

A geographically variable model of hemlock woolly adelgid spread

Alexander M. Evans · Timothy G. Gregoire

Received: 12 January 2006 / Accepted: 3 July 2006 / Published online: 11 November 2006
© Springer Science+Business Media B.V. 2006

Abstract The hemlock woolly adelgid (HWA) (*Adelges tsugae*) has spread through the range of eastern hemlock (*Tsuga canadensis*) leaving dead hemlock trees in its wake. We combined county and previously unavailable township level infestation records with geographic variables including plant hardiness zone, elevation, forest cover type, urbanization, precipitation, temperature, and hemlock range in a geographic information system to build models linking HWA spread to the structure and characteristics of the environment. We explored the connections between site variables and spread rate using quantile regression, Tobit regression, a parametric model for heterogeneous error variance, and classification and regression tree modeling. The results show that since 1990 HWA has spread at a rate closer to 12.5 km/yr rather than the 20–30 km/yr previously estimated, however spread rates vary significantly with environmental variables. County-based data show a faster rate of range expansion in the south, 15.6 km/yr, and very slow spread in the northern part of the current range of HWA. The rate of spread based on township records for Pennsylvania and north is 8.13 km/yr. Our study

suggests that HWA spreads more slowly in colder areas, with a mean minimum temperature of -26.1°C (plant hardiness zone 5B) delineating areas of slower range expansion. HWA also spreads more slowly during cold years, those with lower than average mean January temperatures. Our analysis demonstrates the value of quantile regression, Tobit regression, and classification and regression tree models for understanding the spread of invasive, exotic insects.

Keywords Classification and regression trees · Invasive insect · Hemlock woolly adelgid · Range expansion · Quantile regression · Tobit regression

Introduction

A wide array of pests and pathogens have established in North America and caused hundreds of billions of dollars in damages (Pimentel et al. 2005). The exotic insects and diseases introduced to forests have removed dominant species, reduced diversity, altered disturbance regimes, and affected ecosystem function (Liebhold et al. 1995; Mack et al. 2000). One alien species can facilitate subsequent alien invasions. Positive interactions between alien species can cause the breakdown of natural systems in an “invasion meltdown” (Simberloff and Von Holle 1999). The opportu-

A. M. Evans (✉) · T. G. Gregoire
School of Forestry and Environmental Studies, Yale University, 360 Prospect St., New Haven, CT 06511, USA
e-mail: alexander.evans@aya.yale.edu

nity for exotic invaders to reach new ecosystems grows as foreign trade increases in volume and speed (Work et al. 2005). Global climate change also has potential to aggravate the alien species problem (Simberloff 2000).

An alien organism becomes invasive if it is able to establish in a new region and then spread to suitable habitats (Williamson 1996). Research is helping to identify potential invaders (e.g. Williamson and Fitter 1996; Heger and Trepl 2003), but in general biological invasions are often detected only when they expand their range beyond their initial point of introduction (Liebhold et al. 1995). The rate of range expansion is determined in part by habitat heterogeneity. Some geographic variables facilitate population growth and spread of exotic species while others impede invasion.

The management or mitigation of alien species often relies knowing the rate of range expansion. Information on species spread can allow for earlier detection of invasive species, protection of areas of special concern, and implementation of biological control. Measures of spread rate are crucial for programs aimed at limiting or slowing the spread of exotic invasions (Sharov and Liebhold 1998). Knowledge of spread rates can aid silvicultural responses to exotic forest pests (Waring and O'Hara 2005).

This paper demonstrates a novel application of four statistical techniques to the estimation of the range expansion of a biological invasion. We use quantile regression, Tobit regression, a parametric model for heterogeneous error variance, regression tree, and classification tree modeling to explore the connections between site variables and spread rate. These four techniques are well suited to geographic data on biological invasion, which may not meet the assumptions of common analytical techniques. The use of multiple analytical approaches confirms patterns of range expansion and the importance of geographic factors. The results of this analysis provide an improved estimate of spread for the study organism as well as a more detailed understanding of the geographic factors that influence its rate of range expansion.

Methods

Study system

An invasive exotic insect, the hemlock woolly adelgid (HWA), *Adelges tsugae* Annand, has caused wide-spread mortality of eastern hemlock trees, *Tsuga canadensis* (L.) Carrière and threatens to remove the species from North American forests. HWA-induced mortality is a concern because hemlock provides important forest structure, habitat, economic benefits, and aesthetic values in northeastern forests.

The eastern hemlock's native range extends from the southern Appalachian Mountains in Georgia north to New Brunswick, Canada and as far west as northern Minnesota. Hemlock is a shade tolerant species that can persist in the understory of closed canopy forest for decades, although in more open conditions hemlock can grow quite rapidly (Godman and Lancaster 1990). In many northeastern forests hemlock is one of the few conifers and its presence can augment structural diversity (Beatty 1984). Hardwoods stands with a hemlock component have greater primary productivity than those without hemlock (Kelty 1989). Hemlock shade over streams and wetlands helps maintain a cool and moist microclimate (Snyder et al. 2002). Hemlock provides unique habitat on which numerous bird species depend (Ross et al. 2002; Tingley et al. 2002). Since they maintain needle cover in the winter, hemlock also provide shelter from snowfall on which many animals such as deer, grouse, and turkey rely (DeGraaf et al. 1992 p. 92). Hemlock is an important food resource and deer, hares, and other animals preferentially browse it (White and Mladenoff 1994; Kittredge and Ashton 1995). Without stand-replacing disturbance hemlock becomes part of old growth forests in its range (Godman and Lancaster 1990). While a number of native insects feed on hemlock including hemlock looper (*Lambdina fiscellaria* Guenée) and hemlock borer (*Melanophila fulvoguttata* Harris), none have the impact the HWA does.

Hemlock woolly adelgid

HWA is a tiny (1.5 mm), sap sucking insect native to Asia and named for a woolly wax that covers it. The insect feeds directly on the xylem ray parenchyma cells by inserting a long stylet into the needle cushion (Young et al. 1995). Feeding occurs from the fall through to early summer, which allows HWA to feed unmolested by predators inactive during the winter. Although feeding by an individual causes little damage, populations can grow so large that there are nearly as many HWA as needles on the tree. HWA populations increase rapidly by producing two generations each year and because of a lack of specific predators. In eastern North America, HWA has both a spring and summer generation but reproduces only asexually. In its native range HWA relies on a secondary host for sexual reproduction, however no secondary host has been found in North America (Ward et al. 2004). In China and Japan a number of specialist predators feed on HWA, whereas in North America the available generalists predators fail to keep populations in check (Wallace and Hain 2000).

