land

Article

Patterns and Predictors of Recent Forest Conversion in New England

Alexandra M. Thorn [1,*,†], Jonathan R. Thompson [2] and Joshua S. Plisinski [2]

[1] Earth Systems Research Center, University of New Hampshire, Durham, NH 03824, USA
[2] Harvard Forest, 324 North Main Street, Petersham, MA 01366, USA; jthomps@fas.harvard.edu (J.R.T.); jplisinski@fas.harvard.edu (J.S.P.)
* Correspondence: athorn@alum.wellesley.edu; Tel.: +1-352-262-4870
† Current address: Morse Hall, 8 College Road Durham, NH 03824, USA

Academic Editor: H. Ricardo Grau
Received: 22 July 2016; Accepted: 13 September 2016; Published: 19 September 2016

Abstract: New England forests provide numerous benefits to the region's residents, but are undergoing rapid development. We used boosted regression tree analysis (BRT) to assess geographic predictors of forest loss to development between 2001 and 2011. BRT combines classification and regression trees with machine learning to generate non-parametric statistical models that can capture non-linear relationships. Based on National Land Cover Database (NLCD) maps of land cover change, we assessed the importance of the biophysical and social variables selected for full region coverage and minimal collinearity in predicting forest loss to development, specifically: elevation, slope, distance to roads, density of highways, distance to built land, distance to cities, population density, change in population density, relative change in population density, population per housing unit, median income, state, land ownership categories and county classification as recreation or retirement counties. The resulting models explained 6.9% of the variation for 2001–2011, 4.5% for 2001–2006 and 1.8% for 2006–2011, fairly high values given the complexity of factors predicting land development and the high resolution of the spatial datasets (30-m pixels). The two most important variables in the BRT were "population density" and "distance to road", which together made up 55.5% of the variation for 2001–2011, 49.4% for 2001–2006 and 42.9% for 2006–2011. The lower predictive power for 2006–2011 may reflect reduced development due to the "Great Recession". From our models, we generated high-resolution probability surfaces, which can provide a key input for simulation models of forest and land cover change.

Keywords: forests; land cover change; modeling; spatial analysis; geographic information systems

1. Introduction

New England is a heavily-forested region, and its forests provide numerous benefits to its residents, including water filtration, flood mitigation, aesthetics and recreation, as well as a source of timber [1–7]. The forests also play a role in climate regulation, storing an estimated 2239 Tg of carbon [8] and sequestering an additional 9.7 Tg per year [9,10]. Given the ecosystem services provided by New England forests, planners in the region have a strong interest in understanding the dynamics of changing forest cover, in order to assess the possible impact of these changes and to identify priority areas for conservation (e.g., [11]). Information on changing land cover also informs projections and scenarios for long-term land cover in the region, providing tools to assess possible trajectories for long-term changes in ecosystem services (e.g., [12–14]). To help understand ongoing changes in the New England landscape, we asked: what social and biophysical factors are most strongly associated with forest loss to development in New England?

New England has not always been as heavily forested as today. Following the arrival of European colonists in New England, forests were extensively cleared and cut-over, declining from over 90% forest cover in 1600 to just over 55% in the latter half of the 19th century [2]. During the nineteenth century, the American Midwest was opened up for agriculture, and New England agriculture subsequently declined over the interval from 1870–1970, permitting recovery of forests to 80% cover, even as the population continued to grow [2]. In 1970, loss of forest to sprawled development overtook reforestation of abandoned agricultural areas, and forest area began to decline again [2,15,16].

In recent decades, the primary dynamic of land cover change has been forest loss to development [15], a trend that is expected to continue [13]. National projections from the USDA Forest Service "Forests on the Edge" project identify New England forests as some of the forests at highest risk from development nationally, particularly in watersheds in southern and eastern New England, and many of these watersheds are also nationally important for the role of forests in provisioning timber and clean water [13]. From the perspective of protecting forests and the services they provide, this trend is particularly concerning because, in contrast to the 19th century loss of forest to agriculture, the replacement of forests with buildings and paved surfaces is considered permanent [2].

Numerous analyses ranging from local to global have evaluated predictors of the development of forest and open space. Across studies, population growth is a particularly commonly-cited driver for development rates [17]. Other variables associated with development have included proximity to populated and built areas [18–21], topography [21,22], transportation networks [18] and climate variables (e.g., [22]), as well as economic and political considerations [20,23].

In New England, the prevailing driver for development is generally understood to be the combination of overall population growth and migration to suburban and exurban areas. Mockrin et al. [24] used historical housing densities from 1940–2000 to assess the spatial distribution of residential development in the Northern Forest of the Northeast, including analyses of the New England states of Maine, New Hampshire and Vermont, and demonstrated that areas with lower housing densities have experienced peak development rates more recently than areas with higher housing densities, illustrating a temporal trend of accelerating sprawl. A similar trend was illustrated using Landsat imagery for 1990–2005 to quantify forest loss in southern New England, including Connecticut, Rhode Island, Massachusetts and parts of New Hampshire and Vermont [16]. According to that analysis, the highest rates of development for 2000–2005 were located further from the urban center of Boston than in the 1990–2000 interval. Thus, while population growth appears to be strongly associated with development, the relationship between population and development is expected to vary along urban-rural gradients. The interaction between population and urban-rural gradients is further complicated by the popularity of second homes in counties with high rural amenities, largely in heavily-forested lake and mountain counties of Maine, New Hampshire and Vermont [25].

At finer scales, physical variables, such as slope and access to transportation, emerge as important predictors for development. Local analyses of land cover change in and near New England point to considerable variability in the importance of these variables depending on spatial context. Tyrrell et al. [26] analyzed land cover change in two northeastern locations, the Catskill/Delaware water supply watersheds in New York state and the Thames River watershed in Massachusetts and Connecticut. They found that whereas elevation and proximity to urban areas were the best predictors of development in the Catskill/Delaware area, the model for the development in the Thames River watershed depended on the combined effects of a large number of physical, economic and social variables, each of similar importance. Similarly, an analysis of the entire state of Massachusetts showed that conservation status, distance to built areas and slope were all important predictors of development, as were several metrics of density [19].

We are aware of no comprehensive high resolution analysis for predictors of development across the entire New England region. The research presented here assesses geographic variation in recent trends in forest loss across New England and quantifies the relative importance of multiple social and biophysical variables in predicting the conversion of forest to development. We assembled regional maps of possible predictor variables and used boosted regression trees (BRTs), a type of non-parametric model capable of capturing interactions and non-linear relationships, to quantify the strength of their association with forest conversion, based on land cover maps from the National Land Cover Database (NLCD) for 2001, 2006 and 2011. Using the BRT models, we produced high resolution (30 m) probability surfaces characterizing patterns of development over these time intervals. In addition to other considerations, our analysis inspects the relative importance of population growth as a predictor of development in the region.

2. Materials and Methods

2.1. Study Area

New England is a 186,000-km^2 region in the northeastern United States, located between 40.9 and 47.5°N and between 41.0 and 44.7°W, and comprises six states: Maine, New Hampshire, Vermont, Massachusetts, Connecticut and Rhode Island. The region is temperate in climate, with mean annual temperatures ranging from 3–10 °C and mean annual precipitation ranging from 79–255 cm [27]. New England is nearly 80% forest, including five main vegetation zones: spruce-fir-northern hardwoods, northern hardwoods-hemlock-white pine, transition hardwoods, central hardwoods and pitch pine-oak [2,28]. About ten percent of the land area is developed, primarily in the southern states of Massachusetts, Connecticut and Rhode Island and in southeastern New Hampshire, and rapid development continues throughout the region [2]. Major urban areas include Greater Boston in eastern Massachusetts and New Haven and Hartford in Central Connecticut.

