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