

Modeling urban distributions of host trees for invasive forest insects in the eastern and central USA: A three-step approach using field inventory data



Frank H. Koch^{a,*}, Mark J. Ambrose^b, Denys Yemshanov^c, P. Eric Wiseman^d, F.D. Cowett^e

^a USDA Forest Service, Southern Research Station, Eastern Forest Environmental Threat Assessment Center, 3041 East Cornwallis Road, Research Triangle Park, NC 27709 USA

^b North Carolina State University, Department of Forestry and Environmental Resources, 3041 East Cornwallis Road, Research Triangle Park, NC 27709 USA

^c Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen Street East, Sault Ste. Marie, ON P6A 2E5 Canada

^d Virginia Tech, Department of Forest Resources and Environmental Conservation, 228 Cheatham Hall, 310 West Campus Drive, Blacksburg, VA 24061 USA

^e Cornell University, Horticulture Section, School of Integrative Plant Science, Ithaca, NY 14853 USA

ARTICLE INFO

Keywords:

Forest pests
Invasive alien species
Urban forest inventory
Host distribution modeling
Human-mediated dispersal

ABSTRACT

Despite serving as invasion gateways for non-native forest pests, urban forests are less well understood than natural forests. For example, only a fraction of communities in the USA and Canada have completed urban forest inventories, and most have been limited to street trees; sample-based inventories that provide valid community-wide estimates of urban forest composition are much rarer. As a proof of concept, we devised a three-step approach to model urban tree distributions regionally using available street tree and whole-community inventory data. We illustrate the approach for three tree genera – ash (*Fraxinus* spp.), maple (*Acer* spp.), and oak (*Quercus* spp.) – that are hosts for high-profile insect pests. The objective of the first step was to estimate, for communities with only street tree inventories, the proportion of the community's total basal area (BA) in each host genus. Utilizing data from communities with paired street tree and whole-community inventories, we applied polynomial regression to estimate whole-community BA proportion per genus as a function of a community's street tree BA proportion and its geographic location. The objective of the second step was to estimate per-genus BA proportions for communities in our prediction region (eastern and central USA) with no urban forest inventory. We used stochastic gradient boosting to predict these proportions as a function of environmental and other variables. In the third step, we developed a generalized additive model for estimating the total BA of a community as a function of its canopy cover, geographic location, and area. We then combined the outputs from the second and third steps to estimate ash, maple, and oak BA for the nearly 24,000 communities in our prediction region. By merging these estimates with similar information on natural forests, we can provide more complete representations of host distributions for pest risk modeling, spread modeling, and other applications.

1. Introduction

Invasive species have tremendous impacts globally, including disruption of ecosystem functions, loss of important agricultural crops, declines and extinctions of native species, damage to infrastructure, and direct as well as indirect (e.g., as a vector) effects on human health (Parker et al., 1999; Allen and Humble, 2002; Clavero and García-Berthou, 2005; Bradshaw et al., 2016). These impacts are challenging to specify in economic terms. For example, insects, which comprise one of the largest classes of invasive species, recently were estimated to have an annual impact of US\$77 billion worldwide in terms of direct losses of goods and services, control costs, and associated human health costs

(Bradshaw et al., 2016). However, because there have been few dedicated assessments of the economic impacts of insects, this number likely underestimates the true costs by a large margin (Bradshaw et al., 2016).

Forest pests (i.e., insects and diseases that affect trees) account for a considerable fraction of the impacts of all invasive pests of plants (Liebholt et al., 1995; Kenis et al., 2009; Paini et al., 2016). For instance, more than 450 non-native forest insect species have become established in the continental USA since European settlement (Aukema et al., 2010). Out of these, a subset of 62 high-impact species were estimated to cost nearly US\$1.7 billion annually in government expenditures for management and control, and another US\$830 million in lost residential property values (Aukema et al., 2011). By changing

* Corresponding author.

E-mail addresses: fhkoch@fs.fed.us (F.H. Koch), mambrose@fs.fed.us (M.J. Ambrose), Denys.Yemshanov@canada.ca (D. Yemshanov), pwiseman@vt.edu (P.E. Wiseman), fdc2@cornell.edu (F.D. Cowett).

<https://doi.org/10.1016/j.foreco.2018.03.004>

Received 6 October 2017; Received in revised form 26 February 2018; Accepted 3 March 2018
0378-1127/ © 2018 Published by Elsevier B.V.

forest composition and structure at a regional scale, forest pests also affect critical ecosystem functions such as nutrient cycling and wildlife habitat (Lovett et al., 2016). In urban forests (i.e., forests in cities, suburbs, and other human settlements), extensive tree loss caused by pests has been linked to negative impacts on human cardiovascular and respiratory health (Donovan et al., 2013).

As with other categories of invaders, most forest pest invasions have occurred as a result of human activities, especially the intentional or accidental movement of species by trade or travel (Mack et al., 2000; Perrings et al., 2002; Hulme, 2009). Urban forests play a principal role in facilitating such invasions (US Government Accountability Office, 2006; Colunga-Garcia et al., 2010b, 2010a; Koch et al., 2011). Many invasive species, including forest pests, are discovered initially in urbanized areas, which are frequent destinations for international cargo and passengers (Colunga-Garcia et al., 2010b; Huang et al., 2012; Liebhold et al., 2013). In the USA, insect species such as the European gypsy moth (*Lymantria dispar dispar* L.; EGM), emerald ash borer (*Agrilus planipennis* Fairmaire; EAB), and Asian longhorned beetle (*Anoplophora glabripennis* (Motschulsky); ALB) first emerged as pests in urban forests (Liebhold et al., 1995; Poland and McCullough, 2006; Dodds and Orwig, 2011). Historically, urban forests have high rates of invasive pest introductions (e.g., insects harbored by live plants for landscaping purposes) and may provide habitat that is more conducive to invasive pest establishment than natural forests, with fewer natural enemies, greater abundance of preferred hosts, and favorably altered environmental conditions, for example due to the urban heat island effect (Alvey, 2006; McKinney, 2006). Furthermore, portions of many urban forests in North America are dominated by a single tree species, and the trees are often planted closely together, facilitating the spread of any pest for which that species is a host (Greene and Millward, 2016). Once established in urban forests, invaders may expand into surrounding natural forests (Tait et al., 2005; Alvey, 2006).

Urban forests are not just invasion gateways; for some pests, the most noteworthy impacts actually occur in urban forests. For example, Haack et al. (2010) reported that communities in Illinois, Massachusetts, New Jersey, and New York spent more than US\$373 million on ALB eradication efforts between 1996 and 2008, primarily for identification and removal of infested trees. Based on data from communities across the USA, Hauer and Peterson (2017) estimated the nationwide impact of EAB on municipal forestry budgets to be US\$280.5 ($\pm$ 79) million annually; typically, municipal governments in invaded communities spent twice as much on tree removal as in communities where EAB was not present. Kovacs et al. (2010) predicted that the expansion of EAB in the eastern USA between 2009 and 2019 would necessitate treatment or removal and replacement of more than 17 million ash (*Fraxinus* spp.) trees – both publicly and privately owned – within invaded communities, at a total cost of US\$10.7 billion. Similarly, Sydnor et al. (2007) predicted eventual losses (including losses in landscape value as well as tree removal and replacement costs) of US\$1.8–US\$7.6 billion in Ohio communities as a result of EAB expansion, while combined losses in Illinois, Indiana, Michigan, and Wisconsin were predicted to reach between US\$13.4 and US\$26 billion (Sydnor et al., 2011). The potential economic impact of ALB in the USA could be even larger: Nowak et al. (2001b) projected a total value loss of US\$669 billion (based on compensatory value of trees) if the insect were to spread to communities throughout the country, as the preferred hosts of ALB represent approximately 30% of all urban trees.

The impact projections in these studies were constrained by a lack of information about the distributions of host trees in the communities of interest. Kovacs et al. (2010) asserted that their estimates could be improved markedly through systematic inventories of the communities' forests. The concept of urban forest inventory, performed according to standard protocols, gained momentum in the mid-2000s with advances in mobile data collection and online management of geospatial data (Abd-Elrahman et al., 2010; Miller et al., 2015), as well as the release of urban forestry software applications such as i-Tree Eco (Nowak et al.,

2008b) and i-Tree Streets (Maco and McPherson, 2003; McPherson et al., 2005). (Both applications are included in the freely available i-Tree software suite, <http://www.itreetools.org/>.) Although these applications are promoted as tools to model potential benefits of urban forests such as improved air quality and reduced building energy use (Maco and McPherson, 2003; Nowak and Dwyer, 2007), the basic inventory information that they collect has clear utility for management, including for invasive forest pests. For example, i-Tree Eco data have been applied in regional, continental, and global analyses of urban forest composition and tree species diversity (e.g., Yang et al., 2015; Blood et al., 2016; Jenerette et al., 2016), both of which affect how those forests respond to invasions or other types of disturbances. Integrating urban forest inventory data with similar data for natural forests would provide a stronger foundation for forest health monitoring and early pest detection, mitigation and control efforts, spread modeling, and risk mapping (Dwyer et al., 2000; BenDor and Metcalf, 2006; Cumming et al., 2008; Venette et al., 2010; Hudgins et al., 2017).

Unfortunately, even in a relatively data-rich country like the USA, there are pronounced data gaps with respect to urban forests, especially in comparison to the far more comprehensive data available for natural forests. For instance, the USDA Forest Service's Forest Inventory and Analysis (FIA) Program performs systematic, annualized inventories of rural forestlands throughout the USA. The annualized FIA plot network includes $\approx$ 135,000 forested plots nationwide, which translates to approximately one plot per 2400 ha of natural forest. At this sampling intensity, and because they are distributed systematically across all forestlands (i.e., excluding urban forests), FIA plots can be used to develop regional-scale tree species distribution maps via spatial interpolation and statistical methods (e.g., Iverson et al., 1999; Moisen et al., 2006). In contrast, the FIA Program only recently embarked on urban forest inventories, conducting its first such inventory in 2014 (see Nowak et al., 2016). Furthermore, urban FIA data collection has been proposed for just a limited number (< 100) of relatively large cities (USDA Forest Service, 2016). Upon completion, these efforts will provide consistent measurements of trees in the targeted cities, but will not be sufficiently representative of the thousands of communities, both large and small, across the USA.

Despite its limitations, the urban FIA initiative stands as a significant step forward; prior to this, urban inventory data were not collected systematically in the USA (Roman et al., 2013) or elsewhere, although it is worth noting that communities throughout Sweden recently have begun systematic collection of urban tree data (primarily street tree data) according to standardized inventory protocols (see Östberg, 2013; Östberg et al., 2013). In turn, the lack of systematically collected data has restricted the scale of scientific inquiries regarding urban forests. Indeed, the most extensive analysis of which we are aware (Cowett and Bassuk, 2017) involved street tree data from 275 communities in New Jersey, New York, and Pennsylvania, representing just $\approx$ 8% of the populated places identified by the US Census Bureau in those states. Nevertheless, many communities have completed some type of urban forest inventory for local purposes; for our own work, we have compiled more than 1200 inventory data sets from communities across the USA – especially the eastern USA – and Canada. They differ widely in terms of data collection protocols (e.g., complete inventory vs. sampling across land uses) and target tree populations (e.g., street trees vs. all private and public trees), but they comprise a very large and geographically representative compilation of urban forest inventory data – to our knowledge, the largest such compilation ever assembled. Boyer et al. (2016) identified a need for this sort of compilation to address far-reaching research and management questions at a regional scale. The challenge, of course, is determining how to best make use of the data despite their differences.

We present a model-based approach that utilizes existing inventory data from a sampling of North American communities to characterize urban forests of communities across a large portion of the USA for which inventory data were not available. Specifically, for each