Current biological control efforts include a suite of specialist predators, which in combination may be able to reduce HWA populations below damaging levels (Reardon and Onken 2004). Great effort has been put into biological control because there are no signs of long-term recovery of untreated hemlock trees in eastern North America (Orwig et al. 2002; Ward et al. 2004). All sizes and ages of hemlock are affected by HWA and many die within a few years of being infested. Some hemlock trees can survive 12 years or more, but their prospects without some control of HWA are grim (Orwig et al. 2002; Ward et al. 2004). Short-term hemlock recovery is common after the first year or two of heavy HWA infestation, due hemlock's inability to continuously support high densities of HWA. Unfortunately this recovery is followed by renewed HWA feeding and a decline in tree health (McClure 1991). HWA mortality has been linked to cold winter temperatures (Parker et al. 1998; Cheah and McClure 2002; Shields and Cheah 2005).

Spread

HWA and associated hemlock mortality are spreading at alarming speed through eastern North America. HWA was first reported in eastern North America in 1951 and by 2004 had spread across 21% of eastern hemlock's range (Onken and Reardon 2005) (Fig. 5). Neither the population dynamics of HWA nor its spread in forests is well understood, creating a gap in the information needed for biocontrol efforts and response to the ensuing mortality. Also, there is no co-ordinated monitoring program or an established scientific sampling method to detect HWA presence, abundance, or rate of range expansion. Sampling HWA is complicated by its small size, lack of pheromones, passive dispersal, and tree crown habitat (Evans 2005). Because of this infestation records are typically presence/absence values for towns or counties.

The published estimate of rate of HWA spread is 20–30 km/yr (Ward et al. 2004) but the only published data on dispersal appeared in a 1990 study describing the role of wind, deer, and bird in moving HWA at least 2 km (McClure 1990). This same study points out that dispersal by humans is evident, but has not been quantified. The dispersal and distribution of HWA within tree crowns, between trees, and across forests remain gaps in HWA research. Estimating the rate of HWA spread is an important element in protecting hemlock. Biological control efforts target the leading edge of invasion to ensure that the hemlock has not declined too much in health but there is a population of HWA to feed on. Predictions of HWA spread also affect hemlock logging. Many land managers are harvesting hemlock based on the threat posed by the HWA in advance of its actual impact (Kizlinski et al. 2002). In this study, we use the best available county and township records of infestation to derive a model of HWA spread. Since HWA appears to be spreading unevenly across the range of hemlock, we used regional geographic data to create a model that accounts for landscape effects on spread rates.

Analysis

The infestation records available to measure historic HWA spread are the US Forest Service county records (Onken and Reardon 2005) and township records for Pennsylvania and northward. Since 1990, 201 counties and 344 townships have recorded infestations. The median county in our study was 112,000 ha in size while the median township was 4,000 ha. The sizes of the counties and townships determine the spatial resolution of this study.

Neither the county nor the township dataset is exact, because records of HWA infestation are not exact: arrival of the HWA may at first go unnoticed, and then later go unreported. There are also inconsistencies between the two datasets, perhaps because the methods of recording and reporting infestation vary by state and surveyor. However, these records are the best available information and an improvement on rough guesses. We used data from 1990 through 2004 as it is the most reliable and most relevant to current spread rates. Before 1990, the temporal resolution of infestation records is more coarse and may not be as accurate (Onken and Reardon 2005). HWA was initially reported in Richmond, VA outside the main range of hemlock forest. The rate of HWA spread, previously estimated to 20–30 km/yr (Ward et al. 2004), is likely to have change when it reached a forest environment and should be more consistent since 1990.

We used 12 variables that were likely to influence habitat heterogeneity for the HWA and were available as digital maps. We combined the habitat variables with the county and township records of infestation in a geographic information system (ESRI 2004).

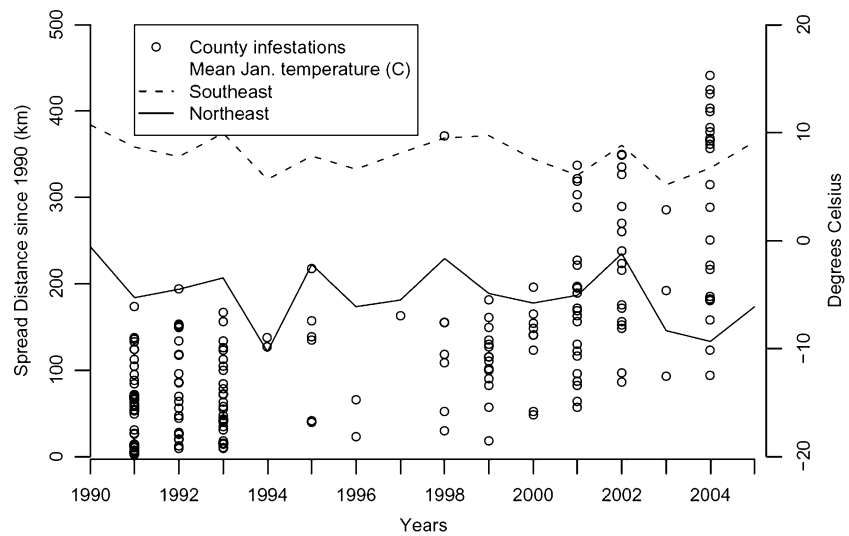
- Hemlock Range (Little 1999)
- USDA Plant Hardiness Zone Map (National Arboretum 1990)
- Average daily minimum temperature, 1971–2000 (Daly et al. 2002)
- Annual precipitation, 1961–1990 (Daly and Taylor 2000)
- Global Land 1-km Base Elevation Digital Elevation Model (National Imagery and Mapping Agency 1996)

- Forest Cover Type (US Geological Survey and US Forest Service 2000)
- The National Atlas urbanization map (US Geological Survey 2000)
- The National Atlas road map (US Geological Survey 2000)
- Population in 1990 per square mile (US Census Bureau 1990)
- Distance from counties infested in 1990
- Northward spread of infestation
- Eastward spread of infestation

Hemlock range determines the potential range of HWA, since its feeding is limited to hemlock in North America. We used the plant hardiness zone map and the map of minimum temperature because of the link between HWA mortality and cold temperatures (Parker et al. 1998; Cheah and McClure 2002; Skinner et al. 2003; Shields and Cheah 2005). Elevation, precipitation, and forest cover are determinants of hemlock habitat suitability and hence may affect HWA. We employed the maps of population, urbanization and roads as a proxy for the human induced movement of HWA. The direction of spread from previous infestation is recorded as the azimuth out of a 360° circle. Since high values of and low values both indicate spread from the north, we transformed the data using an angular transformation to create two variables: “northness of spread” = cosine(aspect) and “eastness of spread” = sine(aspect) (Beers et al. 1966).