2.2. Data

We investigated the spatial distribution of forest conversion to development for the six states of New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont) and assessed the role of biophysical and social variables as predictors of development. Conversion rates were estimated using National Land Cover Database (NLCD) Landsat-based maps of land cover for 2001, 2006 and 2011. NLCD is a 30-m spatial resolution map of land cover in the United States, produced by a collaboration of the U.S. Geological Survey with the National Oceanic and Atmospheric Administration and the U.S. Forest Service. It is based on satellite data from the Landsat Thematic Mapper and since 2001 has been applying a consistent methodology aimed at quantifying land cover change, as well as overall land cover ([29]). We aggregated NLCD categories for deciduous forest (41), evergreen forest (42), mixed forest (43) and shrub/scrub (52) into a single "forest" category and NLCD categories for developed open space (21), developed low intensity (22), developed medium intensity (23) and developed high intensity (24) into a single "developed" category and assessed the rates of forest conversion to development for 2001–2011 and for the two sub-intervals 2001–2006 and 2006–2011.

As geographic predictor variables, we considered elevation, slope, land ownership classes, the distance to nearest major roads, the distance to built areas, New England state, census data (including distance to the center of urban areas), tract-level population density, housing unit density and median household income. We calculated two variables for the rate of population growth: (1) the rate of change in population density from 2000–2010; and (2) the relative rate of change in population density (defined as the positive or negative change in population density, divided by the population density in 2000). To capture the possible role of second homes as a predictor, we also included USDA Economic Research Service County Typology Codes for retirement destination counties and non-metropolitan amenities counties as input for our analysis.

Table 1. Predictor variables.

Predictor variable	**Units**	**Source**	**Processing**
Median Household Income	U.S. dollars	U.S. Census (2000)	Unmodified
Population Density	people per km^2	Calculated from U.S. Census (2000) Population densities in 2000, calculated from U.S. Census tract-level population data	
Population Density Change	people per km^2	Calculated from U.S. Census (2000, difference between the population densities in 2000 and 2010, calculated from U.S. Census tract-level population data	
Population per Housing Unit	people per housing unit	U.S. Census (2000)	Tract-level population density divided by tract-level housing unit density
Distance to Cities	meters	U.S. Census Urban Areas (2013)	Distance raster generated from the center of mass of all urban areas in New England and New York State
Relative Change in Population Density	percent	Calculated from U.S. Census (2000, 2010)	Population density in 2010 decided by population density in 2000
Distance to Built	meters	USGS NLCD (2001)	Distance raster generated from 2001 NLCD developed Classes 22, 23, 24; excludes Class 21 "developed open space"
Distance to Road	meters	Calculated from U.S. Census Topologically Integrated Geographic Encoding and Referencing (TIGER)shapefiles (2013)	Distance raster generated from roads in Classes S1100, S1200 or S1400
Distance to Highway	km of road per km^2	Calculated from U.S. Census TIGER shapefiles (2013)	Distance raster generated from roads in Classes S1100 and S1200
Elevation	meters	National Elevation Database (2007)	30-meter digital elevation model [30]
Slope	degrees	National Elevation Database (2007)	Slope computed from 30-m NED elevation data
State	categorical	U.S. Census (2010)	Rasterized Census Cartographic Boundary Files
Owner Type	categorical	USGS Protected Areas Database (2011)	The categories in the Protected Areas Database were reclassified using the Own_Type field. Public = Own_Type Domain Codes 1,3,4,5. private protected = 6,7,8,9,10 and private = 2, and all other areas.
Retirement County	categorical	USDA Economic Research Service (2004)	Raster map generated from U.S. Census cartographic boundary files and USDA ERS 2004 County Typology Code indicator for retirement destination county
Recreation County	categorical	USDA Economic Research Service (2004)	Raster map generated from U.S. Census cartographic boundary files and USDA ERS 2004 County Typology Code indicator for non-metro recreation county

Base maps for an initial collection of 15 predictor variables (11 continuous and 4 categorical) and of the spatial distribution of forest loss to development were generated using ArcMap 10.3, Quantum GIS (Version 2.2.0) [31] and R (Version 3.2.2) [32]. Once base maps were generated, all analyses were performed using R. For subsequent analysis, values of all variables were sampled using a randomly-positioned uniform 1-km grid of sample points. To avoid redundant correlation among factors, we first assessed the collinearity of the predictor variables. Our protocol was guided by a review of approaches to handling collinearity produced by Dormann and colleagues [33]. This review lists several alternative approaches to reducing collinearity in an input dataset: statistical identification of variable clusters (e.g., by principal component analysis), separation of orthogonal components of related variables by sequential regression or the use of a correlation matrix to identify variables correlated by less than a predetermined threshold (commonly 0.7) and removing correlated variables based on clear differences in ecological importance [33]. To maintain the simplicity of interpretation, we adopted the last approach, producing a correlation matrix relating the values of all variables at sample points to one another, as well as to spatial coordinates and maps of forest conversion to development for each of the two intervals (Table S1). The only pair of input variables with a correlation above 0.7 was housing unit density and population density. To avoid this collinearity, we removed housing unit density as a variable and instead calculated the mean population per housing unit, which was found not to be collinear with any other variables. The final 15 input variables used for our analysis are summarized in Table 1.

2.3. Analysis

To establish general patterns of forest conversion, we first summarized forest loss to development by county, for the intervals 2001–2006, 2006–2011 and 2001–2011.

We then used boosted regression tree (BRT) analysis to estimate the relationships between our predictor variables (Table 1) and forest conversion to development, based on the values at our randomly-positioned grid of sample points. BRTs are machine-learning-based non-parametric statistical models based on regression trees [34]. Regression trees recursively partition datasets to generate a decision tree structure [19], and a BRT is the sum of a large number of simple regression trees used to characterize the overall relationship of the response variable to a collection of predictor variables. The first tree of a BRT is computed to minimize some loss function of predictive performance, and subsequent trees are fit to the residuals in a forward stage-wise fashion, with existing trees left unmodified. In order to improve predictive performance, each tree is generated using a different random subset of the full dataset, as determined by the "bag fraction" parameter [35]. BRTs are generally able to capture both variable interactions (characterized by tree complexity) and non-linear relationships [35].

We selected BRTs for our analysis because of their ability to capture complex non-linear effects and interactions without the need for prior data transformation or the removal of outliers [35]. BRTs generally avoid overfitting [35], can be applied to a variety of spatial analyses, such as the distribution of species and vegetation types [36–42], hydrology [43,44], soil and landform properties [45–47] and natural disturbance [48], as well as quantification of land cover and land use change through human activities [34,49,50]. Across a wide variety of contexts, model comparisons have shown BRTs to perform much better than traditional models and comparably well to other machine-learning models [36,37,43,45,48,51], with some variability in comparative performance with other machine-learning methods depending on context [33,38,44,47].

We performed BRT analyses using the R packages "gbm" [52] and "dismo" [53]. Specific information on the algorithm, equations and assumptions is available in the documentation for those packages and in guides produced by Ridgeway [54] and by Elith and colleagues [35]. Because our response variable is binomial (forest that remains forest or forest that is developed), we used a Bernoulli distribution loss function to model the probability of development at all sample points. We used the default bag fraction (=0.75) and calibrated the parameters "tree complexity" and "learning rate" using 10-fold cross-validation. Tree complexity determines how many branching nodes are included in each tree, and the learning rate determines the relative contribution of each tree to the final model by adjusting the gradient step permitted with each tree added to the model [54]. We used the "dismo" function "gbm.step" to test all combinations of tree complexity from 1–15 with each of six learning rates ranging from 0.01–0.0005 for the full interval 2001–2011 for a randomly-selected 90% of the sample points. The function "gbm.step" automatically determines the number of trees to include in each model, although we did not permit the tree number to exceed 10,000. We then used the resulting model to predict development rates for the remaining 10% and correlated the result with actual development rates to determine which combination produced the best fit.

