International Journal of Geographical Information Science, 2013
http://dx.doi.org/10.1080/13658816.2013.840381

Delineating generalized species boundaries from species distribution data and a species distribution model

Matthew P. Peters[a]*, Stephen N. Matthews[a,b], Louis R. Iverson[a] and Anantha M. Prasad[a]

[a]*Northern Research Station, USDA Forest Service, Delaware, OH, USA;* [b]*Terrestrial Wildlife Ecology Lab, School of Environment and Natural Resources, The Ohio State University, Columbus, OH, USA*

(*Received 3 May 2013; accepted 27 August 2013*)

Species distribution models (SDM) are commonly used to provide information about species ranges or extents, and often are intended to represent the entire area of potential occupancy or suitable habitat in which individuals occur. While SDMs can provide results over various geographic extents, they normally operate within a grid and cannot delimit distinct, smooth boundaries. Additionally, there are instances where a zone of primary occupancy (i.e., a mostly continuous region where species exists, excluding outliers) is better suited for particular analyses, such as when examining source/sink population dynamics or modeling movement into new habitats. We present a semi-automated method to delineate a generalized species boundary (GSB) from SDM output, which provides a practical alternative to digitizing. This preliminary boundary is then manually updated based on inventory data and historical ranges. We used the method to generate contemporary boundaries for 132 tree species of the eastern United States, which are complementary to the ranges generated by Elbert Little for North America during the 1970s, but are not replacements. The methods we present can broadly be applied to other grid-based SDM to generate GSBs.

Keywords: eastern United States; forest inventory and analysis; geoprocessing; R statistical language; tree ranges

Introduction

Current ranges and distributions of many tree species have long been delineated (Munns 1938, Little 1971), or more recently, modeled (Moore *et al*. 1991, Skyes *et al*. 1996, Harrison *et al*. 2006, Rehfeldt *et al*. 2006, McKenney *et al*. 2007, Iverson *et al*. 2008). However, species ranges delineated from large-scale field and herbarium surveys are usually now being superseded by process-based or statistical species distribution models (SDM) to map the potential extent of suitable habitat for a species. SDMs may use environmental data obtained from remote sensing along with field survey records to develop empirical relationships to predict species abundance or occurrence over a landscape (Franklin and Miller 2009). These correlative relationships provide needed

---

*Corresponding author. Email: matthewpeters@fs.fed.us

information about the ways in which species distributions are shaped by their surrounding environment and can then be applied to unsampled locations in a predictive manner to map presence/absence or abundance/importance of a species. While the model output is continuous and highly informative, the resulting maps make it difficult to estimate distinct boundaries associated with the species (MacArthur 1972, Kirkpatrick and Barton 1997). Many processes (e.g., climate, land use, site quality, etc.) govern the area occupied by a species, which result in ranges with unique patterns, especially near range boundaries where conditions often are less than optimal and thresholds of habitat suitability ebb and flow. This dynamic along the range boundaries results from competition among species and disturbance events. Because these interactions vary over time and space, there are numerous ways to define the range boundary, including the use of clustering and edge detection algorithms (Fortin and Drapeau 1995), fuzzy set algorithms (Leung 1987, Burrough and Frank 1996), or Mahalanobis distance algorithms (Sangermano and Eastman 2012). Regardless of the method, species ranges are often generalized to represent the likelihood of the species being present near the edge of the range (Brown *et al*. 1996, Gaston 2009).

In this paper, we define a 'generalized species boundary' (hereafter GSB) as a highly connected region ranging from high to low density of potential suitable habitat. We have generated these boundaries for 132 of the 134 tree species previously modeled for the eastern United States with DISTRIB (Iverson *et al*. 2008). These models are based on species-level data, and as such, they ignore genetic variation within species ranges, such an extensive data set does not exist for most species. We realize that genetic variation plays an important role in determining the successful establishment and survival of a species in a particular place, and recognize the need for further research and data acquisition of genetic spatial variation for many species.