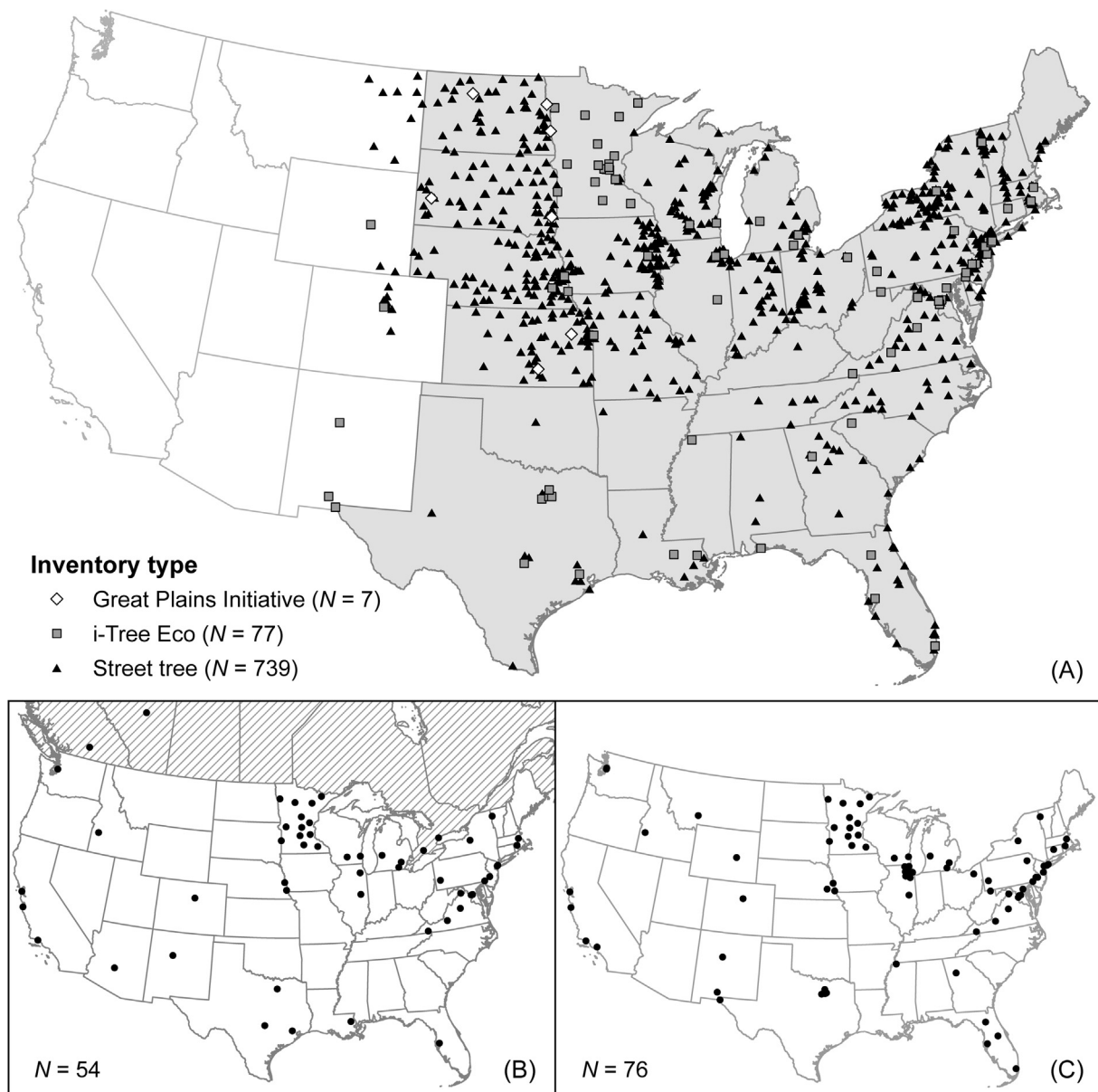


Fig. 1. (A) Communities with urban forest inventories used in the primary training data set, labeled according to inventory type. Target prediction region (eastern and central USA) is shown in gray; (B) “Paired” communities (i.e., communities with both street tree and whole-community inventories) used to model basal area (BA) proportions in communities with only street inventories (model step 1); (C) Communities with total BA estimates, used to model the relationship between canopy cover and BA (model step 3).

community in the eastern and central USA for which we did not have inventories, we used the available data to construct a series of models for estimating the abundance and importance of three tree genera: ash (*Fraxinus* spp.), maple (*Acer* spp.), and oak (*Quercus* spp.). These genera are the preferred hosts for EAB, ALB, and EGM, respectively. Our objective was to fill critical data gaps in the geographic distributions of these host trees in urban areas and thus enable better assessments of the risks and potential impacts posed by each pest. Additionally, our analysis was intended as a proof of concept to show that a multi-step modeling approach, incorporating different types of urban forest inventory data, can fill these data gaps effectively.

2. Material and methods

We define an urban forest as all publicly or privately owned trees growing within the boundaries of a city or similar populated place (Nowak et al., 2001a, 2010). This includes street trees, park trees, and trees on residential or commercial property. Our prediction region

(Fig. 1A) included 37 states and the District of Columbia, and encompassed almost 24,000 populated places (hereafter referred to as “communities”) identified by the US Census Bureau, ranging from very small villages (< 100 residents) to large cities (ESRI, 2014b).

2.1. Urban forest inventory data

We used urban forest inventory data from 842 communities (see Table S1 in Supplement 1). They included all communities in the prediction region for which we had some type of inventory data (see Fig. 1A) as well as additional communities in the western USA and southern Canada that we used in one or more of our modeling steps (see Fig. 1B and C). Fundamentally, the data were a convenience (i.e., non-probabilistic) sample that we compiled from a wide variety of sources. In some cases, we extracted the data from reports published by individual communities, or from open data made available online. In other cases, we obtained data that were collected as part of a regional or statewide initiative to inventory trees in communities according to

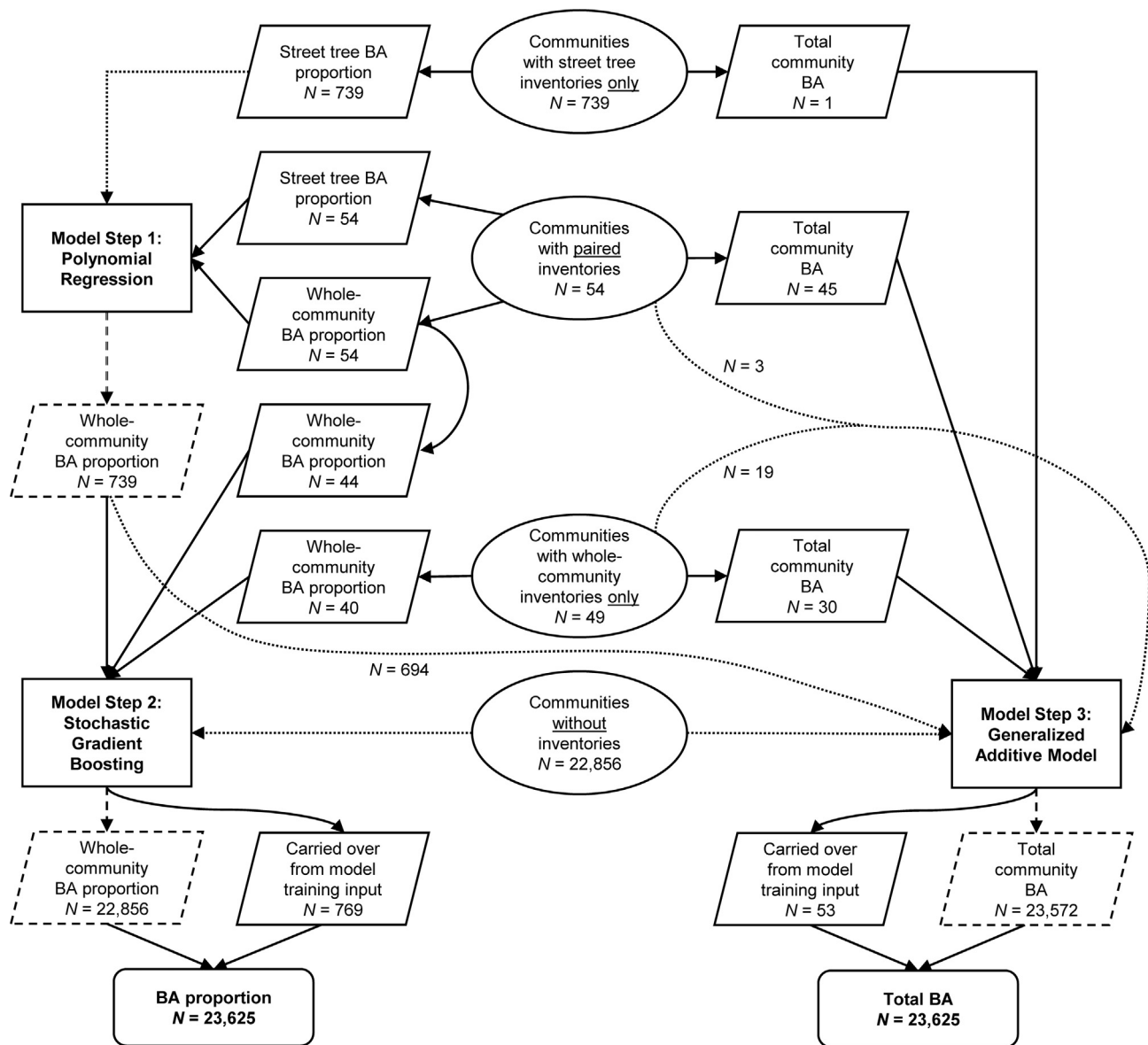


Fig. 2. Conceptual diagram of the three-step modeling approach applied to each host genus. The modeling steps are represented by rectangles, while the final outputs – basal area (BA) proportion and total BA for all communities in the prediction region – are represented by rounded rectangles. Ellipses depict the four types of input community data. Parallelograms indicate a set of values for a metric of interest, either extracted from an input data set (solid border) or generated as output from one of the modeling steps (dashed border). Connector arrows depict the direction of data flow through the modeling process. Solid arrows indicate data used for model training (or appended to one of the final outputs), dotted arrows indicate data used for prediction, and dashed arrows indicate model output. A curved arrow indicates a subset of the data set from which the arrow originated. Sample sizes are shown for each of the diagram elements as appropriate.

standardized methods. Occasionally, we contacted community foresters or arborists directly to get access to unpublished inventory data. Readers are referred to [Supplement 2](#), which is a detailed list of the municipalities, companies, and individuals who provided data for this project.

Regardless of source, the data were broadly from one of two primary inventory types (see [Table S1 in Supplement 1](#) and [Fig. 2](#)): street tree inventories or sample-based inventories that characterized the communities' entire tree populations (hereafter designated as "whole-community" inventories). The large majority were street tree inventories, although a small percentage (6.4%) of communities had both street and whole-community inventories (hereafter designated as "paired" inventories). With respect to both inventory types, there was some variation in how they were implemented, which was likely influenced by community size (i.e., geographic extent) and resources available for implementation. For example, some communities' street

tree inventories included measurements of trees along all streets, while in other communities only a subset of streets were inventoried as a statistical sample (see [Table S1 in Supplement 1](#)). Furthermore, roughly one-third of street tree inventories included measurements of additional publicly owned trees, such as trees growing on the grounds of government buildings or in municipal parks (i.e., managed parks rather than natural areas). For analytical purposes, we treated these "street and public tree" inventories as if they were street tree inventories because they did not extend to any other land uses. There were also some differences between inventories in how they recorded or reported tree measurements. For instance, some communities reported tree diameters (i.e., diameter at breast height, dbh) to the nearest whole unit (typically inches), while others reported them by diameter class. In the latter case, we set the diameter of each inventoried tree to the midpoint of its diameter class. For simplicity, we assumed that tree measurements among the inventories were equally accurate, and that differences in

measurement precision did not affect our results significantly.

Our analyses included whole-community inventory data collected using either the i-Tree Eco approach or the approach adopted by the Great Plains Tree and Forest Invasives Initiative (GPI). Primary aspects of each approach are described in Nowak et al. (2008b) and Lister et al. (2012), respectively. The GPI approach was developed for application to four Great Plains states: North Dakota, South Dakota, Nebraska, and Kansas. Within these states, which are largely non-forested ($\approx 97\%$), the GPI specifically targeted areas and land uses not normally considered forest. Both urban and rural areas in each state were sampled using 0.067-ha circular plots (200 plots in urban and 100 plots in rural areas statewide during the first year of sampling). We only used the urban GPI plots in our analyses and assigned each of those urban plots to the community in which it was located. Unlike the GPI, the typical implementation of i-Tree Eco focuses sampling effort on an individual community. Sampling intensity and plot size vary by community, but most i-Tree Eco inventories include between 100 and 200 plots across the community of interest, and the most common plot size is 0.04 ha, consistent with recommendations from Nowak et al. (2008a).

Most of the inventories were collected during the period 2000–2015 (median year = 2009; see Table S1 in Supplement 1 for a complete listing). Although we planned to omit inventories collected prior to 2000, to maximize the amount of whole-community data that we had available for modeling, we included nine older i-Tree Eco inventories: Atlanta, Georgia (1997); Baltimore, Maryland (1999); Boston, Massachusetts (1996); Brooklyn, New York (1997); Calgary, Alberta (1998); Greenville, South Carolina (1999); Jersey City, New Jersey (1998); New York, New York (1996); and Philadelphia, Pennsylvania (1996). We also used four street tree inventories that were collected prior to 2000: Arlington, Massachusetts (1998); Bordentown, New Jersey (1999); Manchester, New Hampshire (1998); and Mandeville, Louisiana (1996). These inventories helped to fill in portions of our study area where we felt the data were sparse geographically. However, we did not evaluate their geographic representativeness formally.

2.2. Modeling approach

The ultimate goal of our analyses was to estimate basal area (BA) of ash, maple, and oak species in our prediction region using statistical models developed from existing community inventory data. We performed our analyses for entire genera to limit the impact of variable data quality between the inventories with respect to tree species identification; we reasoned that identification errors should be uncommon at the genus level. We chose BA (i.e., the total BA of all trees or a particular subset of the trees in a location, expressed per unit of area) as our primary metric. With respect to the existing data, BA ($\text{m}^2 \text{ha}^{-1}$) per genus can be estimated for communities that have performed whole-community inventories, which provide data on species, size (in terms of dbh), and density (trees ha^{-1}) of inventoried trees, typically summarized by land use category (Nowak et al., 2008b). For this study, we assumed that all BA estimates from whole-community inventories were accurate representations of the true BA values for their corresponding communities (but see Martin et al., 2013 regarding the sufficiency of the standard i-Tree Eco sampling protocol). By contrast, it is seldom possible to compute either BA per genus or total BA (i.e., of all trees) accurately from street tree inventories because they are usually non-representative samples of their communities' forests. In short, a community's street tree population is likely to differ compositionally and biometrically from its full population of trees (Nowak et al., 2001a). We made the fundamental assumption that – despite this heterogeneity – some characteristics of communities' street tree and whole-community tree populations (e.g., the proportion of trees in a particular genus) follow predictable coarse-scale relationships. Based on this assumption, and since the majority of the existing data were from street tree inventories, we developed a three-step modeling approach that allowed us to combine street tree data with whole-

community inventory data, thus creating a substantially larger data set for estimating host tree genera BA across the prediction region. Fig. 2 shows a conceptual diagram of the approach. The modeling steps are described below, and Table S1 (Supplement 1) indicates the step(s) in which each community's data were used.