We found that within this range, correlation generally improved with both increasing tree complexity and decreasing learning rates (Table A1), but that improvements to the correlation coefficient were small with a tree complexity above 4 (Figure A1). Additional cross-validation tests with lower learning rates and a tree complexity of 4 showed that there was not much improvement with learning rates below 0.0005 (Figure A2). For our final BRTs for each time step (2001–2006, 2006–2011 and 2001–2011), we parameterized the model with a tree complexity of 4 and a learning rate of 0.0005 and permitted the function to automatically determine the number of trees in each model. Moran's I correlograms were used to quantify the spatial autocorrelation of the residuals for each model.

The R library gbm reports estimates of the relative influence of each predictor variable, based on the number of times each variable is used in the regression trees and weighted by the square of the improvement to the model produced by the resulting split [35]. We reported these estimates, and for variables with contributions above random chance (100%/15 variables = 6.67%), we further explore their effect on the model using partial dependency plots, which illustrate predicted value as a function of one variable while other variables are held at their mean value [35].

Finally, to illustrate the spatial distribution of each model, we generated probability surfaces from the BRTs from each interval. We used the the function "predict.gbm" from the package gbm to generate rasters of the modeled probability of development for each 30-m cell, using our original rasters of each predictor variable as our input.

3. Results

Our analysis of land cover change from NLCD maps showed a conversion of 35,500 ha (0.0026% of forested area in 2001) from forest to developed between 2001 and 2006 and 33,500 ha (0.0025% of the forested area in 2001) from forest to developed between 2006 and 2011. For both time intervals, the most intense areas of forest loss were in eastern Massachusetts (MA), western Connecticut (CT) and southern New Hampshire (NH) (Figure 1). These counties were close to the urban centers of Boston, MA, Hartford, CT, New Haven, CT, and New York, NY. This pattern was somewhat weaker for 2006–2011, however, with less deforestation in counties close to Boston and more deforestation in New Hampshire and Maine, compared to the earlier time interval.

Boosted regression tree models for development rates for 2001–2011 and both sub-intervals were highly significant ($p < 0.0001$) and explained 6.9% of the geographic variation in development for the full interval, as well as 4.5% of the variation for 2001–2006 and 1.8% of the variation for 2006–2011 (Table 2).

Moran's I correlograms of the residuals for the BRT models show that the models do not account for all spatial autocorrelated variation in development rates. Spatial autocorrelation was significant for distances of up to 150–200 km for all time intervals (Figure A3).

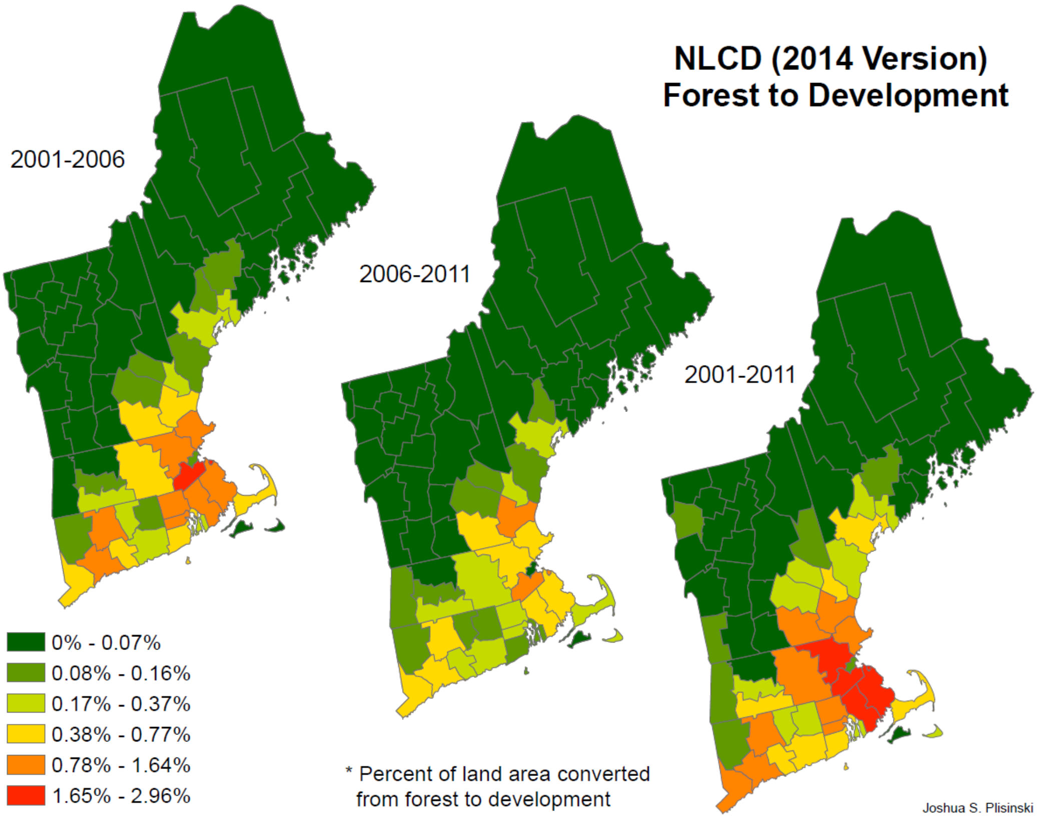

Figure 1. Intensity of forest conversion to developed land in New England by county.

Table 2. Boosted regression tree (BRT) model performance.

	2001–2006	**2006–2011**	**2001–2011**
Tree complexity	5	3	1
Learning rate	0.0005	0.005	0.0005
Number of trees	9150	5300	8700
CV r	0.213	0.136	0.263
CV r^2	0.045	0.018	0.069
Training r	0.374	0.296	0.358
Training r^2	0.140	0.088	0.128
Mean total deviance	0.032	0.018	0.046
Mean residual deviance	0.019	0.012	0.028
CV standard error	0.015	0.022	0.018
Estimated CV deviance	0.021	0.014	0.030
Estimated CV deviance standard error	0.000416	0.000388	0.000678

CV=cross validation; r=coefficient of correlation

The library gbm reports the relative influence of each predictor variable, regardless of whether the relationship between predictor and response variables is positive, negative or positive for some values and negative for others. For all time intervals, the two most important predictive variables by relative influence in the BRTs were "population density" and "distance to road" (Figure 2). These two variables together explained 55.5% of the variation captured by the BRT for the full interval 2001–2011, 49.4% of the variation captured by the BRT for 2001–2006 and 42.9% of the variation captured by the BRT for 2006–2011. The other variables with predictive power above chance were "relative change in population density" for all intervals, "distance to built" for 2001–2006 and 2001–2011, "population density change" for 2001–2006 and 2006–2011, "distance to cities" for 2001–2006 and "median income" for 2006–2011. Relative change in population density was the third most important variable for 2001–2011 and 2006–2011, while "distance to built" was the third most important variable for 2001–2006.

