Forest Ecology and Management 315 (2014) 22–29

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Spatial analysis facilitates invasive species risk assessment

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

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

ARTICLE INFO

Article history:
Received 20 September 2013
Received in revised form 3 December 2013
Accepted 15 December 2013
Available online 9 January 2014

Keywords:
Species distribution modeling
Hotspot analysis
Hemlock woolly adelgid
Risk assessment

ABSTRACT

Regional scale quantitative invasion risk analyses are needed to allow early detection and rapid response in order to effectively control the spread of exotic invasion. Most of the current invasion risk analyses are qualitative and ad hoc based. In this study, we used a spatial statistics based framework to assess the invasion risks of hemlock woolly adelgid (*Adelges tsugae*) with the following major steps: (1) invasion probability was first predicted by two widely used spatial statistics tools, maximum entropy (Maxent) and Mahalanobis distance (MD), based on known adelgid infestation locations and a set of environmental and anthropogenic related factors; (2) an ensemble of the above two models and a multi-threshold approach were employed to reduce prediction uncertainties; and (3) a spatial hotspot analysis were applied to enhance invasion prevention and management decision making. Among the factors investigated, variables representing corridors (e.g., trails and railroads) that are inadvertently spreading adelgid were important for the prediction of adelgid invasion. Large portion of the hemlock forests in the study area had a high adelgid invasion probability. The hotspot analysis based on the ensemble model showed three major clustered areas with high adelgid infestation probability. Our study demonstrated the feasibility of regional-scale quantitative invasion risk assessment with the application of a spatial statistics based framework, which can be used for effective and proactive invasion prevention and management.

1. Introduction

Invasion of exotic insects can be devastating, resulting in significant economic loss (Pimentel et al., 2005) and ecosystem degradation (Simberloff et al., 2012). To minimize the impact and slow or stop the invasion of exotic species, especially at the invasion frontier, an effective early detection and rapid response (EDRR) system is needed (Chornesky et al., 2005; Hulme, 2012). Early detection is critical for increasing the likelihood of eradication or to mitigate the impacts of invasive species. But early detection of exotic invasions can be challenging, especially for rapidly dispersing, cryptic species such as the hemlock woolly adelgid (HWA) (*Adelges tsugae*), a small, highly mobile pest. Monitoring areas of high invasion risk will increase the likelihood of detecting newly invading populations (Lodge et al., 2006). Most often invasion risk analyses are qualitative, and are ad hoc based. The effectiveness of an EDRR system depends profoundly on the accuracy of quantitative prediction of the invasion dispersal process (Lodge et al., 2006; Hulme, 2012). Here we demonstrate the use of spatial analysis to quantitatively assess high risk areas of adelgid invasion in central Appalachia (southeastern Kentucky) for proactive invasive insect management.

The hemlock woolly adelgid is highly invasive in eastern North America, where natural enemies are unable to regulate populations (Wallace and Hain, 2000) and eastern hemlock (*Tsuga Canadensis*) is especially susceptible (McClure, 1992). Following its 1951 introduction in Virginia there was a lag time of approximately 30 years with minimal range expansion. However, in the 1980s infestations expanded northward along the east coast, exploiting the large contiguous tracts of hemlock forest common in the northeast. More recently adelgid range expansion has been southward, where eastern hemlock is more confined to moist coves, higher elevations and north-facing slopes (Godman and Lancaster, 1990; Ward et al., 2004). Although the adelgid was not reported in Kentucky until March 2006, its infestations had been recorded in 22 Kentucky County by 2012, primarily in the southeast (USDA, 2012).

Eastern hemlock is a foundation species in eastern North America and is prominent in riparian areas throughout central and southern Appalachia (Vandermast and Van Lear, 2002; Adkins and Rieske, 2013). Its dense coniferous canopy helps modulate air, soil, and stream temperatures (Godman and Lancaster, 1990; Ford and Vose, 2007). Eastern hemlock helps regulate nutrient cycling

* Corresponding author. Address: PFEN 221E, 715 West State St., West Lafayette, IN 47907, USA. Tel.: +1 765 496 2199; fax: +1 765 494 9461.
E-mail address: sfei@purdue.edu (S. Fei).

0378-1127/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.foreco.2013.12.019

and decomposition rates (Yorks et al., 2003). It is vital for maintaining stream quality, provides habitat for hemlock-dependent birds, and provides seasonal habitat for grouse, turkey, moose, deer, and other wildlife (Shriner, 2001; Snyder et al., 2002; Keller, 2004; Ross et al., 2004; Ford and Vose, 2007). Loss of eastern hemlock will cause changes in vegetation composition and structure; as hemlock trees die, light penetration to the forest floor increases, leading to a larger percent of ground cover to be occupied by vascular plants, including potentially invasive plant species. Increased light penetration also creates suitable habitat for less shade-tolerant tree species such as black birch (*Betula lenta*) and red maple (*Acer rubrum*) (Catovsky and Bazzaz, 2000; Yorks et al., 2003; Spaulding and Rieske, 2010). This shift in vegetative dominance will have serious consequences for native biota, leading to extensive and permanent changes in community composition. Furthermore, the western edge of the contiguous range of eastern hemlock lies in eastern Kentucky (Little Jr., 1971), where it grows in isolated clusters usually confined to moist coves, higher elevations and north-facing slopes (Godman and Lancaster, 1990). These peripheral populations are often critically important in conserving threatened and endangered species (Channell and Lomolino, 2000; van Rossum et al., 2003); an effective monitoring approach may play a crucial role in the preservation of eastern hemlock.

