0	4	5	0	1	0	0	1
Climbing yams	4	1	0	4	1	1	0	3	1	3	4	4	1	1	0	0	0	1	0	2	0	0
English ivy	1	0	0	4	1	1	1	0	2	0	1	5	2	0	2	1	0	0	0	1	0	5
Japanese honeysuckle	3	0	0	3	0	0	3	1	2	2	4	1	1	0	0	3	1	0	0	0	0	1
Kudzu	3	0	1	3	2	1	2	3	1	4	4	3	1	1	2	3	1	0	0	1	0	5
Periwinkles	1	4	0	3	2	2	3	1	0	4	3	4	0	1	0	3	2	2	0	1	0	2
Nonnative wisterias	2	0	0	5	0	0	0	0	4	1	0	4	2	2	2	3	3	0	0	1	0	5
Tall fescue	4	0	0	4	2	4	2	4	0	3	2	0	0	0	0	3	2	0	1	0	0	1
Nepalese browntop	3	0	0	4	2	3	4	4	0	2	4	2	0	0	0	1	1	0	1	0	0	1
Japanese climbing fern	1	0	0	5	5	2	2	3	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Garlic mustard	3	0	0	3	1	2	0	2	2	1	2	2	0	0	0	1	1	0	0	0	0	1
Shrubby lespedeza	4	1	0	4	3	4	3	4	2	2	3	0	0	1	0	2	4	0	1	2	0	0
Chinese lespedeza	3	0	0	4	3	4	1	3	1	1	3	0	0	0	0	1	2	0	4	1	0	1
BRT (%)	73	9	0	95	36	45	23	41	14	45	59	64	27	18	18	50	18	9	5	5	0	18
LR (%)	14	14	0	23	5	0	50	0	45	9	32	23	0	5	0	86	59	5	32	0	0	77
ME (%)	59	5	0	95	14	41	32	59	14	41	41	64	14	5	23	45	32	0	5	27	0	27
RF (%)	100	15	0	100	95	95	65	90	20	90	90	20	5	0	0	10	30	0	0	5	0	25
SVM (%)	50	9	14	95	45	68	27	55	18	36	23	55	14	0	0	14	14	5	5	9	5	23
Overall (%)	58	10	3	81	38	49	39	48	22	44	48	45	12	6	8	42	31	4	9	9	1	34

and when the model agreement was between 25 and 75 percent the area was considered moderately likely to be invaded.

RESULTS

The results are reported in two components: model comparisons, and the influence and relevance of the dependent variables (environmental). Most species had at least one acceptable model as assessed by both test AUC and omission rates with the exception of kudzu, periwinkles, shrubby lespedeza and Chinese lespedeza (Table 1). Only one species, garlic mustard, had five accepted models, and four species (tree of heaven, tallowtree, autumn olive, and Japanese climbing fern) had four acceptable models (logistic regression, maximum entropy algorithms, boosted regression trees, and support vector machine). Overall, boosted regression tree and maximum entropy algorithms produced the strongest models with 59 percent of models considered acceptable (Table 1). Minimum temperature was the most useful of the dependent variables, occurring in 81 percent of the models, followed by elevation (58 percent) and rainfall in the wettest month (49 percent) (Table 2). Every species used minimum temperature in at least three of the models and elevation in at least one of the models (Table 2). Seven variables (distance to roads and rivers, proportion of grass and pine, and change in forest and pine) contributed information to less than 10 percent of the models (Table 2). Tree of heaven was the only species to use the dependent variable change in pine, and distance to river was only used in the support vector machine models for three species. On average, the logistic regression models use the fewest number of variables (five) and random forest models use the highest number (nine). Logistic regression differed from the other methods in the selection of variables, with few logistic models using minimum temperature (23 percent) and elevation (14 percent), but instead being dominated by land use (Table 2). Eleven ensemble models (combining 2 or more models) were developed.

DISCUSSION

The goal of this study was to assess the impact of variable and model selection in SDMs, by comparing the consistency of the independent environmental variables across models, and the consistency of models across species. These issues are fundamental to all SDMs but of particular interest to invasive species. Invasive species often have expanding distributions, and limited information is available on this species, resulting in less defined models. Through identifying agreement between modeling techniques and variables selection, we can have greater confidence in models.