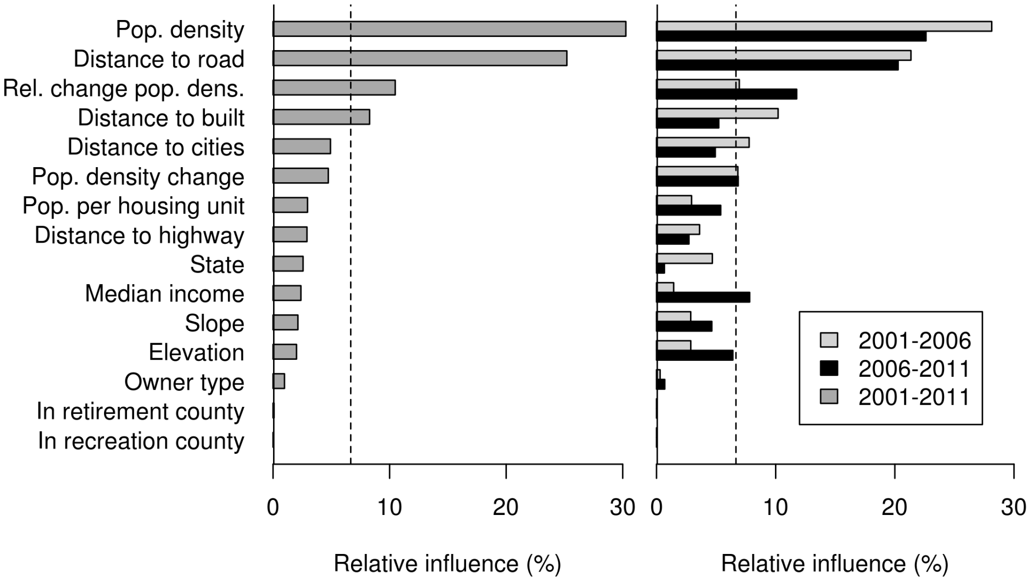

Figure 2. Percent of the relative influence of independent variables in the BRT models of the development probability for the full interval 2001–2011 and for the time intervals 2001–2006 and 2006–2011. The dashed vertical line shows the influence level expected by chance (=6.67%).

Partial dependency plots (PDPs) summarize the overall relationship between input variables and the probability of development, including non-linear relationships, by plotting the modeled relationship between one predictor variable and the response variable when all other predictor variables are held constant at their mean values. In our models for all intervals, PDPs showed a threshold response of development probability to previous population density: probabilities were very low at densities below 60 people per km^2 and increased dramatically for population densities above 100 people per km^2 (Figure 3). In addition, for 2001–2011 and 2001–2006, the development probability decreased for densities above 600 people per km^2. Similarly, the probability of development was consistently higher at shorter distances to roads for all intervals, but there was no effect of road distance for distances greater than 80 m (Figure 3b). On the other hand, the PDPs showed almost no effect of relative population density change on the probability of development when other variables were held constant (Figure 3c). The pattern for the effects of distance to built for 2001–2011 and 2001–2006 was similar to the pattern for distance to roads, with higher probabilities of development at shorter distances to built land, but only out to a distance of 400 m (Figure 3d). For 2001–2006, the probability of development was higher between 1 and 11 km from the center of a city than at smaller or greater distances (Figure 3e). The PDPs indicated that the probability of land conversion was higher at population density changes greater than 20 people per km^2 for 2001–2006 and greater than 390 people per km^2 for 2006–2011 (Figure 3f). The effect of median income shown by the the PDPs was somewhat unclear: for 2006–2011, the rates of development decreased substantially for median income greater than $20,000/year and increased slightly for areas with median income greater than $145,000/year and $189,000/year (Figure 3g).

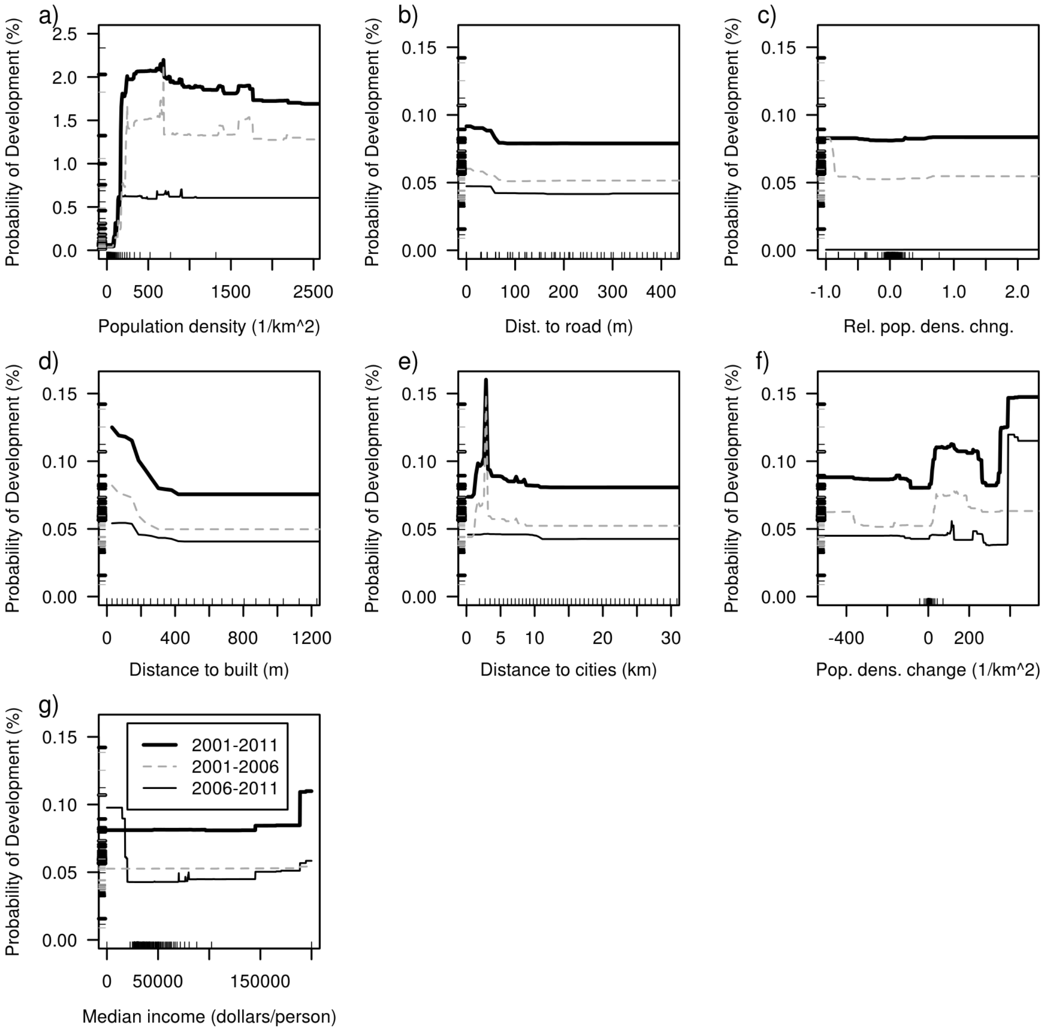

Figure 3. Partial dependency plots of the predicted probability of development with variation in the most important independent variables for BRT models for the full time interval 2001–2011 (thick black) and the two subintervals 2001–2006 (dashed gray) and 2006–2011 (thin black). Plots are shown for (**a**) population density, (**b**) distance to road, (**c**) relative change in population density, (**d**) distance to built, (**e**) distance to cities, (**f**) population density change and (**g**) median income. Rug plots are included on both axes in each plot to show percentiles within the data space for each time interval. X-axis rugs show the distribution of values in the driving variable, and y-axis rugs show the distribution of predicted probabilities for each time interval, using the same style as the graph lines for each interval. Where the percentiles were the same for both time intervals, the rugs overlap. Plots depict only the ranges of the independent variables over which the probability of development showed variation, while percentiles were calculated for the full range of the dataset. Note the greater y-axis range for (**a**) than for the other plots.

PDPs also demonstrate a particularly strong influence of population density on the probability of development with other variables held at their mean values (4a). The probability of development captured by this PDP ranges from 0.06%–2.15% for the full time interval, 0.04%–1.73% for 2001–2006 and 0.03%–0.65% for 2006–2011; for other variables, the effects captured by PDPs were much weaker, with the largest range of values for a partial dependency plot being for population density change for 2006–2011 and 2001–2011 (Figure 3). The weak effects shown in PDPs for variables other than population density reflect the relative lack of forest loss at and below the mean population of 85/km^2; because PDPs by definition hold all variables at their mean value, they do not show whether additional variation might occur in the interaction with population density (or other variables).