Predicting invasion risk for areas that are susceptible to adelgid establishment would be a powerful tool in the battle against this aggressive invader. Since the adelgid is small, highly mobile, and cryptic, detection is difficult. Absence data are not considered reliable, which is a common problem in species distribution modeling and several methods have been developed to address this (Elith et al., 2006). One option is to use a model based on presence-only data, which may be less accurate than presence–absence models, but is found to be robust in species distribution modeling even with a small sample of presence records available (Elith et al., 2006; Pearce and Boyce, 2006). Maximum entropy and Mahalanobis distance are two presence-only models that have shown high performance among the existing model classes (Farber and Kadmon, 2003; Elith et al., 2006; Phillips et al., 2006; Tsoar et al., 2007). These two models have been used to successfully predict the habitat suitability of a wide range of taxa (e.g., Browning et al., 2005; Dudik et al., 2007; Fei et al., 2012; Liang et al., 2013).

The maximum entropy species distribution model (Maxent; Phillips et al., 2006; Elith et al., 2011) uses presence data to produce a continuous probability of relative habitat suitability. Its name refers to the fact that the resulting estimation of the probability distribution is that which is most uniform—in other words, has maximum entropy (Pearson et al., 2007). This program generates randomly selected background environmental samples from the study area. Maxent is similar to generalized linear models (GLMs) and generalized additive models (GAMs), two common techniques which require absence data or background samples that represent true absences, except that Maxent does not interpret randomly selected background samples used in the modeling process as absence data (Phillips et al., 2006).

Mahalanobis distance (MD; Jenness, 2009) is a multivariate statistic based on the ecological niche concept (Hutchinson, 1957) that can be used to map the probability of use or the probability of occupancy of a location by an organism through determining the similarity of habitats (Rotenberry et al., 2002; Tsoar et al., 2007). A hyper-elliptical envelope of variables is calculated using the mean vectors and inverse of the covariance matrix of the variables, the center representing the optimal habitat of the species based on calibration (training) data. The distance from the center of the hyper-ellipsoid to a point representing a geographic location with a particular set of habitat conditions is known as the Mahalanobis distance for that particular location; the shorter the distance, the more likely the location will be suitable for the species (Watrous et al., 2006). MD differs from Maxent in that it does not require background environmental samples to use in the modeling, except for assessing model accuracy.

Utilizing these approaches to generate hemlock woolly adelgid susceptibility maps would create an invaluable tool for land managers to mitigate the impacts of invading adelgid populations. To reduce prediction uncertainties resulted from a specific model algorithm, consensus forecasting with the ensemble of different models is highly recommended because predictions that are consistent across models will be more reliable than any individual model (Araújo and New, 2007; Comte and Grenouillet, 2013). To reduce prediction uncertainties resulted from a single cut-off threshold, we employed a multi-threshold approach to better present the different levels of invasion risks based on the model predictions (Fei et al., 2012). Additionally, application of spatial statistics such as hotspot analysis can further identify and quantify areas with high invasion risks (Fei, 2010; Catford et al., 2011). This information could be used to prioritize conservation measures, e.g., identification of areas to survey for potential new infestations or determining optimal locations for management efforts. We hope that our spatial statistics based framework as demonstrated in this study will be found useful in additional invasive species management and nature resource conservation tasks.

2. Methods

The study area covered approximately 27,006 km^2 of eastern Kentucky (38.29–36.58°N, 81.96–84.83°W; Fig. 1) of which approximately 2300 km^2 were suitable for eastern hemlock cover (Clark et al., 2012). This region lies within the Eastern Coal Field physiographic region of the Cumberland Plateau. This mountainous area is geologically composed of sandstone, shale, and siltstone (McDowell, 1986) and ranges in elevation from 154 to 1259 m. Average monthly temperature ranges from 1.1 °C in January to 23.9 °C in July and average monthly precipitation ranges from 8.1 cm in October to 13.1 cm in May (Jackson Carroll AP, 1971-2000 data; National Oceanic and Atmospheric Administration, 2002). The dominant forest type is mixed mesophytic consisting primarily of pine-oak dominated communities (Braun, 1950; Turner et al., 2008).

Hemlock woolly adelgid infested sites within the study area were surveyed and recorded with global positioning system (GPS) receivers between 2006 and 2011 using both a systematic 2 × 2 km grid survey and opportunistic random surveys. Both approaches involved visual assessment of accessible branches. Sites with observed adelgid presence ($N = 142$) were used in the model construction (Fig. 1). Infestation points were randomly divided into subsets of training ($n = 108$) and testing ($n = 34$) data, respectively, with a partition ratio of 0.24 for the testing subset (Huberty, 1994). The partition ratio was calculated using an empirical formula $[1 + (p-1)^{1/2}]^{-1}$, where p is the number of predictor variables, as prescribed in Huberty (1994). For both Maxent and Mahalanobis distance modeling, 10 replicate runs were made with training/testing data split randomly in the specified ratio each time. The testing data points were withheld from model construction and subsequently used for model accuracy evaluation. Eleven environmental layers covering a range of natural and socioeconomic factors potentially associated with adelgid introductions and spread were derived for use as predictor variables in the models (Table 1). All layers were converted to raster format with a cell-size of 30 m and the projection set to Kentucky Single Zone State Plane. Wind power maps were resampled using bilinear interpolation; slope and aspect were calculated from a digital elevation model (DEM); hemlock distribution was assessed according to probability prediction (Clark et al., 2012). For Maxent modeling we used Maxent

