The Effect of Spatial Interpolation on the Hedonic Model: a Case of Forest Pest Damages

Xiaoshu Li
University of Kentucky
Email: xiaoshu.li@uky.edu

Kevin J. Boyle
Virginia Tech
Email: kjboyle@vt.edu

Evan L. Preisser
University of Rhode Island
Email: preisser@ uri.edu

Thomas P. Holmes
USDA Forest Service
Email: tholmes@fs.fed.us

Klaus Moeltner
Virginia Tech
Email: moeltner@vt.edu

David Orwig
Harvard Forest
Email: orwig@fas.harvard.edu

Abstract:

In the hedonic model, when an investigator wishes to merge property sale data with spatial data on an environmental amenity, one problem encountered in the matching process is that the environmental data are usually limited. A challenging task is how to scale site and time specific environmental data to all property sales within a defined geographical area. Spatial interpolation methods provide us a tool to scale up our economic analysis from specific sites to broad geographical area. In this study, we investigate the effects of these spatial interpolation methods on the estimation of a hedonic model in the context of an invasive forest pest, the hemlock wooly adelgid. Our results indicate a statistically significant relationship between hemlock health and residential property values at the 0.1 km level. Repeat-sale model provides us more robust economic estimation results than tradition hedonic model across different interpolation methods.

Key words:

forest pest infestation, spatial interpolation, hedonic model

1. Introduction

Property value modeling is a commonly employed method to value changes in environmental assets based on the idea that the prices of properties represents the sum of values associated with property attributes (Palmquist 1991). Recent years have seen important developments and refinements in the property value method, which include natural experiments to identify price effects, access to large-scale electronic data on property sale prices and characteristics, and the use of GIS (geographic information system) data to incorporate spatial dimensions of property attributes (Calhoun 2001, Parmeter and Pope 2009, Paterson and Boyle 2002).

When an investigator wishes to merge property sale data with spatial data on an environmental amenity, one problem encountered in the matching process is that the environmental data may be limited (Palmquist and Smith 2001). Although environmental data varies across broad geographical area, these data are usually measured at certain sample sites. For example, if one is interested in how air quality affects property prices, the property data may include all sales in a given area for a specified period of time while the air-quality data is collected at fixed monitoring sites[1]. The air quality data is measured discrete over space, but small increments in time, e.g., hourly. For other environmental media, in addition to being spatially discrete, the data may only be reported for specific points in time and the timing of measurements may vary across sites, e.g., volunteer lake water quality monitoring[2]. A challenging task is how to scale or extrapolate site and time specific environmental data to all property sales within a defined geographical area.

[1] http://www.epa.gov/airquality/airdata/, accessed December 12, 2013.
[2] http://water.epa.gov/type/watersheds/monitoring/vol.cfm, accessed December 12, 2013

The method commonly conducted to enlarge the spatial analysis scale under limited information is benefit transfer method (Troy and Wilson 2006, plummer 2009, Nelson et al. 2009, Richardson et al. 2014). It conducts the property value model just using limited environmental information, and then applies the model estimates to new study areas under the assumption that there exist similar impacts over geographical area. Or instead of directly transferring the estimates, they would consider to make adjustments based on the socioeconomic and environmental conditions (Loomis 1992, Shrestha and Loomis 2003). However, even if the environmental conditions of those specific study sites are representative of the overall geographical area, the property sales data at these sampled sites may not be representative of all the property values over space. When we only consider the impact at these specific sites, we would lose the information from property sales data over large spatial area. Then it would introduce prediction errors by applying the benefit transfer methods to enlarge the spatial analysis scale.

Another method we could consider is to first employ geostatistical methods to interpolate the environmental conditions from specific sites to larger geographical study area. And then we could utilize all the property sale data and match them with the interpolated environmental condition to conduct the economic impact analysis. However, there may exist statistical interpolation errors at the first step because of the environmental distribution uncertainty over space. These interpolation errors would be introduced into property value model as measurement errors and also influence the prediction accuracy of economic impacts.

Here we investigate the effect of spatial data extrapolation on the estimation of property sale model in the context of an invasive forest pest, the hemlock wooly adelgid

(Brush, 1979, McClure 1991, Orwig et al., 2002). Here the data on the infestation are measured at discrete spatial locations and points in time like for the water and air quality data, but the host (hemlock trees) does not provide for continuous dispersal through space. Hemlock trees are clustered in discrete stands and, thus, the infestation must jump from stand to stand as the infestation spreads spatially through time. There could exist larger spatial distribution uncertainty over space than water and air quality data.

We employ the spatial interpolation methods to predict the hemlock health conditions from sampled stands to all the hemlock stands over study area. We investigate the use of three spatial data interpolation methods: inverse distance weighting (IDW), splines, and Kriging. We also check the spatial interpolation errors from spatial distribution uncertainty and make adjustments of the interpolation errors based on cross-validation results. We check the robustness of estimation results from both a hedonic property value model and a repeat-sale model and investigate the effects of spatial data interpolation on the property sale models.

Our results indicate that the invasion of HWA has caused dramatic losses of healthy hemlock stands in the study area. Both the hedonic model and repeated sale model show that there exist substantial accompanying losses in property value for the households located nearby. Spatial interpolation methods provides us useful tools to scale up our economic analysis. The repeat sale model gives us robust estimation results across different spatial interpolation methodologies.

2. Literature Review

Hedonic models are employed to investigate environmental issues based on spatial proximity to environmental amenities and dis-amenities. Some studies just match property data directly with data from environmental database, e.g. water quality studies. Boyle et al. (1999) measured the water clarity of 25 lakes in Maine and found that the water quality would significantly impact the residential property value located nearby. Poor et al. (2007) acquired the water quality data from twenty-two water-monitoring stations located throughout the watershed of the St. Mary's River, Maryland. And they assigned the water quality measures to each sale property from the closest monitoring station.

Other studies use ArcGIS tools to match property data with spatially explicit information on environmental amenities and dis-amenities. e.g. proximity to open space. Geoghegan et al. (1997) calculated measures of percent open space around household properties. They found that the land uses surrounding a parcel have an influence on the price. Cho et al. (2008) calculated the distance to nearest evergreen, deciduous and mixed forest patches for each household. Proximities to evergreen forest are valued positively in the rural–urban interfaces, while the proximities to deciduous forest and mixed forest are valued positively in the urban area.

These spatial interpolation methods has not commonly employed in economic research except for the evaluation analysis of air pollution or water quality. The reason is that the air pollution or water quality are only measured at several monitoring stations, while the households are located across the space. Both air and water quality, within a specific area, such as an air quality basin or coastal bay might be consider ubiquitous, in that pollutants disperse out from emission sites and perhaps experience some rate of decay over distance and/or time. Previous researches employed spatial statistical method to interpolate the environmental condition measurements across study area and then match them with the household sales value based on their locations (Leggett and Bockstael 2000, Kim et al. 2003, Beron et al. 2004, Anselin and Lozano-Gracia 2008, Fernandez-Aviles et al. 2012). Ara et al. (2006) interpolated beach water quality data over space and time. The water quality along 18 beaches was originally measured at different place and at different time. They also employed geo-statistical method to interpolate the water quality and matched the household properties with the interpolated water quality of the nearest beach.

However, there are seldom studies which investigate how spatial interpolation will affect our hedonic economic analysis when we use prediction results rather than the true measurements in the property vale models. Anselin and Le Gallo (2006) have compared different spatial interpolation techniques (Thiessen polygons, inverse distance weighting, Kriging and splines) when their prediction results are used as the measurement of air quality in the hedonic models. Their results showed that Kriging provides the best results for interpretation.

In our case, we try to investigate the economic losses from hemlock mortality caused by HWA infestation. Holmes et al. (2010a) found that severely-defoliated hemlocks in northern New Jersey reduced the values of residential parcels with the stricken trees and reduced the value of nearby (up to 0.5km) properties. In our study, we employ the interpolation approach (inverse distance weighting, Kriging and splines) to scale up the analysis from specific area (e.g., county) to state level. We also investigate how different spatial interpolation methodologies would affect the hedonic model and repeated sale model analysis.