Probability surfaces generated from the BRT models for 2001–2011 and the two sub-intervals show that overall patterns of development projected by BRTs were generally consistent among intervals: the areas of highest development probability were the areas closest to urban areas, especially evident for the major urban centers (Boston, Hartford and New Haven; Figure 4). The major difference between the probability surfaces for the two sub-intervals was that the areas of elevated probability of development extended somewhat further from the urban centers for 2001–2006 than for 2006–2011. The probability surfaces also illustrate the higher probability of forest conversion for southern New Hampshire counties north of Boston for 2006–2011 than for 2001–2006, consistent with Figure 1.

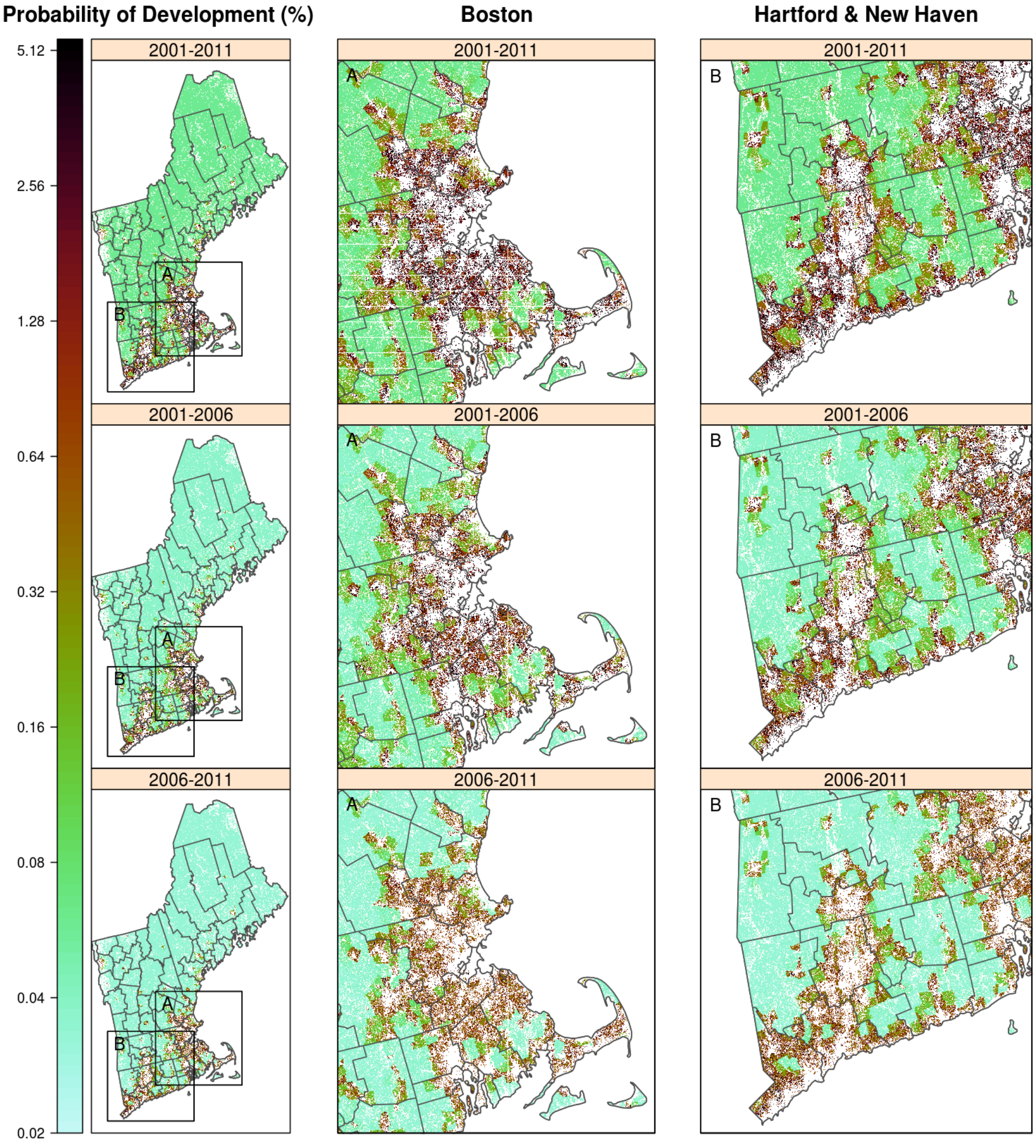

Figure 4. Probability of development surfaces for BRT models for 2001–2011 (**top**) and for the two sub-intervals, 2001–2006 (**middle**) and 2006–2011 (**bottom**). The second and third columns show closeups of the areas of highest development probability, the area around Boston (A) and the area around New Haven and Hartford (B). For each time interval, only pixels containing forest in 2011 are shown in color. Other land cover types are shown in white. The color scheme was generated using David Green's "cubeHelix" implemented in the R package "rje" (Green 2011, Evans 2014). Non-forest land cover is shown in white, and county boundaries are shown in dark gray.

4. Discussion

Our analysis indicates that across New England, the key predictors of forest conversion to development are population and built infrastructure: population density was the most important predictor of forest conversion to development, followed by distance to roads. Distance from built land cover and relative change in population also emerged as predictor variables. Variation captured by the models ranged from 2%–7%, which are high r^2 values, given the low percentage of forest developed during each time interval, the high spatial resolution of our analysis and the complexity of processes underlying human decisions of where to develop: it has been shown that, in general, land change models have a low rate of success in predicting locations of development, particularly at high resolutions [55]. The difference in model performance between the 2001–2011 and 2006–2011 time intervals, however, strongly suggests a qualitative difference in development patterns between the two time intervals. For 2006–2011, less variation was captured, indicating that variables not included in our model were more important in that time interval.

Our results provide unique detail of the spatial trends in forest conversion between 2001 and 2011. First, our qualitative analysis illustrates that, at the county-level resolution, the highest rates of forest loss were further from Boston's urban core in 2006–2011 than in 2001–2006 (Figure 1), suggesting a continuation of the trend observed by [16]. In particular, Rockingham county in southeastern New Hampshire and Berkshire county in western Massachusetts showed an increase in development rates despite decreased development for the region as a whole. Second, our quantitative analysis provides additional detail to the relationship between development and urbanization gradients: the most intense areas of forest loss occurred at intermediate population densities (100–600/km^2) and, for the 2001–2006 interval, at moderate distances (1–5 km) from city centers (Figure 3a,e). The range of population densities with the highest development rates bridges the U.S. Census definition of urban areas (>386 people per km^2), as well as Theobald's definition of the boundary between suburban and exurban (147 housing units per km^2; roughly 367.5 people per km^2) [56]. Overall, our models illustrate with unprecedented specificity that the most intense areas of development in the region are at low to moderate density and on the urban fringes. If maintained over time, this pattern of development is expected to produce a shifting "front" of maximum development rates at an increasing distance from urban centers, consistent with the pattern noted by [16] and with historical analysis of housing records for the region [24].

The differences between 2001–2006 and 2006–2011 demonstrated by our analysis provide hints to the effect of economic environment on development patterns. We found less overall development in 2006–2011 than 2001–2006, a pattern correlated with, and likely driven by, the economic downturn that began in the latter half of the 2000s, sometimes called the "Great Recession." Both nationally and in the northeastern United States, housing starts declined steadily from 2005–2009, with 2009 housing starts the lowest in decades, and construction has only partially recovered in the years since ([57], http://www.census.gov/construction/nrc/historical_data/). We found that the spatial distribution in development differed, as well. The high resolution probability surfaces in Figure 4 show that for both intervals, the greatest development was in the vicinity of major urban areas, but for 2001–2006, the zone of elevated development extended further from urban centers than for 2006–2011. This fine-granularity trend suggests a pattern of reduced exurban development near each urban center, counter to the county-level pattern of regional development shifting further from the densest urban areas that is illustrated in Figure 1. Another intriguing difference between the two time intervals is the increased relative importance of median income in the model of the development probability for the 2006–2011 interval. Interestingly, the partial dependency plot indicates that the highest development rates were for areas with lower median income for 2006–2011 (Figure 3g). The economic reason for this pattern is not clear. It is possible that the high rate of development for low income areas relates to amenities development in previously rural areas. Alternatively, the larger portion of the variation for median income may simply not be captured in the "slice" of the overall trend captured by the PDP.

