
Volume 261, issue 1, 1 January 2011
ISSN 0378-1127
ELSEVIER
Forest Ecology and Management

Forest Ecology and Management 266 (2012) 216–222

Contents lists available at SciVerse ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Mapping eastern hemlock: Comparing classification techniques to evaluate susceptibility of a fragmented and valued resource to an exotic invader, the hemlock woolly adelgid

Joshua T. Clark [a], Songlin Fei [b,*], Liang Liang [c], Lynne K. Rieske [a]

[a] University of Kentucky, Department of Entomology, Lexington, KY 40546, USA
[b] Purdue University, Department of Forestry and Natural Resources, West Lafayette, IN 47907, USA
[c] University of Kentucky, Department of Geography, Lexington, KY 40506, USA

ARTICLE INFO

Article history:
Received 25 February 2011
Received in revised form 15 November 2011
Accepted 18 November 2011

Keywords:
Eastern hemlock
Hemlock woolly adelgid
Invasive species
Decision tree
MaxEnt

ABSTRACT

Eastern hemlock (*Tsuga canadensis* Carriére), an ecologically important foundation species in forests of eastern North America, is currently threatened by the hemlock woolly adelgid (HWA, *Adelges tsugae* Annand, Hemiptera: Adelgidae), an aggressive invasive insect herbivore. HWA colonization of eastern hemlock results in rapid tree mortality. There is a pressing need to accurately determine eastern hemlock distribution in the face of expanding HWA populations to preserve this important forest species. However, efficient modeling of large geographic extents of eastern hemlock habitats to facilitate state-wide HWA management is lacking. We employ two modeling approaches, decision tree classification (based on presence–absence data) and maximum entropy (MaxEnt, based on presence-only data) method, to map eastern hemlock distribution in eastern Kentucky using a comprehensive suite of environmental parameters as predictor variables. Results demonstrate moderate model accuracies around 70%, supporting the practicality of mapping hemlock distribution over extensive regions. Comparison of the two modeling techniques suggests that decision tree classification has higher overall accuracies, while MaxEnt method was more efficient in model construction. In comparison to the decision tree method, MaxEnt suffered from possibly over-fitting as indicated by increased producer's accuracies yet lower user's accuracies. Our study provides useful references for selecting optimized approaches in accordance with study region characteristics and end user's preferences.

1. Introduction

Eastern hemlock (*Tsuga canadensis* Carriére) has suffered widespread mortality from an exotic invasive herbivore, the hemlock woolly adelgid (HWA, *Adelges tsugae* Annand, Hemiptera: Adelgidae) (McClure et al., 2001). HWA was first reported in the eastern United States (US) in Virginia in the 1950s, and since the 1980s populations have expanded northward and westward (Havill et al., 2006), exploiting the large contiguous tracts of hemlock forest common in the northeast. More recently, HWA has expanded its geographic range southward along the southern Appalachian Mountains (Ward et al., 2004). HWA was first discovered in Kentucky in March 2006 (Kentucky Forest Health Task Force, 2006), and since then, infestations have been steadily expanding in the state. The continued spread of HWA through the region's hemlock forests is imminent, but data concerning how that spread will develop is lacking, due in part to a poor understanding of the spatial distribution of the host tree species.

Eastern hemlock is an essential component of forests in eastern North America. It plays a vital role maintaining stream quality by regulating air and soil temperatures, soil moisture, hydrologic discharge, and amplitude of stream flow (Ford and Vose, 2007). Disruption of these processes can lead to loss of aquatic invertebrate and vertebrate biodiversity (Ross et al., 2003; Snyder et al., 2002). Loss of eastern hemlock will also cause changes in vegetation composition and structure; increased light penetration after hemlock death creates suitable habitat for less shade-tolerant tree species such as red maple (*Acer rubrum*) and black birch (*Betula lenta*) (Daley et al., 2007; Spaulding and Rieske, 2010; Yorks et al., 2003).

In spite of the ecological importance of eastern hemlock, resource managers have only limited information about the scope of its spatial distribution. The lack of detailed hemlock-specific maps is due in part to hemlock's low economic importance, low density and basal area, and confined distribution to the study area (Little, 1971 and Fig. 1). Better mapping of eastern hemlock is

* Corresponding author. Address: FORS 111, 195 Marsteller St., West Lafayette, IN 47907-2033, USA. Tel.: +1 765 496 2199.
E-mail address: sfei@purdue.edu (S. Fei).