As the hemlock stands are spread discrete over space, and the outbreak of HWA infestation may not always happen continuous through space. There are much more variation of the sampled HWA damage data compared to the measurement of air pollution. Different with Anselin and Le Gallo (2006), we also employed cross-validation results to make adjustments of kriging interpolation methods. Because of our data structure, we can also compare the property value analysis results based only on sampled hemlock damage data (true measurements) with the results based on interpolated data (prediction results). And here we will check how the different spatial interpolation methods would affect the inference from property value models in this data structure.

3. Application

In the research here, our application is to estimate the economic consequences of the spatial and temporal expansion of HWA through central Connecticut and Massachusetts. The HWA was first introduced into Virginia from Japan in the early 1950s; in the past half-century, it has spread to hemlock forests along the east coast of the U.S and became a threat to the eastern hemlock forests of New England (McClure 1991). The population growth of HWA is sensitive to temperature and precipitation, and climate change is expected to favor the spread of HWA (Orwig et al., 2002).

It is thought that changes in climate may increase the frequency and severity of forest fires, insect and disease outbreaks, droughts and storms that can affect trees (Dale et al., 2001; Bentz, 2008; Frankel, 2008). The subsequent losses of trees can impact property values via reductions in shading to reduce heat impacts, reductions in the scenic aesthetics of an area and other consequences. Dying and dead trees can pose risks to residents and their homes. The hemlock wooly adelgid, as an example case study, causes death of affected trees within about five years. Forward-looking communities can adapt their tree planting and protection efforts to lessen these climate-induced impacts and information on the economic value of tree canopy cover can be used to help justify such efforts. Thus, exploring the impacts of data extrapolation to investigate the economic effects of forest impacts is important to support forest and climate policy.

In an effort to understand and characterize hemlock stands at the local and landscape levels in New England, ecologists at the Harvard Forest identified, mapped, and characterized hemlock stands within a 7,500 km^2 area covers central portions of Connecticut and Massachusetts (see Figure 1). In both states, all stands of eastern hemlock >1.3 ha in area were identified using high-resolution aerial photographs that were then scanned and digitally transferred into a GIS overlay. A total of 6,126 hemlock stands were identified in the study area using this method.

The field surveys were conducted at 142 hemlock stands (red dots in Figure 4.2), and hemlock health characteristics were documented in these sampled stands. The sampled stands were distributed as evenly as possible over the study area. Hemlock vigor and live basal area are two key hemlock damage characteristics recorded in 2007, 2009, and 2011. Live basal area (m^2/ha) is the area of a given section of land that is occupied by the cross-section of tree trunks and stems at their basal. Vigor was measured on the basis of the amount of retained foliage in each stand. There are four vigor categories; 1 = 76 – 99% foliar loss, 2= 51-75% foliar loss, 3 = 26-50% foliar loss and 4 = 0 – 25% foliar loss.

Table 1 shows the summary descriptive statistics of live basal area and frequency distribution of hemlock vigor for the sampled hemlock stands. Both the mean and maximum value of hemlock live basal area decreased between 2007 and 2011. Except for the severely damaged hemlock stands (vigor=1), the number of damaged hemlock stands (vigor=2 or vigor=3) is increasing through time while the number of healthy hemlock stands (vigor=4) is decreasing. It is likely that the number of severely damaged hemlock stands dropped in 2011 because dead trees either fell over or were removed, and were no longer included in the survey.

To investigate the effect of HWA infestation on residential property, we would need to match the environmental attributes with household sales. In this study we assume only when the hemlock stands locate within certain distance of households, they will affect the property value. We select three groups of household properties which are located around the hemlock stands within a distance of 0.1km, 0.5km or 1km.Then, there are four types of locations between household properties and hemlock stands (Figure 2). In case A, the buffer only intersects with sampled hemlock stands. In case B, the buffer intersects with both the sampled and non-sampled hemlock stands. In case C, the buffer intersects only with the non-sampled hemlock stands; we do not have the hemlock damage information. In case D, the buffer does not intersect with any hemlock stands; these properties are assumed to not be affected by the HWA infestation.

Based on the sampled data, we only have the full hemlock damage information for case A. However, as the number of household properties which are only influenced by sampled hemlock stands is small, it may not be representative of the whole study area. To make the maximum use of household sale data, we would want to interpolate the environmental damage to the whole study area and match them with the sales data. In our case, using interpolated hemlock damage data, we would be able to enlarge our economic analysis to all the household sales which are influenced by HWA infestation (case A, B and C in Figure 2).

DataQuick provides the dataset which contains all the house attributes and sales price during the study years (2007-2011). Lot size, living area, number of bath rooms, number of bed rooms, house age and distance to highway are included as housing characteristics. We also introduce dummy variable to indicate whether the house has air conditioning and fireplace. In Appendix Table A, we list the summary statistics of housing characteristic variables for properties both located nearby the sampled hemlock stands and all the hemlock stands (<0.1km). Comparing the summary statistics, we can see most of the housing characteristics are similar between the two groups except that the households near the sampled stands are also located near to highway and developed area. These differences could introduce different economic impacts from the HWA damages.

Land cover in the neighborhood can also influence the household property value (Irwin 2002; Patterson and Boyle 2002). We constructed land cover variables based on National Land Cover Data (2006) using raster of $30m^2$ pixels. The six different types of land cover variables constructed are water, open space, high developed district, forest, agricultural land and wetland. They are calculated as the percentage of the buffer area around household property which is covered by each land type, while the size of the buffer is respectively 0.1km, 0.5km, or 1km corresponding to each group.

4. Spatial Interpolation

4.1 Interpolation Methods

We first employ the geo-statistic interpolation methodologies to scale up our hedonic analysis and predict the spatial distribution of forest damage data; they are based on the assumption that the values should be more similar when the points are near to each other. Inverse distance weighting (IDW), Spline and kriging are three interpolation methodologies commonly applied in the forestry studies; they are readily available in the geo-statistical wizard of Geostatistical Analyst Tool in ARCGIS 10.1.

Inverse distance weighting (IDW) assigns values of hemlock damages to unknown points with a weighted average of the values observed at the locations in the neighbor. It gives greater weights to points closest to the prediction location, and the weights diminish as a function of distance (Shepard 1968). An interpolated value u at non-sampled point x using IDW is:

$$u(x) = \frac{\sum_i w_i(x) u_i}{\sum_i w_i(x)} \tag{1}$$

Here u_i denotes the hemlock damages at sampled point x_i, $i = 0,1, \cdots N$. $w_i = \frac{1}{{d_i}^p}$ while d_i is a given distance from the known point x_i to the unknown point x. p is the power parameter which is set equal to 1 here.

The Spline tool estimates values of non-sampled points using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points (Franke 1982; Mitas and Mitasova 1988). The kernel parameter and kernel function employed will determine the extent to which any given point influences the fitted surface and specify the smoothness of the surface. They are calculated by minimizing the root mean square error during cross validation.

Kriging predicts the value at a non-sampled location using the weighted average of the known values of its neighbors while the weights are determined by a semivariogram model of spatial autocorrelation (Isaaks and Srivastava 1989; Cressie 1993; Goovaerts 1997; Schabenberger and Gotway 2005). The semivariance (Stein 1999) is

$$\rho(h) = \frac{1}{2n} \sum_{i=1}^{n} (u(x_i) - u(x_i + h))^2 \quad (2)$$

where x_i represents any sampled hemlock stand, $(x_i + h)$ is a sampled stand distance h from x_i, $u(\cdot)$ is the observational value at the sampled location. Based on the semivariance, the parameters of function form, range, nugget effect and sill are estimated which define the empirical semivariogram model (Cressie 1985; Chilès and Delfiner 1999).