We also included a measure of temperature for each year between 1990 and 2004 (National Climate Data Center 2005). Since extreme cold appears to have the largest effect on HWA mortality (Parker et al. 1998), we used the mean January temperature. Figure 1 displays the January temperature in both the Northeast and Southeast regions along with the county HWA spread data. Figure 1 suggests that it is the temperature in the year that infestation was reported that has an impact. This is likely to be due to using January temperatures and HWA surveys from the summer later in the same year. We measured spread as the distance from the center of counties (or townships) infested in 1990 to the center of counties (or townships) newly infested. Other analyses of exotic insect range expansion

Fig. 1 The distance of spread from 1990 to the center of counties reporting new infestations of HWA with mean January temperature for the northeast and southeast. The highest spread value in 1998 is due to human transport of infested seedlings to western New York



have also used the distance from previously infested areas to newly infested areas as the spread distance (e.g. Sharov et al. 1999). Another common approach to measuring range expansion, the square root of area infested, was problematic because of the coarse grain of the HWA data (Shigesada et al. 1995). Because current monitoring does not record confirmed absences of HWA, no years of zero spread are included in this analysis. Each infestation record after 1990 is one observation in our analysis so there are 201 observations in the county data and 344 in the township data.

Many efforts at analyzing range expansion of invading organisms rely on process-based models to predict spread. For example, the Skellam model is a commonly applied representation of simple diffusion, which uses estimates of population growth and a diffusion coefficient to predict the asymptotic rate of range expansion (Skellam 1951). In the case of the HWA, there are no estimates of the diffusion coefficient, or adequate sampling methods for measuring it. In addition, Grosholz's (1996) review found that diffusion models are not likely to predict spread rates accurately, in part because of the difficulty of estimating the diffusion coefficient. Therefore, we have taken advantage of the records of infestation to empirically calculate the rate of range expansion. We use three different models to examine the relationship between the spread distance from

1990 and time. We investigated an unusually distant infestation in western New York in 1998, evident as the highest spread value in 1998 in Fig. 1, where infested nursery stock caused HWA movement. However in the following analyses its impact was minor, so we did not exclude it.

A scatter plot of spread distance versus time (Fig. 1) reveals that the variation of distance increases with time. Also, it is reasonable to assume that at least some of the observed data are spatially correlated. A regression of distance as a function of time can be fitted to these data with the ordinary least squares method. However the OLS fit will provide inefficient estimates of model parameters if the heteroscedasticity and spatial correlation are ignored in the modeling. If the intercept were constrained to pass through the origin heteroscedasticity would cause the underestimation of the variance of the slope estimate. Forcing the intercept through zero would also be an extrapolation of the straight-line trend between 0 and year 1, which there is no data to support. In some respects, a more salient aspect of these data is the truncation of values at zero distance, which calls into question the meaningfulness of modeling the mean, or expected spread distance as a function of time. Thus an OLS fit of the regression line for the mean will be biased owing to the truncation at zero.

In view of these statistical issues and their consequential impact on the estimation and

interpretability of model parameters, e.g., the slope of the line as a measure of rate of spread (sensu Liebhold et al. 1992), we investigated three alternative techniques to fit the regression: quantile regression, a Tobit model for truncated response, and a linear model for the mean response coupled with a parametric model for the heterogeneous error variance.

Quantile regression fits a model to portray the change in the 50th (median) or any other quantile of the distribution of distance as a function of time. Because it tracks the quantile trend, there is no presumption of uniform variance around the quantile regression line. Quantile regression makes no distributional assumptions of the model error term and so creates reliable models for a wide variety of data (SAS Institute Inc. 2002). Estimates from a range of quantiles (e.g. 5th and 95th) provide an excellent way to track heteroscedasticity as the difference between them would increase as the variance increased (Fig. 2 is an example of this) (Koenker and Hallock 2001). In fact, quantile regression is championed as both more informative and more general than OLS regression because it is free from strict distributional assumptions and multiple quantiles more fully describe the distribution of the response variable (Gannoun et al. 2005). Quantile regression has been used in ecology to analyze heteroscedastic data and investigate limiting factors (Cade et al. 1999; Cade and Noon 2003).

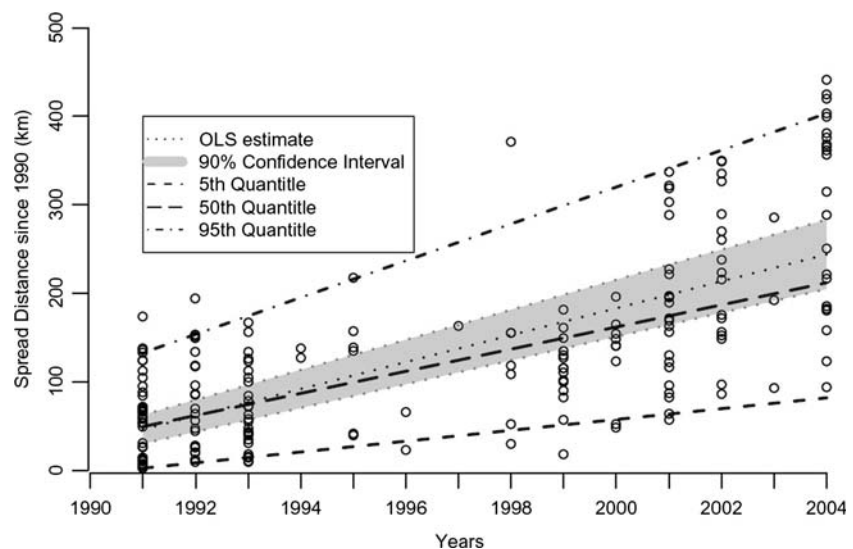
Tobit regression techniques for data where the response variable is truncated are most easily explained via an example. Suppose X and Y are positively correlated and related in the following manner:

$$Y = \beta_0 + \beta_1 * X + u$$

where u is the random deviation of Y from its expected value, $E[Y] = \beta_0 + \beta_1 * X$. If the values where $Y \leq 0$ are removed, or set to zero, a Tobit model is appropriate. The Tobit uses a conditional distribution to combine a probit and linear regression. For those values of X where $E[Y|u > -\beta_0 - \beta_1 * X] < E[Y]$ a probit model is used to model the probability that Y is positive. For those values where $E[Y|u > -\beta_0 - \beta_1 * X] = E[Y]$ a standard regression is used. In this manner, Tobit models account for the negative correlation between X and u in the truncated data (Amemiya 1984). The Tobit model is useful because of the truncation of distance at 0 km, which would cause the OLS estimated slope to be biased upwards.