The use of a GSB, or the primary zone of occupancy for a particular species, has the benefit of excluding potential outlier populations when exploring species migration potentials (Ordonez and Williams 2013, Prasad *et al*. 2013), or when defining range limits (Purves 2009). Analyses that utilize the full species' range as defined by an unmodified SDM are susceptible to introducing potential errors, or at least numerous outlier 'islands', at or near the edge of the range boundary where greater uncertainty exists at the tail of a distribution (Jiménez-Valverde and Lobo 2007). This error is particularly problematic for binary range boundaries (e.g., Little's range boundaries, Little 1971), where no abundance data are available, or when small sample sizes are used. To further reduce error propagation along the boundary edges, as described by Brown *et al*. (1996), for tree species using our DISTRIB models, we set out to define the zone of primary occupancy of suitable habitat for each species to generate GSBs.

Range boundaries occur naturally as a result of various interacting phenomena, and there are several ways to define a species range (Gaston 1991, Brown *et al*. 1996). However, there is little literature on how to create a boundary via digital processes (Purves 2009). This deficit may be due to the fact that in recent decades, most researchers are using grid-based SDMs to identify potential suitable habitat (Graham and Hijmans 2006), thus replacing the need for maps and vector boundaries digitized from survey data. However, each SDM has its own set of caveats [e.g., environmental equilibrium, inclusion of habitat characteristics, and relevant scale (spatial and temporal), Franklin and Miller 2009], and it might be undesirable to include the entire extent of suitable habitat from a SDM for certain analyses (e.g., those considering the core or primary region of a species). In this paper, we describe and present a new method for creating a boundary of primary occupancy from a grid-based SDM output of potential species abundance.

Methodology

Data

During the 1970s, Elbert Little published a five-volume atlas of tree species ranges in North America delineated from published tree distributions, herbarium records, and field surveys. In general, Little's boundaries define a generous extent for the species, with lines drawn around the greatest extent of the species, and thus there are often large areas within the boundaries that have sparse occupancy. The final ranges were reviewed by botanists, foresters, and other experts (Little 1971), and today digital copies of the range boundaries are available from the US Geological Survey (http://esp.cr.usgs.gov/data/atlas/little/) for 679 tree and shrub species. Little's ranges are widely used, including by us for this paper, to define a historical extent for a species' distribution throughout the United States. It is important to emphasize that the work we present in this paper is not intended to replace Little's boundaries, but to complement them in that we focus on the 'zone of primary occupancy', rather than the extent of the species. Thus, we would expect that for most species, Little's ranges would be larger than the GSBs.

The US Forest Service's Forest Inventory and Analysis (FIA) program is a long-term effort mandated by the US Congress with the principal purpose to survey the extent, condition, volume, growth, and depletions of timber on US-forested lands (Smith 2006). Plot-level data were used to indicate the presence or absence of species as a visual check while editing the GSBs. Plot locations with truncated coordinates (Lister *et al*. 2005) from periodic inventories during 1980–1993 (hereafter 'old', Smith 2006) provided information related to the distribution of tree species of the current model. Additionally, newer annualized inventories during 1999–2010 (Woudenberg *et al*. 2010) were combined with the earlier periodic records (1980–2010, hereafter 'newer') to help with the delineation of core boundaries. At each FIA plot, the most recent inventory records were used to calculate importance values for 132 modeled tree species (Iverson *et al*. 1998).

Predicted importance values (IV, i.e., relative abundance) from the DISTRIB model (Iverson and Prasad 2008) use 38 environmental predictor variables and FIA data to statistically model potential suitable habitat for tree species. An ensemble technique using RandomForest, bagging, and regression tree analysis (Prasad *et al*. 2006) is employed to relate IVs at plot locations with environmental predictors. Output from DISTRIB (http://www.nrs.fs.fed.us/atlas/) has a 20 × 20 km resolution across the eastern United States and predicts importance values representing suitable habitat, even where few forested FIA plots existed, under both current (1961–1990) and future (2071–2100) climate scenarios. Each species model has a potential value of 0–100 for each grid and is rated for reliability based on a series of statistical procedures, so that users know how to evaluate the predicted IVs (Iverson *et al*. 2008). The models thus range from high to low reliability, with less reliability where FIA data for the species are sparse or wide-ranging. DISTRIB differs from many other SDMs in that it produces a grid of predicted IVs (i.e., potential abundance), rather than a probability of presence/absence, thus low values indicate areas where a species' habitat is potentially suitable but where individuals would be found only in low numbers.