0378-1127/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.foreco.2011.11.030

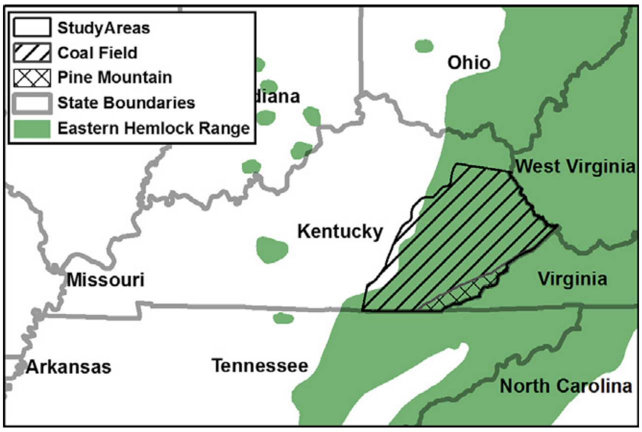

Fig. 1. Study area in relation to the range of eastern hemlock (Little, 1971), with the two physiographic regions, Coal Field and Pine Mountain, shown.

critically needed to slow the spread of HWA by strategically monitoring and treating certain areas that can serve as HWA pathways or hotspots. To develop a detailed hemlock distribution map at the regional level, ground-based mapping of hemlock is impractical. Existing satellite image based maps such as the National Land Cover Dataset (NLCD, http://www.mrlc.gov/) provide general distribution information for evergreen forests, but do not provide any spatial information at the species level. The Kentucky Land Cover Dataset (KLCD, http://kygeonet.ky.gov/) does include species specific information for eastern hemlock, however their application to hemlock management is limited for the following two reasons. First, KLCD is intended for land cover and land type classification not for species classification, therefore, its accuracy at species level is limited. Second, due to its low density and sparse distribution, hemlock is often classified into other types/groups that do not have a defined hemlock component. Other products such as Forest Inventory and Analysis (FIA, Bechtold and Patterson, 2005) data may provide rich ecological information on sampled plots and unbiased statistical estimation of the overall abundance of hemlock distribution, but FIA data alone cannot produce continuous coverage that is necessary for hemlock management spatial prioritization and strategic planning, given their limited sampling density. Hence, we consider species distribution modeling as an effective way to bridge this gap and provide more physiologically coherent approximation of hemlock distribution with continuous coverage.

Species distribution models are frequently used to predict species distributions across landscapes or to gain insights into species ecological traits. Two types of species distribution models, presence–absence data based models and presence-only data based models, are widely used in various ecosystems (Elith and Leathwick, 2009). The decision tree classification method, a presence–absence data based method, is popular in species distribution prediction. Maingi and Luhn (2005) used decision tree classification to map conifer tree distribution with Landsat TM imagery and ancillary data in the Daniel Boone National Forest in Kentucky. Koch et al. (2005) used a decision tree classification aided with remotely sensed data to successfully map eastern hemlock in other regions in response to HWA-induced eastern hemlock decline. MaxEnt (Phillips et al., 2006), a presence-only data based method, is also a commonly used method in species distribution modeling and prediction. Elith et al. (2006, 2011) found that MaxEnt provides relatively consistent and reliable results compared to other presence-only data based methods.

In this study, both decision tree classification and MaxEnt were employed to map the distribution of eastern hemlock in Kentucky. The primary objective of our study was to explore effective ways to map eastern hemlock on a regional scale to facilitate state level management and conservation efforts. For comparison purposes, classification accuracy, processing complexity, and time to completion for the two techniques were evaluated.

2. Material and methods

2.1. Study area

Our study area covered 27,006 km^2 of eastern Kentucky (38.29–36.58°N, 81.96–84.83°W) and was partitioned into two physiographic regions: Upper Eastern Coal Field (hereafter, Coal Field) and Pine Mountain (Fig. 1). All models were constrained to the same geographic extent, shown as "Study Area" on the map. This mountainous region is composed of sandstone, shale, and siltstone and ranges in elevation from 154 to 1259 m (McDowell, 1986). Monthly average temperature ranges from 1.1 °C in January to 23.9 °C in July and monthly average precipitation ranges from 8.1 cm in October to 13.1 cm in May (NOAA, 2002). The forest type is mixed mesophytic, consisting primarily of pine-oak (Braun, 1950).

2.2. Modeling approach overview

We modeled eastern hemlock distribution using two modeling techniques in each physiographic region, for a total of four models. In the first approach we performed a decision tree analysis against environmental variables. A decision tree approach requires at least two classes of dependent variables, necessitating the collection of both presence and absence points. In a large study area this can be exceedingly time-consuming. For that reason, the second method we tested was maximum entropy species distribution modeling (MaxEnt), which requires only presence data (Phillips et al., 2006). Given the differences between the two physiographic regions with respect to elevation, slope, and geologic substrate, modeling was performed for each region separately.