Here the spatial correlation is assumed as isotropic over the study area, i.e. that the spatial correlation is only depending on the distance between two points but not the direction of their separation. As the infestation of HWA has the potential to move from southwest to northeast, we make trend analysis at first. A second-order polynomial trend is removed at first if there exists significant trend over space, while the kriging analysis is performed on the residuals. Ordinary kriging interpolation methods are applied.

4.2 Kriging Adjustment Methods

Cross validation is used to evaluate which model provides the best predictions. Cross validation is performed by removing one point from data set and then predicting its value using the data at the rest locations. Comparing the predicted value to the observed value across all the points, we can obtain the measurements for evaluation. The cross validation results for live basal area based on Kriging interpolation method is shown in Figure 3[3]. For a good fit, the points in the cross validation results (Figure 3) should scattered around 45 degree line. However, these spatial interpolation estimates present a serious drawback well known by geo-statisticians as the smoothing effect. This prediction error may exacerbate the computation of HWA damages in terms of maximum live basal area. Similarly for vigor, with the small frequency of observing severely damaged stands, there are relatively large prediction errors for stands with low vigor. In our study, from year 2007 to year 2011, with the increased damages from HWA, these smoothing effects of live basal area increase over years.

Approaches have been investigated to make corrections of the smoothing effects from ordinary Kriging based on cross validation results. [4]Olea and Pawlowsky (1996) have proposed a procedure called compensated kriging, which is carried out in four steps. (1) Make an estimation from ordinary Kriging to get the best fitted models. (2) Use the cross-validation results to run a regression of the predicted value on the true values. $Z^*(x)$ is the predicted value and $Z(x)$ is the true value. The regression provides the values of the slope a and the intercept b. The regression equation would be $Z^*(x) = aZ(x) + b$; (3) Employ Equation (3) to make adjustment of the predicted value $Z^*(x)$

[3] The cross validation results from Spline and IDW are similar with Kriging.
[4] Correction approaches are conducted based on ordinary kriging interpolation method as it not only gives expected mean but also variance at each location.

$$Z_c^*(x) = Z^*(x) - \frac{\varepsilon^2(x)}{\epsilon_{max}^2}(Z^*(x) - \frac{Z^*(x)-b}{a}) \quad (3)$$

$\varepsilon^2(x)$ is the kriging estimation variance and ϵ_{max}^2 is the maximum value of the kriging estimation variance. $Z_c^*(x)$ is the adjusted value.

(4) Examine the extreme values for $Z_c^*(x)$ and compare them with the extreme values of the sample, in case that straight compensation may overdo the job.[5]

Another approach is introduced by Yamamoto (2005). It also included four steps. (1) Run the cross-validation procedure in order to achieve for each data point the interpolation standard deviation ε and the true error. These variables are then transformed into another variable named number of interpolation standard deviations as follows: $N_{S_0} = \frac{-\text{TrueError}}{\varepsilon}$; (2) Run the ordinary kriging procedure to determine the number of interpolation standard deviations (N_{S_0}) at all locations to be corrected. (3). Run the ordinary kriging procedure to make estimates $Z^*(x)$ and the interpolation standard deviation (ε) at all locations; (4) Correct the ordinary kriging estimates as follows:

$$Z_c^*(x) = Z^*(x) + N_{S_0}(x) * \varepsilon \quad (4)$$

The correction provided by expression (4) must be checked if the corrected value is within the data range of the subset of neighboring points. [6]

In our study, we employed both correction methods to adjust the Kriging interpolation results and check how they will influence the estimation of economic models.

[5] For a detailed introduction of method, see Olea and Pawlowsky (1996).
[6] For a detailed introduction of method, see Yamamoto (1996).

5. Model Specification

Live basal area and vigor are two key hemlock damage variables that have the potential to induce economic losses for the residential property nearby. Based on the interpolation methodologies described above (IDW, Spline and Kriging), we interpolated both hemlock live basal area and vigor separately for each year over the whole study area. Both the live basal area and vigor are interpolated directly as continuous variables. For the kriging method, we also make adjustments of the predicted value based on Olea and Pawlowsky (1996) and Yamamoto (2005).

The value of live basal area and vigor for the 6,126 hemlock stands are then extracted from the interpolated space based on 30m $\times$30m grid. They are calculated as the mean of interpolated value for covered spatial area. Then for each household property, the measurements for live basal area and vigor are calculated by the mean of the hemlock stands which intersects with the buffers (0.1km, 0.5km, or 1km) around the household properties. For 2008 and 2010, they are calculated as the mean value of the previous year and following year. Live basal area (lba_{it}) and the interaction between live basal area and vigor ($lba_{it} * vigor_{it}$) are included in the model as the measurements for environmental attributes.

For the traditional hedonic model, the household property value would be affected by different attributes, house-specific characteristics Z_i, land cover characteristics L_i, and environmental characteristics E_{it} (lba_{it} and $lba_{it} * vigor_{it}$). The fixed effect panel models are commonly used to handle the problem with other spatially correlated omitted variables as Equation (5).

$$\ln P_{it} = Z_i\alpha + +L_i\beta + lba_{it}\gamma + lba_{it} * vigor_{it}\theta + \tau_t + \omega_j + \varepsilon_{it} \quad (5)$$

Here P_{it} is the sale price for property i at time t in the semi-logarithmic function form. τ_t is the time effect, and ω_j is the spatial effect. We only used the most recent sale for each household property in the estimation equation to avoid autocorrelation between observations for the same property.

The house-specific characteristics Z_i include lot size, living area, number of bath rooms, number of bed rooms, house age, air conditioning, fireplace and distance to highway. The land cover characteristics L_i include the percentage of buffer area covered by water, open space, high developed district, forest, agricultural land and wetland. The time effect is set as dummy variables from 2007 to 2011, while the spatial fixed effect is set based on zip code[7].

According to the hedonic estimation result in equation (5), the marginal effect of live basal area on property value can be calculated as in equation (6), which depends on the value of hemlock vigor.

$$\frac{\partial lnp}{\partial lba} = \gamma + vigor_{it}\theta \quad (6)$$

One concern with the traditional hedonic model is that the property value is affected by lots of characteristics. When the missing attributes in the error term are correlated with the attribute variables in the model, the coefficients could be biased. One way to deal with this problem is to conduct a quasi-experiment design using repeated sale model. The repeated sale method starts with the assumption that the change of price value

[7] The number of the spatial fixed effect variables varies with the changes of buffer sizes.

is only introduced by the change of household characteristics. Then by taking difference between two sales from one property, the effects from fixed attributes would be cancelled out, which are less susceptible to omitted variable bias (Kuminoff et al. 2010). Then the model turns to be

$$\ln P_{it} - \ln P_{it-1} = (lba_{it} - lba_{it-1})\gamma + (lba_{it} * vigor_{it} - lba_{it-1} * vigor_{it-1})\theta + (\tau_t - \tau_{t-1}) + \omega_j + \varepsilon_{it} \quad (7)$$

Here ω_j are employed to capture the spatial effect. τ_t is the year dummy variable of the most recent sale, while τ_{t-1} is the year dummy variable of the previous sale.

6. Results

6.1 Spatial Interpolation

Live basal area and vigor are separately interpolated over space for the sample years of 2007, 2009 and 2011. The interpolated live basal area over the study area for each year based on kriging is shown in Figure 4. Through time we see the live basal area has declined. In year 2007, the mean of interpolated live basal area is about 40 m^2/ha. In year 2011, the mean of interpolated live basal area over all the study area is about 15 m^2/ha.

The interpolated probability of healthy hemlock stands (2007-2011) based on kriging is shown in Figure 5. The change of the hemlock vigor follows the same pattern as live basal area. Overall, the data reveals the HWA infestation in 2007 was primarily in southern Connecticut and the hemlocks in Massachusetts were generally healthy. By 2011, a period of four years, the HWA infestation had spread substantially into Massachusetts.