Procedures

A digital processing method, using R statistical language (R Development Core Team 2010), and custom geoprocessing tools, was used to delineate a representation of the

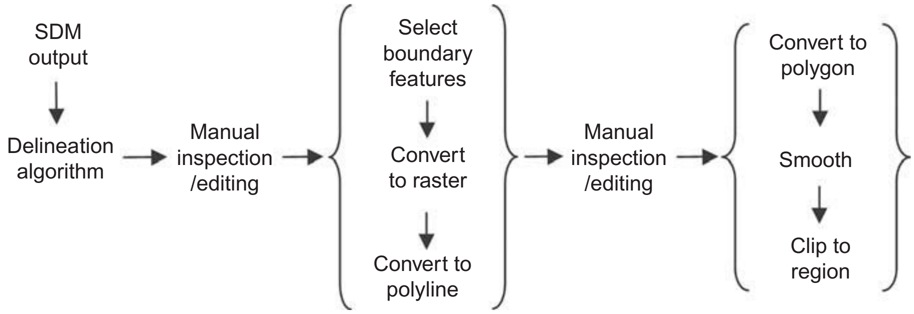

Figure 1. Diagram of the processes used to generate a generalized species boundary.

species' modeled distribution boundary (Figure 1). The contourLines function from the grDevices library of R was used to define preliminary range boundaries based on the output of DISTRIB, following the method used by Matthews *et al*. (2011). To delineate the modeled boundary, a binary contour was identified for SDM values representing presence (IV $\geq$1) or absence. Because IVs represent potential suitable habitat related to species abundance, it is reasonable to include low values to initially define the species GSB. Outliers were removed if contour lines contained $<$ 12 segments, typically 3 adjacent DISTRIB cells optimized for the minimum number of cells, defined by the number of unique segments of the vectorized grid across all species modeled by DISTRIB. By removing small isolated polygons, irrespective of their potential habitat suitability, we identify the larger continuous polygons of the species' distribution. Without implementing a segment threshold, the contours contained numerous small segments, both inside and outside the species' modeled extent. The resulting 20 × 20 km vectorized grid consisted of cells designating the rough boundary of the species' current modeled habitat.

Shapefiles of line segments were then generated from the delineation algorithm representing the outline of a species' boundary in a presence/absence format, followed by additional GIS processing to edit and finalize the GSB. Several custom tools were developed in ArcGIS 9.3 ModelBuilder (ESRI 2009), to reduce the set-up time of individual processes. Although much of the work was automated, users must interact with the intermediate products generated throughout the boundary generation process by manually inspecting and editing input/output files as explained below, and setting up the model parameters.

The initial step is to visualize the output file from the delineation algorithm containing the boundary outline (Figure 2a) and determine how the primary zone of occupancy for the species should encompass the inventory data. The algorithm identifies the boundary based on neighboring cell values from the DISTRIB modeled current output (Figure 2b), which can cause 'islands', 'gaps', 'coves', and 'peninsulas' to be present in the initial boundary. Islands are defined as smaller boundaries detached from the larger GSB distribution, and conversely, gaps are voids within the GSB (Figure 3). Coves are defined as small areas along the GSB edge that contain 1–2 cells indicating a gap, and peninsulas occur when two or more cells are aligned in a straight line (Figure 3). Depending on the purpose of the boundary delineation, it is often necessary to edit the initial file to simplify the final boundary, which we performed by manually selecting cells and changing the boundary value (presence or absence), within the attribute table so that the GSB follows the desired path. Small alterations to the boundary files were thus accomplished subjectively, but with