We developed matching point and polygon geospatial data layers that depicted the 842 communities listed in Table S1 (Supplement 1). We extracted point features for most of the communities from a point data layer of populated places developed from US Census data (ESRI, 2014b). A large majority of the communities were also represented in a corresponding data layer of populated place areas (ESRI, 2014a), which consisted of polygons defining the communities' administrative boundaries. If a community's inventory documentation showed that the inventoried area differed significantly from its defined boundaries, then we edited its polygon (or polygons) to correspond to the inventoried area described in the documentation. If a community did not have a record in the populated place areas layer (e.g., communities in Canada), we obtained its polygon(s) from other publicly accessible geospatial data sources, such as a municipal GIS database. Since these communities also did not have records in the populated place point layer, we generated point features based on the inner centroids (i.e., the internal geometric centers) of their polygons.

2.2.1. Estimating BA proportions for communities with only street tree inventories

Although we could not calculate accurate community-level BA estimates directly using the street tree inventories, these inventories did include data on tree species, number, and dbh. From these data, we were able to calculate the total BA of all inventoried street trees, and subsequently, the proportion of the total street tree BA in each of the three host genera. In turn, our first modeling step was to relate street tree BA proportion for each host genus to a corresponding whole-community BA proportion, utilizing data from communities for which we had paired street tree and whole-community inventories. There were 54 such communities distributed across the USA and Canada (Fig. 1B and Fig. 2; Table S1 in Supplement 1). Twelve of these communities – four in Canada and eight distributed across the western USA – were outside our prediction region. We included them primarily to maximize the number of observations in our paired data set for model-building. As we assumed would be the case, preliminary analysis revealed a strong correlation between street tree BA proportion and whole-community BA proportion. This emphasized the value of using as many observations as possible to model this relationship, even if some of the observations were remote geographically.

For this paired data set, we employed polynomial regression in the R statistical software environment (R Core Team, 2016) to estimate whole-community BA proportion as our response variable. We used polynomial regression because we suspected that, for some tree genera, the relationship between street tree BA proportion and whole-community BA proportion is non-linear. We fitted a small set of models separately for each host genus: a full model including the street tree BA proportion of the genus of interest and its second-order polynomial (i.e., its square) as well as the x- and y-coordinates of each community, plus the reduced models that could be constructed from these explanatory variables. We did not include models with higher than second-order polynomials. We applied the arcsine square root transformation to the BA proportion variables (both explanatory and response) prior to modeling.

We used the *car* package in R to perform diagnostics (Fox and Weisberg, 2011; Fox et al., 2016). We tested the models for multicollinearity, outliers, influential observations, normality, and heteroscedasticity. We evaluated and compared models for each host genus in terms of statistical significance (overall and of individual parameters), adjusted R^2 , the pattern of the fitted values versus the measured values and the residuals, and the results of ten-fold cross-validation using the DAAG package in R (Maindonald and Braun, 2015). We then utilized

the best-performing model for each genus to estimate whole-community BA proportions for communities in our second-step data set for which only street tree inventory data were available ($N = 739$; see Fig. 2 and Section 2.2.2, below).

2.2.2. Estimating BA proportions for communities without inventories

The training data set for our second modeling step included 84 communities with whole-community inventories (77 i-Tree Eco inventories and 7 GPI inventories) and 739 communities with street tree inventories (Fig. 1A and Fig. 2). We included communities beyond the western boundary of the prediction region to ensure that the training data had sufficient geographic coverage with respect to communities near but inside this boundary. For the communities with street tree inventories, we applied the regression equations from our first modeling step (Section 2.2.1) to estimate whole-community BA proportions for the three host genera. We assumed that these model-estimated BA proportions were as valid as the proportions derived directly from the whole-community inventory data.

The objective of this second modeling step was to estimate BA proportions of ash, maple, and oak for each community in our training data set as a function of environmental, geographic, land cover, and demographic variables. We did this so we could estimate host BA proportions for communities without any sort of urban forest inventory, based on the assumption that, collectively, these other variables could serve as effective predictors. We chose stochastic gradient boosting (also known as boosted trees) as our modeling method. Details about this machine learning method and some of its applications are available elsewhere (Friedman, 2002; Moisen et al., 2006; De'ath, 2007). We implemented stochastic gradient boosting models for the three genera in DTReg predictive modeling software (Sherrod, 2007). We set the maximum number of trees in the modeling sequence to 400 and the maximum number of splitting levels in each tree to five. We also specified no splitting of nodes with fewer than 10 observations. We chose to hold out a 20% random sample of the training data for model validation and pruning, which is recommended over k -fold cross-validation when working with larger data sets (Sherrod, 2014). We tested other software parameter settings, but these provided the best model performance. Table 1 lists the explanatory variables used in the models, and several of the variables are described further in Supplement 3.

2.2.3. Estimating total BA for communities without inventories

The objective of this third step was to devise a practical way to estimate total BA (i.e., of all trees, whether in one of the host genera or not) for communities without urban forest inventories. Although field inventories of community trees are uncommon, remote sensing has been used extensively to map key characteristics of urban forests at broad spatial scales (Heynen and Lindsey, 2003). For example, one of the derivative products of the 2011 National Land Cover Database (NLCD) is a nationwide, 30-m resolution raster map of percent tree canopy cover, developed in cooperation with the USDA Forest Service. Methods used to create the map, which covers both urban and rural areas of the United States, are discussed in Coulston et al. (2012) and Coulston et al. (2013). We constructed a model based on the assumption that the total BA of a community could be estimated adequately from a measure of its canopy cover, perhaps in combination with a small set of additional covariates.

Our training data set consisted of 76 communities across the USA (Fig. 1C and Fig. 2; Table S1 in Supplement 1). Similar to our first modeling step, we included twelve communities from outside our prediction region in order to maximize the number of observations in our data set for model-building. Each community had a total BA value derived from a whole-community inventory (i.e., an i-Tree Eco inventory) as well as an estimate of its degree of forested canopy cover, which we determined via spatial overlay of the communities' polygon features (from the geospatial data layer described previously; see Section 2.1) and the 2011 NLCD percent tree canopy cover map. Briefly, we

calculated a mean canopy proportion value for each community as the average of the proportion values (converted from percentages) for each 30-m cell from the canopy cover map that fell within the community's polygon(s). We assumed that these values provided accurate representations of each community's canopy cover. We applied the arcsine square root transformation to the mean canopy proportion values prior to modeling.

We used the *mgcv* package in R (Wood, 2006, 2016) to build generalized additive models (GAMs). GAMs are semi-parametric extensions of generalized linear models (GLMs) that have been used widely in ecological analyses, including in modeling of species distributions (Guisan et al., 2002; Moisen et al., 2006). Like GLMs, GAMs are based on an assumed relationship between the mean of the response variable and a function of the explanatory variables. In GLMs, this function is a linear combination of the explanatory variables, although the data and errors may be assumed to follow other probability distributions besides the normal distribution, making GLMs more flexible than classical linear models (Guisan et al., 2002). GAMs take this flexibility further by assuming a relationship between the mean of the response variable and a smooth (and additive) function of the explanatory variables (Hastie and Tibshirani, 1986; Guisan et al., 2002); the form of the smooth function is determined from the data directly rather than depending on a linear or other parametric relationship between the response and explanatory variables (Hastie and Tibshirani, 1987; Guisan et al., 2002; Wood, 2006). With respect to estimation techniques, Wood (2006) outlined the use of penalized regression splines to represent smooth functions in GAMs, as well as the application of generalized cross-validation to estimate the appropriate degree of smoothness.

Our response variable was the log-transformed total BA ($\text{m}^2 \text{ha}^{-1}$) of each community. In terms of explanatory variables, we tested several other candidates besides mean canopy proportion. For instance, forest cover tends to be highest in communities within predominately forested regions (Nowak et al., 1996; Heynen and Lindsey, 2003). Therefore, we included annual precipitation (i.e., mean annual precipitation for the period 1961–1990; see Table 1) as a possible explanatory variable because it is strongly correlated with broad vegetation type (e.g., forest or grassland), abundance, and density, and thus can serve as a simple surrogate measure of regional forest cover. We also included the x - and y -coordinates of the communities (i.e., their inner centroid coordinates) as potential explanatory variables, with the idea that they might further capture variation between geographic regions. In addition, we tested two demographic variables, population density (people ha^{-1}) and housing density ($\text{housing units ha}^{-1}$), that have been linked to the amount of urban forest cover (Dwyer et al., 2000; Nowak et al., 2001a; Heynen and Lindsey, 2003), as well as the area (ha) of each community. Communities with high total land area often have a disproportionately large amount of open space that can be inhabited by trees (Nowak et al., 1996; Heynen and Lindsey, 2003).

We tested a large set of GAMs that utilized different combinations of the explanatory variables. The explanatory variables were represented by individual smooth terms (i.e., smooth functions estimated using penalized regression splines). To select a best-performing GAM, we focused initially on those models with the lowest generalized cross-validation (GCV) scores, as recommended by Wood (2016). Ultimately, we also considered other performance measures (e.g., explained deviance) as well as model parsimony; we ruled out models where one or more of the smooth-term estimates was not at least moderately significant (approx. $P < 0.1$). In addition, we evaluated whether the models' predictions were plausible and realistic (e.g., never exceeding total BA values seen in natural forest stands). We performed ten-fold cross-validation to estimate the expected out-of-sample prediction error for the selected GAM.

2.2.4. Computing and evaluating the combined BA estimates

For each of the 22,856 communities in our prediction region without any sort of urban forest inventory (see Fig. 2), we multiplied

Table 1

Explanatory variables used in stochastic gradient boosting models of basal area (BA) proportion of three host tree genera (*Acer* spp., *Fraxinus* spp. and *Quercus* spp.) in urbanized areas of the eastern and central USA.

Variable	Description	Reference/data source
Province	USFS ecoregion province	Cleland et al. (2007)
X	X-coordinate of community (m)	Calculated in GIS from community point feature
Y	Y-coordinate of community (m)	Calculated in GIS from community point feature
Elevation	Elevation (m)	US Geological Survey National Elevation Dataset
Area	Total area of community (ha)	Calculated in GIS from community polygon feature
Population	Census population of community as of 2010	US Census Bureau
Housing units	Number of housing units in community as of 2010	US Census Bureau
Hardiness zone	USDA plant hardiness zone	USDA Agricultural Research Service (2012)
Extreme minimum temperature	Mean extreme annual minimum temperature (°C) ^a	PRISM Climate Group, Oregon State University
Maximum temperature	Mean annual maximum temperature (°C) ^b	The Climate Source (PRISM)
Summer maximum temperature	Mean maximum temperature across June, July, and August (°C) ^b	The Climate Source (PRISM)
Growing degree days	Growing degree days (base 10 °C) ^c	The Climate Source (PRISM)
Last freeze	Median (Julian) date of last temperature ≤ 0 °C in spring ^d	The Climate Source (PRISM)
First freeze	Median (Julian) date of first temperature ≤ 0 °C in autumn ^d	The Climate Source (PRISM)
Freeze-free days	Median number of days between last spring and first autumn temperature ≤ 0 °C ^d	The Climate Source (PRISM)
Wet days	Mean number of days annually with measurable precipitation ^c	The Climate Source (PRISM)
Precipitation	Mean annual precipitation (mm) ^b	The Climate Source (PRISM)
Moisture index	Balance between precipitation and potential evapotranspiration; scaled between -1 and 1	Willmott and Feddema (1992); see main text and Supplement 2 for additional details
Road density	Log-transformed road length (km) in a 1-km resolution map cell	US Geological Survey, Fort Collins Science Center
Agriculture proportion	Proportion agricultural land cover	National Land Cover Database 2011 (Homer et al., 2015)
Developed proportion	Proportion developed land cover	National Land Cover Database 2011 (Homer et al., 2015)
Natural proportion	Proportion natural land cover	National Land Cover Database 2011 (Homer et al., 2015)
Forest proportion	Proportion forested land cover	National Land Cover Database 2011 (Homer et al., 2015)

^a Mean of the lowest temperature recorded each year 1976–2005.

^b Calculated from mean monthly values for the period 1971–2000.

^c Calculated from mean monthly values for the period 1961–1990.

^d Based on mean minimum monthly temperatures for 1961–1990; spring = March–June, autumn = September–December.

the per-genus BA proportions estimated via stochastic gradient boosting (Section 2.2.2) by the total BA value estimated with the best-performing GAM (Section 2.2.3). This yielded estimates of ash, maple, and oak BA in m² ha⁻¹. Alternatively, for the 716 communities in our prediction region with some sort of urban forest inventory but without measured total BA values, we multiplied either their measured ($N = 22$) or model-estimated ($N = 694$) BA proportions by their GAM-estimated total BA values (see Fig. 2). Another 53 communities in our prediction region had measured BA proportions from whole-community inventories as well as measured total BA values, which allowed them to serve as training data for the stochastic gradient boosting models as well as the GAM. Because we had insufficient data to set aside beforehand for model validation, we opted for a simple assessment of the combined performance of our models, in which we computed and compared ash, maple, and oak BA using both measured and model-estimated values for these 53 communities. This assessment was also intended to provide approximations of the propagated errors from the full modeling process.

3. Results

3.1. BA proportions for communities with only street tree inventories

The best-fit polynomial regression model for each host tree genus is summarized in Table 2. None of the models exhibited significant multicollinearity, influential observations, or outliers. Non-constant error variance tests showed no significant heteroscedasticity, and errors were normally and independently distributed (i.e., did not exhibit significant autocorrelation). According to basic diagnostic measures, all three models fit reasonably well. The model for maple (Table 2B) had the best overall fit based on adjusted R^2 and the F-statistic. The estimated parameters of the maple model (i.e., other than the intercept) were highly significant. The model for ash (Table 2A) fit nearly as well according to basic measures, although its primary parameter estimates were less significant than those for maple. (Note that it is standard

Table 2

Best-fit models for estimating whole-community basal area (BA) proportion of three host tree genera: (A) *Fraxinus* spp., (B) *Acer* spp., and (C) *Quercus* spp. We developed the models from data for communities ($N = 54$) with paired street tree and whole-community inventories. We applied arcsine square root transformations to the BA proportion variables (explanatory and response) prior to modeling. X and Y explanatory variables are the geographic coordinates of the communities. Significance codes: $P < 0.001 = ***$, $P < 0.01 = **$, $P < 0.05 = *$.