The summary descriptive statistics of interpolated live basal area, vigor and their interaction terms in the hedonic model (<0.1km) are shown in Appendix Table B. Comparing three interpolation methods, we can see the descriptive statistics are similar across three interpolation methods. The predicted standard deviations turn to be relatively smaller than the sampled data. After the adjustment based on the cross validation results, the descriptive statistics of live basal area, vigor and their interaction terms got changed and the predicted standard deviation got larger.

6.2 Basic Estimation Results

We estimated the economic impact from HWA only based on sampled data, including 0.1km, 0.5km, and 1km buffer (Table 2). Because of the small sample problem, we could only estimate the traditional hedonic model. From the results, we can see that 0.1 km buffer shows out the significance at 10% level for both the coefficients of live basal area and its interaction with vigor. If vigor=1 (seriously damaged hemlock stand), when the live basal area increases by 1 m^2/ha the house price will decrease by 0.38%. If vigor=4 (healthy hemlock stand), when the live basal area increases by 1 m^2/ha the house price will increase by 0.23% (Table 5). The coefficients of live basal area and its interaction with vigor are not significant in 0.5km buffer. Although they are significant in 1km buffer, their signs are objective with the intuition.

We first employed the Kriging interpolation methods to predict the hemlock damage variable for all the hemlock stands. As the spatial interpolation methods enlarged our observation sample set, we can employ both the hedonic model and the repeated sale model. Although the results for hedonic model are significant at all three buffer size, only 0.1 km buffer shows the significant results in the repeat sale model (Table 2). Compared with the repeat sale estimation results, the significances in the hedonic model with buffer 0.5km and 1km could be caused by the correlation with the missing variable in the error term.

The magnitudes of the coefficients are decreasing with the increase of the buffer size. The magnitudes of the coefficients based on hedonic model are similar with the sampled stands, while they are much larger in the repeat sale model for 0.1km buffer. Based on the repeat sale model of 0.1km buffer, if vigor=1 (seriously damaged hemlock stand), when the live basal area increases by 1 m^2/ha the house price will decrease by 4.32 %. If vigor=4 (healthy hemlock stand), when the live basal area increases by 1 m^2/ha the house price will increase by 1.13% (Table 5).

To check the reason for the differences between hedonic model and repeat sale model results, we estimate the hedonic model based only on the repeat sale sample (Table 2). The hedonic models with repeat sale sample show significance at both 0.1km buffer and 0.5 km buffer. The differences between hedonic model and repeat sale model are not just introduced by different sample. Different model specification is also a major factor.

6.3 Interpolation Robustness Analysis

We also employed Spline and IDW methods to make interpolations and compared their results with the Kriging (Table 3). In the hedonic model (0.1 km buffer), the coefficients of the interaction term is significant at 5% level for IDW methods, while none is significant for the spline interpolation methods. In the repeat sale model (0.1 km buffer), the coefficients of live basal area and its interaction are both significant. The magnitude of the coefficients are similar between IDW and Spline method, however they are smaller compared with the Kriging interpolation method.

We have also checked the robustness of the estimation results after we conducted the adjustment of kriging interpolation methods based on Olea & Pawlowsky (1996) and Yamamoto (2005). Based on the adjustment method of Olea & Pawlowsky (1996), the coefficients of live basal area and its interaction in the hedonic model are still significant at 5% level. Based on the adjustment method of Yamamoto (2005), the coefficients are no longer significant. However, in the repeat sale model, both adjustment methods show significant estimation results at 1% level. And their magnitudes are relatively similar and even smaller with the IDW and Spline interpolation methods.

We estimate the marginal impact of live basal area based on the adjustment method of Yamamoto (2005). If vigor=1 (seriously damaged hemlock stand), when the live basal area increases by 1 m^2/ha the house price will decrease by 2.24%. If vigor=4 (healthy hemlock stand), when the live basal area increases by 1 m^2/ha the house price will increase by 1.15% (Table 5).

7. Conclusion and Discussion

The results of this study indicate that HWA has caused dramatic damages to hemlock stands in central Connecticut and central Massachusetts during the period 2007-2011. This landscape change causes the decrease of the sales price for properties residing in the study area. The repeat sale model give relatively consistent estimation results that the hemlock damage caused by HWA infestation will decrease the value of residential properties which locate inside 0.1km buffer area.

Because of the small sample size, we can see that the inference based only on the sampled stands may lose property sale information. And we could not correctly estimate the effect of hemlock damage using repeat sale model. Spatial interpolation methods provide us useful tools to enlarge the scale of our analysis and lead to consistent inference. After spatial interpolation (Kriging, Spline and IDW), we could utilize quasi-experimental design method to avoid the missing variable problem. The results show the robustness in the repeat sale model.

However, there will exist measurement errors when we used the prediction results from spatial interpolation rather than the true value in hedonic model analysis. We attempt to adjust the smoothing effects of Kriging by two approaches proposed by Olea & Pawlowsky (1996) and Yamamoto (2005). These adjustments are supposed to improve the accuracy of HWA damage variables in the hedonic model estimation. These adjustments changed the magnitudes of the coefficients while the significance levels keep the same for the repeat sale model. Adjustments reduce repeat sale effects and bring it close to IDW and spline. These also imply the repeat sale provide the robust estimation results compared with the tradition hedonic model.

The implicit price changed significantly from hedonic model to repeat sale model. Moving from sampled to interpolated estimation for hedonic model has minimum effects. Moving from sampled hedonic model to interpolated estimation for repeat sale model has big effects. This is due to repeat sales rather than the sample of repeat sales.

In our analysis, the live basal area and vigor were measured at specific locations which did not match with the property sale data. This spatial misalignment would result in similar problems as a set of environmental epidemiology studies (Madsen et al. 2008, Gryparis et al. 2009, Szpiro et al. 2010). The measurement error of hemlock health variables from spatial interpolation might cause the estimates of their coefficients and the standard errors in the hedonic model to be biased. Several studies have investigated this spatial misalignment issue and proposed different approaches to adjust the estimation bias. Madsen et al. (2008) suggested adjusted krige and regress approach which employed a generalized least square estimator to correct the correlated error structure. Gryparis et al. (2009) pointed out the Berkson-type error from spatial smoothing. They compared different estimation methods and suggested that both the out-of-sample regression calibration and Bayesian models provided good performance. Szpiro et al. (2008, 2010) proposed three different bootstrap approaches that account for both the classic prediction error and the Berkson error.

However, in all the studies proposed above there is only one predicted covariate which would bring extra uncertainty in the model estimation. In our hedonic model, we included both predicted covariates of live basal area and its interaction with vigor in the estimation model, and both hemlock health variables are interpolated separately for each year. These data structure made the covariance structure more complex to make adjustment of the additional spatial correlation. It is hard to check how the predicted covariates of hemlock health variable will impact the estimation results. Besides, the previous mentioned studies addressed the point to point spatial misalignment issue while our study is under point to area spatial misalignment settings. The household property values are assumed to be associated with the hemlock health of the nearby forest stands which are the average of the spatial interpolated value over the stand area. In the point to area settings, Lopiano et al. (2010) and Young et al. (2012) compared the estimation results from directly using the predicted covariates with the adjustment methods proposed by Madsen et al. (2008) and Szpiro et al. (2009). They found that the coefficients of the interpolated covariates are expected to be unbiased in the point to area misalignment setting. And the standard errors of the covariate coefficients using the basic methods performs well in certain cases, although it is not clear yet whether this would be the general case. Besides, the trend of the spatial interpolation over study area could also impact the biasness and standard errors of the covariate coefficients (Young et al. 2012). The coefficients of the predicted covariates would not change dramatically, while the variability of the coefficients will increase. Therefore, in this paper we just employed the interpolated value without further adjustment of the complex covariance error structure. Future researches will need to show its performance.