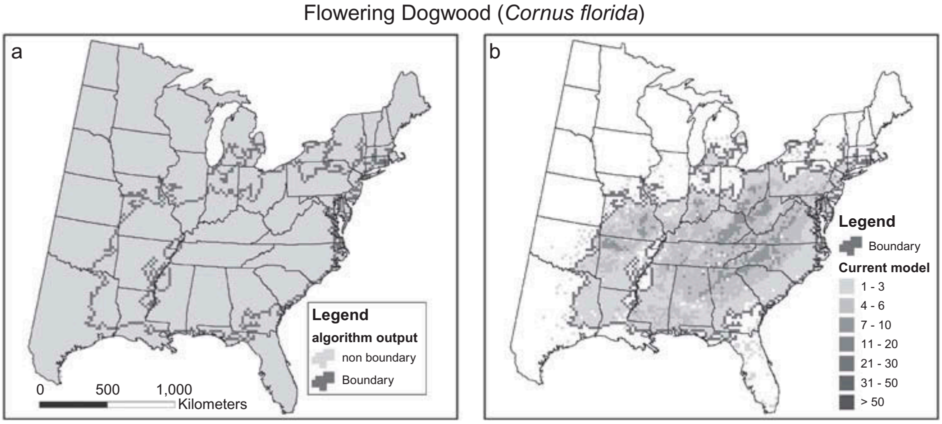

Figure 2. Unedited results from the delineation algorithm with A) the generalized species boundary and B) the generalized species boundary with the current modeled DISTRIB importance values.

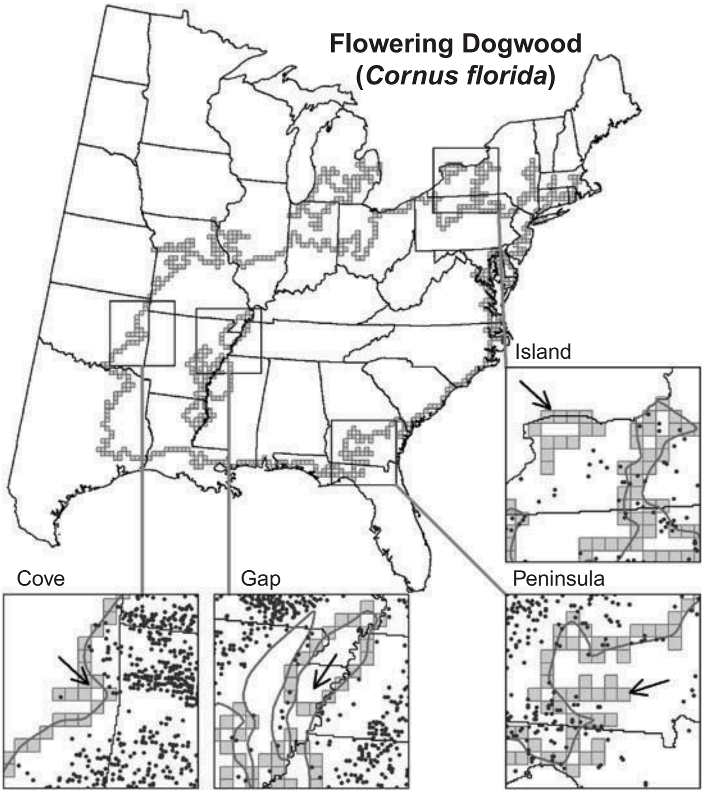

Figure 3. The gray boundary derived from the delineation algorithm may contain various features (coves, gaps, islands, and peninsulas) which often need to be edited when creating the generalized species boundary. The dark line within the insets is the final smoothed boundary and points are the FIA plots reporting presence.

ample underlying information consisting of the current modeled DISTRIB values (IVs $\geq$ 1 interpreted as present), the density of FIA points (presence vs. absence), and Little's historical range, simultaneously viewed in a GIS. We created holes, or areas of absence, within the primary zone of occupancy when sufficient FIA sampling (1 plot per $\sim$6 km^2) indicated that the species was not found among many plots (usually $<$1.5% tolerance of presence).

Once this initial editing was completed, a rudimentary polyline outlining the extent of the primary zone of occupancy was created. This was an automated step where cells identified as 'present' in the edited boundary shapefile were converted to a temporary raster file with the same dimensions as the shapefile, by using the cell center option for cell assignment type and setting the cell size equal to the SDM output (20 $\times$ 20 km in our case). The raster file was then converted to a polyline with a minimum dangle length set to 1.25 times the cell size and 'simplify polyline' set to true. The dangle length parameter is used to remove lines that do not connect or are noncontinuous, thus setting the minimum length $\geq$ 1 cell width ensured that each included line could be present in a single cell.