The third technique for modeling spread rate we examined is similar to other efforts to model heterogeneous variance with a parametric function, e.g. Gregoire and Dyer (1989), but in the present effort the parameter describing the heterogeneity in the error distribution when $Y > E[Y]$ (above the line) differed from that for the

Fig. 2 County infestations of HWA with the quantile and OLS regression lines. The slope of each regression line provides an estimate of the rate of range expansion



other half of the distribution, i.e. when $Y < E[Y]$. We used the NLMIXED procedure in SAS (2002) to iteratively search for optimal solution to the following equation:

$$Y = \beta_0 + \beta_1 * X + e * X^{(\gamma_1 \text{ if residuals } > 0 \text{ and } \gamma_2 \text{ if residuals } < 0)},$$

where Y is spread distance, X is years since infestation, e is error, and γ_1 is used if residuals > 0 and γ_2 is used if residuals < 0 .

A geographically variable model of spread

County data of HWA infestations and anecdotal reports do not suggest an isotropic spread (Ward et al. 2004). Spatially variable range expansion, which includes spatial heterogeneity, may be a more appropriate model for HWA spread. We examined the geographic influence on HWA spread rate using multiple quantile regression and both regression and classification tree analyses. Multiple quantile regression is an extension of quantile regression that includes more explanatory variables, just as ordinary multiple regression does. Unlike ordinary regression, quantile regression can provide information about the range of values for a covariate and identify how the effects of covariates change across the distribution of the independent variable (Koenker and Hallock 2001). Multiple quantile regression also remains free of the strict distributional assumptions of ordinary multiple regression (Koenker and Hallock 2001).

We used regression tree modeling to identify the factors that determine the rate of HWA spread. Regression tree analysis uses training data to assign estimates of the mean of the dependent variable to new data. The training data defines a decision tree that then categorizes new data based on a series of attributes. Unlike OLS regression with categorical variables, regression trees use a set of logical conditions to split the data rather than a single predictive equation. The regression tree methodology is computationally intensive since the algorithm tests the improved classification accuracy of all possible splits at each node. Splits are locally but not necessarily globally optimal. To avoid over fitting the model, limits are set on the minimum number of stands that can

be split at each node. Allowing splits of small groups encourages the algorithm to tailor the model too exactly to the training data, which limits applicability to data not included in the model creation. Pruning the decision tree corrects over fitting. We pruned our models based on a 10-fold cross validation. We determined the best model size by selecting the tree that was within one standard deviation of the minimum misclassification error (Breiman et al. 1984; Therneau and Atkinson 1997).

Our classification tree analysis uses the same basic modeling technique, but assigns a class membership, rather than the value of a continuous variable as a regression tree does. To build classes we split the dataset into those counties that were infested without having an infested neighbor (fast), those counties that became infested after 1 year of adjacency to an infested county (normal), and those counties that remained uninfested after one or more years adjacent to an infested county (slow). We measure classification accuracy of a potential split with the Gini index, which calculates the cost of misclassification based on the probability of class membership (Breiman et al. 1984; Therneau and Atkinson 1997). Both the misclassification percentage and the Kappa statistic are commonly used metrics for comparing classification accuracy. Unlike the misclassification percentage, the Kappa statistic accounts for the chance agreement of model and test data (Monserud and Leemans 1992).

Some of the advantages of classification and regression tree modeling are that they do not make any assumptions about the distribution of the data, they are robust to outliers and misclassifications in the training data, they combine both categorical and continuous data, and the interpretation of results is straightforward (Breiman et al. 1984). We used the *rpart* package in the statistical language *R* that implements a recursive partitioning algorithm according to the methodology of Breiman et al. (1984; R Development Core Team 2004).

Results

An ordinary least squares fit of the regression provides a useful benchmark for comparison.

OLS estimates HWA spread at about 15 km/yr (Spread distance = $31.1 + 15.2 \times$ years since 1990; adjusted $R^2 = 0.50$). Quantile regression provides results for both the median as well as the maximum spread rate. The median estimate is 12.5 km/yr and maximum estimate from the 95th quantile is 20.8 km/yr.

5th quantile: Spread distance = $-3.79 + 6.12 \times$ years since 1990

50th quantile: Spread distance = $36.9 + 12.5 \times$ years since 1990

95th quantile: Spread distance = $112 + 20.8 \times$ years since 1990

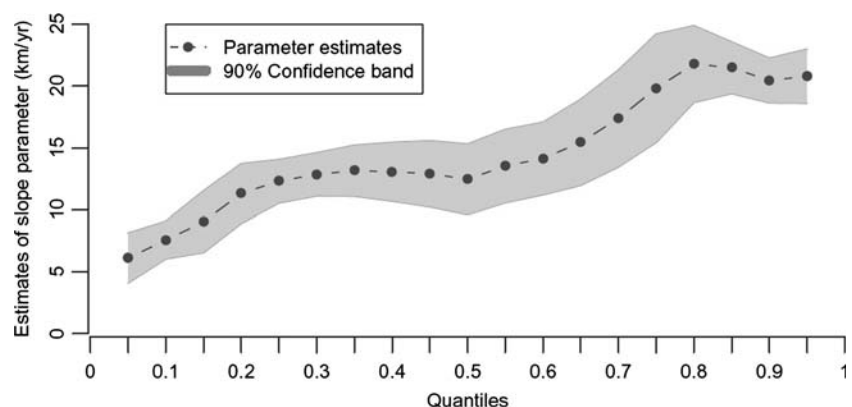
Figure 2 is a graph of the simple linear regression estimate as well as the 5th, 50th, and 95th quantile estimates. Confidence bands were calculated by inverting rank score tests because of the heterogeneous variance (SAS Institute Inc. 2002, see details of the QUANTREG procedure at <http://www.support.sas.com/rnd/app/da/quantreg.html>). Estimates of the slope parameters for all quantiles from the 5th to the 95th are shown in Fig. 3. The result from the Tobit regression is an estimate of 30.2 km/yr

Spread distance = $-149 + 30.2 \times$ years since 1990.