Our results show that the hemlock damages will decrease the property value about 3% in 0.1km buffer. Holmes et al. (2010b) estimated that hemlock defoliation and mortality resulted in a 1-1.6% decrease in residential property values of parcels that had hemlocks on the property. The estimated marginal effect for vigor is relatively large here compared with Holmes et al. (2010b). One reason could be that we employed different formation of hemlock damage measurements here. As the interaction of vigor and live basal area is introduced into the model, their effects on the property price are interdependent. Another reason could be that the property markets in Connecticut and Massachusetts were still in adjustments to the hemlock damage while the property market of New Jersey was in different market equilibrium.

According to the impact of HWA on residential property values, the aggregate economic losses have likely rapidly accelerated in the study area during the past several years. It can cause even larger damages when the infestation of HWA moves further into northern area where there are more hemlock forests. Although forest management tools are not currently available to either slow the spread or to protect naturally regenerated hemlock forests from HWA, the economic benefits of developing such tools could be substantial. Slowing the advance of HWA into residential forests could convey substantial benefits to homeowners, and may substantially exceed the cost of such programs. Protecting or delaying the onset of HWA in such areas may be a smart investment of public and private funds.

References

Anselin, L. and Le Gallo, J. 2006. Interpolation of air quality measures in hedonic house price models: spatial aspects. Spatial Economic Analysis 1:32-52.

Anselin, L. and Lozano-Gracia, N. 2008. Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics 34:5-34.

Ara, S., Irwin, E., and Haab, T. 2006. The influence of water quality on the housing prices around Lake Erie. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California, July 23-26.

Bentz, B. 2008. Western U.S. Bark Beetles and Climate Change. (May 20, 2008). U.S. Department of Agriculture, Forest Service, Climate Change Resource Center. http://www.fs.fed.us/ccrc/topics/bark-beetles.shtml

Beron, K.J., Murdoch, J.C., and Thayer, M.A. 2001. The benefits of visibility improvement: new evidence from the Los Angeles metropolitan area. Jounal of Real Estate Finance and Economics 22:319-337.

Biondi, F., Myers, D.E., and Avery, C.C. 1994. Geostatistically modeling stem size and increment in an old-growth forest. Canadian Journal of Forest Research 24:1354-1368.

Brantley, S., Ford, C.R., and Vose, J.M. 2013. Future species composition will affect forest water use after loss of eastern hemlock from southern Appalachian forests. Ecological Applications, 23:777-790.

Brush, R.O. 1979. The attractiveness of woodlands: perception of forest landowners in Massachusetts. Forest Science 25:495–506.

Chiles, J.P., and Delfiner, P. 1999. Geostatistics: Modeling Spatial uncertainty. Wiley, New York.

Cressie, N. 1993. Statistics for spatial data. Wiley, New York.

Cressie, N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17:563-586.

Dale, V.H., Joyce, L.A., McNulty, S., Neilson, R.P., Ayres, M.P., Flannigan, M.D., Hanson, P.J., Irland, L.C., Lugo, A.E., Peterson, C.J., Simberloff, D., Swanson, F.J., Stocks, B.J., and Wotton, B.M. 2001. Climate change and forest disturbances. BioScience 51(9): 723-734.

Earnhart, D. 2006. Using contingent-pricing analysis to value open space and its duration at residential locations. Land Economics 82:17-35

Fausold, C. J., and Lilieholm, R.J. 1999. The economic value of open space: A review and synthesis. Environmental Management 23:307-320.

Fernandez-Aviles, G., Minguez, R., and Montero, J. 2012. Geostatistical air pollution indices and spatial hedonic and models: the case of Madrid Spain. Journal of real estate research 34:243-274.

Figueiredo-Filho, A., Borders, B., and Hitch, K.L. 1996. Number of diameters required to represent stem profiles using interpolated cubic splines. Canadian Journal of Forest Research 26:1113–1121.

Franke, R. 1982. Smooth interpolation of scattered data by local thin plate splines. Computer and Mathematics with Applications 8: 273–281.

Frankel, S. J. 2008. Forest Plant Diseases and Climate Change. (May 20, 2008). U.S. Department of Agriculture, Forest Service, Climate Change Resource Center (http://www.fs.fed.us/ccrc/topics/plant-diseases.shtml).

Freeman, E. A., and Moisen, G.G. 2007. Evaluating Kriging as a tool to improve moderate resolution maps of forest biomass. Environmental monitoring and assessment 128:395-410.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation, Oxford University Press, New York. 483 pp.

Gunnarsson, F., Holm, S., Holmgren, P., and Thuresson, T. 1998. On the potential of Kriging for forest management planning. Scandinavian Journal of Forest Research 13:237-245.

Holmes, T.P., Murphy, E.A., Bell, K., and Royle, D. 2010a. Property value impacts of hemlock woolly adelgid in residential forests. Forest Science 56:529-540.

Holmes, T.P., Liebhold, A., Kovacs, K., and Von Holle, B. 2010b. A spatial-dynamic value transfer model of economic losses from a biological invasion. Ecological Economics 70:86-95.

Isaaks, E. H., and Srivastava, R.M. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York. 561 pp.

Irwin, E. G. 2002. The effects of open space on residential property values. Land economics 78:465-480.

Jansen, M., Judas, M., and Saborowski, J. 2002. Spatial Modelling in Forest Ecology and Management: A Case Study. Springer, Berlin. 225 pp.

Köhl, M., and Gertner, G. 1997. Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. Forest Ecology and Management 95:131-140.

Malhi, Y., Wood, D., Baker, T.R., Wright, J., Phillips, O.L., Cochrane, T., and Meir, P., et al. 2006. The regional variation of aboveground live biomass in old-growth Amazonian forests. Global Change Biology 12:1107-38.

McClure, M. 1990. Role of wind, birds, deer, and humans in the dispersal of hemlock woolly adelgid (Homoptera: Adelgidae). Environmental Entomology 19:36-43.

McClure, M. 1991. Density-dependent feedback and population cycles in Adelges tsugae (Homoptera: Adelgidae) on Tsuga canadensis. Environmental Entomology 20:258-264.

McConnell, V., and Walls, M.A. 2005. The value of open space: Evidence from studies of nonmarket benefits. Resources for the Future, Washington, DC. 78 pp.

Mitas, L., and Mitasova, H. 1988. General variational approach to the interpolation problem. Computer and Mathematics with Applications 16: 983–992.

Moore, C. C., Holmes, T.P., and Bell, K.P. 2011. An attribute-based approach to contingent valuation of forest protection programs. Journal of Forest Economics 17:35-52.

Orwig, D. and Foster, D. 1998. Forest response to the introduced hemlock woolly adelgid in southern New England, USA. Journal of the Torrey Botanical Society 125:60-73.

Orwig, D., Foster, D., and Mausel., D. 2002. Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. Journal of Biogeography 29:1475-1487.

Orwig, D., Thompson, J., Povak, N., Manner, M., Niebyl, D., and Foster, D. 2012. A foundation tree at the precipice: Tsuga canadensis health after the arrival of Adelges tsugae in central New England. Ecosphere 3:10.

Palmquist, R.B. 1991. Hedonic methods, in: Measuring the Demand for Envrionemntal Quality. Elsevier, New York, NY. 77-120.

Patterson, R.W., and Boyle, K.J. 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic propertyvalue models. Land Economics 78:417– 425.

Preisser, E., Miller-Pierce, M., Vansant, J., and Orwig, D. 2011. Eastern hemlock (Tsuga canadensis) regeneration in the presence of hemlock woolly adelgid (Adelges tsugae) and elongate hemlock scale (Fiorinia externa). Canadian Journal of Forest Research 41:2433-2439.

Schabenberger, O., and Gotway, C. A. 2005. Statistical Methods for Spatial Data Analysis. Chapman& Hall/CRC, Boca Raton, FL. 504pp.

Shepard, D. 1968. A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 ACM National Conference. pp. 517–524.