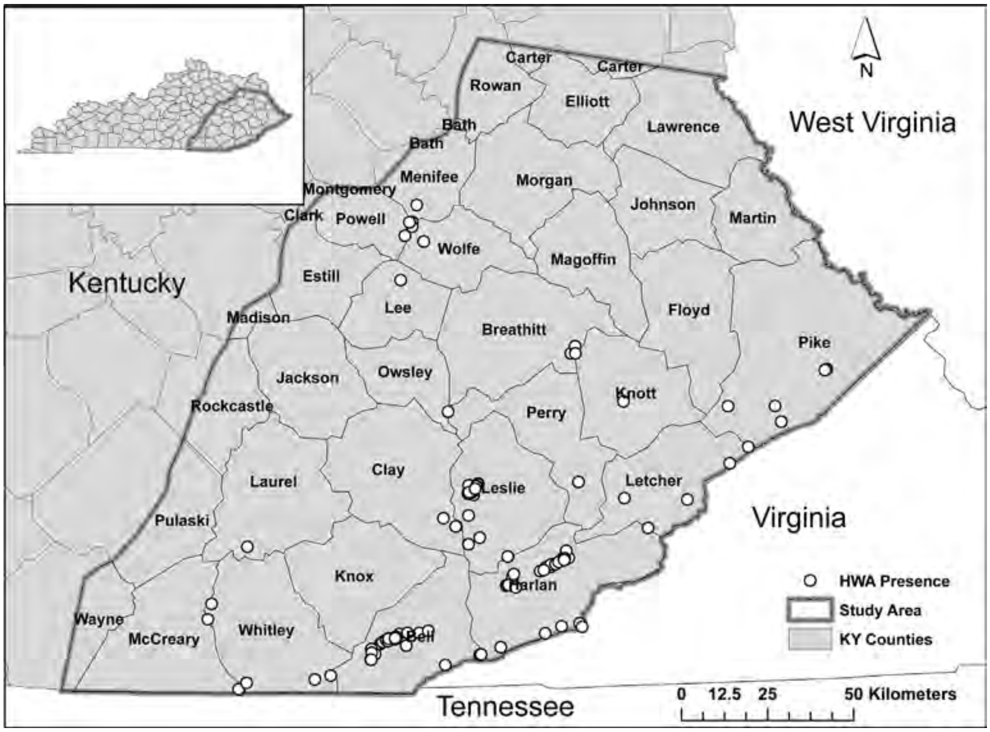

Fig. 1. Study area in eastern Kentucky with hemlock woolly adelgid (HWA) infestation locations observed. Geographic extent is 38.43°N – 36.58°N, 81.96°W – 84.83°W.

Table 1
Environmental layers used as predictor variables in the species distribution models.

Environmental layer	Description	Source
Abandoned railroads	Euclidean distance (m) from nearest abandoned railroad	
Active railroads	Euclidean distance (m) from nearest active railroad	Kentucky geography network
Aspect	Aspect (compass bearing 0–360°) derived from DEM	Calculated using ArcGIS aspect tool
Electric transmission lines	Euclidean distance (m) from nearest electric transmission line	Kentucky public service commission
Populated places	Euclidean distance (m) from nearest feature listed in the USGS Geographic Names Information System (GNIS)	Kentucky geography network
Roads	Euclidean distance (m) from nearest road	Kentucky geography network
Slope	Slope derived from DEM (Degree)	Calculated using ArcGIS slope tool
Streams	Euclidean distance (m) from nearest stream	Kentucky geography network
Trails	Euclidean distance (m) from nearest recreational trail	Kentucky geography network
Wind power	Average of wind power (W/m^2) densities at hub heights of 30, 50, and 70 m	Kentucky geography network
Hemlock	Probability (0–1) of eastern hemlock presence	Prediction as from previous work (Clark et al., 2012)

3.3.3 k (Phillips et al., 2006; Phillips and Dudík, 2008). The Maxent program provided calculations of percentage contribution of variables to models as well as Jackknife analysis using subsets of variables. Therefore the evaluation of relative importance of environmental predictors was for Maxent models only. The Maxent model was run with the maximum number of iterations and convergence threshold parameters set to 500 and 0.00001, respectively. The Mahalanobis Distance (MD) model was generated using the MD function embedded in the Land Facet Corridor Tools (Jenness et al., 2011) with ArcGIS 10.1 (ESRI, Redlands, CA, USA).

The logistic outputs of Maxent models represent a probability estimate (0–1) of habitat suitability for each pixel. For MD models we converted MD surface grids to p-values by fitting raw MD values to a chi-square distribution using a function available within the Land Facet Corridor Tools (Jenness et al., 2011). The p-values of MD surface provided 0–1 estimates of the probability of habitat suitability similar to the logistic output of Maxent (Jenness et al., 2010). The predictive performance of these models was then assessed using the area under the curve (AUC) metric (the area under a receiver operator characteristic [ROC] curve) which is typically adopted for threshold-independent model prediction assessment (Zweig and Campbell, 1993; Fielding and Bell, 1997). The Maxent program provided an AUC value using testing dataset and 10,000 background points for each model. For each MD model we calculated AUC using a similar approach with 10,000 randomly selected background points and the testing dataset using SDMTools package within R, a free statistical modeling program. Because projections that are consistent across models will be more reliable than any individual model (Araújo and New, 2007; Comte and Grenouillet, 2013), we employed an ensemble of two models to predict adelgid invasion risk. The Maxent and MD models with the highest AUC values were used to produce the ensemble model through weighted committee averaging of predictions from the two selected models (Araújo and New, 2007). The respective AUC values were used as weights in the committee averaging.

Table 2
Mean (μ) and standard error (*se*) of environmental variables for observed hemlock woolly adelgid (HWA) infestation locations ($N = 142$) and the entire study area.

Variable*	Study area		HWA infested area		Dissimilarity (%)**
	μ	*se*	μ	*se*	
Abandoned railroads (km)	8.45	0.01	8.03	0.76	5.0
Active railroads (km)	10.80	0	9.07	0.62	16.0
Aspect (degree)	182.03	0.01	163.59	9.22	10.1
Electric transmission lines (km)	3.28	0	3.34	0.22	1.8
Populated places (km)	0.84	0	0.62	0.04	26.2
Roads (km)	0.44	0	0.39	0.04	11.4
Slope (degree)	16.88	0	17.55	0.73	4.0
Streams (km)	0.22	0	0.09	0.01	59.1
Trails (km)	10.59	0	5.21	0.37	50.8
Wind power (W/m^2)	68.8	0	52.83	2.32	23.2
Hemlock (probability ratio)	0.25	0	0.64	0.02	156.0

* See Table 1 for variable description.
** Dissimilarity was calculated as the absolute value of (mean of HWA infested area – mean of overall study area)/mean of overall study area $\times$100%.