2.3. Field data

Eastern hemlock presence and absence data were acquired from a variety of sources (Table 1); the majority of these data are a result of systematic sampling efforts. A total of 2624 points were used in this study. All points were georeferenced using a GPS unit in the field. Forest Inventory and Analysis data represent ground samples collected systematically at a rate of one plot per 2428 ha (6000 acre) hexagon. Actual FIA plot coordinates were used in this study. A plot is designated as a hemlock presence plot if any hemlock tree with >2.54 cm in diameter at breast height (d.b.h.) is present. The second source of systematic sampling was derived from an ongoing gypsy moth survey program. The traps for this monitoring effort were placed on a 2 $\times$ 2 km grid. The individuals setting and monitoring the traps were asked to record presence or absence of eastern hemlock within 30 m of the trap site. The remaining field data were derived from HWA monitoring by individuals representing various agencies, including the Kentucky Division of Forestry, the Kentucky Office of the State Entomologist, and the Kentucky State Nature Preserves Commission. Hemlock data from these sources were collected in an opportunistic manner. These field data were compiled into a database representing hemlock or non-hemlock categories; for all sampling efforts, any survey location with at least one hemlock tree was classified as hemlock. This liberal classification scheme reflects the overall goals of this research, to map hemlocks which may serve as a corridor for the spread of HWA. The data points were split by physiographic

Table 1
Sources of hemlock and non-hemlock data points used in model building and accuracy assessments.

Source		Hemlock	Non-hemlock	Total
Forest Inventory and Analysis	Training	88	703	791
	Test	22	207	229
Gypsy Moth Survey	Training	7	349	356
	Test	0	39	39
HWA Monitoring	Training	749	204	953
	Test	202	54	256
Total		1068	1556	2624

region, and then one-fifth of the points were randomly selected and set aside to be used as test data for accuracy assessments. The remaining data were used in the modeling process.

2.4. Environmental data

In both modeling approaches, 11 environmental layers were used as independent variables representing physiographic and climatic characteristics. These environmental data were acquired or derived from various sources as detailed in Table 2. To maintain consistency across all models and for comparison to the NLCD (Homer et al., 2004) and KLCD (Commonwealth Office of Technology, 2004), each variable was converted to a raster layer with cell size of 30 $\times$ 30 m using ArcGIS 9.3 (ESRI, Redlands, CA, US).

2.5. Classification tree

We chose a decision tree classification model because it is non-parametric, robust to noisy or missing data, easy to interpret, and can utilize both continuous and categorical data (Huang et al., 2003; Koch et al., 2005; Nelson et al., 2003). Similar methods have produced maps for eastern hemlock at a landscape scale (Koch et al., 2005) and for other evergreen trees at regional scales (Landenburger et al., 2008; McDermid and Smith, 2008). Using PASW Statistics software (SPSS Inc., Chicago, IL, US), decision tree analysis was performed to classify areas with presence or absence of eastern hemlock.

For the decision tree in each physiographic region a Chi-squared Automatic Interaction Detection (CHAID) growing method was applied with a Pearson chi-square statistic, maximum tree depth set to 10 levels, and a significance level of 0.1 for splitting nodes. The CHAID method was used in order to take advantage of multiple splitting pathways for each node, as opposed to binary splitting. This multi-splitting at each node in the CHAID method allows variables to be partitioned in a more biologically meaningful way. All environmental variables were designated as continuous variables except geologic formations and soil types, which were specified as categorical. The significance level was set to 0.1 in order to further divide nodes to minimize overestimation or underestimation. Node response was used as the threshold to assign the area represented by that node to either the presence or absence class; response values greater than 50% were designated as hemlock presence.

2.6. MaxEnt modeling

We employed maximum entropy modeling using the MaxEnt species distribution program, version 3.2.19 (Phillips et al., 2006). The MaxEnt algorithm predicts the probability distribution of species occurrence based on the environmental constraints estimated from known locations.

All environmental variables were again designated as continuous variables except geologic formations and soil types (categorical) as for CHAID. Eighty percent of hemlock presence points were used to build the model, leaving one-fifth that were randomly selected by the program to be used in model validation procedures. The model was executed for 10 iterations with refreshed random partitions of training and validation data. We chose the best model to use based on highest area under the curve value (Boubli and de Lima, 2009; Pauchard and Alaback, 2004). The jackknife option was used to evaluate the importance of each environmental variable in the model. The raw model output was a continuous probability map with values between 0 and 1; therefore, a threshold was assigned in order to reclassify the output into dichotomous classes representing either suitable or unsuitable areas for hemlock habitat. The threshold was determined using a method based on the maximum difference of cumulative frequency of probabilities for observed hemlock locations and random points (Browning et al., 2005; Fei et al., 2007; Thompson et al., 2006). The geographic areas with probabilities above the threshold were designated as hemlock and those areas below the threshold were designated as non-hemlock.

2.7. Accuracy assessments

For both modeling methods, accuracy assessments were performed for predicted eastern hemlock distribution using surveyed reference points. A total of 419 ground truthing reference points were used for Coal Field and 105 points for Pine Mountain. Overall accuracies with user's and producer's accuracies were calculated for each model.

Table 2
Environmental variables used in models with respective data sources.