Smith, D. 1997. The Practice of Silviculture: Applied Forest Ecology. John Wiley & Sons, New York. 560 pp.

Stein, M. L. 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer, NY. 247 pp.

Turner, J.L., Fitzpatrick, M.C., and Preisser, E.L. 2011. Simulating the dispersal of hemlock woolly adelgid in the temperate forest understory. Entomologia Experimentalis et Applicata 141:216-223.

Wulff, S., Hansson, P., and Witzell, J. 2006. The applicability of national forest inventories for estimating forest damage outbreaks-Experiences from a Gremmeniella outbreak in Sweden. Canadian Journal of Forest Research 36:2605-2613.

Table 1. Live Basal Area and Vigor for Sampled Hemlock Stands

	Year	2007	2009	2011
Live Basal Area	Mean	38.23	27.83	15.31
(m^2/ha)	SD	27.59	16.29	11.89
	Min	0	0	0
	Max	125.45	73.34	54.04
	N[a]	140	138	122
Vigor Class[b]	1	8	11	9
(stand count)	2	18	19	23
	3	33	37	44
	4	82	71	47
	N	141	138	123

[a] The sample size decreased over years as some hemlock stands disappeared or were not allowed for access again.
[b] For vigor classes: 1 = 76 – 99% foliar loss, 2= 51-75% foliar loss, 3 = 26-50% foliar loss and 4 = 0 – 25% foliar loss.

Table 2. Estimation Results Based on Sampled Data and Kriging Interpolated Data

	<0.1km	<0.5km	<1km
	Sample Data		
Hedonic Model			
lba ($\times 10^{-3}$)	-5.856*	1.699	0.217*
	(3.442)	(2.250)	(1.291)
lba*vigor ($\times 10^{-3}$)	2.039*	-0.459	-0.649*
	(1.042)	(0.745)	(0.373)
N	148	484	1,655
	Kriging Results		
Hedonic Model			
lba ($\times 10^{-3}$)	-8.396**	-5.825**	-2.935*
	(4.026)	(2.327)	(1.702)
lba*vigor ($\times 10^{-3}$)	2.174**	1.591***	0.831**
	(0.914)	(0.493)	(0.367)
N	2,762	13,087	23,267
Repeat-sale Model			
lba ($\times 10^{-3}$)	-61.306***	-9.359	0.354
	(15.767)	(8.399)	(6.576)
lba*vigor ($\times 10^{-3}$)	18.144***	1.968	-1.028
	(5.022)	(2.224)	(1.654)
N	180	818	1,425
Hedonic Model with Repeat-sale Sample			
lba ($\times 10^{-3}$)	-7.558	-19.640***	-3.632
	(21.876)	(6.453)	(5.126)
lba*vigor ($\times 10^{-3}$)	9.046*	4.984***	1.275
	(4.657)	(1.538)	(1.444)
N	180	818	1,425

Note: *** denotes significant at the 1% level, ** denotes significant at the 5% level, * denotes significant at the 10% level; standard deviations in parentheses.

Table 3. Estimation Results Based on Different Interpolation Methods

	Kriging	IDW	Spline
Hedonic Model			
lba ($\times 10^{-3}$)	-8.396**	-4.032	-3.364
	(4.026)	(2.858)	(2.701)
lba*vigor ($\times 10^{-3}$)	2.174**	1.232**	1.064
	(0.914)	(0.722)	(0.683)
N	2,762	2,762	2,762
Repeat-sale Model			
lba ($\times 10^{-3}$)	-61.306***	-39.03*	-39.276**
	(15.767)	(21.651)	(19.383)
lba*vigor ($\times 10^{-3}$)	18.144***	13.14**	13.334**
	(5.022)	(5.915)	(5.500)
N	180	180	180

Note: a. *** denotes significant at the 1% level, ** denotes significant at the 5% level, * denotes significant at the 10% level; Standard deviations in parentheses.

b. The results listed here is for 0.1km buffer; 0.5 and 1km buffers excluded because they are not significant in the estimation results of repeat sale models.

Table 4. Estimation Results Based on Cross Validation Adjustment

	Kriging	Olea & Pawlowsky Adjustment	Yamamoto Adjustment
Hedonic Model			
lba ($\times 10^{-3}$)	-8.396**	-5.652**	-2.625
	(0.022)	(2.426)	(2.186)
lba*vigor ($\times 10^{-3}$)	2.174***	1.374**	0.802
	(0.009)	(0.543)	(0.586)
N	2,762	2,762	2,762
Repeat-Sale Model			
lba ($\times 10^{-3}$)	-61.306***	-35.782***	-33.654***
	(15.767)	(11.348)	(11.414)
lba*vigor ($\times 10^{-3}$)	18.144***	9.735***	11.285***
	(5.022)	(3.412)	(4.010)
N	180	180	180

Note: a. *** denotes significant at the 1% level, ** denotes significant at the 5% level, * denotes significant at the 10% level; Standard deviations in parentheses.

b. The results listed here is for 0.1km buffer; 0.5 and 1km buffers excluded because they are not significant in the estimation results of repeat sale models.

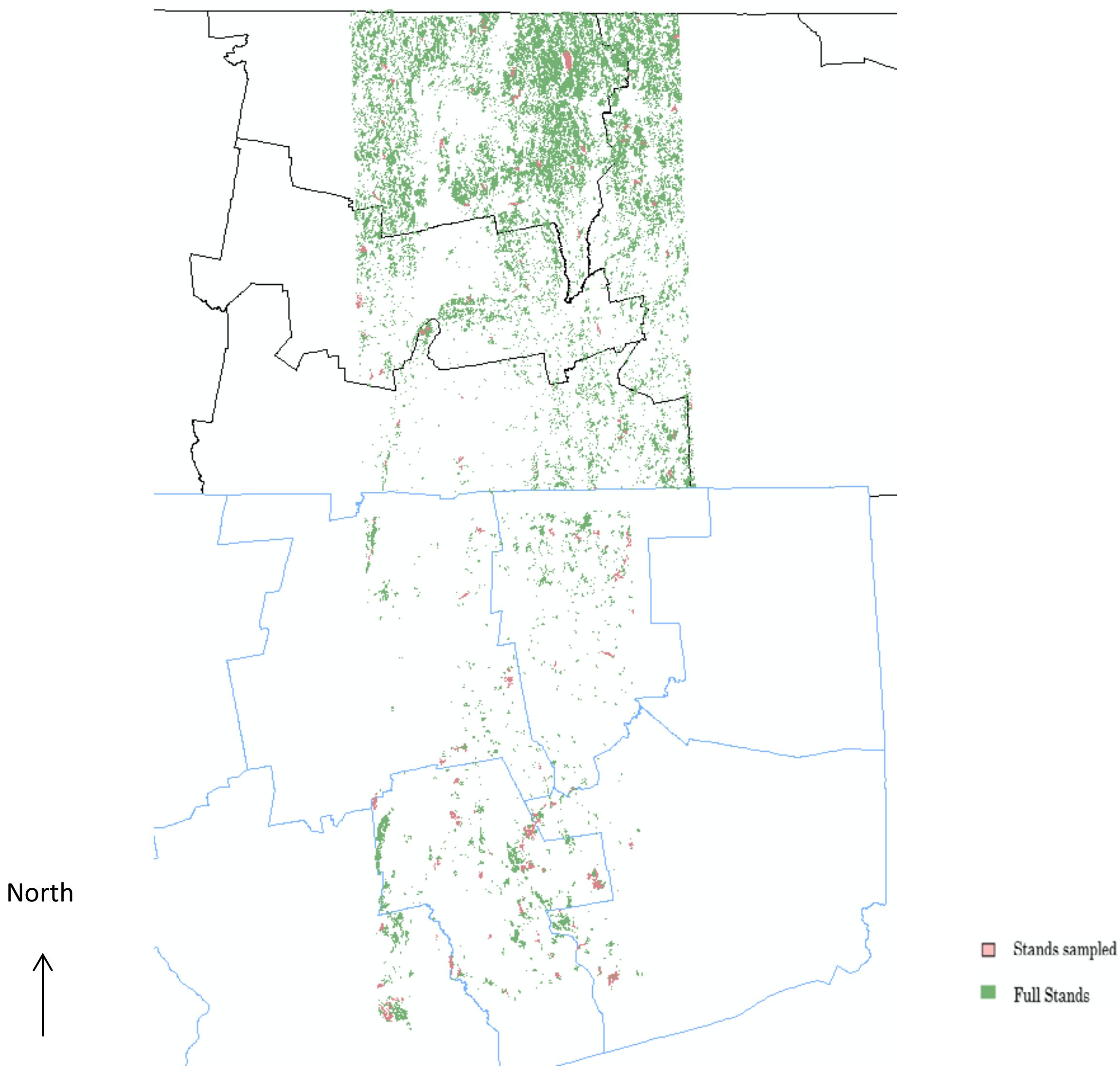
Table 5. Implicit Price for Live Basal Area