The resulting boundary line then needed to be inspected and manually edited to remove unwanted lines (Figure 4a) or to correct the shape of lines that should follow another path (Figure 4b). The process used to convert the raster grid cells to a line places the line along the cell centroid; thus cells along the coast can underestimate the extent by not including the coastline by roughly one-half of a cell width. Where data support the presence of a species to the coast (or within 1 cell of the coast), the rough boundary line was manually digitized outward to encompass the coastline (to be later clipped by the mapped coastline). Once satisfied that the lines conform to the primary zone of occupancy for a particular species (based on the modified presence/absence vectorized grid), a second custom tool was used to finalize the shape of the boundary. Here, the edited boundary polyline was converted to a polygon, after which the outline was smoothed using the Polynomial Approximation with Exponential Kernel (PAEK) algorithm with a smoothing tolerance set, by trial and error to 1.75 times the cell size (35 km in our case). The smoothing provides a continuous boundary that reduces artifacts imposed by coarse grid cells and creates more biologically realistic and esthetically reasonable boundaries. The PAEK algorithm was the default option for

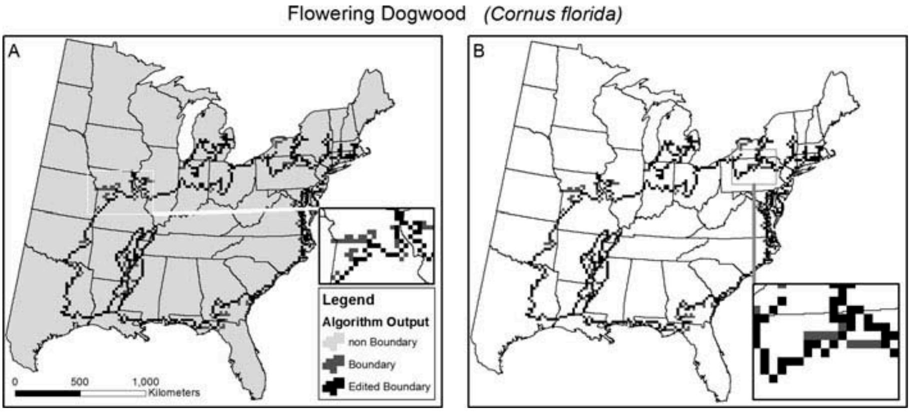

Figure 4. The original boundary from the delineation algorithm (light and medium gray) with the edited boundary (black). The inset in panel A shows removed cells (medium gray) from the edited boundary (black). Panel B shows the original boundary (medium gray) with the edited boundary (black) and the inset shows an alternative path where the few black cells in the center have been added while the gray cells are the original path.

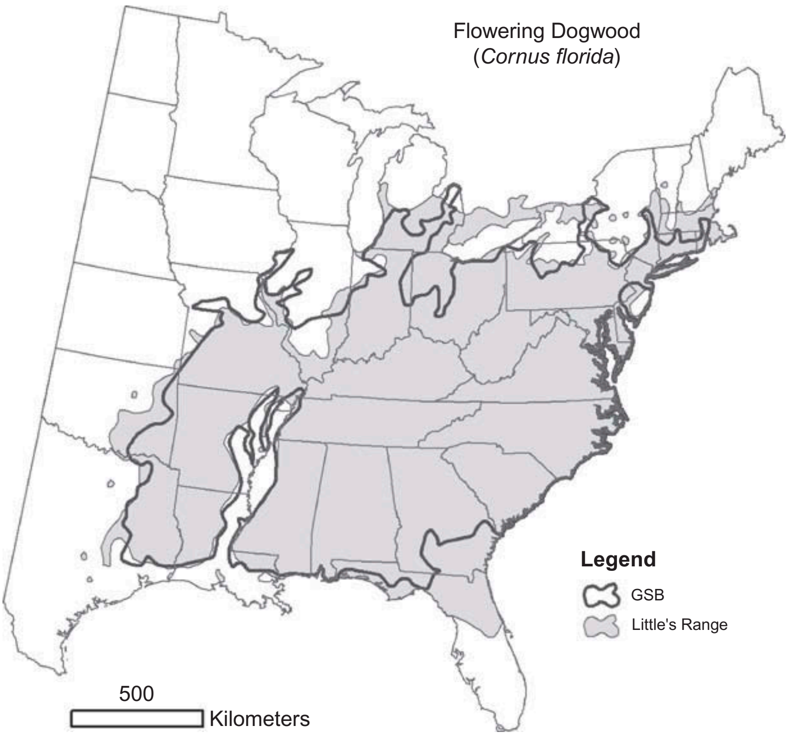

Figure 5. Final GSB along with Little's range.