Table 3
Percent contribution of environmental variables to hemlock woolly adelgid susceptibility prediction using the Maxent model.

Variable	Percent contribution
Hemlock	41.2
Slope	10.8
Trails	9.9
Abandoned railroads	9.5
Active railroads	8.9
Aspect	5.2
Roads	5.0
Populated places	3.2
Streams	2.8
Electric transmission lines	2.3
Wind power	1.3

We also performed hotspot analysis at the small watershed level (Hydrological Unit Code – HUC 14) based on the ensemble prediction with Getis-Ord Gi* statistic using the corresponding tool in Spatial Statistics Tools of ArcGIS 10.1 (ESRI, Redlands, CA, USA). This technique helped to better reveal areas with clustered higher susceptibility (hotspots) or lower susceptibility (coldspots) for adelgid establishment and spread.

For the selected Maxent, MD and ensemble models, we further implemented a multi-threshold approach that better incorporated the prediction uncertainty versus using a single-threshold for binary classification of the results (Fei et al., 2012). First, the sum of model sensitivity (proportion of correctly predicted presence) and specificity (proportion of correctly predicted absence—using background points as pseudo-absence) (Jiménez-Valverde and Lobo, 2007) were calculated using a whole range of different testing thresholds (0–1 with a 0.01 interval). Then, the change of the sum of sensitivity and specificity (SSS) was evaluated using standard deviation. We identified thresholds corresponding to the maximum SSS and where standard deviations of SSS from the max showed abrupt increases (c.f., Fei et al., 2012). The three thresholds chosen were then assigned with relative meanings to partition different levels of HWA susceptibility (i.e., none, low, medium, and high). This was intended to provide more informative presentations of modeling results to facilitate better management practices.

3. Results

Of the environmental variables compared between adelgid-infested sites and the overall study area, relatively large differences are evident with respect to distance to streams, distance to trails and hemlock presence (Table 2). Adelgid-infested locations appear closer to streams and trails, and where there is a high probability of eastern hemlock presence. Further, according to the variable percentage contribution estimates (Table 3) and Jackknife test results from the Maxent model (Fig. 2), hemlock presence has the greatest predictive power, followed by distance to streams, distance to

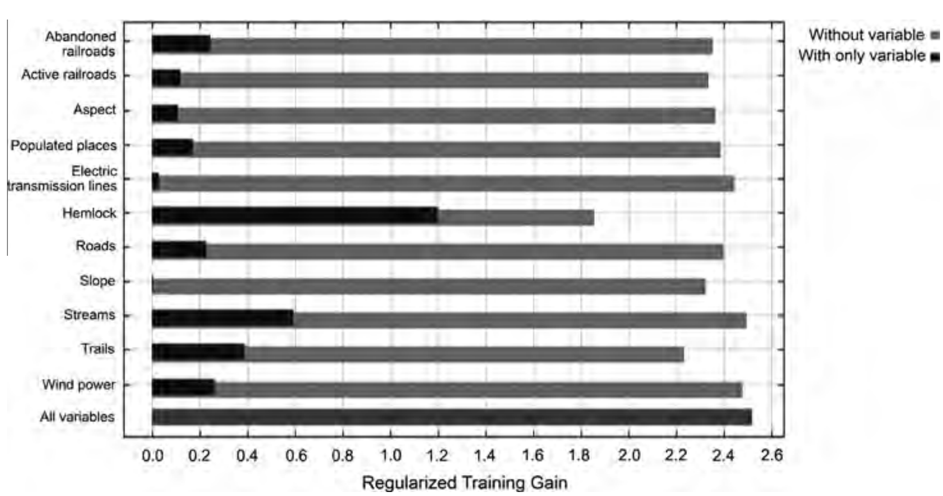

Fig. 2. Jackknife test of Maxent model training gain for environmental variables. Light bars represent training gain when each variable is removed from the model and dark bars when only each variable is used to build the model.