Environmental Layer	Abbreviation	Description	Source
DEM	DEM	Elevation	USGS
Slope	Slope	Slope derived from DEM	Calculated using ArcGIS slope tool
Aspect	Aspect	Aspect derived from DEM	Calculated using ArcGIS aspect tool
Soil type	Soil	Soil Survey Geographic (SSURGO) data	USDA Geospatial Data Gateway
Moisture index	TRMIM	Topographic relative moisture index (modified)	TRMIM.aml script downloaded from http://earth.gis.usu.edu/swgap/landform.html
Stream distance	Streams	Euclidean distance from nearest stream	Calculated using ArcGIS Euclidean distance tool
Topographic position	TPI	Topographic position index	Calculated using TPI extension (Jenness, 2006) for ArcView (ESRI, Redlands, CA, US)
Curvature	Curvature	Curvature of the terrain derived from DEM	Calculated using ArcGIS curvature tool
Minimum temperature	Min temp	Average January minimum temperature (1971–2000)	The Geospatial Data Gateway http://datagateway.nrcs.usda.gov/
Maximum temperature	Max temp	Average July maximum temperature (1971–2000)	The Geospatial Data Gateway http://datagateway.nrcs.usda.gov/
Geologic formation	Geo form	Geologic formations of Kentucky	Kentucky Geological Survey

3. Results

Both types of models generated maps of eastern hemlock distribution with moderate accuracies (Fig. 2, Table 3). Across models, overall accuracies were fairly close, ranging from 65.7% to 74.7%, but the total area classified as eastern hemlock differed between models by up to thousands of square kilometers (Fig. 3). Decision tree models appeared to yield slightly higher accuracies than the MaxEnt models for both regions (74.7% and 72.4%, respectively, for Coal Field and Pine Mountain). MaxEnt models had higher producer's accuracies than the decision tree models in both regions.

The combined area (Coal Field and Pine Mountain) predicted to be eastern hemlock by decision tree models was 11.4% of the total area, approximately 2300 km^2 less than the MaxEnt models, which predicted 19.9% of the total area as eastern hemlock (Fig. 3). The environmental variable ranking (Table 4) for the decision trees revealed proximity to nearest stream to be a prominent classifier.

The MaxEnt models, in particular, resulted in test AUC values (>0.80) that indicated good model performance (Fig. 4). The threshold values for an area to be designated as hemlock are 0.42 and 0.35 for Coal Field and Pine Mountain physiographic regions, respectively (i.e., areas with probability value higher than the threshold value will be classified as hemlock). The resulting binary map has a fairly high overall classification accuracy (72.8%) in the Coal Field region and slightly lower overall accuracy (65.7%) in the Pine Mountain region (Table 3). The MaxEnt models were associated with relatively large commission errors (37.6% and 39.0% for Coal Field and Pine Mountain, respectively) and small omission errors (16.4% and 11.3%). The predictor variables with the highest percent contributions were proximity to nearest stream, soil type, and geologic formation, which together made over 85% contributions to the models. The rank of these three variables varied between the two physiographic regions (Table 5).

We generated a synthesized intersection map (Fig. 2) which shows areas predicted as eastern hemlock by both modeling approaches. This intersection map was used to compare environmental data with the background because it likely indicates areas with a very high probability of finding hemlock. Hemlock coverage predicted by both methods is about 2300 km^2, approximately 56% of the predicted hemlock area by the decision tree models and 32% by MaxEnt models (Fig. 3). Considering that the study area lies on the edge of the distribution range for eastern hemlock, the

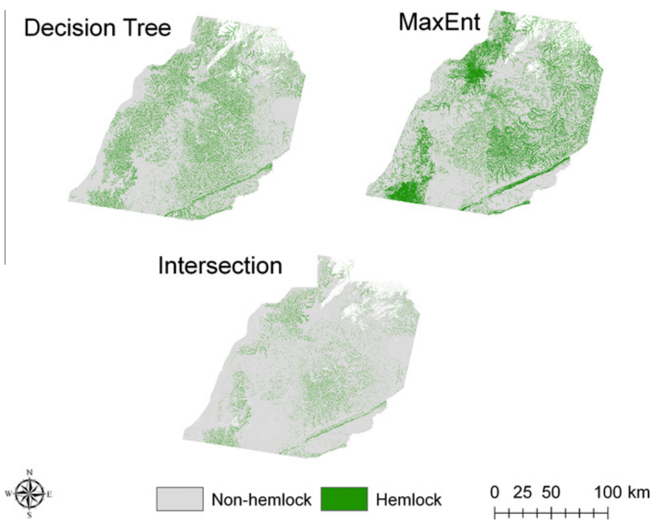

Fig. 2. Predicted distribution of eastern hemlock resulting from two modeling approaches, decision tree and MaxEnt (see text for full explanation). Intersection denotes the areas in which both models predict the presence of eastern hemlock.

Table 3
Accuracy assessment for two distribution models (decision tree and MaxEnt) by respective physiographic regions (Coal Field and Pine Mountain).