When we only included the years between 1990 and 2000 in the analysis, after which the data do not appear truncated by zero, the Tobit regression estimates a less steep slope:

1990 through 2000: Spread distance
= $-1.6 + 13.1 \times$ years since 1990.

Fig. 3 Estimates of the slope parameter from the quantile regression, which is the estimate of the rate of range expansion, and the 90% confidence band. The estimates of the slope parameter change across the quantiles of the data and plateau at around the unconstrained rate of spread



Our model with different parameters for positive and negative residuals supplies yet another estimate of the spread rate at 7.65 km/yr.

Spread distance = $56.9 + 7.65 \times$ years since 1990 + error* years since 1990** (1.81 if residuals > 0 and 2.90 if residuals < 0).

Township based results

We also tested three methods for calculating the spread rate of HWA on the township data, which is only available for Pennsylvania and north. For the township data the ordinary least squares estimate of spread is about 7 km/yr (Spread distance = $22.6 + 7.12 \times$ years since 1990; adjusted $R^2 = 0.31$). The results of the quantile regression on the township data are:

5th quantile: Spread distance = $-9.00 + 3.00 \times$ years since 1990

50th quantile: Spread distance = $10.9 + 8.13 \times$ years since 1990

95th quantile: Spread distance = $105 + 6.40 \times$ years since 1990

Analysis of geographic variables

We use quantile regression to test the influence of physiographic variables on the spread of HWA. We tested the 12 variables available from the digital maps listed in the methods section. Four variables have significant impact on the distance HWA spread at the 90% confidence level, in addition to the years since 1990. In addition an indicator variable for

areas colder than frost hardiness zone 5B (average annual minimum temperature -23.4 to -26.1°C) was also significant. Estimates of their coefficients for the 5th, 50th, and 95th quantiles appear in Table 1. The estimates for the 50th quantile are very similar to the OLS estimates from the multiple regression. In Table 1, the 5th and 95th quantiles coefficient estimates are different from those of the 50th quantile, which shows that the impact of precipitation, elevation, and hemlock range change across the distribution of distances from 1990.

The regression tree assigned faster spread rates to counties with greater precipitation and lower elevations with an R^2 of 0.55. The estimated spread rates are 14.8 km/yr (> 146 cm average annual precipitation), 11.1 km/yr (< 146 cm precipitation and < 554 m median elevation), 11.2 km/yr (< 146 cm, < 554 m, and $> 5.6^{\circ}\text{C}$ average annual daily minimum temperature), 12.5 km/yr (< 146 cm, < 554 m, and $< 5.6^{\circ}\text{C}$).

Our classification tree correctly predicts about 63% of the counties as “fast”, “normal” or “slow” spread. The kappa value for the classification model is 0.31. A more realistic test of the ability of a classification tree to categorize the HWA data is the construction of the model from 80% of the data and testing with the remaining 20%. After 1,000 iterations of random selection, model construction, and validation the classification trees correctly classify an average of 52% of the data not used in model construction, with an average Kappa statistic of 0.17. The model performs much

better on those counties infested either “fast” or “slow” while those counties infested at the “normal” rate are most difficult to classify. In the tree shown in Fig. 4 no nodes are classified as “normal”.

The variables used to categorize the counties by speed of infestation shown in Fig. 4 are similar to the quantile regression and the regression tree. At the first split 78% of the counties within 5 km of an urban area (built-up areas with large populations as defined in USGS 2000) fall in the “fast” infestation group. Those counties > 5 km from urban areas and in USDA plant hardiness zone 5B (average annual minimum temperature -23.4 to -26.1°C) or colder are likely to be slowly infested. For those counties > 5 km from urban areas and warmer than plant hardiness zone 5B, the mean temperature in January in the year of spread was the next determinant of class. If the mean temperature in January is above average for the period 1990–2004, then counties infested in that year are likely to be in the “fast” category. For the remaining counties (> 5 km from urban areas, $> -26.1^{\circ}\text{C}$, and infested in a colder than average year) average annual precipitation greater than 159 cm resulted in a “fast” infestation assignment.

The median quantile estimate of slope for those counties in plant hardiness zone 5B or colder is not significantly different from zero at the 90% confidence level with 31 counties included in the quantile regression. The median quantile estimate for the warmer half of the counties is 15.6 km/yr.

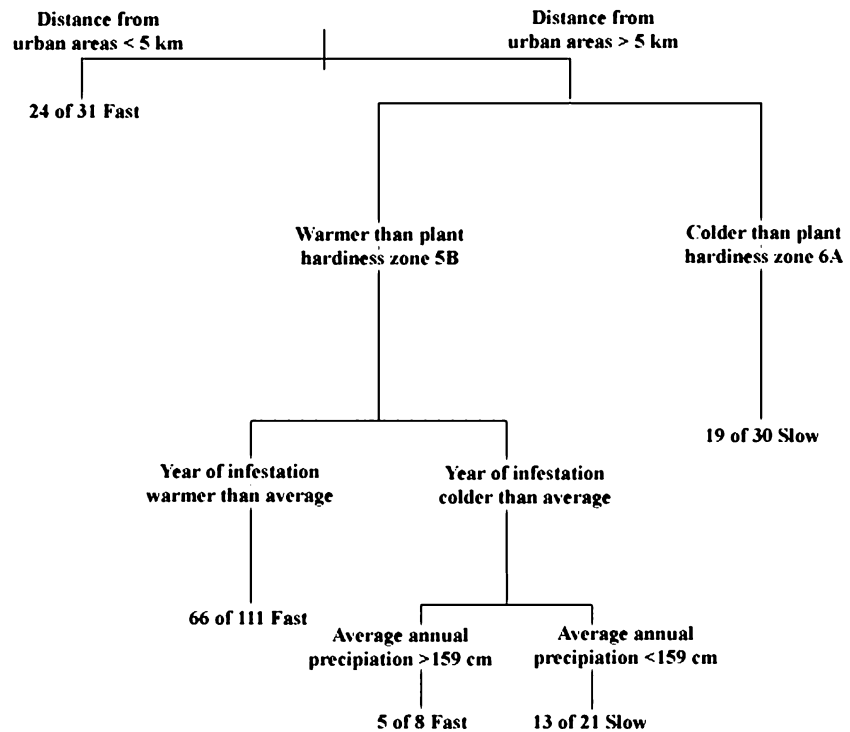
Table 1 Variables with significant influence on 50th quantile of HWA spread

Variable name	Coefficient estimates		
	5th quantile	50th quantile	95th quantile
Years since 1990	7.90**	14.9**	18.8**
Annual precipitation	2.29	4.70**	2.58**
Median elevation	-0.01	-0.08**	-0.14**
Hemlock range	46.2**	86.2**	121**
hardiness zone 5B and colder	55.8**	59.6*	-8.44
zone 5B * years since 1990	-5.45*	-10.4**	-9.30**

**Significant at the 95% confidence level

*Significant at the 90% confidence level

Fig. 4 The classification tree shows how the data is split into counties that were infested without having an infested neighbor (fast) and those counties that remained uninfested after one or more years adjacent to an infested county (slow). The numbers at the each final node refer to the number of counties correctly classified out of the total number of counties assigned to the node



Discussion

Our analysis using four different modeling techniques and two different datasets all suggest that HWA spread varies geographically and is much slower than previously reported. Results from the quantile, Tobit, and regression and classification trees point to slower range expansion in colder areas and during colder years.