Parameter	Estimate	Std. Error	t	Prob(> t)
<i>(A) Fraxinus spp., whole-community BA proportion</i>				
(Intercept)	0.111	0.072	1.55	0.1285
Street BA	0.299	0.195	1.53	0.1313
proportion				
Street BA	0.224	0.111	2.03	0.0479*
proportion ²				
X	4.883×10^{-8}	2.190×10^{-8}	2.23	0.0303*
Adjusted $R^2 = 0.698$; F = 41.9 on 3 and 50 degrees of freedom, $P = 1.11 \times 10^{-13}$				
<i>(B) Acer spp., whole-community BA proportion</i>				
(Intercept)	-0.098	0.110	-0.89	0.3765
Street BA	0.356	0.074	4.80	< 0.0001***
proportion				
Street BA	2.498×10^{-7}	6.340×10^{-8}	3.94	0.0003***
X	1.299×10^{-7}	3.010×10^{-8}	4.32	0.0001***
Adjusted $R^2 = 0.763$; F = 57.9 on 3 and 50 degrees of freedom, $P = 2.7 \times 10^{-16}$				
<i>(C) Quercus spp., whole-community BA proportion</i>				
(Intercept)	0.299	0.053	5.63	< 0.0001***
Street BA	0.502	0.075	6.70	< 0.0001***
proportion				
X	7.680×10^{-8}	3.020×10^{-8}	2.54	0.0140*
Adjusted $R^2 = 0.565$; F = 35.4 on 2 and 51 degrees of freedom, $P = 2.26 \times 10^{-10}$				

practice to include the first-order polynomial term in the model, even if not significant, when the second-order term is significant.) According to basic measures, the oak model (Table 2C) did not fit as well as the other two, although this may be partially explained by it being a simpler model with only two explanatory variables. The oak model was also the

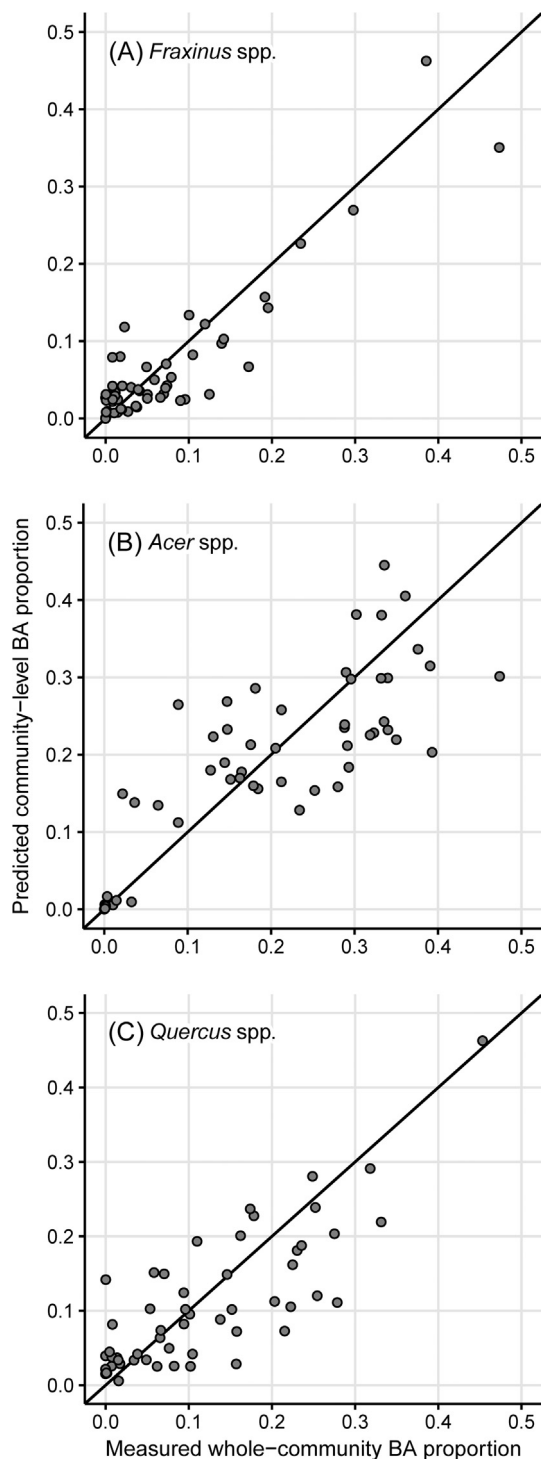


Fig. 3. Plots showing predicted versus measured whole-community basal area (BA) proportion of three host tree genera: (A) *Fraxinus* spp., (B) *Acer* spp., and (C) *Quercus* spp. The plotted points correspond to communities ($N = 54$) with paired street tree and whole-community inventories, which served as training data for the models described in Table 1.

only one where the intercept was significantly different from zero, which means that the predicted whole-community BA proportion will always be greater than zero, even when the street tree BA proportion is zero. Arguably, this is realistic: oaks are a significant component of most natural forests in the central and eastern USA, yet they may be excluded as street trees in some communities due to space constraints, imposing size, and cost of establishment and maintenance. Therefore,

we might expect to see some oak presence in every community, at least in parks or other open spaces, despite their absence as street trees.

Plots of measured versus predicted (i.e., back-transformed) BA proportion for each host genus (Fig. 3) suggest that all three models performed reasonably well. Nevertheless, it can also be seen from the plots that the ash BA proportions (Fig. 3A) in the training data communities were typically much lower than the proportions for the other two genera: the mean measured BA proportion for ash was 0.07, while it was 0.20 for maple (Fig. 3B) and 0.12 for oak (Fig. 3C). At the same time, the mean absolute error (MAE) for ash was 0.03, while it was 0.06 for maple and 0.05 for oak. Thus, while the prediction errors for ash were smaller on average than those for maple or oak, often they represented larger percentages of the measured values.

The results described above document in-sample performance. Ten-fold cross-validation is a reasonable approximation of the expected out-of-sample error when there are insufficient data for true validation (Davison and Hinkley, 1997; Borra and Di Ciaccio, 2010; Rodríguez et al., 2010). The normalized root mean square error (RMSE) from ten-fold cross-validation (i.e., the RMSE expressed as a percentage of the mean of the measured values) was 43.6% for ash, 26.3% for maple, and 40.8% for oak.

Maps of the model-adjusted BA proportions for communities with street tree inventories (Figs. S1, S2, and S3 in Supplement 4) exhibit clear geographic trends. Ash presence was estimated to be highest in communities in the northern part of the central USA and declined quickly going both south and east from that portion of the prediction region. Maple presence was estimated to be highest in communities in the northeastern USA and declined gradually moving west and south from there. Oak presence was estimated to be lowest in the north-western portion of the central USA and comparatively higher moving east and south from there.

3.2. BA proportions for communities without inventories

We estimated BA proportions of the three host genera for communities without urban forest inventories using stochastic gradient boosting (Figs. 4–6). Of the three genera, ash (Fig. 4) was the most difficult to model. This is illustrated by a plot of the input ash BA proportion – either modeled with polynomial regression or measured directly, depending on the inventory type – versus the BA proportion predicted by the stochastic gradient boosting model (Fig. 4A). Although the difference between the input and predicted BA proportion values was usually small, a handful of communities were under-predicted substantially. Notably, the two communities with the largest underestimates (Minot and Grand Forks, both in North Dakota) had whole-community inventories where the measured ash BA proportions exceeded 0.8. These proportions may seem high, but green ash (*F. pennsylvanica*) is the most abundant tree species in North Dakota, which is < 2% forested (Haugen and Pugh, 2014); in a community where, for example, much of the limited urban forest is contained in riparian areas, it is conceivable that green ash, which is frequently found along riparian corridors (Poland and McCullough, 2006), could dominate an inventory of the community's trees. Furthermore, ash, and green ash in particular, was widely planted in urban areas of the north central USA as a replacement for elms (*Ulmus* spp.) killed by Dutch elm disease (*Ophiostoma ulmi* and *O. novo-ulmi*) (MacFarlane and Meyer, 2005).

Regardless, the difficulty of modeling ash BA proportion is further demonstrated by the disparity between the proportion of the variance in the training data (0.70) explained by the model versus the proportion explained in the held-out validation data (0.47). This is much larger than the difference in explained variance between the training and validation data for maple (0.92 and 0.82, respectively) and oak (0.78 and 0.71, respectively) and suggests that the ash model was the least robust of the three models. The smaller disparities for maple and oak with respect to the training and validation data are echoed in their plots of input versus predicted BA proportion (Fig. 5A and Fig. 6A,

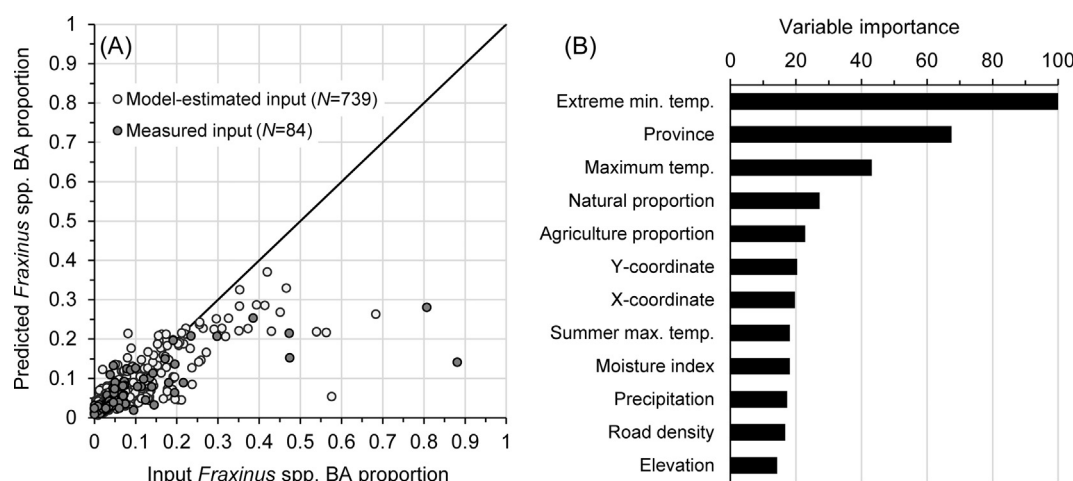


Fig. 4. Stochastic gradient boosting results for basal area (BA) proportion of *Fraxinus* spp.: (A) plot showing input versus predicted ash proportion for the training data, where the input proportions were either measured from a whole-community inventory ($N = 84$) or estimated via polynomial regression ($N = 739$); (B) relative importance of the explanatory variables in the stochastic gradient boosting model. The twelve most important variables are shown. See Table 1 for full description of explanatory variables.

respectively). The plot for the maple model suggests an excellent fit, with most observation points (i.e., communities) falling close to the 1:1 line. The plot for the oak model indicates a good fit, with some communities under-predicted, but a better overall performance than the ash model.

Ecoregion province was one of the most important explanatory variables for all three stochastic gradient boosting models. For ash (Fig. 4B), ecoregion province was second to extreme minimum temperature, while for maple (Fig. 5B) it was second to moisture index. It was the most important explanatory variable for oak (Fig. 6B), ahead of precipitation. These results are unsurprising: temperature and available moisture are key factors governing the geographic distributions of trees, including urban trees (Kendal et al., 2012a, b; Ramage et al., 2013), while ecoregion province captures variation between large geographic areas (mean province size in the continental USA $\approx 216,000 \text{ km}^2$) in terms of the most prominent forest types and their constituent species.

3.3. Total BA for communities without inventories

We estimated total BA (i.e., of all trees) for communities without urban forest inventories using a GAM. The GAM we identified as the best-performing model (Table 3) included smooth terms for four explanatory variables: canopy proportion, community x-coordinate,

community y-coordinate, and community land area. During the testing process, models that included a smooth term for another variable, precipitation, outperformed the selected model according to standard statistical measures, but yielded unrealistic predictions (e.g., predicting some of the highest total BA values for communities with relatively intense development and therefore limited forest cover).

With respect to the best-performing GAM, the smooth function for canopy proportion displayed a positive and essentially linear relationship with the original (i.e., transformed) variable values (Fig. 7A). Relationships between the smooth function and original values were less straightforward for the other explanatory variables, but none were unusual for a GAM. A plot of the measured total BA values versus the predicted (and back-transformed) values for the communities in the training data (Fig. 7B) suggests a good fit overall, although the two communities with the highest measured values (Moorestown, New Jersey and Orlando, Florida) were also the most under-predicted. The mean of the measured total BA values was $5.48 \text{ m}^2 \text{ ha}^{-1}$, while the mean of the predicted (i.e., back-transformed) total BA values was $5.18 \text{ m}^2 \text{ ha}^{-1}$. The MAE was $1.33 \text{ m}^2 \text{ ha}^{-1}$. The normalized RMSE from ten-fold cross-validation was 20.2%, suggesting that the model should perform reasonably well out-of-sample.