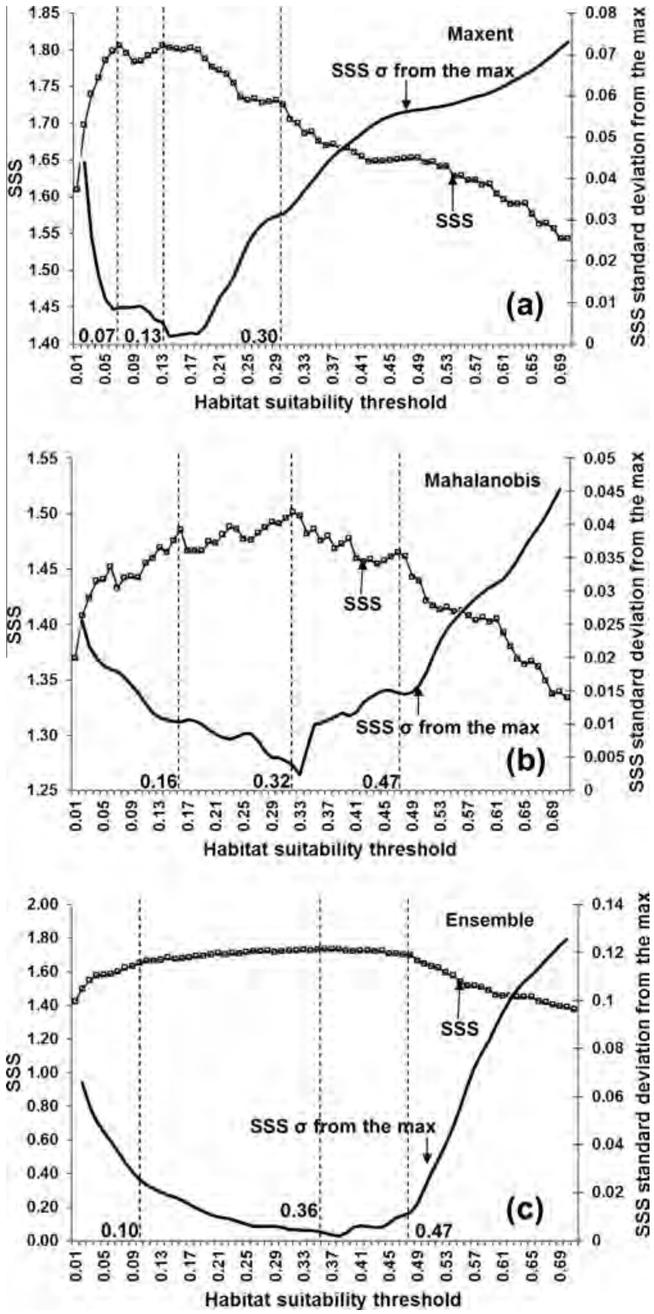

Fig. 3. Sum of sensitivity and specificity (SSS) and standard deviation (σ) from the maximum SSS relative to testing threshold (0–0.70, at 0.01 intervals). SSS values were calculated using hemlock wooly adelgid testing points and 10,000 randomly selected background locations.

trails, distance to abandoned railroads, and distance to active railroads. Slope ranked high for percent contribution to the model, but alone it did not predict adelgid presence well, as indicated in the results of the Jackknife analysis. Wind power ranked lowest in percent contribution, but alone did show some power to predict adelgid presence (Fig. 2).

Maxent models yielded testing AUC values ranging from 0.88 to 0.94; and MD models from 0.81 to 0.86. These results suggest relatively high predictive performance for all models (relative to a random occasion with AUC = 0.5) with Maxent models offering slightly closer estimates. We used the models with the highest AUC values for Maxent and MD respectively, and merged the two selected models into an ensemble prediction that retains information from both models (Fig. 3). While choosing the boundary thresholds (low limit and high limit) was a relatively arbitrary exercise when the points with abrupt standard deviation changes were hard to identify, we included these additional levels to provide a broader range of possibilities as being more informative than a single threshold approach. The resultant map showing areas with HWA susceptibility prediction demonstrated overall agreement between the models which reveal the same general areas under risk (Fig. 4). The upper north region and a northeast-southwest oriented corridor in the middle of the region show the least susceptibility to adelgid infestation. The remaining areas are predicted with varying degrees of susceptibility which contribute to most of the inter-model differences. The Maxent model predicts less adelgid susceptible area than the MD model (Fig. 5). The ensemble model covers the largest area with low adelgid susceptibility and the lowest area for medium and high susceptibility.

The hotspot analysis using the ensemble model pointed to high susceptibility areas that are consistent with patterns shown in the multi-level predictions, but with more explicit details (Fig. 6; c.f. Fig. 1). Three major hotspots with a Getis-Ord Gi * Z-score greater than 1.96 (p-value < 0.05) were identified in the study area. The largest hotspot is located in the northeast portion of the study area, extending from areas bordering Virginia where HWA initially occurred in Kentucky. This hotspot covers the Harland County, central part of Leslie County, most part of Letcher, Knott, Floyd, and Johnson County, and southeastern part of Pike County. The second hotspot is in the southwest of the study area, primarily in McCreary County. The third major hotspot is in the northwest portion of the study area focused on the borders between the Powell, Menifee, Wolfe, and Lee County, near the Natural Bridge State Resort Park and the Red River Gorge National Geological Area, where the HWA infestation is relatively new. A coldspot corridor appeared to separate the high susceptibility areas on the east and on the west which are connected to a limited extent in the southern edge of the study area.

4. Discussion

We employed two species distribution models to predict the spread of the highly invasive hemlock woolly adelgid through southeastern Kentucky, which may be typical of invading populations in central and Southern Appalachia. The adelgid invasion in Kentucky is still relatively early, but the area has extensive eastern hemlock (Clark et al., 2012) capable of supporting a broad scale invasion. The predictions of the distribution models should not be understood as limits of the invasion. Eastern hemlock is widely used as an ornamental and in green space plantings; all are susceptible to colonization. This is confirmed by the paramount importance of hemlock presence as a predictor of adelgid presence. However, trees that are already colonized must be associated with a dispersal pathway. Our models reveal areas of eastern hemlock highly susceptible to invasion based on dispersal paths to known infestations. Variables representing corridors that animals or humans may use (e.g., distance to trails and distance to abandoned railroads) are most important.

Recreational trails that may be in close contact with hemlock trees offer ideal pathways for birds, deer and humans. Abandoned and active railroads may provide similar dispersal corridors for invading populations. Surprisingly, wind, which is an important mechanism of adelgid dispersal (McClure, 1990), played a very minor role in our models. This may be a function of eastern hemlock distribution in the central and southern Appalachian region, where it occurs in discrete patches and further suggests the importance of corridors for adelgid dispersal. It is likely that wind is in