The differences in the estimates of rate of spread from the 5th, 50th, and 95th quantiles shown in Fig. 2 demonstrate the heteroscedasticity in the HWA data, and hence importance of modeling methods that account for this. The 95th quantile can be considered the rate of change when the number of years since infestation is the active limiting constraint (Cade et al. 1999; Cade and Noon 2003). This unconstrained spread rate estimated by the 95th quantile (20.8 km/yr) is a good estimate where other factors do not retard spread. The unconstrained spread rate is similar to the 20–30 km/yr estimate from literature (Ward et al. 2004). Choosing the quantile to define the unconstrained spread rate is straightforward in this case because the estimate of the slope

plateaus at about 20 km/yr after the 80th quantile (Fig. 3) (Cade et al. 1999; Cade and Noon 2003). The 95th quantile is the highest quantile with sufficient data for accurate estimation. The estimates for the slope coefficient are significant at the 90% confidence level since none of the confidence bands, shown as shaded regions in Fig. 3, included zero.

The quantile regression estimated rate of spread across the entire HWA range of 12.5 km/yr obscures a faster rate in the south, 15.6 km/yr, and very slow spread in northern areas. The township data is confirmation that spread rates are 8.13 km/yr or less for Pennsylvania and north.

The Tobit regression appears to function best for the section of the data where truncation occurs. Within the period 1990–2000, where obvious truncation below zero affects the data, the estimate from the Tobit regression (13.1 km/yr) fits well with the estimates from the quantile regression. The spread rate estimate from the model with the error estimate conditioned on the sign of the residual is more conservative at 7.35 km/yr. This low estimate is caused by giving great emphasis to short spread distances in order to

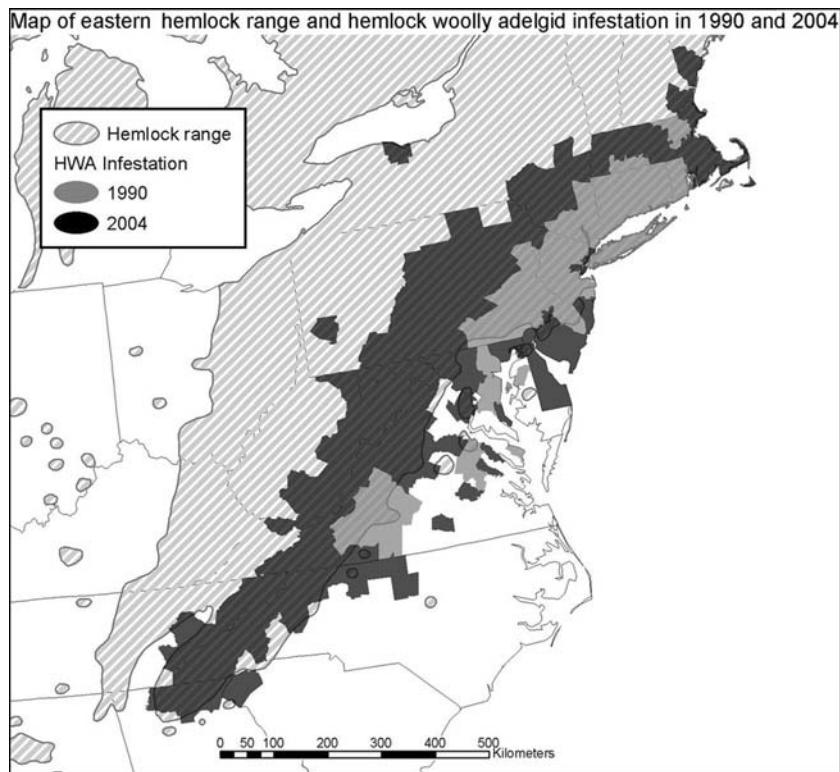


Fig. 5 Map of eastern hemlock range and hemlock woolly adelgid, *A. tsugae*, infestation in 1990 and 2004 from USDA Forest Service records at the county scale

balance the lack of negative distances. This model does not appear to fit the data well. It may be a more appropriate estimate in areas where counties fail to become infested, essentially becoming zero or negative spread, such as the northern frontier of HWA spread. Our estimated rates of HWA range expansion are well within the range of spread rates reported for other invading insects (Grosholz 1996).

The substantial difference between the unconstrained spread rate and the 50th quantile rate reflects the geographic and interannual differences in range expansion. Each of the multivariate techniques uses a slightly different set of variables to estimate rate of spread. Multiple quantile regression employs hemlock range, median elevation, and annual precipitation to predict rates of spread. The regression tree identifies higher elevations and lower average annual daily minimum temperature as variables that reduce spread rates. The classification tree highlights four variables that classify HWA spread as

“fast” or “slow”: distance from urban areas, plant hardiness zones, mean January temperature in the year of spread, and average annual precipitation. A common thread among the techniques is that cold temperatures inhibit spread. The importance of higher elevations, lower average annual daily minimum temperature, colder plant hardiness zones, and lower mean January temperature in the year of spread implies that low temperatures reduce the speed of HWA range expansion. Cold temperatures have been shown to influence HWA survival so it is not surprising that they also inhibit spread (Parker et al. 1998; Cheah and McClure 2002). Cold temperature may be the factor that reduces HWA spread from the unconstrained spread rate, 20.8 km/yr, represented by the 95th quantile. In addition to geographic variation in temperature, interannual fluctuations in mean January temperature also appear to affect rates of spread. Counties were infested slowly if HWA arrived in their vicinity in a relatively cold year.