A map of total BA for more than 23,000 communities in the prediction region (Fig. S4 in Supplement 4) shows generally higher values

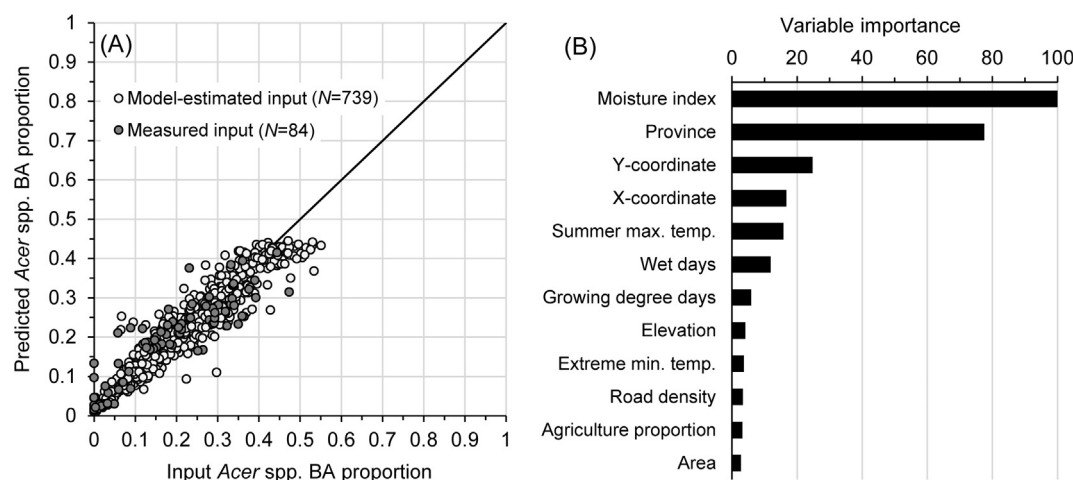


Fig. 5. Stochastic gradient boosting results for basal area (BA) proportion of *Acer* spp.: (A) plot showing input versus predicted proportion for the training data, where the input proportions were either measured from a whole-community inventory ($N = 84$) or estimated via polynomial regression ($N = 739$); (B) relative importance of the explanatory variables in the stochastic gradient boosting model. The twelve most important variables are shown. See Table 1 for full description of explanatory variables.

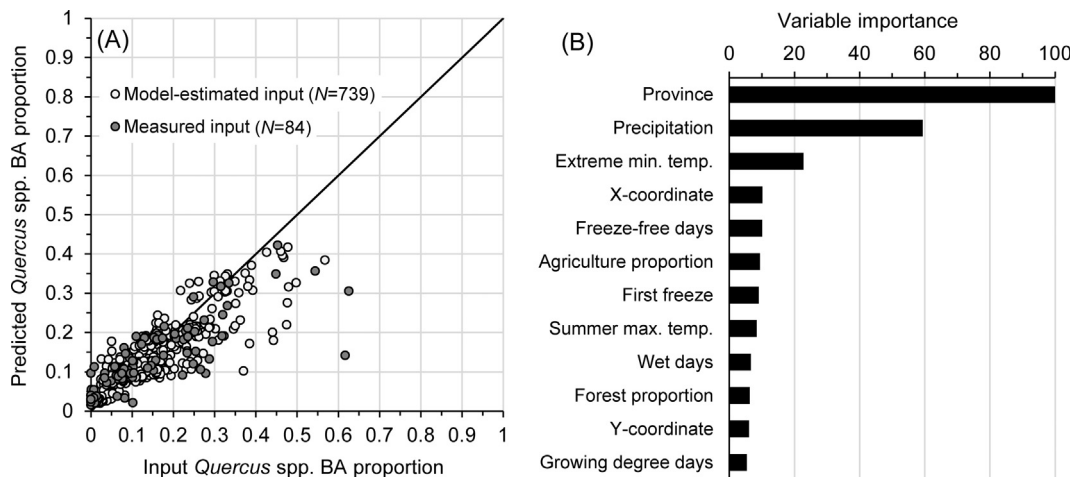


Fig. 6. Stochastic gradient boosting results for basal area (BA) proportion of *Quercus* spp.: (A) plot showing input versus predicted proportion for the training data, where the input proportions were either measured from a whole-community inventory ($N = 84$) or estimated via polynomial regression ($N = 739$); (B) relative importance of the explanatory variables in the stochastic gradient boosting model. The twelve most important variables are shown. See Table 1 for full description of explanatory variables.

Table 3

Best-performing generalized additive model (GAM) for estimating total community basal area (BA) of all trees in urbanized areas of the eastern and central USA. We developed the model from data for communities with whole-community inventories for which we could calculate valid total BA estimates ($N = 76$). Prior to modeling, we applied the log transformation to the response variable total BA ($\text{m}^2 \text{ha}^{-1}$) and the arcsine square root transformation to one explanatory variable, canopy proportion. Each explanatory variable z was represented by a smooth term, $s(z)$. Note that in the Gaussian case, the scale estimate is the square of the residual standard error. X and Y explanatory variables are the geographic coordinates of the communities. Significance codes: $P < 0.001 = ***$, $P < 0.01 = **$, $P < 0.05 = *$.

Parametric coefficients:				
	Estimate	Std. Error	<i>t</i>	Prob(> <i>t</i>)
(Intercept)	3.746	0.057	65.17	< 0.0001 ***
Approximate significance of smooth terms:				
	Estimated d.f.	Reference d.f.	<i>F</i>	<i>P</i> -value
$s(\text{Canopy proportion})$	0.894	9	0.905	0.0035**
$s(X)$	7.504	9	3.065	0.0013**
$s(Y)$	8.193	9	5.137	< 0.0001***
$s(\text{Area})$	6.187	9	1.735	0.0211*

Adjusted $R^2 = 0.752$; deviance explained = 82.7%; scale estimate = 0.25116.

in the eastern USA than in the central USA. In particular, the map shows a large contiguous area of high total BA values extending from the Florida Panhandle to southern Indiana, Ohio, and West Virginia. This area included numerous communities with greater-than-average forest canopy proportions, which – when fed into the best-performing GAM – yielded high total BA estimates.

3.4. Combined model estimates

We combined the results from our modeling steps to estimate and map ash, maple, and oak BA for all communities in the prediction region (Fig. 8; see Section 2.2.4 for details about combining values from the different types of inventories). With the exception of some communities in the north central USA ($\approx 2\%$ of all communities), ash was estimated to be a lesser component of most urban forests than either maple or oak. The mean ash BA estimate was $0.23 \text{ m}^2 \text{ha}^{-1}$ (SE 0.001), while the maximum estimate was $4.06 \text{ m}^2 \text{ha}^{-1}$. By comparison, the mean and maximum maple BA estimates were $1.34 \text{ m}^2 \text{ha}^{-1}$ (SE 0.008) and $8.61 \text{ m}^2 \text{ha}^{-1}$, respectively, while the mean and maximum oak BA estimates were $1.29 \text{ m}^2 \text{ha}^{-1}$ (SE 0.008) and $12.37 \text{ m}^2 \text{ha}^{-1}$. Maple comprised the largest component of urban forests in $\approx 62\%$ of communities, mostly in the northern portion of the prediction region. Oak

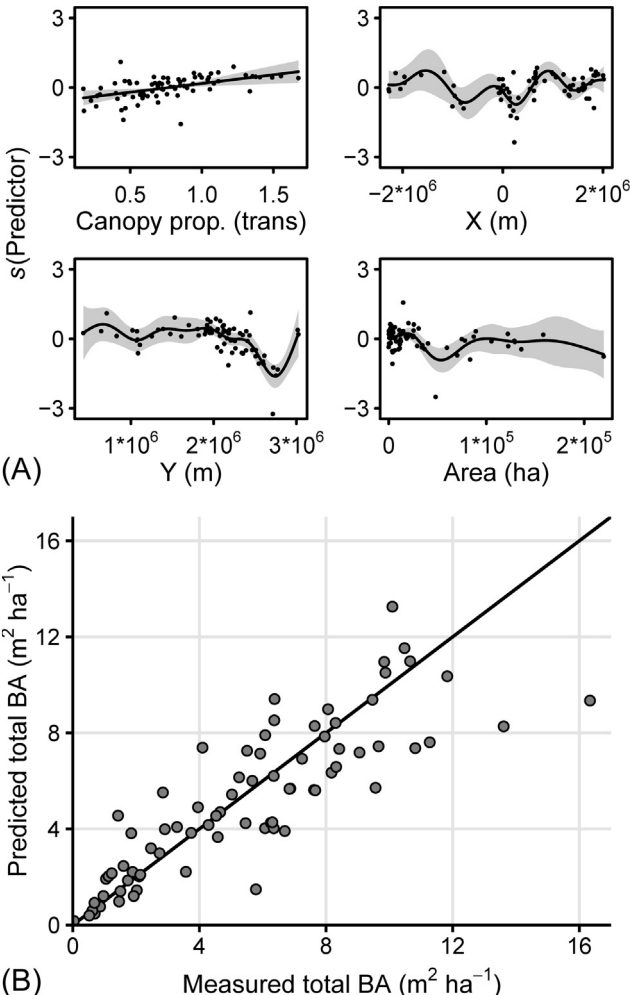


Fig. 7. Generalized additive model (GAM) results for total basal area (BA) of all trees: (A) relationship between original and smooth function values for each explanatory variable in the best-performing GAM: canopy proportion (with arcsine square root transformation), x -coordinate of the community, y -coordinate of the community, and total area of the community. Black dots represent the training data points, and the shaded area is the approximate 95% confidence region around the smooth function (black line); (B) plot showing measured versus predicted total BA for the training data ($N = 76$).

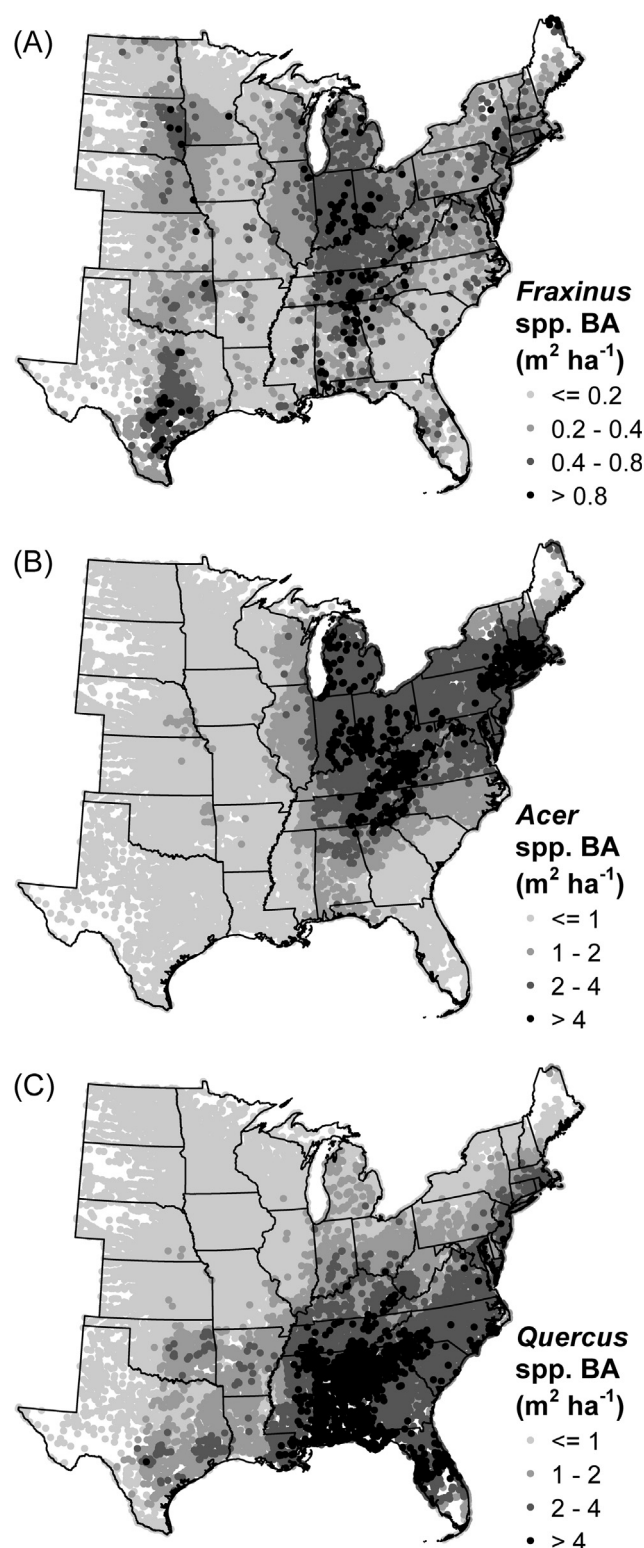


Fig. 8. Maps of estimated basal area (BA) by host genus for all communities ($N = 23,625$) in the prediction region: (A) *Fraxinus* spp.; (B) *Acer* spp.; (C) *Quercus* spp.

was the largest component of urban forests in $\approx 36\%$ of communities, primarily in the southern portion of the region. All three genera exhibited significant spatial clustering of high (and low) urban BA values ($P < 0.0001$ in global Moran's I tests), but clustering was perhaps most apparent for oak (see Fig. 8C), including a contiguous area of high BA values extending across Mississippi, Alabama, Florida, Georgia, and South Carolina. By comparison, the geographic distribution of urban

Table 4

Error summary for the subset of 53 communities used to assess combined model performance in estimating basal area (BA) of three host tree genera (*Fraxinus* spp., *Acer* spp., and *Quercus* spp.) in urbanized areas of the eastern and central USA. Bias is the mean signed difference between the measured and predicted BA values. All values are in $\text{m}^2 \text{ha}^{-1}$.