Classification data		Reference data			
		Hemlock	Non-Hemlock	Row total	User's
Coal Field Decision Tree overall accuracy – 74.7%					
Map data	Hemlock	115	50	165	69.7%
	Non-Hemlock	56	198	254	78.0%
	Column total	171	248	419	
	Producer's	67.3%	79.8%		
Coal Field MaxEnt overall accuracy – 72.8%					
Map data	Hemlock	143	86	229	62.4%
	Non-Hemlock	28	162	190	85.3%
	Column total	171	248	419	
	Producer's	83.6%	65.3%		
Pine Mountain Decision Tree overall accuracy – 72.4%					
Map data	Hemlock	40	16	56	71.4%
	Non-Hemlock	13	36	49	73.5%
	Column total	53	52	105	
	Producer's	75.5%	69.2%		
Pine Mountain MaxEnt overall accuracy – 65.7%					
Map data	Hemlock	47	30	77	61.0%
	Non-Hemlock	6	22	28	78.6%
	Column total	53	52	105	
	Producer's	88.7%	42.3%		

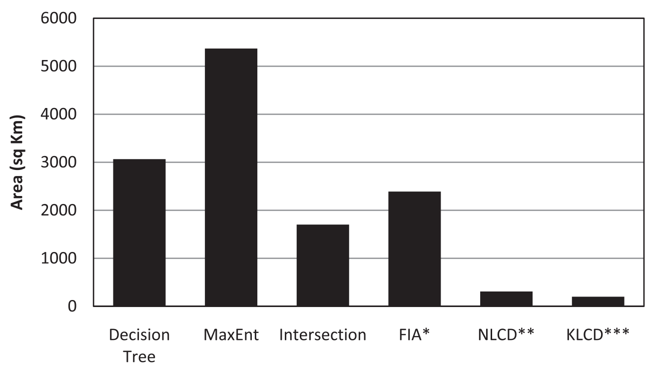

Fig. 3. Area (km^2) predicted as eastern hemlock by each modeling technique. Intersection denotes the areas in which all three models predict the presence of eastern hemlock. *Total forested area on which one or more hemlock trees occur based on Forest Inventory and Analysis data. **National Land Cover Dataset (NLCD) shows the total area classified as evergreen forest. ***Kentucky Anderson Level III Land Cover Dataset (KLCD) shows the total area classified as hemlock forest and hemlock-mixed deciduous forest.

Table 4
Results of decision tree analyses showing variable rankings and the frequencies of each variable used as a splitting condition at each tree level. Decision trees were generated for each physiographic region separately, then numbers of nodes and tree level frequencies of the two regions were combined.

Rank	Variable*	# of nodes	Level of tree						
			1	2	3	4	5	6	7
1	Streams	5	2			2	1		
2	Min temp	3		2		1			
3	Geo form	3		1	1				1
4	TRMIM	3		1		1		1	
5	Aspect	3			2	1			
6	Curvature	3			2	1			
7	TPI	3				1	1	1	
8	Slope	2		1		1			
9	DEM	2		1			1		
10	Max temp	2			2				

* Detailed descriptions of variables can be found in Table 2.

intersection map reveals important environmental conditions that better support hemlock growth within the state (Table 6). The

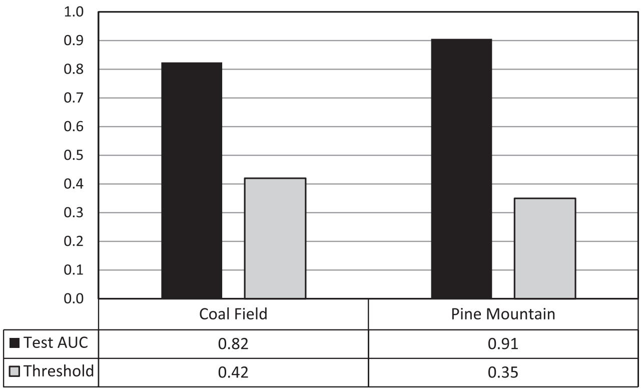

Fig. 4. Area under the curve (AUC) and presence–absence threshold determined for hemlock distribution models in two physiographic regions generated from MaxEnt.

intersection areas predicted as eastern hemlock by both classification methods indicate that 95% of hemlock coverage is located less than 296 m from streams and between 189 and 433 m in elevation in the Coal Field region and less than 930 m from streams and between 238 and 779 m in elevation in the Pine Mountain region. Further, the majority of the eastern hemlock area is situated on northeast to southeast facing slopes and on geologic formations composed of shale, sandstone, or alluvium. The most abundant soils in the predicted hemlock areas are well drained, loamy, acidic ultisols and inceptisols including shelocta, helechawa, or grigsby complexes.