A few differences between our results and previous studies should be noted. In contrast with more local analyses within New England [19,26], we did not find that topographic variables were important predictors of development. We also did not find evidence that recreation counties were a predictor of development, despite the high levels of growth in recreation counties and the importance of second homes suggested by the national analysis by [25]. The lack of topographic effects might be explained by masking by correlated variables associated with existing developed land. In particular, the variable "distance to road" is a major contributor to our models and is weakly collinear with both elevation and slope (Table S1). The implications of the lack of an effect of recreation counties in our analysis is less clear. It might be that there is no significant effect of recreation and retirement counties on development when our other variables are taken into account. The role of recreation as a driver of development warrants further investigation.

Our results need to be interpreted within the context of the intrinsic limitations of the NLCD. The accuracy assessment for the NLCD 2001–2006 change indicates that the accuracy of detecting land cover change in the Northeast is relatively poor compared to other regions [58]. In particular, Wickham and coauthors found that user accuracy for change detection was 69% for the Northeast, compared to 84% nationally; for forest loss, it was 62% compared to 82% nationally; and for urban gain, it was 58% compared to 72% nationally [58]. Incorrect classification in the NLCD dataset is particularly common in heterogeneous landscapes [58].

Nevertheless, our results are broadly consistent with existing land cover change models, such as the Spatially Explicit Regional Growth Model (SERGoM)development model [56]. SERGoM has been used to simulate development in the United States in several contexts including the USDA "Forests on the Edge" project [59,60] and the EPA "Integrated Climate and Land Use Scenarios" (ICLUS) project [12]. In the SERGoM model, the development of new housing units is assumed to be driven first by county-level population growth and second by existing housing unit density and travel time to urban cores [56]. Our county-level analysis shows a pattern of forest loss that is indeed generally correlated with data on changing population density during the 2000–2010 time interval [61]. In our statistical models, however, changing population was a relatively minor predictor, probably reflecting the high spatial resolution of our analysis compared to county-level trends. The variables of greatest importance in our models do, however, reflect similar allocation patterns to the assumptions of SERGoM. Our metrics of "distance to road" and "distance to cities" are closely related to travel time to urban cores, and "population density" is related to both housing unit density and degree of urbanization. This consistency reinforces the validity of both our approach and the SERGoM approach and points to the future usefulness of BRT models like ours as a novel statistical tool for capturing highly specific predictors of land cover change.

Our analysis makes use of region-wide datasets to produce the first high resolution models of geographic predictors of development for all of New England. Our results demonstrate the capabilities of the BRT methodology for statistically modeling land cover change and for illustrating temporal variation in the importance of different predictor variables. In the future, our region-wide probability surfaces will provide input for simulation models of land cover change in the region, and the methods presented here will translate well to land cover change analyses in other regions and are expected to increase in usefulness as new NLCD maps and additional geographical datasets become available.

Supplementary Materials: The following is available online at http://www.mdpi.com/2073-445X/5/3/30/s1, Table S1. Correlation matrix of all variables evaluated.

Acknowledgments: We thank Scott Ollinger for ideas in the early development of this project, Meghan Blumstein for the contribution of scripts and Christian Levers for help with the interpretation of the boosted regression tree analyses. This research was funded by grants from the National Science Foundation (NSF-DEB 12-37491, EPS-1101245).

Author Contributions: A.M.T., J.R.T. and J.S.P. conceived of and designed the analysis and collaboratively selected input variables. J.S.P. compiled input data. A.M.T. performed the analysis and produced summary statistics and graphics. A.M.T. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; nor in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BRT: boosted regression tree

NLCD: National Land Cover Database

Appendix

Table A1. Parameter calibration for boosted regression trees for models of forest conversion to development, 2001–2011. Ten-fold cross-validation correlation coefficients were determined for all combinations of tree complexity from 1–15 with six learning rates ranging from 0.01–0.0005. The mean correlation coefficients across learning rates are shown.

Tree Complexity	**Learning Rate**						
	0.0100	**0.0075**	**0.0050**	**0.0025**	**0.0010**	**0.0005**	**Row Mean**
1	0.2150	0.2171	0.2212	0.2329	0.2396	0.2416	0.2279
2	0.2167	0.2273	0.2291	0.2353	0.2488	0.2529	0.2350
3	0.2447	0.2539	0.2522	0.2556	0.2615	0.2609	0.2548
4	0.2642	0.2679	0.2590	0.2644	0.2744	0.2766	0.2678
5	0.2585	0.2461	0.2648	0.2706	0.2762	0.2805	0.2661
6	0.2514	0.2610	0.2679	0.2699	0.2744	0.2794	0.2673
7	0.2688	0.2602	0.2658	0.2680	0.2688	0.2843	0.2693
8	0.2620	0.2654	0.2570	0.2753	0.2822	0.2860	0.2713
9	0.2723	0.2673	0.2684	0.2679	0.2782	0.2842	0.2730
10	0.2657	0.2536	0.2700	0.2752	0.2876	0.2899	0.2736
11	0.2569	0.2749	0.2787	0.2757	0.2856	0.2899	0.2770
12	0.2766	0.2892	0.2640	0.2752	0.2840	0.2920	0.2802
13	0.2659	0.2658	0.2821	0.2756	0.2847	0.2854	0.2766
14	0.2662	0.2777	0.2682	0.2719	0.2840	0.2857	0.2756
15	0.2735	0.2732	0.2664	0.2721	0.2827	0.2882	0.2760

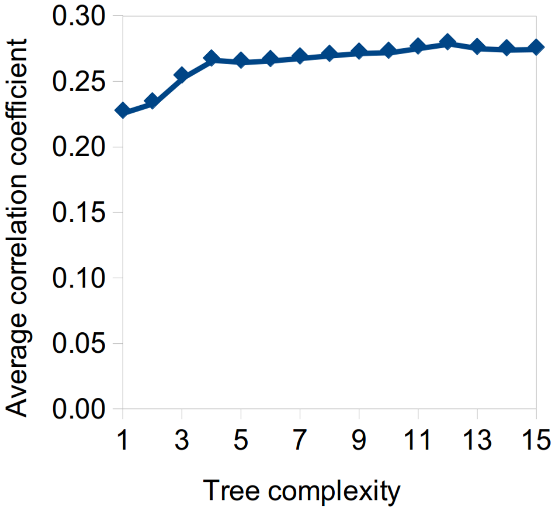

Figure A1. Mean correlation coefficient across learning rates (0.01–0.0005) for 10-fold cross-validation of the BRT model for forest conversion in the interval 2001–2011 evaluating all learning rates from 0.01–0.0005 paired with all tree complexities from 1–15.