Host genus	Mean measured BA	Mean predicted BA	Bias	Mean absolute error (MAE)	Root mean square error (RMSE)
<i>Fraxinus</i> spp.	0.307	0.219	−0.089	0.180	0.313
<i>Acer</i> spp.	1.139	1.092	−0.048	0.375	0.577
<i>Quercus</i> spp.	0.916	0.878	−0.038	0.323	0.472

ash BA was relatively diffuse (Fig. 8A).

Error estimates (Table 4) from our assessment of the combined performance of the models were based on a small subset of communities. Nevertheless, they should be passable approximations of the propagated errors from the full three-step modeling process. Comparing the RMSE to the mean measured BA for each host genus (Table 4), it is clear that ash was modeled least successfully at the scale of a large prediction region, although the mean absolute error (MAE), which is less sensitive to outliers than RMSE, indicates that the modeling approach had some predictive value for ash as well as maple and oak. The bias estimates suggest a tendency toward under-prediction regardless of genus, although the degree of under-prediction was small for maple and oak. Furthermore, a larger, and thus more representative, subset of communities may have indicated less negative prediction bias (i.e., a mean signed difference closer to zero) for all three host genera.

4. Discussion

One of the key elements that determine whether a location is vulnerable to invasion by a plant-feeding pest is host availability (Bartell and Nair, 2004). Fundamentally, a location is suitable if it has sufficient hosts to support establishment of a viable invader population. However, the particular characteristics of the hosts (e.g., density or health) in a given location may also influence an invasion's overall trajectory by affecting, among other aspects, when (i.e., under what conditions) an invader spreads to other locations as well as its pattern and rate of dispersal (Burdon, 1982; Mercader et al., 2011). This highlights the importance of representing spatial heterogeneity in host resources when analyzing or modeling plant pest invasions at a regional scale (Hastings et al., 2005).

We maintain that depicting this heterogeneity appropriately for forest pests requires integration of urban and natural forest distributions of host tree species. As a practical matter, the lack of urban forest inventory data would not be problematic if urban forests resembled nearby natural forests in terms of species composition. In that case, data from natural forests (i.e., FIA plot data) could reasonably be extended to represent their urban forest counterparts. However, other research (e.g., McKinney, 2006; Kendal et al., 2012b; Ramage et al., 2013; Aronson et al., 2015, 2016; Blood et al., 2016) has shown that proximate urban and natural forests can be quite different in terms of species composition. This disparity is due to a variety of factors, perhaps most importantly the phenomenon of ecological homogenization, wherein human modification of urbanized landscapes makes them resemble other urban systems – including geographically distant urban systems – more closely than their neighboring native ecosystems (Groffman et al., 2014). Regardless, this disparity reemphasizes the importance of inventories dedicated to urban forests.

For the foreseeable future, modeling will remain the only practical way to generate comprehensive (i.e., across all communities) urban forest information for large geographic regions, which can then be integrated with FIA or other data describing the nation's natural forests.

Our three-step modeling approach utilizes urban forest inventories whether they are whole-community or include only street trees. We believe this is a key positive aspect of our approach: it avoids omitting street tree data sets that, despite being partial inventories, are certain to contain useful information about urban trees. The capacity to include such data is critical since so many of the existing urban inventories in the USA (and elsewhere) are street tree inventories. Furthermore, we expect that many future inventories (i.e., outside of urban FIA data collection or similar efforts) will also be limited to street trees, either because community governments seek to minimize inventory costs or to prioritize just those trees for which they are directly responsible and require the highest degree of management intervention. Fortunately, as long as some of the input data are from whole-community inventories – or more properly, as long as there are some paired street tree and whole-community inventories – then it should be possible to estimate BA values at the genus level as we did in this study.

Indeed, the availability of paired inventory data is our approach's principal constraint. Paired data are instrumental to the first modeling step (i.e., predicting whole-community BA proportion from street tree BA proportion), and in turn, predictions from this first step serve as input data in the second step. Unfortunately, relatively few communities have both street tree and whole-community inventories. In our case, we deemed a sample of 56 communities across the USA and Canada to be suitable for model-building, yet not large enough that we could afford to set aside a portion of the communities for model validation. Nevertheless, based on typical model performance measures, we feel confident in concluding that the first-step models were successful for all three host tree genera. While we recognize that the results from our second modeling step (i.e., predicting community-level BA proportion using a suite of variables) were shaped by the first-step results, we think they further support these conclusions.

Another constraint is the number of communities with total BA values that can serve as input data for the third modeling step (i.e., predicting total BA in communities based on canopy cover and other factors). Theoretically, it is possible to compute total BA for any community with a whole-community inventory, as long as the total area in each sampled land use category is known, and the inventory was performed according to standard protocols. In our case, just under 80 communities across the USA had useable estimates of total BA. This is a larger sample than in our first modeling step, but again we did not set aside any of these communities for model validation because we did not want to omit potentially important geographic variation or size differences (i.e., in total land area) between communities during the model-building process. Consequently, we had to rely on common statistical measures to assess performance, and according to these, the chosen model performed fairly well. Still, one of our eventual goals is to reassess the models via proper validation, and unless a centralized repository for urban forest data becomes available for the USA (as recommended by Boyer et al., 2016), we will continue to gather inventory data sets from additional communities to achieve this goal.

Ultimately, the data we used were samples of convenience. We treated them as random for inferential purposes, but realize there were inevitably some geographic and other kinds of biases. Furthermore, we did not have a practical way to ascertain the relative data quality of the inventories for some aspects of interest. For example, our assumption that tree genus identification errors are uncommon seems reasonable for inventories conducted by certified arborists or comparable professionals, but non-professionals (e.g., students and citizen scientists) are demonstrably less skilled at identification (Bloniarz and Ryan, 1996; Roman et al., 2017a). Unfortunately, we seldom had information about the expertise of the field personnel who conducted the inventories. Faced with such uncertainties, we chose to focus on the utility of the various models for prediction rather than hypothesis testing. In this regard, we were most concerned about the sets of predictions being reasonable at a regional scale, rather than the accuracy of any individual prediction, especially since a moderate level of prediction

error can be anticipated when modeling phenomena that are relatively high-frequency in space. Based on somewhat lower performance expectations, we feel comfortable in asserting that, overall, our modeling approach was very successful for maple and oak, but only marginally successful for ash. By extension, we believe our results also demonstrate that our primary assumptions were reasonable for maple and oak, but less so for ash.

Why was ash so difficult to model? We believe it is due to large regional differences in the frequency and abundance of ash in urban settings. In the original inventory data as well as the modeled outputs, maple and oak (see Figs. S2 and S3 in Supplement 4) were especially prevalent in certain portions of the prediction region, but were also fairly common elsewhere. In contrast, ash (Fig. S1 in Supplement 4) was prevalent in communities in the north central USA, but only a minor component of urban forests throughout most of the prediction region, particularly in the southeastern USA. Moreover, in communities where ash was prevalent, the total BA (actual or predicted) was typically less than in communities dominated by either oak or maple (Fig. S4 in Supplement 4). This is probably because ash was most prominent in areas with continental (i.e., relatively dry) climates, where tree cover, both in urban and natural settings, is fairly sparse. The high degree of regional variation and low ash levels in general would be challenging for any statistical modeling effort. A case can be made that urban distributions of ash might be modeled best using smaller prediction regions, which then could be combined if necessary.

One factor to consider with respect to ash is the timing of urban forest inventories in relation to the pattern of EAB expansion. We could not always determine whether communities in already-invaded areas had performed their inventories before or after the arrival of EAB. Likewise, for communities in the path of invasion, we usually could not ascertain whether there had been any pre-emptive removal of ash in anticipation of the pest's arrival. Furthermore, ash was not the only one of our target genera subjected to removals because of a pest; for instance, approximately 30,000 maple trees were removed from a quarantine zone established after the discovery of ALB in central Massachusetts in 2008 (Hostetler et al., 2013). Thus, we acknowledge that uncertainty about the timing of the inventories was a potential source of error in our modeling efforts, especially for ash, but note that this does not diminish the validity of the modeling approach.

Another consideration that is relevant for all three host genera, but perhaps especially for ash, is that the extent and composition of a community's urban forest are shaped by its tree planting history. Sometimes, this history may have been driven by regional-scale factors. For example, we have noted that many communities throughout the USA planted ash to replace elms killed by Dutch elm disease (MacFarlane and Meyer, 2005). In other cases, the trees planted in a community may have been limited by the nursery stock available at the time (e.g., Sydnor et al., 2010). Practical considerations (e.g., growth rate, habit, or ease of care) likely also played a part in some past planting decisions, as did cultural preferences for certain species (Johnston, 2015; Jonnes, 2016). Regrettably, communities' tree planting histories are not always documented. Even when historical data are available, they are seldom linked to the current composition and pattern of communities' forests, other than in a qualitative way (e.g., Dorney et al., 1984; McPherson and Luttinger, 1998; Roman et al., 2017b). A potentially fruitful area of research would be to identify and integrate data describing historical events and trends – particularly if they are understood to have affected urban forests at a regional scale – into future modeling efforts directed at urban forests.

Our analysis was intended as a proof of concept. Regardless of the constraints and uncertainties, we believe our results demonstrate the promise of the approach. We also recognize that it may need to be refined or applied differently in some circumstances. For instance, we chose our modeling methods after a lengthy trial-and-error process with this particular set of host genera. Other methods might work as well or better for different hosts. However, we maintain that, at least until

whole-community and relatively consistent data become widely available, a three-step approach remains the most practical way to deal with the diversity in format and content of current urban forest inventories.

4.1. Potential applications

Because our modeling approach yields estimates of host presence and abundance in terms of a commonly used forest metric (BA), it can be integrated easily with similar information from inventories of natural forests. This feature is useful for broad-scale assessments of the risks and potential impacts of invasive forest pests (Venette et al., 2010). For example, the Forest Service's 2012 National Insect and Disease Risk Map (NIDRM) depicts, for each 240-m map cell, the risk of significant forest mortality (i.e., expected loss of at least 25% total live BA) due to insects and diseases over a 15-year period, 2013–2027 (Krist et al., 2014). The 2012 version of the NIDRM improved on previous iterations by incorporating simple representations of host distributions in urban areas based primarily on land cover data. Nonetheless, Krist et al. (2014) acknowledged the lack of detailed urban inventory data as a limitation. This limitation could be addressed immediately for EAB, ALB, and EGM (as well as other pests of ash, maple, and oak trees) using the outputs from our study. Our approach could also be applied to other potential hosts in the prediction region, and with some additional effort, could be extended to the western USA.

Moreover, our approach could be extended to countries other than the USA. Canada is an obvious candidate, but urban forest inventories have been conducted in many parts of the world, including in South America, Australia, Europe, and Asia (Nielsen et al., 2014). As urban forest inventories become more commonplace, an increasing number of countries (or groups of countries) should have enough data to implement our three-step modeling approach with a fair degree of success.

Whereas we have presented our modeling approach in the context of invasive forest pests (specifically insect pests), we see utility of the approach and its outputs in a variety of other contexts. For instance, it may facilitate estimation of ecosystem services (e.g., carbon sequestration) provided by trees in communities without urban forest inventory data. Our approach may also enable better estimation of the impacts of hurricanes, wildfires, and other disturbances that affect both natural and urban forests. Model-based estimates of urban forest characteristics may not be ideal in these cases – just as they are not ideal in the invasive species context – but we maintain that they are a practical substitute in the absence of numbers derived directly from urban inventories.

Regarding the latter point, it is important to recognize the strides that have been made in remote sensing of urban forest characteristics, especially in the last decade. For example, by combining hyperspectral imagery and lidar, it is possible to distinguish some tree species in urban environments with reasonable accuracy (Zhang and Qiu, 2012; Alonzo et al., 2014), and automated approaches continue to improve for distinguishing individual tree crowns in high-spatial-resolution imagery (Lee et al., 2016). Of course, there are limitations: tree metrics such as dbh cannot be measured directly from remotely sensed images (Lee et al., 2016), and smaller trees can be completely obscured by dominant canopy trees (Alonzo et al., 2014). Still, remote sensing provides the capacity to collect some urban forest information at far less cost than field-based inventories, and analytical strategies for extracting that information are evolving constantly. We see great potential for integration of remote sensing and field-based measurements as a way to enhance modeling efforts such as ours, particularly as certain types of remotely sensed data (e.g., high-spatial-resolution hyperspectral imagery) become more universally available.

Acknowledgements

We wish to thank all those municipalities, companies, agencies, and individuals who provided urban forest inventory data for this project.

There are far too many to recognize here, so we refer you to Supplement 2 for the complete list. This work was supported by the USDA Forest Service, Southern Research Station [agreements 11-JV-11330146-090, 12-JV-11330146-063, 13-JV-11330110-043, 14-CS-11330110-042, 15-CS-11330110-067, and 16-CS-11330110-038].

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foreco.2018.03.004>.