4. Discussion and conclusions

Both of our modeling approaches produced maps of eastern hemlock distribution in the Coal Field and Pine Mountain physiographic regions with moderate accuracy. The overall accuracy (~70%) was lower than that of similar studies (~80%) to map eastern hemlock in other regions (Koch et al., 2005; Maingi and Luhn, 2005). However, our study evaluated a much larger region with more complex topographic and physiographic characteristics, and greater vegetative variations. The reduced accuracies may be a consequence of the extensive spatial coverage which included greater landscape heterogeneity. We initially attempted a second decision tree classifier which included Landsat spectral bands as independent variables, but this method was abandoned. Reduced spectral resolution, cloud cover, and altered brightness values caused by the rugged terrain of eastern Kentucky, especially in the Pine Mountain area, affected the reliability of the satellite imagery. In addition, image processing time in areas with rugged terrain, such as our study area, can be extremely time consuming.

In general, MaxEnt modeling appeared to provide predictions with reasonable accuracy. Multiple model runs allowed for selecting optimum results. However, overall accuracies of MaxEnt models were a balance of relatively low user's accuracy and high producer's accuracy, which may imply over-fitting. The MaxEnt approach is known to be robust for presence-only models, but its generally high accuracy may be compromised by lacking absence data, as indicated by the low user's accuracy in this study and suggested by related investigations (Elith et al., 2011). In addition, omission of absence data when building the model, a stipulation of the MaxEnt method, may also compromise the accuracy in comparison to the decision tree based method.

The difficulty and time required to complete each model varied considerably. Since the decision tree method required final maps to be built in GIS using the node-splitting conditions, the decision trees were significantly more time-intensive than the MaxEnt models. The existing MaxEnt program was straightforward with a simple user interface, and the processing was relatively efficient. While estimation of the thresholds added to the time committed, the MaxEnt model was the less complicated and quicker to execute.

The environmental characteristics associated with hemlock habitat prediction affirmed previous descriptions of eastern hemlock habitat in the southern Appalachian Mountains, which suggest that hemlock grows on north to east facing slopes in moist riparian zones and in neutral to acidic soils (Godman and Lancaster, 1990; Quimby, 1996). Soil was a significant contributor in the MaxEnt models in predicting the presence of hemlock, while it did not contribute in node-splitting in the decision trees. This could be partially due to the tight association between geologic formation and soil type and different statistical mechanisms used in the two methods in dealing with categorical variables. Eastern hemlock is commonly reported between 610 and 1520 m in this region (Southern Appalachians, Godman and Lancaster, 1990), much higher than predicted by our study (Table 6), which potentially reflects the regional difference of the river valley reliefs, as eastern Kentucky is located on the western edge of hemlock distribution with a relatively lower elevation range.

The total area predicted as eastern hemlock by each model is larger than national and state land cover datasets and FIA data (Fig. 3). The reason for these differences is twofold. First, areas predicted by the two models used in this study are potential niche/habitat, while the land cover datasets and FIA estimate the realized niche/habitat, which is always smaller than the former estimation. Second, classifications in the land cover datasets are often too broad and neglect the minor components in areas with mixed species. The total area predicted as hemlock by the decision tree model is slightly higher than the statistical estimation from FIA data. The intersected areas of the two models are more conservative than the FIA data based estimation.

Our results suggest that the most advantageous procedure for mapping hemlock resources in the central Appalachians may depend on the priorities set forth by the end users. This study demonstrates that a relatively accurate map can be produced with environmental data at the regional scale. In situations where absence data are unreliable or unavailable, a presence-only species distribution model such as MaxEnt is most promising. However, if absence data have been collected and resource managers desire to achieve a more accurate classification, the decision tree classifier may be a better option. Moreover, an 'ensemble modeling' approach could be considered to assist land management decision making (Jones-Farrand et al., 2011). A union of both predicted maps can be used if land managers want to be inclusive about all possible hemlock areas, or an intersection of predicted maps can be used if high certainty is needed.

Table 5
Percent contribution of environmental variables used in MaxEnt models for the Coal Field and Pine Mountain physiographic regions.

Variable*	Percent contribution (%)	
	Coal Field	Pine Mountain
Streams	41.0	21.1
Soil	37.1	43.8
Geo form	12.0	21.2
DEM	2.4	2.3
TPI	2.0	1.8
Aspect	1.6	1.2
Min temp	1.5	3.1
Slope	1.0	1.6
TRMIM	0.5	1.1
Curvature	0.5	2.3
Max temp	0.4	0.5

* Detailed descriptions of variables can be found in Table 2.

Table 6
Statistics for elevation (DEM) and proximity to nearest stream for the entire study area (Background), area predicted as hemlock by the decision tree classification (Decision tree hemlock) and MaxEnt (MaxEnt hemlock), and areas predicted as hemlock by both models (Intersection). Minimum, maximum, mean, and standard deviation are shown in meters for each area in the Coal Field and Pine Mountain regions.