The area of distribution of a species is determined by its ecological and evolutionary history. Many factors affect species distribution, but the most important are the limits of the species' tolerances and needs for certain abiotic conditions, the suite of other species with which it interacts, and the potential for dispersal and colonization within a given time period (Soberón and Peterson 2005). Aboitic conditions can be used to define the potential distribution (the focus of our work), with species interactions and dispersal constraints defining the realized distribution. Many studies have found large-scale environmental factors can produce strong SDMs (Franklin 2009). Overall we found similar results, with the environmental variables used in this study useful in predicting the species potential distribution with 40 percent of the models considered good (test AUC > 0.75 and test omission rate < 0.25). Physiographic variables dominated the model over land use variables, suggesting these distributions are driven by species tolerances. Minimum temperature was the dominate variable suggesting many of these species are limited by the extreme temperatures of winter or length of growing season and competition with other species in that niche. Elevation, the second most dominate variable, has some correlation with temperature and was selected over temperature for inclusion in some models, while in others it was selected in conjunction with temperature. By selecting in conjunction with temperature, it may assist in more narrowly defining climatic conditions associated with the species or forest communities that occupy the area. When used without temperature, elevation is likely a representation of climatic conditions, with elevation integrating both aspects of temperature (high elevation, cooler temperature) and rainfall. Some of the finer-scale characteristics such as slope and distance to rivers were not widely used in models, suggesting the models may apply across a regional scale but not necessary at a local scale. Many of these relationships are nonlinear, with species having preferences for the intermediate temperatures and elevations. The two models that gave the strongest results (boosted regression and maximum entropy algorithms) capture nonlinear relationships well. Logistic regression is not designed to assess bimodal relationships, and as such, many of the models do not integrate temperature, rainfall and elevation-based variables, instead focusing on land cover characteristics. Our results were similar to other studies (Elith et al. 2006), with maximum entropy algorithms coming out as one of the strongest modeling techniques.

ACKNOWLEDGMENTS

We wish to thank the U.S. Department of Agriculture, Forest Service for supporting this work (cooperative agreement 10-DG-11330101-107). Also, we thank Kathy Roberts for her review and suggestions on the manuscript.

LITERATURE CITED

Elith, J.; Graham, C.H.; Anderson, R.P.; Dudik, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; Li, J.; Lohmann, L.G.; Loiselle, B.A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; Overton, J.M.; Townsend Peterson, A.; Phillips, S.J.; Richardson, K.; Scachetti-pereira, R.; Schapire, R.E.; Soberon, J.; Williams, S.; Wisz, M.S.; Zimmermann, N.E. 2006. **Novel methods improve prediction of species' distributions from occurrence data.** Ecography. 29: 129-151.

Franklin, J. 2009. **Mapping species distributions: spatial inference and prediction.** 3rd ed. New York, NY: Cambridge University Press.

Gallien, L.; Munkemuller, T.; Albert, C.H.; Boulangeat, I.; Thuiller, W. 2010. **Predicting potential distributions of invasive species: Where to go from here?** Diversity and Distributions. 16: 331-342.

Lemke, D.; Hulme, P.E.; Brown, J.A.; Tadesse, W. 2011. **Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA.** Forest Ecology and Management. 262: 139-149.

Miller, J.H.; Lemke, D.; Coulston, J. In press. **The invasion of southern forests by nonnative plants: current and future occupation with impacts, management strategies, and mitigation approaches.** In: The southern forest futures project. Technical Report. Chapter 15.

R Core Team. 2012. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing. Available at www.R-project.org. (Accessed June 1, 2012).

Soberon, J.M.; Peterson, A.T. 2005. **Interpretation of models of fundamental ecological niches and species' distributional areas.** Biodiversity Informatics. 2: 1-10.

U.S. Department of Agriculture, Forest Service (USDA FS). 2007. **The forest inventory and analysis database: database description and users guide version 3.0.** Washington, DC: U.S. Department of Agriculture, Forest Service.

Vitousek, P.M.; D'Antonio, C.M.; Loope, L.L.; Rejmanek, M.; Westbrooks, R. 1997. **Introduced species: aignificant component of human-caused global change.** New Zealand Journal of Ecology. 21: 1-16.