References

- Abd-Elrahman, A.H., Thornhill, M.E., Andreu, M.G., Escobedo, F., 2010. A community-based urban forest inventory using online mapping services and consumer-grade digital images. *Int. J. Appl. Earth Obs.* 12, 249–260. <http://dx.doi.org/10.1016/j.jag.2010.03.003>.
- Allen, E.A., Humble, L.M., 2002. Nonindigenous species introductions: a threat to Canada's forests and forest economy. *Can. J. Plant Pathol.* 24, 103–110. <http://dx.doi.org/10.1080/07060660309506983>.
- Alonzo, M., Bookhagen, B., Roberts, D.A., 2014. Urban tree species mapping using hyperspectral and lidar data fusion. *Remote Sens. Environ.* 148, 70–83. <http://dx.doi.org/10.1016/j.rse.2014.03.018>.
- Alvey, A.A., 2006. Promoting and preserving biodiversity in the urban forest. *Urban For. Urban Green.* 5, 195–201. <http://dx.doi.org/10.1016/j.ufug.2006.09.003>.
- Aronson, M.F., Handel, S.N., La Puma, I.P., Clemants, S.E., 2015. Urbanization promotes non-native woody species and diverse plant assemblages in the New York metropolitan region. *Urban Ecosyst.* 18, 31–45. <http://dx.doi.org/10.1007/s11252-014-0382-z>.
- Aronson, M.F.J., Nilon, C.H., Lepczyk, C.A., Parker, T.S., Warren, P.S., Cilliers, S.S., Goddard, M.A., Hahs, A.K., Herzog, C., Katti, M., La Sorte, F.A., Williams, N.S.G., Zipperer, W., 2016. Hierarchical filters determine community assembly of urban species pools. *Ecology* 97, 2952–2963. <http://dx.doi.org/10.1002/ecy.1535>.
- Aukema, J.E., McCullough, D.G., Von Holle, B., Liebhold, A.M., Britton, K., Frankel, S.J., 2010. Historical accumulation of nonindigenous forest pests in the continental United States. *Bioscience* 60, 886–897. <http://dx.doi.org/10.1525/bio.2010.60.11.5>.
- Aukema, J.E., Leung, B., Kovacs, K., Chivers, C., Britton, K.O., Englin, J., Frankel, S.J., Haight, R.G., Holmes, T.P., Liebhold, A.M., McCullough, D.G., Von Holle, B., 2011. Economic impacts of non-native forest insects in the continental United States. *PLoS ONE* 6, e24587. <http://dx.doi.org/10.1371/journal.pone.0024587>.
- Bartell, S.M., Nair, S.K., 2004. Establishment risks for invasive species. *Risk Anal.* 24, 833–845. <http://dx.doi.org/10.1111/j.0272-4332.2004.00482.x>.
- BenDor, T.K., Metcalf, S.S., 2006. The spatial dynamics of invasive species spread. *Syst. Dyn. Rev.* 22, 27–50. <http://dx.doi.org/10.1002/sdr.328>.
- Bloniarz, D.V., Ryan, H.D.P., 1996. The use of volunteer initiatives in conducting urban forest resource inventories. *J. Arboric.* 22, 75–82.
- Blood, A., Starr, G., Escobedo, F., Chappella, A., Staudhammer, C., 2016. How do urban forests compare? Tree diversity in urban and periurban forests of the southeastern US. *Forests* 7, 120. <http://dx.doi.org/10.3390/f7060120>.
- Borra, S., Di Ciaccio, A., 2010. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Comput. Stat. Data Anal.* 54, 2976–2989. <http://dx.doi.org/10.1016/j.csda.2010.03.004>.
- Boyer, D.J., Roman, L.A., Henning, J.G., McFarland, M., Dentice, D., Low, S.C., Thomas, C., Abrams, G., 2016. Data Management for Urban Tree Monitoring – Software Requirements. Azavea, Philadelphia, PA.
- Bradshaw, C.J.A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., Courchamp, F., 2016. Massive yet grossly underestimated global costs of invasive insects. *Nat. Commun.* 7, 12986. <http://dx.doi.org/10.1038/ncomms12986>.
- Burdon, J.J., 1982. Host density as a factor in plant disease ecology. *Annu. Rev. Phytopathol.* 20, 143–166. <http://dx.doi.org/10.1146/annurev.py.20.090182.001043>.
- Clavero, M., García-Berthou, E., 2005. Invasive species are a leading cause of animal extinctions. *Trends Ecol. Evol.* 20, 110. <http://dx.doi.org/10.1016/j.tree.2005.01.003>.
- Cleland, D.T., Freeouf, J.A., Keys, J.E., Jr., Nowacki, G.J., Carpenter, C.A., McNab, W.H., 2007. Ecological subregions: sections and subsections for the conterminous United States. In: Sloan, A.M., (tech. ed.), 2007. Gen. Tech. Rep. WO-76D. [Map, presentation scale 1:3,500,000; colored]. Also on CD-ROM as a GIS coverage in ArcInfo format. US Department of Agriculture, Forest Service, Washington, DC.
- Colunga-Garcia, M., Magarey, R.A., Haack, R.A., Gage, S.H., Qi, J., 2010a. Enhancing early detection of exotic pests in agricultural and forest ecosystems using an urban-gradient framework. *Ecol. Appl.* 20, 303–310. <http://dx.doi.org/10.1890/09-0193.1>.
- Colunga-Garcia, M., Haack, R.A., Magarey, R.A., Margosian, M.L., 2010b. Modeling spatial establishment patterns of exotic forest insects in urban areas in relation to tree cover and propagule pressure. *J. Econ. Entomol.* 103, 108–118. <http://dx.doi.org/10.1603/ec09203>.
- Coulston, J.W., Moisen, G.G., Wilson, B.T., Finco, M.V., Cohen, W.B., Brewer, C.K., 2012. Modeling percent tree canopy cover: a pilot study. *Photogramm. Eng. Rem. S.* 78, 715–727. <http://dx.doi.org/10.14358/PERS.78.7.715>.
- Coulston, J.W., Jacobs, D.M., King, C.R., Elmore, I.C., 2013. The influence of multi-season imagery on models of canopy cover: a case study. *Photogramm. Eng. Rem. S.* 79,

- 469–477. <http://dx.doi.org/10.14358/PERS.79.5.469>.
- Cowett, F.D., Bassuk, N., 2017. Street tree diversity in three northeastern U.S. states. *Arboric. Urban For.* 43, 1–14.
- Cumming, A.B., Twardus, D.B., Nowak, D.J., 2008. Urban forest health monitoring: large-scale assessments in the United States. *Arboric. Urban For.* 34, 341–346.
- Davison, A.C., Hinkley, D.V., 1997. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge, UK.
- De'ath, G., 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88, 243–251. [http://dx.doi.org/10.1890/0012-9658\(2007\)88\[243:BTFFEMA\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2007)88[243:BTFFEMA]2.0.CO;2).
- Dodds, K.J., Orwig, D.A., 2011. An invasive urban forest pest invades natural environments—Asian longhorned beetle in northeastern US hardwood forests. *Can. J. For. Res.* 41, 1729–1742. <http://dx.doi.org/10.1139/x11-097>.
- Donovan, G.H., Butry, D.T., Michael, Y.L., Prestemon, J.P., Liebhold, A.M., Gatzliolis, D., Mao, M.Y., 2013. The relationship between trees and human health: evidence from the spread of the emerald ash borer. *Am. J. Prev. Med.* 44, 139–145. <http://dx.doi.org/10.1016/j.amepre.2012.09.066>.
- Dorney, J.R., Guntenspergen, G.R., Keough, J.R., Stearns, F., 1984. Composition and structure of an urban woody plant community. *Urban Ecol.* 8, 69–90. [http://dx.doi.org/10.1016/0304-4009\(84\)90007-X](http://dx.doi.org/10.1016/0304-4009(84)90007-X).
- Dwyer, J.F., Nowak, D.J., Noble, M.H., Sisinni, S.M., 2000. Connecting people with ecosystems in the 21st century: an assessment of our nation's urban forests. Gen. Tech. Rep. PNW-GTR-490. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.
- ESRI, 2014a. *US Populated Place Areas, 2014 edition [Geospatial Data in Vector Format]*. ESRI, Redlands, CA.
- ESRI, 2014b. *US Populated Place Points, 2014 edition [Geospatial Data in Vector Format]*. ESRI, Redlands, CA.
- Fox, J., Weisberg, S., 2011. *An R Companion to Applied Regression*, second ed. SAGE Publications, Thousand Oaks, CA.
- Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Laboissiere, R., Monette, G., Murdoch, D., Nilsson, H., Ogle, D., Ripley, B., Venables, W., Winsemius, D., Zeileis, A., R-Core, 2016. *Car: Companion to Applied Regression*, version 2.1-4. R Foundation for Statistical Computing, Vienna, Austria. < <https://cran.r-project.org/package=car> >.
- Friedman, J.H., 2002. Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 367–378. [http://dx.doi.org/10.1016/S0167-9473\(01\)00065-2](http://dx.doi.org/10.1016/S0167-9473(01)00065-2).
- Greene, C.S., Millward, A.A., 2016. The legacy of past tree planting decisions for a city confronting emerald ash borer (*Agrilus planipennis*) invasion. *Front. Ecol. Evol.* 4. <http://dx.doi.org/10.3389/fevo.2016.00027>.
- Groffman, P.M., Cavender-Bares, J., Bettez, N.D., Grove, J.M., Hall, S.J., Heffernan, J.B., Hobbie, S.E., Larson, K.L., Morse, J.L., Neill, C., Nelson, K., O'Neil-Dunne, J., Ogden, L., Pataki, D.E., Polsky, C., Chowdhury, R.R., Steele, M.K., 2014. Ecological homogenization of urban USA. *Front. Ecol. Environ.* 12, 74–81. <http://dx.doi.org/10.1890/120374>.
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.* 157, 89–100. [http://dx.doi.org/10.1016/S0304-3800\(02\)00204-1](http://dx.doi.org/10.1016/S0304-3800(02)00204-1).
- Haack, R.A., Hérard, F., Sun, J., Turgeon, J.J., 2010. Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. *Annu. Rev. Entomol.* 55, 521–546. <http://dx.doi.org/10.1146/annurev-ento-112408-085427>.
- Hastie, T., Tibshirani, R., 1986. Generalized additive models. *Stat. Sci.* 1, 297–318. <http://dx.doi.org/10.1214/ss/1177013604>.
- Hastie, T., Tibshirani, R., 1987. Generalized additive models: some applications. *J. Am. Stat. Assoc.* 82, 371–386. <http://dx.doi.org/10.1080/01621459.1987.10478440>.
- Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K., Taylor, C., Thomson, D., 2005. The spatial spread of invasions: new developments in theory and evidence. *Ecol. Lett.* 8, 91–101. <http://dx.doi.org/10.1111/j.1461-0248.2004.00687.x>.
- Hauer, R.J., Peterson, W.D., 2017. Effects of emerald ash borer on municipal forestry budgets. *Landsc. Urban Plan.* 157, 98–105. <http://dx.doi.org/10.1016/j.landurbplan.2016.05.023>.
- Haugen, D.E., Pugh, S.A., 2014. *Forests of North Dakota, 2014. Resource Update FS-33*. US Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA.
- Heynen, N.C., Lindsey, G., 2003. Correlates of urban forest canopy cover: implications for local public works. *Public Works Manag. Policy* 8, 33–47. <http://dx.doi.org/10.1177/1087724X03008001004>.
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States – representing a decade of land cover change information. *Photogramm. Eng. Rem. S.* 81, 345–354. <http://dx.doi.org/10.14358/PERS.81.5.345>.
- Hostetler, A.E., Rogan, J., Martin, D., DeLauer, V., O'Neil-Dunne, J., 2013. Characterizing tree canopy loss using multi-source GIS data in Central Massachusetts, USA. *Remote Sens. Lett.* 4, 1137–1146. <http://dx.doi.org/10.1080/2150704X.2013.852704>.
- Huang, D., Zhang, R., Kim, K.C., Suarez, A.V., 2012. Spatial pattern and determinants of the first detection locations of invasive alien species in mainland China. *PLoS ONE* 7, e31734. <http://dx.doi.org/10.1371/journal.pone.0031734>.
- Hudgins, E.J., Liebhold, A.M., Leung, B., 2017. Predicting the spread of all invasive forest pests in the United States. *Ecol. Lett.* 20, 426–435. <http://dx.doi.org/10.1111/ele.12741>.
- Hulme, P.E., 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* 46, 10–18. <http://dx.doi.org/10.1111/j.1365-2664.2008.01600.x>.
- Iverson, L.R., Prasad, A.M., Hale, B.J., Sutherland, E.K., 1999. Atlas of current and potential future distributions of common trees of the eastern United States. Gen. Tech. Rep. NE-265. US Department of Agriculture, Forest Service, Northeastern Research Station, Radnor, PA.
- Jenerette, G.D., Clarke, L.W., Avolio, M.L., Pataki, D.E., Gillespie, T.W., Pincetl, S., Nowak, D.J., Hutrya, L.R., McHale, M., McFadden, J.P., Alonzo, M., 2016. Climate tolerances and trait choices shape continental patterns of urban tree biodiversity. *Global Ecol. Biogeogr.* 25, 1367–1376. <http://dx.doi.org/10.1111/geb.12499>.
- Johnston, M., 2015. *Trees in Towns and Cities: A History of British Urban Arboriculture*. Windgather Press, Oxford, UK.
- Jonnes, J., 2016. *Urban Forests: A Natural History of Trees and People in the American Cityscape*. Penguin, New York, NY.
- Kendal, D., Williams, N.S.G., Williams, K.J.H., 2012a. A cultivated environment: exploring the global distribution of plants in gardens, parks and streetscapes. *Urban Ecosyst.* 15, 637–652. <http://dx.doi.org/10.1007/s11252-011-0215-2>.
- Kendal, D., Williams, N.S.G., Williams, K.J.H., 2012b. Drivers of diversity and tree cover in gardens, parks and streetscapes in an Australian city. *Urban For. Urban Green.* 11, 257–265. <http://dx.doi.org/10.1016/j.ufug.2012.03.005>.
- Kenis, M., Auger-Rozenberg, M.-A., Roques, A., Timms, L., Péré, C., Cock, M.J.W., Settele, J., Augustin, S., Lopez-Vaamonde, C., 2009. Ecological effects of invasive alien insects. *Biol. Invasions* 11, 21–45. <http://dx.doi.org/10.1007/s10530-008-9318-y>.
- Koch, F.H., Yemshanov, D., Colunga-Garcia, M., Magarey, R.D., Smith, W.D., 2011. Potential establishment of alien-invasive forest insect species in the United States: where and how many? *Biol. Invasions* 13, 969–985. <http://dx.doi.org/10.1007/s10530-010-9883-8>.
- Kovacs, K.F., Haight, R.G., McCullough, D.G., Mercader, R.J., Siegert, N.W., Liebhold, A.M., 2010. Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. *Ecol. Econ.* 69, 569–578. <http://dx.doi.org/10.1016/j.ecolecon.2009.09.004>.
- Krist, F.J., Ellenwood, J.R., Woods, M.E., McMahon, A.J., Coward, J.P., Ryerson, D.E., Sapiro, F.J., Zweifler, M.O., Romero, S.A., 2014. 2013–2027 national insect and disease forest risk assessment. FHTET-14-01. US Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team, Fort Collins, CO.
- Lee, J.-H., Ko, Y., McPherson, E.G., 2016. The feasibility of remotely sensed data to estimate urban tree dimensions and biomass. *Urban For. Urban Green.* 16, 208–220. <http://dx.doi.org/10.1016/j.ufug.2016.02.010>.
- Liebhold, A.M., MacDonald, W.L., Bergdahl, D., Mastro, V.C., 1995. Invasion by exotic forest pests: a threat to forest ecosystems. *Forest Sci. Monograph* 30, 1–49.
- Liebhold, A.M., McCullough, D.G., Blackburn, L.M., Frankel, S.J., Von Holle, B., Aukema, J.E., 2013. A highly aggregated geographical distribution of forest pest invasions in the USA. *Divers. Distrib.* 19, 1208–1216. <http://dx.doi.org/10.1111/ddi.12112>.
- Lister, A.J., Scott, C.T., Rasmussen, S., 2012. Inventory methods for trees in nonforest areas in the great plains states. *Environ. Monit. Assess.* 184, 2465–2474. <http://dx.doi.org/10.1007/s10661-011-2131-6>.
- Lovett, G.M., Weiss, M., Liebhold, A.M., Holmes, T.P., Leung, B., Lambert, K.F., Orwig, D.A., Campbell, F.T., Rosenthal, J., McCullough, D.G., Wildova, R., Ayres, M.P., Canham, C.D., Foster, D.R., LaDeau, S.L., Weldy, T., 2016. Nonnative forest insects and pathogens in the United States: impacts and policy options. *Ecol. Appl.* 26, 1437–1455. <http://dx.doi.org/10.1890/15-1176>.
- MacFarlane, D.W., Meyer, S.P., 2005. Characteristics and distribution of potential ash tree hosts for emerald ash borer. *For. Ecol. Manage.* 213, 15–24. <http://dx.doi.org/10.1016/j.foreco.2005.03.013>.
- Mack, R.N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., Bazzaz, F.A., 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* 10, 689–710. [http://dx.doi.org/10.1890/1051-0761\(2000\)010\[0689:BICEGC\]2.0.CO;2](http://dx.doi.org/10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2).
- Maco, S.E., McPherson, E.G., 2003. A practical approach to assessing structure, function, and value of street tree populations in small communities. *J. Arboric.* 29, 84–97.
- Mainton, J.H., Braun, W.J., 2015. DAAG: data analysis and graphics data and functions, version 1.22. R Foundation for Statistical Computing, Vienna, Austria.
- Martin, N.A., Chappellka, A.H., Somers, G., Loewenstein, E.F., Kever, G.J., 2013. Evaluation of sampling protocol for i-Tree Eco: a case study in predicting ecosystem services at Auburn University. *Arboric. Urban For.* 39, 56–61.
- McKinney, M.L., 2006. Urbanization as a major cause of biotic homogenization. *Biol. Conserv.* 127, 247–260. <http://dx.doi.org/10.1016/j.biocon.2005.09.005>.
- McPherson, E.G., Luttinger, N., 1998. From nature to nurture: the history of Sacramento's urban forest. *J. Arboric.* 24, 72–88.
- McPherson, G., Simpson, J.R., Peper, P.J., Maco, S.E., Xiao, Q., 2005. Municipal forest benefits and costs in five U.S. cities. *J. Forest.* 103, 411–416.
- Mercader, R.J., Siegert, N.W., Liebhold, A.M., McCullough, D.G., 2011. Influence of foraging behavior and host spatial distribution on the localized spread of the emerald ash borer, *Agrilus planipennis*. *Popul. Ecol.* 53, 271–285. <http://dx.doi.org/10.1007/s10144-010-0233-6>.
- Miller, R.W., Hauer, R.J., Werner, L.P., 2015. *Urban Forestry: Planning and Managing Urban Greenspaces*, third ed. Waveland Press, Long Grove, IL.
- Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards Jr, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecol. Modell.* 199, 176–187. <http://dx.doi.org/10.1016/j.ecolmodel.2006.05.021>.
- Nielsen, A.B., Østberg, J., Delshammar, T., 2014. Review of urban tree inventory methods used to collect data at single-tree level. *Arboric. Urban For.* 40, 96–111.
- Nowak, D.J., Rowntree, R.A., McPherson, E.G., Sisinni, S.M., Kerkmann, E.R., Stevens, J.C., 1996. Measuring and analyzing urban tree cover. *Landsc. Urban Plan.* 36, 49–57. [http://dx.doi.org/10.1016/S0169-2046\(96\)00324-6](http://dx.doi.org/10.1016/S0169-2046(96)00324-6).
- Nowak, D.J., Noble, M.H., Sisinni, S.M., Dwyer, J.F., 2001a. People and trees: assessing