New research on the home range of HWA adds support to cold temperatures as a constraint on spread. DNA studies indicate that the race of HWA in eastern North America originated in southern Japan on *T. sieboldii* (Havill 2005). The average temperatures within the range of *T. sieboldii* more closely resembles the variation that exists within the southern portion of the eastern hemlock's range, which might explain the why – 26.1°C and mean January temperature divide “slow” from “fast” infestations. In addition, Skinner and colleagues (2003) showed that – 25°C in January or February was a critical temperature that significantly reduced HWA cold hardiness.

In addition to temperature, three other geographic factors influence the rate of HWA spread in our study. The hemlock range variable used in the multiple quantile regression is associated with increases spread rates because the presence of the main host for HWA assists range expansion. Annual precipitation is included in the multiple quantile regression as well as the regression and classification trees because the most recent spread in the southern end of HWA range has occurred quickly, and in an area of high annual precipitation. The increased speed of infestation within 5 km of urban areas supports McClure's (1990) suggestion that humans facilitate the dispersal of HWA. A more complete description of the influence of these and other geographic variables on HWA spread requires higher resolution infestation records and further research.

The four different modeling approaches we used provide multiple views of the HWA invasion. The combination of the quantile regression with the classification and regression trees offers a consistent and informative analysis, which has great potential for the study of the range expansion of other alien species. While either model approach is useful on its own, the combination of the two makes for a stronger examination of the data because their agreement is confirmation of range expansion patterns. The Tobit regression and the heterogeneous variance models appear less useful for the HWA data, but they may be useful in future analyses of invasive species spread. No modeling technique is appropriate for all data and use of multiple approaches is

important, particularly in biological invasion where new ecological interactions can create novel patterns and processes.

Our ability to identify the geographic variability in the spread of HWA demonstrates the utility of coarse, county-scale invasive species monitoring. Each modeling application will have difference performance criteria, which depend on the purpose of the model. In the HWA case, the limitations of the county and township data meant model performance was weaker than would be expected with higher quality monitoring data. Nevertheless, the measures of model performance (misclassification rates, Kappa values, and R^2 values) fall within the range common to ecological studies.

Conclusion

The historic record of HWA infestations provides an opportunity to better understand, and hence combat this invasive insect. Quantile regression, Tobit regression, and classification and regression tree analyses are techniques particularly well suited to the complex range expansion data and supply insight into rates of spread and the geographic factors that moderate these rates. Unconstrained HWA is able to spread at a rate of 20.8 km/yr, while under less favorable conditions 12.5 km/yr is more likely. If HWA continues to spread northward it will do so at a very slow rate, advancing less than 12.5 km even in warm years. The main constraint on HWA spread is temperature, with a mean minimum temperature of – 26.1°C (plant hardiness zone 5B) delineating areas of slower range expansion. The January temperature in the year of spread also reduces spread distance. These new estimates of HWA range expansion fit within the context of previous studies of HWA and other insect spread, as well as with new research into the home range of HWA.

Acknowledgements We are grateful to Jennifer Bofinger, Charlie Burnham, Jason Denham, Don Ouellette, Alan Sior, and Brent Teillon for providing township records of HWA infestation as well as the John F. Enders Fund and the Connecticut Forest and Parks Association for financial support. R. Talbot Trotter III, and Brian Cade provided insightful comments on the manuscript.

References

- Amemiya T (1984) Tobit models: a survey. *J Econom* 24:3–61
- Beatty SW (1984) Influence of microtopography and canopy species on spatial patterns of forest understory plants. *Ecology* 65:1406–1419
- Beers TW, Dress PE, Wensel LC (1966) Aspect transformation in site productivity research. *J For* 64:691–692
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman and Hall/CRC, Boca Raton, FL
- Cade B, Noon B (2003) A gentle introduction to quantile regression for ecologists. *Front Ecol Environ* 1:412–420
- Cade BS, Terrell JW, Schroeder RL (1999) Estimating effects of limiting factors with regression quantiles. *Ecology* 80:311–323
- Cheah CAS-J, McClure MS (2002) Pseudoscymnus tsugae in Connecticut forests: the first five years. In: Onken B, Reardon R, Lashomb J (eds) Hemlock woolly adelgid symposium. USDA Forest Service and Rutgers University, East Brunswick, NJ, pp 150–165
- Daly C, Taylor G (2000) United States average annual precipitation, 1961–90. Spatial Climate Analysis Service at Oregon State University (SCAS/OSU), Corvallis, Oregon
- Daly C, Gibson W, Taylor G (2002) Average daily minimum temperature: annual 1971–2000. 103-year high-resolution temperature climate data set for the conterminous United States. Spatial Climate Analysis Service, Oregon, Corvallis, OR
- DeGraaf RM, Yamasaki M, Leak WB, Lanier JW (1992) New England wildlife: management forested habitats. General Technical Report NE-144, USDA Forest Service, Northeastern Forest Experiment Station, Radnor, PA, p 271
- ESRI (2004) ArcGIS. version 9.1. Environmental Systems Research Institute (ESRI), Redlands CA
- Evans A (2005) A randomized branch sampling method for hemlock woolly adelgid. In: Onken B, Reardon R (eds) Proceedings of the 3rd symposium on hemlock woolly adelgid. USDA Forest Service, Ashville, NC, USA, pp 63–72
- Gannoun A, Saracco J, Yuan A, Bonney G (2005) Non-parametric quantile regression with censored data. *Scand J Stat* 32:527–550
- Godman RM, Lancaster K (1990) Eastern hemlock. In: Burns RM, Honkala BH (eds) *Silvics of North America. Agriculture Handbook 654*, USDA, Forest Service, Washington, DC, p 877
- Gregoire TG, Dyer ME (1989) Model fitting under patterned heterogeneity of variance. *For Sci* 35:105–125
- Grosholz ED (1996) Contrasting rates of spread for introduced species in terrestrial and marine systems. *Ecology* 77:1680–1686
- Havill N (2005) Using mitochondrial DNA to determine the native range of the hemlock woolly adelgid. In: Onken B, Reardon R (eds) Proceedings of the 3rd symposium on hemlock woolly adelgid. USDA Forest Service, Ashville, NC, USA, p 211
- Heger T, Trepl L (2003) Predicting biological invasions. *Biol Invasions* 5:313–321
- Kelty MJ (1989) Productivity of New England hemlock/hardwood stands as affected by species composition and canopy structure. *For Ecol Manage* 28:237–257
- Kittredge DB, Ashton PMS (1995) Impacts of deer browse on regeneration in mixed-species stands in southern New England. *North J Appl For* 12:141–145
- Kizlinski ML, Orwig DA, Cobb RC, Foster DR (2002) Direct and indirect ecosystem consequences of an invasive pest on forests dominated by eastern hemlock. *J Biogeogr* 29:1489–1503
- Koenker R, Hallock KF (2001) Quantile regression. *J Econ Perspect* 15:143–156
- Liebholt A, Luzader E, Halverson J, Elmes G (1992) The spatial dynamics of invasions by exotic forest pests. General Technical Report NE-175, USDA Forest Service, Northeastern Forest Experiment Station, Radnor, PA, pp 125–132
- Liebholt AM, MacDonald WL, Bergdahl D, Mastro VC (1995) Invasion by exotic forest pests: a threat to forest ecosystems. *For Sci Monogr* 30:1–38
- Little EL (1999) Digital representation of the atlas of United States trees. Digital Edition. US Geological Survey, Reston, VA
- Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol Appl* 10:689–710
- McClure MS (1990) Role of winds, birds, deer, and humans in the dispersal of hemlock woolly adelgid (Homoptera: Adelgidae). *Environ Entomol* 19:36–43
- McClure MS (1991) Density-dependent feedback and population cycles in Adelges tsugae (Homoptera: Adelgidae) on Tsuga canadensis. *Environ Entomol* 20:258–264
- Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecol Model* 62:275–293
- National Arboretum (1990) USDA plant hardiness zone map. No. 1475. USDA National Arboretum, Washington, DC
- National Climate Data Center (2005) Area-weighted state, regional, and national monthly and seasonal temperature and precipitation. Asheville, NC
- National Imagery and Mapping Agency (1996) Digital terrain elevation data level 0. National Imagery and Mapping Agency, Fairfax, Virginia
- Onken B, Reardon R (2005) Proceedings of the 3rd symposium on hemlock woolly adelgid. FHTET-2005-01, USDA Forest Service, Ashville, NC, p 363
- Orwig DA, Foster DR, Mausel DL (2002) Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. *J Biogeogr* 29:1475–1488
- Parker BI, Skinner M, Gouli S, Ashikaga T, Teillon HB (1998) Survival of hemlock woolly adelgid (Homoptera: Adelgidae) at low temperatures. *For Sci* 44:414–420