Coal Field								
	Background				Decision tree hemlock			
	Min	Max	Mean	SD	Min	Max	Mean	SD
DEM	160.4	843.2	349.8	86.2	158.5	795.5	316.4	65.3
Streams	0.0	5238.8	699.6	525.4	0.0	2362.2	138.6	89.1
	MaxEnt hemlock				Intersection			
	Min	Max	Mean	SD	Min	Max	Mean	SD
DEM	160.8	806.8	318.0	67.9	173.2	696.3	311.0	61.1
Streams	0.0	4922.2	297.9	309.0	0.0	354.9	119.9	88.0
Pine Mountain								
	Background				Decision tree hemlock			
	Min	Max	Mean	SD	Min	Max	Mean	SD
DEM	266.5	1258.4	607.9	178.2	266.5	1197.2	535.1	156.9
Streams	0.0	4953.6	897.2	681.9	0.0	4734.6	376.7	530.6
	MaxEnt hemlock				Intersection			
	Min	Max	Mean	SD	Min	Max	Mean	SD
DEM	266.5	1218.0	561.2	156.1	266.5	1107.5	508.2	135.2
Streams	0.0	4871.8	525.1	602.0	0.0	4734.6	208.5	361.0

As non-native invaders continue to proliferate and threaten valued resources, land managers must utilize diverse approaches to develop increasingly aggressive tools for management and mitigation. The hemlock woolly adelgid is particularly threatening, since eastern hemlock is considered an essential foundation species in forests of eastern North America (Godman and Lancaster, 1990), and it is especially susceptible to adelgid feeding (Souto et al., 1996). The hemlock woolly adelgid threatens the persistence of the hemlock forest type (Spaulding and Rieske, 2010), which has tremendous consequences for ecosystem function (Orwig and Foster, 1998). Our results demonstrate that spatial habitat modeling using approaches such as the decision tree classification or the maximum entropy method is a feasible means of determining high-risk areas of invasion at the state level, and could aid in mitigating the effects of the hemlock woolly adelgid invasion.

Acknowledgments

We thank C. Harper, J.D. Loan and J. Obrycki for support and data collection and S.G. Lambert for the access of FIA plot data. This study is supported by the United States Department of Agriculture, Cooperative State Research, Education, and Extension Service (Proposal No. 2006-06176), and is published as Kentucky Agricultural Experiment Station Project 11-08-002.

References

Bechtold, W., Patterson, P., 2005. The Enhanced Forest Inventory and Analysis Program: National Sampling Design and Estimation Procedures. US Department of Agriculture Forest Service, Southern Research Station.
Boubli, J., de Lima, M., 2009. Modeling the geographical distribution and fundamental niches of *Cacajao* spp. and *Chiropotes israelita* in Northwestern Amazonia via a maximum entropy algorithm. International Journal of Primatology 30, 217–228.
Braun, E.L., 1950. Deciduous Forests of Eastern North America. Hafner Press, New York.
Browning, D., Beaupre, S., Duncan, L., 2005. Using partitioned Mahalanobis D^2 (k) to formulate a GIS-based model of timber rattlesnake hibernacula. Journal of Wildlife Management 69, 33–44.
Commonwealth Office of Technology – KY Division of Geographic Information, 2004. Kentucky 2001 Anderson Level III Land Cover. <ftp://ftp.kymartian.ky.gov/kls/KY_KLCD01.zip>.
Daley, M., Phillips, N., Pettijohn, C., Hadley, J., 2007. Water use by eastern hemlock (*Tsuga canadensis*) and black birch (*Betula lenta*): implications of effects of the hemlock woolly adelgid. Canadian Journal of Forest Research 37, 2031–2040.
Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology Evolution and Systematics 40, 677–697.
Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of maxent for ecologists. Diversity and Distributions 17, 43–57.
Fei, S., Schibig, J., Vance, M., 2007. Spatial habitat modeling of American chestnut at Mammoth Cave National Park. Forest Ecology and Management 252, 201–207.
Ford, C., Vose, J., 2007. *Tsuga canadensis* (L.) Carr. mortality will impact hydrologic processes in southern Appalachian forest ecosystems. Ecological Applications 17, 1156–1167.
Godman, R., Lancaster, K., 1990. Eastern hemlock. Silvics of North America 1, 604–612.
Havill, N., Montgomery, M., Yu, G., Shiyake, S., Caccone, A., 2006. Mitochondrial DNA from hemlock woolly adelgid (Hemiptera: Adelgidae) suggests cryptic speciation and pinpoints the source of the introduction to eastern North America. Annals of the Entomological Society of America 99, 195–203.
Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national landcover database for the United States. Photogrammetric Engineering and Remote Sensing 70, 829–840.
Huang, C., Homer, C., Yang, L., 2003. Regional forest land cover characterisation using medium spatial resolution satellite data. In: Wulder, M.A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments: Concepts and Case Studies. Kluwer Academic Publishers, Boston.
Jenness, J., 2006. Topographic Position Index (http://www.tpi_jen.avx) extension for ArcView 3.x, v. 1.2. Jenness Enterprises. <http://www.jennessent.com/arcview/tpi.htm>.
Jones-Farrand, D.T., Fearer, T.M., Thogmartin, W.E., Thompson, F.R., Nelson, M.D., Tirpak, J.M., 2011. Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble prediction. Ecological Applications 21, 2269–2282.
Kentucky Forest Health Task Force, 2006. Forest health: an assessment of the health of Kentucky's forests and threats to maintaining the health of our forests. <http://www.kyforesthealth.org/2006report.pdf>.
Koch, F.H., Cheshire, H.M., Devine, H.A., 2005. Mapping hemlocks via tree-based classification of satellite imagery and environmental data. In: US Department of Agriculture, Forest Service, FHTET-2005-01.
Landenburger, L., Lawrence, R.L., Podruzny, S., Schwartz, C.C., 2008. Mapping regional distribution of a single tree species: whitebark pine in the greater Yellowstone ecosystem. Sensors 4983, 4994.
Little Jr., E., 1971. Atlas of United States Trees. Vol. 1. Conifers and Important Hardwoods. Miscellaneous Publication 1146. US Department of Agriculture, Forest Service, Washington, DC.
Maingi, J., Luhn, W., 2005. Mapping insect-induced pine mortality in the Daniel Boone National Forest, Kentucky using Landsat TM and ETM+ data. GIScience & Remote Sensing 42, 224–250.