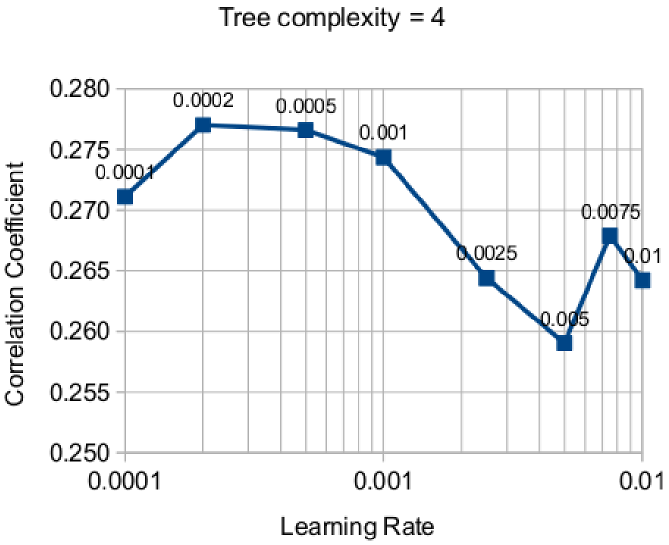

Figure A2. Semilog plot of the correlation coefficients for 10-fold cross-validation of the BRT model for forest conversion in the interval 2001–2011, with tree complexity equal to four and learning rates from 0.01–0.0001.

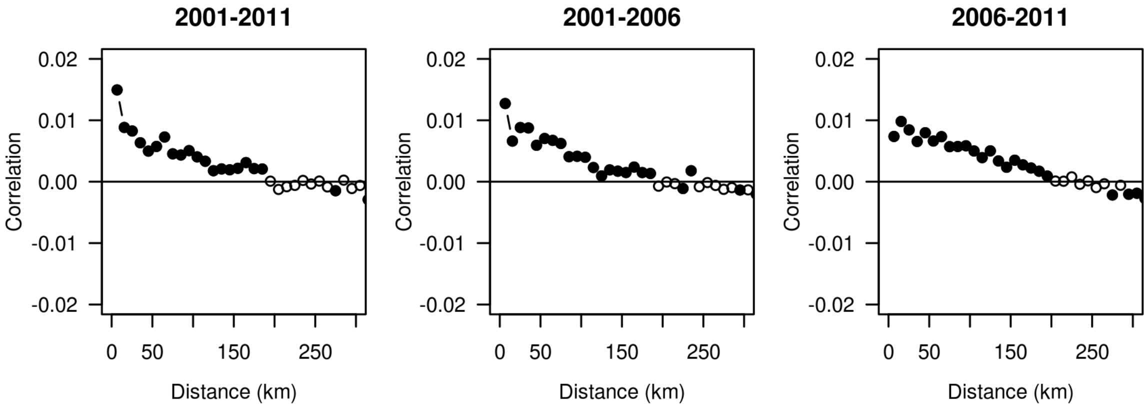

Figure A3. Correlograms of Moran's I spatial autocorrelation of residuals from BRT models of the probability of development for the time intervals 2001–2011, 2001–2006 and 2006–2011. Points with significance at the α=0.05 level are filled.

References

1. Breunig, K. *Losing Ground: At What Cost? Changes in Land Use and Their Impacts on Habitat, Biodiversity and Ecosystem Services in Massachusetts*; Summary Report; Advocacy Department, Mass Audubon: Lincoln, MA, USA, 2003.
2. Foster, D.; Donahue, B.M.; Kittredge, D.B.; Lambert, K.F.; Hunter, M.L.; Hall, B.R.; Irland, L.C.; Lilieholm, R.J.; Orwig, D.A.; D'Amato, A.W.; et al. *Wildlands and Woodlands: A Vision for the New England Landscape*; Harvard Forest; Harvard University Press: Cambridge, MA, USA, 2010.
3. Levesque, C.A. *New Hampshire Statewide Forest Resources Assessment—2010: Important Data and Information about New Hampshire's Forests*; New Hampshire Department of Resources and Economic Development Division of Forests and Lands: Concord, NH, USA, 2010.
4. Zheng, D.; Heath, L.S.; Ducey, M.J.; Butler, B. Relationships between major ownerships, forest aboveground biomass distributions, and landscape dynamics in the New England Region of USA. *Environ. Manag.* **2010**, *45*, 377–386.

5. Zheng, D.; Heath, L.S.; Ducey, M.J. Potential overestimation of carbon sequestration in the forested wildland-urban interface in Northern New England. *J. For.* **2012**, *110*, 105–111.
6. Zheng, D.; Ducey, M.J.; Heath, L.S. Assessing net carbon sequestration on urban and community forests of northern New England, USA. *Urban For. Urban Green.* **2013**, *12*, 61–68.
7. Blumstein, M.; Thompson, J.R. Land-use impacts on the quantity and configuration of ecosystem service provisioning in Massachusetts, USA. *J. Appl. Ecol.* **2015**, *52*, 1009–1019.
8. USDA Forest Service. *Standard Tables of Forest Carbon Stock Estimates by State*; USDA Forest Service: Washington, DC, USA, 2014.
9. Smith, J.E.; Heath, L.S.; Nichols, M.C. *US Forest Carbon Calculation Tool: Forest-Land Carbon Stocks and Net Annual Stock Change*; General Technical Report NRS-13; USDA Forest Service, Northern Research Station: Durham, NH, USA, 2007.
10. U.S. Environmental Protection Agency. Land Use, Land-Use Change, and Forestry. In *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2013*; Number EPA 430-R- 15 -004; U.S. Environmental Protection Agency: Washington, DC, USA, 2015.
11. Meyer, S.R.; Johnson, M.L.; Lilieholm, R.J.; Cronan, C.S. Development of a stakeholder-driven spatial modeling framework for strategic landscape planning using Bayesian networks across two urban-rural gradients in Maine, USA. *Ecol. Model.* **2014**, *291*, 42–57.
12. Bierwagen, B.G.; Theobald, D.M.; Pyke, C.R.; Choate, A.; Groth, P.; Thomas, J.V.; Morefield, P. National housing and impervious surface scenarios for integrated climate impact assessments. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20887–20892.
13. Stein, S.M.; McRoberts, R.E.; Mahal, L.G.; Carr, M.A.; Alig, R.J.; Comas, S.J.; Theobald, D.M.; Cundiff, A. *Private Forests, Public Benefits: Increased Housing Density and Other Pressures on Private Forests Contributions*; General Technical Report PNW-GTR-795; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 2009.
14. Thompson, J.; Fallon Lambert, K.; Foster, D.; Blumstein, M.; Broadbent, E.; Almeyda Zambrano, A. *Changes to the Land: Four Scenarios for the Future of the Massachusetts Landscape*; Technical Report; Harvard Forest: Petersham, MA, USA, 2014.
15. Drummond, M.A.; Loveland, T.R. Land-use pressure and a transition to forest-cover loss in the Eastern United States. *BioScience* **2010**, *60*, 286–298.
16. Jeon, S.B.; Olofsson, P.; Woodcock, C.E. Land use change in New England: A reversal of the forest transition. *J. Land Use Sci.* **2014**, *9*, 105–130.
17. Alcamo, J.; Kok, K.; Busch, G.; Priess, J.A.; Eickhout, B.; Rounsevell, M.; Rothman, D.S.; Heistermann, M. Searching for the future of land: Scenarios from the local to global scale. In *Land-Use and Land-Cover Change*; Springer: Berlin, Germany, 2006; pp. 137–155.
18. Tayyebi, A.; Pijanowski, B.C.; Linderman, M.; Gratton, C. Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world. *Environ. Model. Softw.* **2014**, *59*, 202–221.
19. Thompson, J.R.; Foster, D.R.; Scheller, R.; Kittredge, D. The influence of land use and climate change on forest biomass and composition in Massachusetts, USA. *Ecol. Appl.* **2011**, *21*, 2425–2444.
20. Rounsevell, M.; Reginster, I.; Araújo, M.; Carter, T.; Dendoncker, N.; Ewert, F.; House, J.; Kankaanpää, S.; Leemans, R.; Metzger, M.; et al. A coherent set of future land use change scenarios for Europe. *Agric. Ecosyst. Environ.* **2006**, *114*, 57–68.
21. Schneider, L.C.; Pontius, R.G. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agric. Ecosyst. Environ.* **2001**, *85*, 83–94.
22. Sohl, T.; Sayler, K. Using the FORE-SCE model to project land-cover change in the southeastern United States. *Ecol. Model.* **2008**, *219*, 49–65.
23. Radeloff, V.C.; Nelson, E.; Plantinga, A.J.; Lewis, D.J.; Helmers, D.; Lawler, J.J.; Withey, J.C.; Beaudry, F.; Martinuzzi, S.; Butsic, V.; et al. Economic-based projections of future land-use under alternative economic policy scenarios in the conterminous US. *Ecol. Appl.* **2012**, *22*, 1036–1049.
24. Mockrin, M.H.; Stewart, S.I.; Radeloff, V.C.; Hammer, R.B.; Johnson, K.M. Spatial and temporal residential density patterns from 1940 to 2000 in and around the Northern Forest of the Northeastern United States. *Popul. Environ.* **2012**, *34*, 400–419.