- Pimentel D, Zuniga R, Monison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ* 52:273–288
- Reardon R, Onken B (2004) Biological control of hemlock woolly adelgid in the eastern United States. FHTET-2004-04, USDA, Forest Service, Forest Health Technology Enterprise Team, Morgantown, WV, p 24
- Ross R, Redell L, Bennett R (2002) Mesohabitat use of threatened hemlock forest by breeding birds of the Delaware Water Gap National Recreation Area. In: Onken B, Reardon R, Lashomb J (eds) Hemlock woolly adelgid symposium. USDA Forest Service and Rutgers University, East Brunswick, NJ, pp 353–354
- SAS Institute Inc. (2002) SAS OnlineDoc. version 9. SAS Institute Inc., Cary, NC
- Sharov AA, Liebhold AM (1998) Model of slowing the spread of gypsy moth (Lepidoptera: Lymantriidae) with a barrier zone. *Ecol Appl* 8:1170–1179
- Sharov AA, Pijanowski BC, Liebhold AM, Gage SH (1999) What affects the rate of gypsy moth (Lepidoptera: Lymantriidae) spread: winter temperature or forest susceptibility? *Agric For Entomol* 1:37–45
- Shields K, Cheah CAS-J (2005) Winter mortality in *Adeles tsugae* populations in 2003 and 2004. In: Onken B, Reardon R (eds) Proceedings of the 3rd symposium on hemlock woolly adelgid. USDA Forest Service, Ashville, NC, USA, pp 354–356
- Shigesada N, Kawasaki K, Takeda Y (1995) Modeling stratified diffusion in biological invasions. *Am Nat* 146:229–251
- Simberloff D (2000) Global climate change and introduced species in United States forests. *Sci Total Environ* 262:253–261
- Simberloff D, Von Holle B (1999) Positive interactions of nonindigenous species: invasional meltdown? *Biol Invasions* 1:21–32
- Skellam JG (1951) Random dispersal in theoretical populations. *Biometrika* 38:196–218
- Skinner M, Parker B, Gouli S, Ashikaga T (2003) Regional responses of hemlock woolly adelgid (Homoptera: Adelgidae) to low temperatures. *Environ Entomol* 32:523–528
- Snyder CD, Young JA, Lemarie DP, Smith DR (2002) Influence of eastern hemlock (*Tsuga canadensis*) forests on aquatic invertebrate assemblages in head-water streams. *Can J Fish Aquat Sci* 59:262–275
- Therneau TM, Atkinson EJ (1997) An introduction to recursive partitioning using the RPART routine. Technical Report 61, Mayo Clinic, Section of Statistics, p 52
- Tingley MW, Orwig DA, Field R, Motzkin G (2002) Avian response to removal of a forest dominant: consequences of hemlock woolly adelgid infestations. *J Biogeogr* 29:1505–1516
- US Census Bureau (1990) Census 1990, Summary File 3
- US Geological Survey (2000) Urbanization. National Atlas web edition. US Geological Survey, Reston, VA
- US Geological Survey and US Forest Service (2000) Forest cover types. US Geological Survey and USDA Forest Service, Reston, VA
- Wallace MS, Hain FP (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 JS, Montgomery ME, Cheah CAS-J, Onken BP, Cowles RS (2004) Eastern hemlock forests: guidelines to minimize the impacts of hemlock woolly adelgid. NA-TP-03-04, USDA Forest Service, Northeastern Area State and Private Forestry, Morgantown, WV, p 32
- Waring KM, O'Hara KL (2005) Silvicultural strategies in forest ecosystems affected by introduced pests. *For Ecol Manage* 209:27–41
- White MA, Mladenoff DJ (1994) Old-growth forest landscape transitions from pre-European settlement to present. *Landsc Ecol* 9:191–205
- Williamson MH (1996) Biological invasions. Chapman and Hall, London
- Williamson M, Fitter A (1996) The character of successful invaders. *Biol Conserv* 78:163–170
- Work TT, McCullough DG, Cavey JF, Komsa R (2005) Arrival rate of non-indigenous insect species into the United States through foreign trade. *Biol Invasions* 7:323–332
- Young RF, Shields KS, Berlyn GP (1995) Hemlock woolly adelgid (Homoptera: Adelgidae): stylet bundle insertion and feeding sites. *Ann Entomol Soc Am* 88:827–835