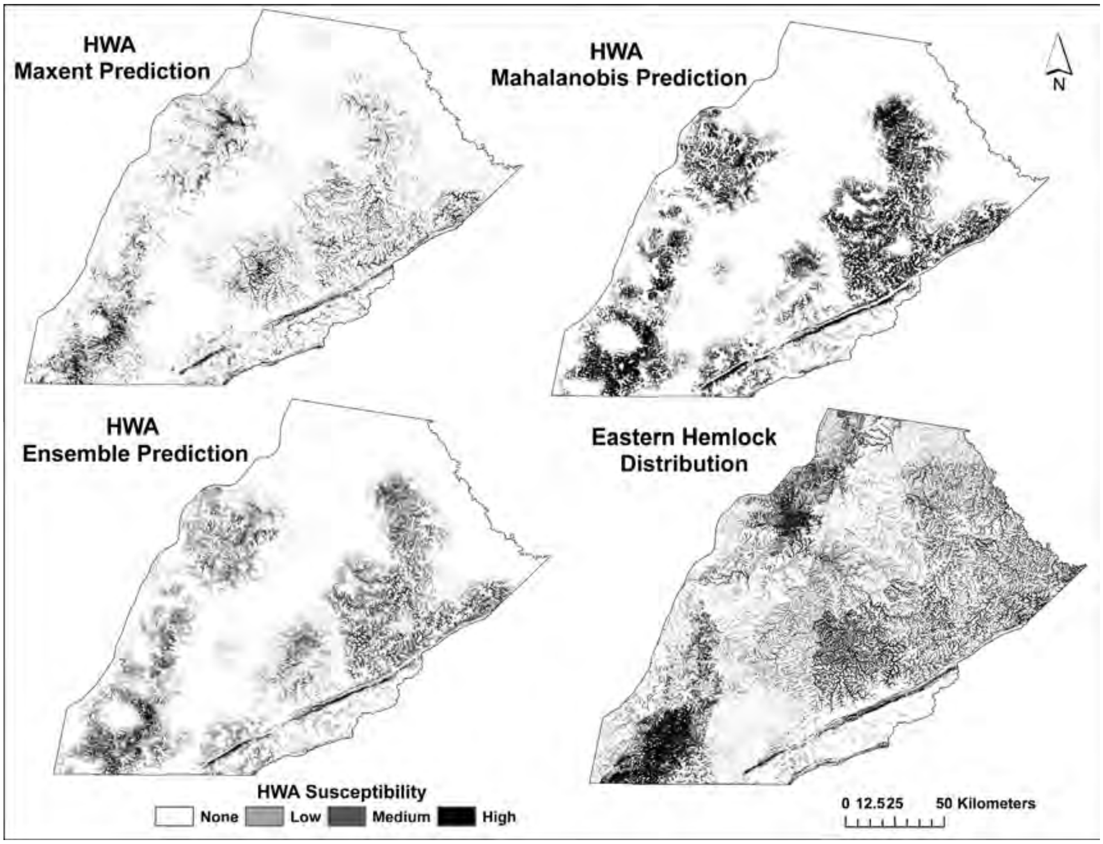

Fig. 4. Multilevel prediction of areas with hemlock wooly adelgid (HWA) susceptibility using maxent, Mahalanobis distance and ensemble models (c.f. Fig. 5). An estimated eastern hemlock distribution map is provided based on our previous work (Clark et al., 2012).

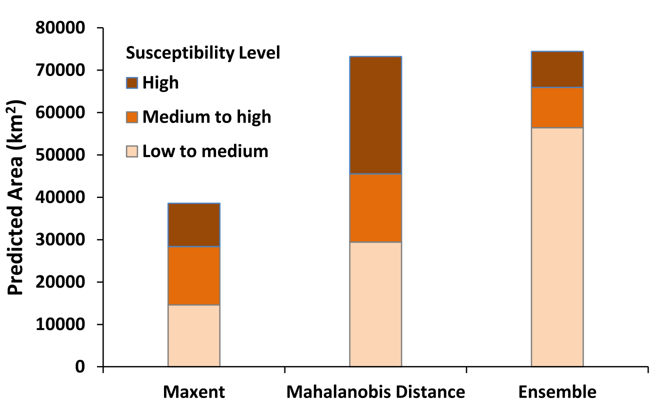

Fig. 5. Predicted areas according to different levels of hemlock woolly adelgid susceptibility prediction for Maxent, Mahalanobis and Ensemble models.

fact dispersing adelgids in the region, but the lack of nearby hemlock to sustain dispersing crawlers minimizes colonization success. Our results suggest that birds, deer, and to a lesser extent, humans and other mammals, are the leading dispersal vectors for hemlock woolly adelgid in the region. Other physiological factors such as slope and aspect may mainly reflect the habitat of eastern hemlock trees.

According to the AUC measurements, Maxent models performed better than MD models, with the latter likely overestimating the area of adelgid susceptibility. This might be due to the lack of absence data. Since Maxent uses background points, this typically leads to higher model accuracy (Vaclavik and Meentemeyer, 2009). The uneven distribution of sampling points, which were concentrated in the southeastern area of study region, may have also contributed to the model uncertainty. Despite these modeling limitations, the predictions from different models demonstrate consistent spatial patterns pointing to the same general areas where adelgid are likely to further invade. The ensemble model predicts greater adelgid susceptible area than Maxent and MD, thus providing a conservative estimate that is useful to inform forest managers of potential locations under risk of invasion. On the other hand, management efforts can be prioritized by targeting medium to high susceptibility areas, which are relatively small in the ensemble prediction. The hotspot analysis based on the ensemble model showed that the clustered areas with high adelgid susceptibility extends to the north and south from a part of Harlan County (adjacent to Virginia) on Pine Mountains where the infestation was first reported in Kentucky (Kentucky Forest Health Task Force, 2006). The infestation in the southwest edge of the study area appears connected to the infestation in the east via dispersal pathways (see Fig. 1). Similarly, the infestation in the northwest portion of the study area may have occurred via dispersal corridors from the south, originating in Owsley and Leslie County, or via the southwest, originating in Laurel County. Regardless, this area represents a hotspot (Fig. 6) and so deserves greater attention by land managers. A single HWA case found on the border of Owsley and Perry County seemed to indicate the direct connectivity between the western hotspot of infestation and the east, but the predicted coldspot corridor lying between the west and the east is to make this potential dispersal pathway unlikely. In fact this particular case was reported in 2010, yet the western hotspot began to establish as early as 2008, suggesting that the adelgid dispersal most likely has detoured through the southern edge of the study area.