25. Johnson, K.M.; Beale, C.L. Nonmetro recreation counties: Their identification and rapid growth. *Rural Am.* **2002**, *17*, 4.
26. Tyrrell, M.; Hall, M.; Sampson, R. *Dynamic Models of Land Use Change in Northeastern USA: Deveoping Tools Techniques, and Talents for Effective Conservation Action*; Yale University School of Forestry and Environmental Studies: New Haven, CT, USA, 2004.
27. Duveneck, M.J.; Thompson, J.R.; Wilson, B.T. An imputed forest composition map for New England screened by species range boundaries. *For. Ecol. Manag.* **2015**, *347*, 107–115.
28. Foster, D.R., Aber, J.D., Eds. *Forests in Time: The Environmental Consequences of 1000 Years of Change in New England*; Yale University Press: New Haven, CT, USA, 2004.
29. Homer, C.G.; Dewitz, J.A.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.; Coulston, J.; Herold, N.D.; Wickham, J.D.; Megown, K. Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 345–354
30. Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. The National Elevation Dataset. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–11
31. QGIS Development Team. *QGIS Geographic Information System*; Open Source Geospatial Foundation Project; Open Source Geospatial Foundation: Beaverton, OR, USA, 2015.
32. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.
33. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46.
34. Levers, C.; Verkerk, P.J.; Müller, D.; Verburg, P.H.; Butsic, V.; Leitão, P.J.; Lindner, M.; Kuemmerle, T. Drivers of forest harvesting intensity patterns in Europe. *For. Ecol. Manag.* **2014**, *315*, 160–172.
35. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813.
36. Elith, J.; Graham, C.H.; Anderson, R.P.; Dudík, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **2006**, *29*, 129–151.
37. Leathwick, J.R.; Elith, J.; Francis, M.P.; Hastie, T.; Taylor, P. Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* **2006**, *321*, 267–281.
38. Marmion, M.; Luoto, M.; Heikkinen, R.K.; Thuiller, W. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecol. Model.* **2009**, *220*, 3512–3520.
39. Thomaes, A.; Kervyn, T.; Maes, D. Applying species distribution modelling for the conservation of the threatened saproxylic Stag Beetle (*Lucanus cervus*). *Biol. Conserv.* **2008**, *141*, 1400–1410.
40. Li, X.; Wang, Y. Applying various algorithms for species distribution modelling. *Integr. Zool.* **2013**, *8*, 124–135.
41. Johnstone, J.F.; Hollingsworth, T.N.; Chapin, F.S.; Mack, M.C. Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. *Glob. Chang. Biol.* **2010**, *16*, 1281–1295.
42. Greve, M.; Lykke, A.M.; Blach-Overgaard, A.; Svenning, J.C. Environmental and anthropogenic determinants of vegetation distribution across Africa: Determinants of African vegetation distribution. *Glob. Ecol. Biogeogr.* **2011**, *20*, 661–674.
43. Snelder, T.H.; Lamouroux, N.; Leathwick, J.R.; Pella, H.; Sauquet, E.; Shankar, U. Predictive mapping of the natural flow regimes of France. *J. Hydrol.* **2009**, *373*, 57–67.
44. Nolan, B.T.; Fienen, M.N.; Lorenz, D.L. A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *J. Hydrol.* **2015**, *531*, 902–911.
45. Marmion, M.; Hjort, J.; Thuiller, W.; Luoto, M. A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surf. Process. Landf.* **2008**, *33*, 2241–2254.
46. Martin, M.P.; Wattenbach, M.; Smith, P.; Meersmans, J.; Jolivet, C.; Boulonne, L.; Arrouays, D. Spatial distribution of soil organic carbon stocks in France. *Biogeosciences* **2011**, *8*, 1053–1065.
47. Mosleh, Z.; Salehi, M.H.; Jafari, A.; Borujeni, I.E.; Mehnatkesh, A. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. *Environ. Monit. Assess.* **2016**, *188*, 195.

48. Parisien, M.A.; Moritz, M.A. Environmental controls on the distribution of wildfire at multiple spatial scales. *Ecol. Monogr.* **2009**, *79*, 127–154.
49. Müller, D.; Leitão, P.J.; Sikor, T. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. *Agric. Syst.* **2013**, *117*, 66–77.
50. Linard, C.; Tatem, A.J.; Gilbert, M. Modelling spatial patterns of urban growth in Africa. *Appl. Geogr.* **2013**, *44*, 23–32.
51. Aertsen, W.; Kint, V.; van Orshoven, J.; Özkan, K.; Muys, B. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol. Model.* **2010**, *221*, 1119–1130.
52. Ridgeway, G. *GBM: Generalized Boosted Regression Models*; R Package Version 2.1.1; R Foundation for Statistical Computing: Vienna, Austria, 2015.
53. Hijmans, R.J.; Phillips, S.; Leathwick, J.; Elith, J. *dismo: Species Distribution Modeling*; R Package Version 1.0-12; R Foundation for Statistical Computing: Vienna, Austria, 2015.
54. Ridgeway, G. *Generalized Boosted Regression Models: A Guide to the GBM Package*; 2012.
55. Pontius, R.G.; Boersma, W.; Castella, J.C.; Clarke, K.; de Nijs, T.; Dietzel, C.; Duan, Z.; Fotsing, E.; Goldstein, N.; Kok, K.; et al. Comparing the input, output, and validation maps for several models of land change. *Annal. Reg. Sci.* **2008**, *42*, 11–37.
56. Theobald, D.M. Landscape patterns of exurban growth in the USA from 1980 to 2020. *Ecol. Soc.* **2005**, *10*, 32.
57. Woodall, C.W. An overview of the forest products sector downturn in the United States. *For. Prod. J.* **2011**, *61*, 595–603.
58. Wickham, J.D.; Stehman, S.V.; Gass, L.; Dewitz, J.; Fry, J.A.; Wade, T.G. Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sens. Environ.* **2013**, *130*, 294–304.
59. Stein, S.; McRoberts, R.; Alig, R.J.; Nelson, M.; Theobald, D.; Eley, M.; Dechter, M.; Carr, M. *Forests on the Edge: Housing Development on America's Private Forests*; General Technical Report PNW-GTR-636; U.S. Department of Agriculture, Forest Service, Pacific Northwest REsearch STation: Portland, OR, USA, 2005.
60. Stein, S.M.; Alig, R.J.; White, E.M.; Comas, S.J.; Carr, M.; Eley, M.; Elverum, K.; O'Donnell, M.; Theobald, D.M.; Cordell, K. *National Forests on the Edge: Development Pressures on America's National Forests and Grasslands*; General Technical Report PNW-GTR-728; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 2007.
61. Mackun, P.J.; Wilson, S.; Fischetti, T.R.; Goworowska, J. *Population Distribution and Change: 2000 to 2010*; U.S. Census Briefs; US Department of Commerce, Economics and Statistics Administration, US Census Bureau: Washington, DC, USA, 2011.