	Sampled Model (hedonic model)	Kriging (hedonic model with repeat sale sample)	Kriging (repeated sale)	Adjustment Yamamoto (repeated sale)
Vigor=1	-0.38%	0.149%	-4.32%	-2.24%
	(0.2515)	(1.794)	(1.171)	(0.777)
Vigor=4	0.23%	2.863%	1.13%	1.15%
	(0.1671)	(1.060)	(1.026)	(0.659)
n	148	180	180	180

Note: a. *** denotes significant at the 1% level, ** denotes significant at the 5% level, * denotes significant at the 10% level; Standard deviations in parentheses.

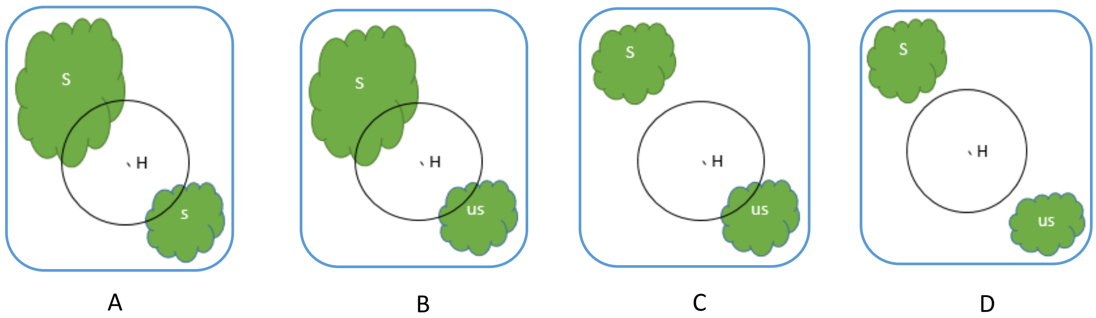
b. The results listed here is for 0.1km buffer

Figure 1. Hemlock Stands in the Study Area



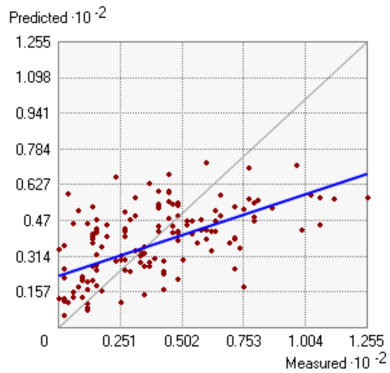
Note: The blue line represents the border of the counties in Connecticut; while the black line represents the border of the counties in Massachusetts.

Figure 2. Locations between Household Properties and Hemlock Stands

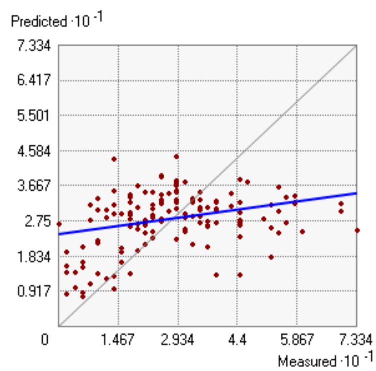


Note: H represents household property; S represents sampled hemlock stand; US represents non-sampled hemlock stand.

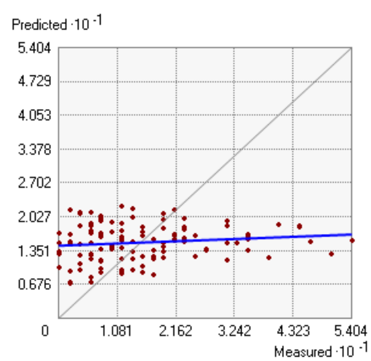
Figure 3. Cross Validation Results for live Basal Area (m^2/ha) by Kriging



2007



2009



2011

Figure 4. Interpolated Live Basal Area (m^2/ha) by Kriging in the Study Area (2007-2011)