- the US urban forest resource. *J. Forest.* 99, 37–42.
- Nowak, D.J., Pasek, J.E., Sequeira, R.A., Crane, D.E., Mastro, V.C., 2001b. Potential effect of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) on urban trees in the United States. *J. Econ. Entomol.* 94, 116–122. <http://dx.doi.org/10.1603/0022-0493-94.1.116>.
- Nowak, D.J., Dwyer, J.F., 2007. Understanding the benefits and costs of urban forest ecosystems. In: Kuser, J.E. (Ed.), *Urban and Community Forestry in the Northeast*, second ed. Springer, New York, pp. 25–46.
- Nowak, D.J., Walton, J.T., Stevens, J.C., Crane, D.E., Hoehn, R.E., 2008a. Effect of plot and sample size on timing and precision of urban forest assessments. *Arboric. Urban For.* 34, 386–390.
- Nowak, D.J., Crane, D.E., Stevens, J.C., Hoehn, R.E., Walton, J.T., Bond, J., 2008b. A ground-based method of assessing urban forest structure and ecosystem services. *Arboric. Urban For.* 34, 347–358.
- Nowak, D.J., Stein, S.M., Randler, P.B., Greenfield, E.J., Comas, S.J., Carr, M.A., Alig, R. J., 2010. Sustaining America's urban trees and forests. Gen. Tech. Rep. NRS-62. US Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA.
- Nowak, D.J., Bodine, A.R., Hoehn, R.E., III, Edgar, C.B., Hartel, D.R., Lister, T.W., Brandeis, T.J., 2016. Austin's urban forest, 2016. Resource Bulletin NRS-100. US Department of Agriculture, Forest Service, Newtown Square, PA.
- Östberg, J., Delshamm, T., Wiström, B., Nielsen, A.B., 2013. Grading of parameters for urban tree inventories by city officials, arborists, and academics using the Delphi method. *Environ. Manage.* 51, 694–708. <http://dx.doi.org/10.1007/s00267-012-9973-8>.
- Östberg, J., 2013. Tree inventories in the urban environment: methodological development and new applications. Ph.D. thesis. Swedish University of Agricultural Sciences, Alnarp, Sweden.
- Paini, D.R., Sheppard, A.W., Cook, D.C., De Barro, P.J., Worner, S.P., Thomas, M.B., 2016. Global threat to agriculture from invasive species. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7575–7579. <http://dx.doi.org/10.1073/pnas.1602205113>.
- Parker, I.M., Simberloff, D., Lonsdale, W.M., Goodell, K., Wonham, M., Kareiva, P.M., Williamson, M.H., Von Holle, B., Byers, J.E., Goldwasser, L., 1999. Impact: toward a framework for understanding the ecological effects of invaders. *Biol. Invasions* 1, 3–19. <http://dx.doi.org/10.1023/A:1010034312781>.
- Perrings, C., Williamson, M., Barbier, E.B., Delfino, D., Dalmazzone, S., Shogren, J., Simmons, P., Watkinson, A., 2002. Biological invasion risks and the public good: an economic perspective. *Conserv. Ecol.* 6 (1), 1. <http://dx.doi.org/10.5751/ES-00396-060101>.
- Poland, T.M., McCullough, D.G., 2006. Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *J. Forest.* 104, 118–124.
- R Core Team, 2016. R: a language and environment for statistical computing, version 3.3.3 ("Another Canoe"). R Foundation for Statistical Computing, Vienna, Austria. < <http://www.R-project.org> > .
- Ramage, B.S., Roman, L.A., Dukes, J.S., 2013. Relationships between urban tree communities and the biomes in which they reside. *Appl. Veg. Sci.* 16, 8–20. <http://dx.doi.org/10.1111/j.1654-109X.2012.01205.x>.
- Rodríguez, J.D., Pérez, A., Lozano, J.A., 2010. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal.* 32, 569–575. <http://dx.doi.org/10.1109/TPAMI.2009.187>.
- Roman, L.A., McPherson, E.G., Scharenbroch, B.C., Bartens, J., 2013. Identifying common practices and challenges for local urban tree monitoring programs across the United States. *Arboric. Urban For.* 39, 292–299.
- Roman, L.A., Scharenbroch, B.C., Östberg, J.P.A., Mueller, L.S., Henning, J.G., Koeser, A.K., Sanders, J.R., Betz, D.R., Jordan, R.C., 2017a. Data quality in citizen science urban tree inventories. *Urban For. Urban Green.* 22, 124–135. <http://dx.doi.org/10.1016/j.ufug.2017.02.001>.
- Roman, L.A., Fristensky, J.P., Eisenman, T.S., Greenfield, E.J., Lundgren, R.E., Cerwinka, C.E., Hewitt, D.A., Welsh, C.C., 2017b. Growing canopy on a college campus: understanding urban forest change through archival records and aerial photography. *Environ. Manage.* 60, 1042–1061. <http://dx.doi.org/10.1007/s00267-017-0934-0>.
- Sherrod, P.H., 2007. DTREG, version 6.0 ("Advanced Version"). P. H. Sherrod, Brentwood, Tennessee, USA. < <http://www.dtreg.com> > .
- Sherrod, P.H., 2014. DTREG predictive modeling software – users manual. P. H. Sherrod, < <http://www.dtreg.com> > .
- Sydnor, T.D., Bumgardner, M., Todd, A., 2007. The potential economic impacts of emerald ash borer (*Agrilus planipennis*) on Ohio, U.S., communities. *Arboric. Urban For.* 33, 48–54.
- Sydnor, T.D., Subburayalu, S., Bumgardner, M., 2010. Contrasting Ohio nursery stock availability with community planting needs. *Arboric. Urban For.* 36, 47–54.
- Sydnor, T.D., Bumgardner, M., Subburayalu, S., 2011. Community ash densities and economic impact potential of emerald ash borer (*Agrilus planipennis*) in four mid-western states. *Arboric. Urban For.* 37, 84–89.
- Tait, C.J., Daniels, C.B., Hill, R.S., 2005. Changes in species assemblages within the Adelaide metropolitan area, Australia, 1836–2002. *Ecol. Appl.* 15, 346–359. <http://dx.doi.org/10.1890/04-0920>.
- US Government Accountability Office, 2006. Invasive forest pests: lessons learned from three recent infestations may aid in managing future efforts. GAO-06-353. US Government Accountability Office, Washington, DC.
- USDA Agricultural Research Service, 2012. USDA plant hardiness zone map. < <http://planthardiness.ars.usda.gov> > (accessed March 23 2017).
- USDA Forest Service, 2016. US Forest Service – Forest Inventory & Analysis (FIA) cities. < <https://www.nrs.fs.fed.us/fia/urban/docs/Urban%20FIA%20Implementation%20Map%2012-2016.pdf> > (accessed May 17, 2017).
- Venette, R.C., Kriticos, D.J., Magarey, R.D., Koch, F.H., Baker, R.H.A., Worner, S.P., Gómez Raboteaux, N.N., McKenney, D.W., Dobesberger, E.J., Yemshanov, D., De Barro, P.J., Hutchison, W.D., Fowler, G., Kalaris, T.M., Pedlar, J., 2010. Pest risk maps for invasive alien species: a roadmap for improvement. *Bioscience* 60, 349–362. <http://dx.doi.org/10.1525/bio.2010.60.5.5>.
- Willmott, C.J., Feddema, J.J., 1992. A more rational climatic moisture index. *Prof. Geogr.* 44, 84–87. <http://dx.doi.org/10.1111/j.0033-0124.1992.00084.x>.
- Wood, S.N., 2006. Generalized Additive Models: An Introduction With R. Chapman and Hall/CRC, Boca Raton, FL.
- Wood, S.N., 2016. mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation, version 1.8-16. R Foundation for Statistical Computing, Vienna, Austria. < <https://cran.r-project.org/package=mgcv> > .
- Yang, J., La Sorte, F.A., Pyšek, P., Yan, P., Nowak, D., McBride, J., 2015. The compositional similarity of urban forests among the world's cities is scale dependent. *Global Ecol. Biogeogr.* 24, 1413–1423. <http://dx.doi.org/10.1111/geb.12376>.
- Zhang, C., Qiu, F., 2012. Mapping individual tree species in an urban forest using airborne lidar data and hyperspectral imagery. *Photogramm. Eng. Rem. S.* 78, 1079–1087. <http://dx.doi.org/10.14358/PERS.78.10.1079>.