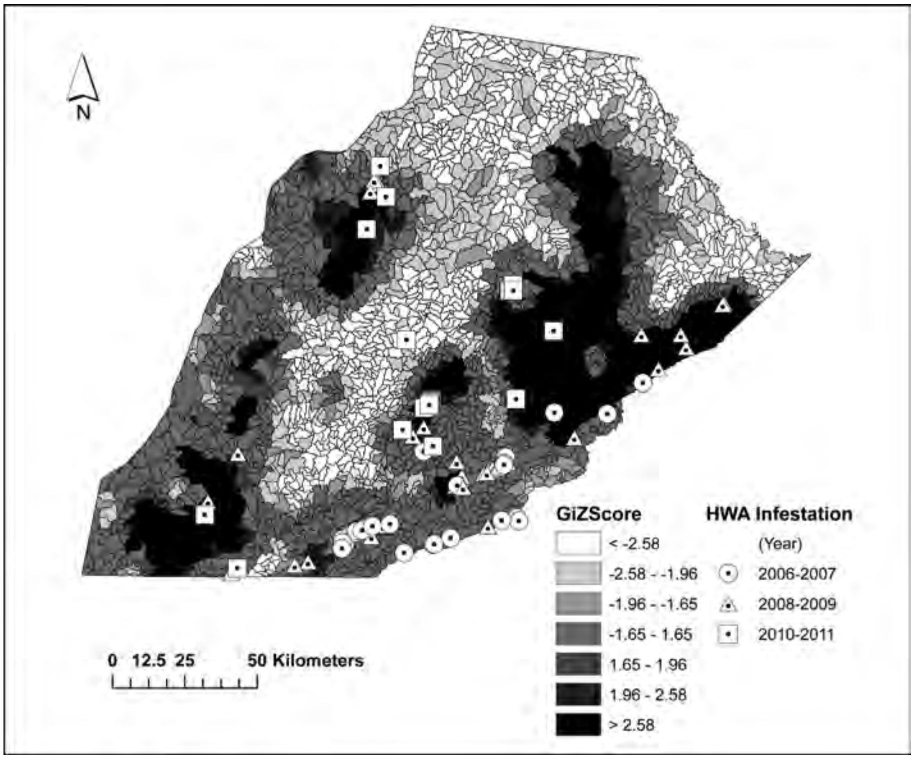

Fig. 6. Hotspot distribution of high (and low) hemlock wooly adelgid (HWA) susceptibility areas by small watersheds (Hydrological Unit Code [HUC] 14). The Getis-Ord Gi * Z-scores (GiZScores) are provided to indicate different levels of clustering of either high values (Z-score positive) or low values (Z-score negative). The Z-scores are the numbers of standard deviation and correspond to respective p-values (e.g., ZScore >1.96 or <−1.96, p-value <0.05; ZScore >2.58 or <−2.58, p-value <0.01). Adelgid locations are presented according to infestation years and the three primary hotspot areas are labeled with numbers (1–3).

5. Conclusions

This study contributes to an EDRR system focusing on the highly invasive hemlock woolly adelgid in Kentucky with a spatial statistics based framework. The species distribution models not only predicted the invasion probability of adelgid for the region, but also revealed the environmental factors that are associated with adelgid invasion. The ensemble of two models and multi-threshold approach facilitated the reduction of model uncertainties. The spatial hotspot analysis further highlighted the areas with high risk of adelgid invasion and revealed the potential dispersal pathways. Our results will facilitate more precise monitoring, and allow development of focused mitigation strategies that could more effectively contribute to eastern hemlock conservation. Our study demonstrated the feasibility of regional-scale quantitative invasion risk assessment, which can be used for more effective and proactive invasion prevention and management.

Acknowledgments

The authors thank C. Harper, J.D. Loan and J.J. Obrycki for support and data collection. This study was supported in part by the USDA, Cooperative State Research, Education, and Extension Service (Project No. 06176) and by the National Science Foundation (Project No. 1241932).

References

Adkins, J.K., Rieske, L.K., 2013. Loss of a foundation forest species due to an exotic invader impacts terrestrial arthropod communities. For. Ecol. Manag. 295, 126–135.
Araújo, M., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47.
Braun, E.L., 1950. Deciduous Forests of Eastern North America. Blakiston, Philadelphia.
Browning, D.M., Beaupre, S.J., Duncan, L., 2005. Using partitioned Mahalanobis D-2(K) to formulate a GIS-based model of timber rattlesnake hibernacula. J. Wildl. Manag. 69, 33–44.
Catford, J.A., Vesk, P.A., White, M.D., Wintle, B.A., 2011. Hotspots of plant invasion predicted by propagule pressure and ecosystem characteristics. Divers. Distrib. 17, 1099–1110.
Catovsky, S., Bazzaz, F.A., 2000. The role of resource interactions and seedling regeneration in maintaining a positive feedback in hemlock stands. J. Ecol. 88, 100–112.
Channell, R., Lomolino, M.V., 2000. Dynamic biogeography and conservation of endangered species. Nature 403, 84–86.
Chornesky, E., Bartuska, A., Aplet, G., Britton, K., Cummings-Carlson, J., Davis, F., Eskow, J., Gordon, D., Gottschalk, K., Haack, R., 2005. Science priorities for reducing the threat of invasive species to sustainable forestry. Bioscience 55, 335–348.
Clark, J.T., Fei, S., Liang, L., Rieske, L.K., 2012. Mapping eastern hemlock: comparing classification techniques to evaluate susceptibility of a fragmented and valued resource to an exotic invader, the hemlock woolly adelgid. For. Ecol. Manag. 266, 216–222.
Comte, L., Grenouillet, G., 2013. Species distribution modelling and imperfect detection: comparing occupancy versus consensus methods. Divers. Distrib. 19, 996–1007.
Dudik, M., Phillips, S.J., Schapire, R.E., 2007. Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. J. Mach. Learn. Res. 8, 1217–1260.
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., Peterson, A.T., 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.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.
Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. Ecol. Model. 160, 115–130.
Fei, S., 2010. Applying hotspot detection methods in forestry: a case study of chestnut oak regeneration. Int. J. Forestry Res.. http://dx.doi.org/10.1155/2010/815292.

Fei, S., Liang, L., Paillet, F.L., Steiner, K.C., Fang, J., Shen, Z., Wang, Z., Hebard, F.V., 2012. Modelling chestnut biogeography for American chestnut restoration. Divers. Distrib., 754–768.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Ford, C.R., Vose, J.M., 2007. *Tsuga canadensis* (L.) Carr. mortality will impact hydrologic processes in southern appalachian forest ecosystems. Ecol. Appl. 17, 1156–1167.