Mcclure, M.S., Salom, S.M., Shields, K.S., 2001. Hemlock woolly adelgid. In: US Department of Agriculture, Forest Service, FHTET-2001-03.
McDermid, G.J., Smith, I.U., 2008. Mapping the distribution of whitebark pine (*Pinus albicaulis*) in Waterton Lakes National Park using logistic regression and classification tree analysis. Canadian Journal of Remote Sensing 34, 356–366.
McDowell, R.C., 1986. The Geology of Kentucky – A Text to Accompany the Geologic Map of Kentucky. United States Government Printing Office, Washington DC.
Nelson, B., Runger, G., Si, J., 2003. An error rate comparison of classification methods with continuous explanatory variables. IIE Transactions 35, 557–566.
NOAA, 2002. Climatography of the United States. In: Administration, N.O.a.A. (Ed.). URL: <http://cdo.ncdc.noaa.gov/climatenormals/clim81/KYnorm.pdf>.
Orwig, D.A., Foster, D.R., 1998. Forest response to the introduced hemlock woolly adelgid in southern New England, USA. Journal of Torrey Botanical Society 125, 60–73.
Pauchard, A., Alaback, P.B., 2004. Influence of elevation, land use, and landscape context on patterns of alien plant invasions along roadsides in protected areas of south-central Chile. Conservation Biology 18, 238–248.
Phillips, S., Anderson, R., Schapire, R., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, 231–259.
Quimby, J.W., 1996. Value and importance of hemlock ecosystems in the Eastern United States. In: Salom, S.M., Tigner, T.C., Reardon, R.C. (Eds.), Proceedings of the First Hemlock Woolly Adelgid Review. USDA Forest Service, FHTET 96-10, pp. 1–8.
Ross, R., Bennett, R., Snyder, C., Young, J., Smith, D., Lemarie, D., 2003. Influence of eastern hemlock (*Tsuga canadensis* L.) on fish community structure and function in headwater streams of the Delaware River basin. Ecology of Freshwater Fish 12, 60–65.
Snyder, C., Young, J., Lemarié, D., Smith, D., 2002. Influence of eastern hemlock (*Tsuga canadensis*) forests on aquatic invertebrate assemblages in headwater streams. Canadian Journal of Fisheries and Aquatic Sciences 59, 262–275.
Souto, D., Luther, T., Chianese, B., 1996. Past and current status of HWA in eastern and Carolina hemlock stands. In: Salom, S.M., Tignor, T.C., Reardon, R.C. (Eds.), Proceedings of the First Hemlock Woolly Adelgid Review. USDA Forest Service, Morgantown, WV, pp. 9–15.
Spaulding, H., Rieske, L., 2010. The aftermath of an invasion: structure and composition of Central Appalachian hemlock forests following establishment of the hemlock woolly adelgid, *Adelges tsugae*. Biological Invasions 12, 3135–3143.
Thompson, L., Van Manen, F., Schlarbaum, S., DePoy, M., 2006. A spatial modeling approach to identify potential butternut restoration sites in Mammoth Cave National Park. Restoration Ecology 14, 289–296.
Ward, J.S., Montgomery, M.E., Cheah, C.A.S.-J., Onken, B.P., Cowles, R.S., 2004. Eastern Hemlock Forests: guidelines to minimize the impacts of hemlock woolly adelgid. In: USDA Forest Service NA-TP-03-04.
Yorks, T., Leopold, D., Raynal, D., 2003. Effects of *Tsuga canadensis* mortality on soil water chemistry and understory vegetation: possible consequences of an invasive insect herbivore. Canadian Journal of Forest Research 33, 1525–1537.