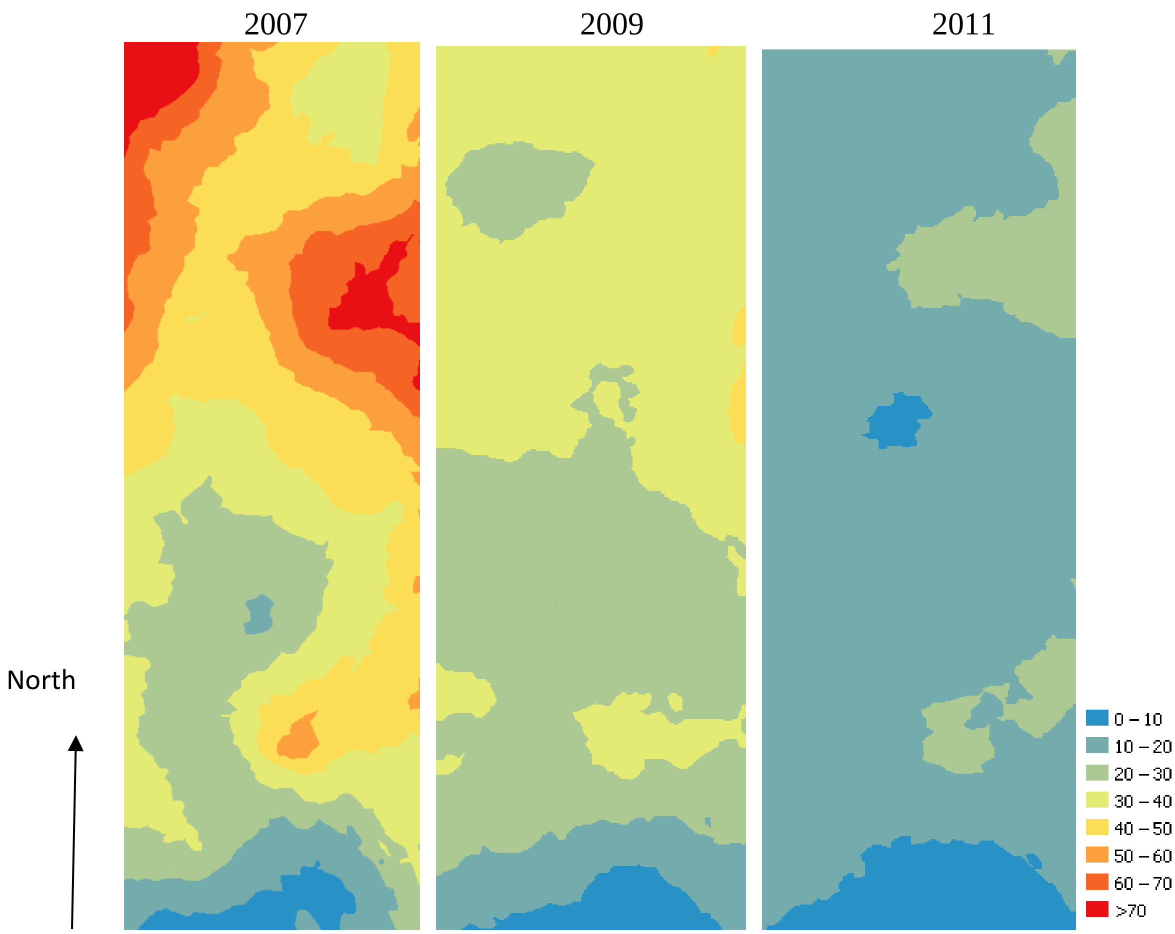
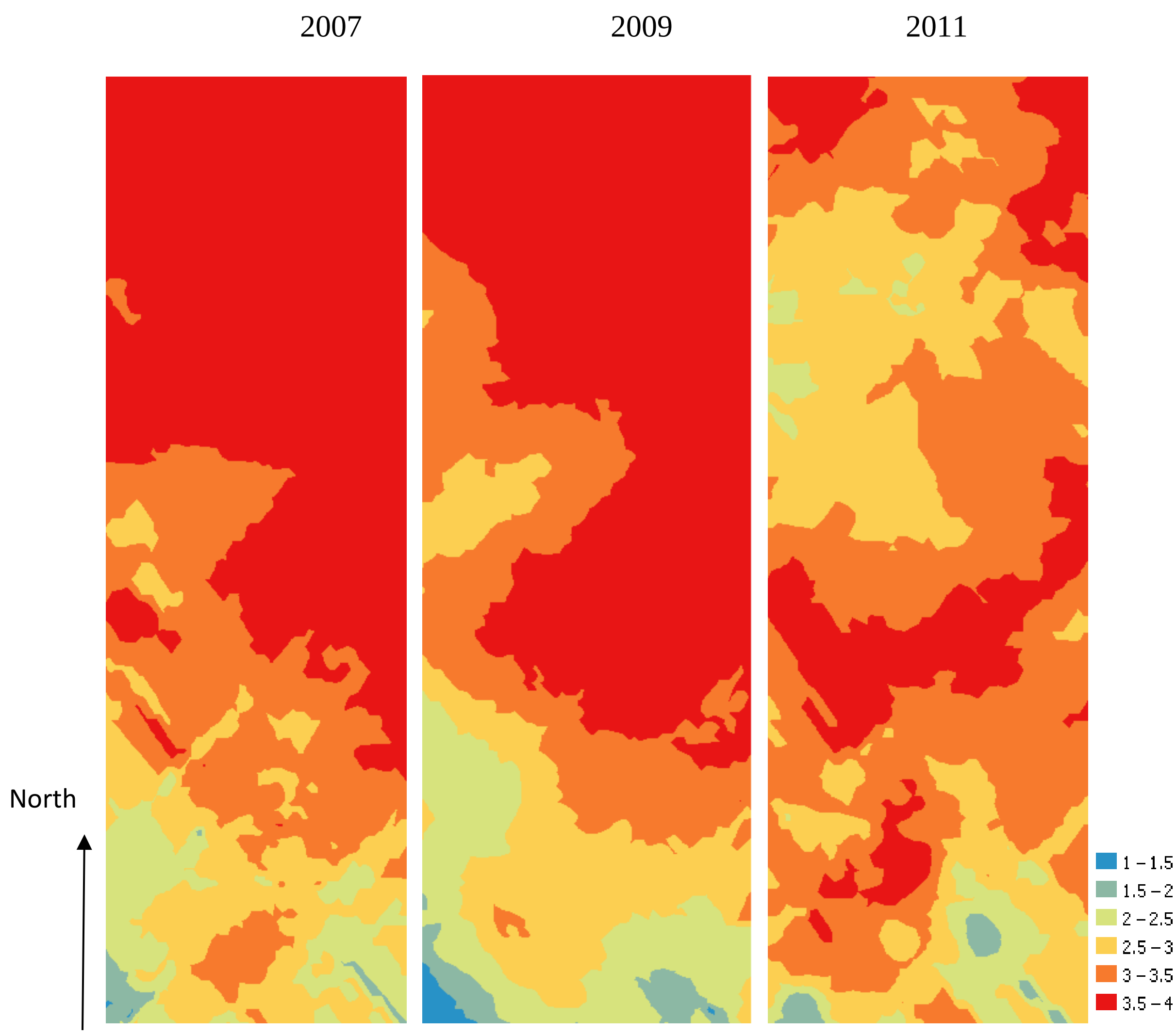


Figure 5. Interpolated Vigor by Kriging in the Study Area (2007-2011)



Appendix

Table A. Descriptive Statistics of Housing Characteristics Variables

	Hedonic Model		Repeat Sale Model	
	Mean	SD	Mean	SD
Sampled Model				
House price ($2003)	384,246	250,960		
Living area (ft^2)	2,289	910		
Lot size (ft^2)	86,936	133,246		
Baths	2.00	0.77		
Bedrooms	3.40	0.88		
Age	1,966	49		
Air conditioning (%)	0.39	0.49		
Fireplace (%)	0.54	0.50		
Distance to highway (m)	937	845		
Water (%)	0.29	2.46		
Open space (%)	23.58	21.58		
Developed area (%)	0.41	3.22		
Forest (%)	52.06	30.41		
Agricultural (%)	2.29	8.71		
Wetland (%)	1.80	6.61		
n	148	148		
Interpolated Model				
House price ($2003)	320,188	315,590	330,318	343,568
Living area (ft^2)	1,914	848	1,865	870
Lot size (ft^2)	93,244	213,209	80,985	148,524
Baths	1.79	0.76	1.76	0.70
Bedrooms	3.17	0.83	3.15	0.79
Age	1,970	41	1,969	40
Air conditioning (%)	0.31	0.46	0.28	0.45
Fireplace (%)	0.48	0.50	0.42	0.49
Distance to highway (m)	1,068	1,068	1,017	1,156
Water (%)	0.54	3.42	0.68	3.40
Open space (%)	22.12	19.54	22.16	18.44
Developed area (%)	0.08	1.24	0.002	0.03
Forest (%)	47.47	29.87	47.66	29.25
Agricultural (%)	4.92	12.87	4.75	11.59
Wetland (%)	2.72	8.40	1.83	6.11
n	2,762	2,762	180	180

Note: The samples are based on 0.1km buffer.

Table B. Descriptive Statistics of Hemlock Health Variables in Hedonic Analysis

	Mean	SD	Min	Max
Sampled Data (n=148)				
Live Basal Area (m^2/ha)	22.886	18.435	0	98.43
Vigor	2.723	0.995	1	4
Lba*vigor	65.542	55.910	0	295.29
Interpolated Data				
Kriging (n=2,762)				
Live Basal Area (m^2/ha)	30.247	12.789	5.797	79.57546
Vigor	3.378	0.545	1.369	4.125
Lba*vigor	106.344	55.364	13.121	314.147
IDW (n=2,762)				
Live Basal Area (m^2/ha)	30.293	13.020	3.013	87.592
Vigor	3.366	0.582	1.127	4
Lba*vigor	106.299	55.483	5.973	346.141
Spline (n=2,762)				
Live Basal Area (m^2/ha)	30.462	13.253	3.771	83.438
Vigor	3.376	0.577	1.071	4.589
Lba*vigor	107.091	56.493	6.479	329.149
Interpolated Data with Cross Validation Adjustment				
Olea & Pawlowsky Adjustment (n=2,762)				
Live Basal Area (m^2/ha)	32.653	19.960	0	137.444
Vigor	3.430	0.692	0.907	4.521
Lba*vigor	120.380	87.252	0	585.272
Yamamoto Adjustment (n=2,762)				
Live Basal Area (m^2/ha)	28.830	15.542	0.221	109.589
Vigor	3.278	0.550	1.192	4
Lba*vigor	98.750	62.164	0.506	419.330

Note: The samples are based on hedonic model in 0.1km buffer.