Godman, R.M., Lancaster, K., 1990. Eastern hemlock. In: Burns, R.M., Honkala, B.H. (Eds.), Silvics of North America. USDA Forest Service, Washington, DC, pp. 1238–1255.

Huberty, C., 1994. Applied Discriminant Analysis. Wiley-Interscience, New York.

Hulme, P.E., 2012. Weed risk assessment: a way forward or a waste of time? J. Appl. Ecol. 49, 10–19.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harbor Symp. Quant. Biol. 22, 415–427.

Jenness, J., 2009. Mahalanobis distances (mahalanobis.exe) extension for ArcGIS 9.x. In. Jenness Enterprises.

Jenness, J., Brost, B., Beier, P., 2010. Land Facet Corridor Designer. In. Manual.

Jenness, J., Brost, B., Beier, P., 2011. Land Facet Corridor Designer: extension for ArcGIS. In. Jenness Enterprises. <http://www.jennessent.com/arcgis/land_facets.htm>.

Jiménez-Valverde, A., Lobo, J., 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecol. 31, 361–369.

Keller, D.A., 2004. Associations between eastern hemlock (Tsuga canadensis) and avian occurrence and nest success in the Southern Appalachians. Wildlife and Fisheries Science. University of Tennessee, Knoxville. p. 114.

Kentucky Forest Health Task Force, 2006. Forest Health Annual Report. <http://www.kyforesthealth.org/2006report.pdf>.

Liang, L., Fei, S., Ripy, J.B., Blandford, B.L., Grossardt, T., 2013. Stream habitat modelling for conserving a threatened headwater fish in the Upper Cumberland River. Kentucky. River Res. Appl. 29, 1207–1214.

Little Jr., E., 1971. Atlas of United States trees. Conifers and important hardwoods. Miscellaneous publication 1146, vol. 1. U.S. Department of Agriculture, Forest Service, Washington, DC.

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

McClure, M.S., 1990. Role of wind, birds, deer, and humans in the dispersal of hemlock woolly adelgid (homoptera: adelgidae). Environ. Entomol. 19, 36–43.

McClure, M.S., 1992. Hemlock woolly adelgid. Am. Nurseryman 175, 82–89.

McDowell, R.C., 1986. The geology of Kentucky – a text to accompany the geologic map of Kentucky. United States Government Printing Office, Washington D.C.

National Oceanic and Atmospheric Administration, 2002. Climatography of the United States No. 81. In.

Pearce, J.L., Boyce, M.S., 2006. Modelling distribution and abundance with presence-only data. J. Appl. Ecol. 43, 405–412.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., Peterson, A.T., 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. 34, 102–117.

Phillips, S., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

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

Ross, R.M., Redell, L.A., Bennett, R.M., Young, J.A., 2004. Mesohabitat use of threatened hemlock forests by breeding birds of the Delaware river basin in northeastern United States. Nat. Areas J. 24, 307–315.

Rotenberry, J.T., Knick, S.T., Dunn, J.E., 2002. A minimalist approach to mapping species' habitat: pearson's planes of closest fit. In: Scott, J.M., Heglund, P.J., Morrison, M.L. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 281–289.

Shriner, S.A., 2001. Distribution of Breeding Birds in Great Smoky Mountains National Park. In: Zoology. North Carolina State University, Raleigh, p. 198.

Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D.A., Aronson, J., Courchamp, F., Galil, B., García-Berthou, E., Pascal, M., 2012. Impacts of biological invasions: What is what and the way forward. Trends Ecol. Evol. 28, 58–66.

Snyder, C.D., Young, J.A., Lemarie, D.P., Smith, D.R., 2002. Influence of eastern hemlock (*Tsuga canadensis*) forests on aquatic invertebrate assemblages in headwater streams. Can. J. Fish. Aquat. Sci. 59, 262–275.

Spaulding, H.L., Rieske, L.K., 2010. The aftermath of an invasion: structure and composition of Central Appalachian hemlock forests following establishment of the hemlock woolly adelgid, *Adelges tsugae*. Biol. Invasions.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007. A comparative evaluation of presence-only methods for modelling species distribution. Divers. Distrib. 13, 397–405.

Turner, J.A., Oswalt, C.M., Chamberlain, J.L., Conner, R.C., Johnson, T.G., Oswalt, S.N., Randolph, K.C., 2008. Kentucky's Forests. Asheville, NC.

USDA, 2012. <http://na.fs.fed.us/fhp/hwa/maps/2012.pdf.In>.

Vaclavik, T., Meentemeyer, R.K., 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecol. Model. 220, 3248–3258.

van Rossum, F., Vekemans, X., Gratia, E., Meerts, P., 2003. A comparative study of allozyme variation of peripheral and central populations of *Silene nutans* L. (Caryophyllaceae) from Western Europe: implications for conservation. Plant Syst. Evol. 242, 49–61.

Vandermast, D.B., Van Lear, D.H., 2002. Riparian vegetation in the southern Appalachian mountains (USA) following chestnut blight. For. Ecol. Manag. 155, 97–106.

Wallace, M.S., Hain, F.P., 2000. Field surveys and evaluation of native and established predators of the hemlock woolly adelgid (Homoptera: Adelgidae) in the southeastern United States. Environ. Entomol. 29, 638–644.

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 (Ed.), Forest Service.

Watrous, K.S., Donovan, T.M., Mickey, R.M., Darling, S.R., Hicks, A.C., Von Oettingen, S.L., 2006. Predicting minimum habitat characteristics for the Indiana bat in the Champlain Valley. J. Wildl. Manag. 70, 1228–1237.

Yorks, T.E., Leopold, D.J., Raynal, D.J., 2003. Effects of Tsuga canadensis mortality on soil water chemistry and understory vegetation: possible consequences of an invasive insect herbivore. Can. J. For. Res. – Rev. Can. Rech. For. 33, 1525–1537.

Zweig, M., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577.