smoothing a polygon and does not generalize past the original extent. Lastly, the smoothed boundary was clipped to the outline of the coastline to remove any generalization that may have included areas outside the continental land mass (Figure 5).

The final boundary polygon represents the primary zone of occupancy for a species based on the current model of a SDM and recent inventory data and should be evaluated for accuracy. Boundaries that include gaps should especially be examined, as they were found to contain errors when overlapping polygons did not share the same vertices. Again this is where manual editing, at a fine scale, may need to be performed to snap vertices of one feature to the adjoining feature. One final examination and reality check is then required, and at times, decisions need to be made and remade (with expert opinion), regarding where the 'zone of primary occupancy' should lie. Species of rare but extensive distribution are especially difficult to settle on a satisfactory boundary; it will be the case for these species that many possibilities exist for how 'primary occupancy' is defined and where the GSB is finally drawn.

Analyses

GSBs were generated for 132 tree species modeled with DISTRIB, though we focus on flowering dogwood (*Cornus florida*), to describe some processes of creating the GSBs. To analyze trends within the new GSBs, we conducted three sets of GIS overlays with resulting calculations: (1) the area of Little's range (within the eastern United States), as compared to the area of the GSB representing presence; (2) the portions of each range extending beyond the other (e.g., percent of GSB's area beyond Little's range and the percent of Little's range beyond GSB boundaries); and (3) the percent of FIA plots, from two sets of FIA data, reporting presence contained within each boundary (Table 1).

Using combinations of values based on the proportion of area shared among FIA 'presence' points, Little's ranges, and the GSBs, the species were assigned to a general class indicating the extent of the GSB as compared to Little's range (i.e., sparse, no difference, broader, and narrower, see Table 2) to identify general trends among each group. Sparse indicates the species is widely found in low densities and a single zone of occupancy is not readily defined; no difference means GSB and Little are similar; broader indicates GSB has a greater area than Little, and narrower indicates Little has a greater area than GSB. To interpret the general classes of Table 2, the species were classified as large for Little's range if the area of Little's range shared $\geq$ 75% of its area with the GSB; otherwise, it was considered small. Likewise, the species had a GSB registered as small if $<$ 75% of its area, with $<$ 15% extending beyond Little's range, was shared; otherwise, it was considered large. Finally, the species had percent FIA classed as high if the GSB contained $\geq$ 75% of all FIA presence records, medium if 50–75%, and low if $<$ 50%.

Results

A comparison of the area defined by Little's range and the GSB for a select number of species is provided in Table 1. Overall, 7 of the 132 tree species' GSBs were larger than Little's range (Supplementary Table 1), providing evidence that the GSBs provide a 'primary zone of occupancy', rather than the full extent of the species. This result also emphasizes that we are not replicating Little's ranges, but producing a complementary product. Among all GSBs, the percentage of all old (1980–1993) and newer (1980–2010) FIA plots reporting presence within each GSB was roughly similar (±5%) for 107 species, indicating a good fit of actual species presence to the GSB. Of the remaining 25 GSBs with a larger ($\geq \pm$5%) difference between old and newer FIA plots, only one species (*Quercus ilicifolia*) had a lower proportion of newer FIA plots reporting presence, while the remaining 24 species had higher proportions of newer FIA plots reporting presence than the older plots. These statistics indicate that the GSBs do well to account for the newest FIA information available.

Classifying the species as having a sparse, no difference, broader, or narrower extent for GSB as compared to Little's range resulted in 15, 39, 5, and 73 species, respectively (Supplementary Table 1). Of the sparse GSBs, each had a relatively larger range defined by Little, with only 9–50% of their area shared with GSB, and capturing only 52–73% (median 64.6%) of the old and 39–73% (median 68.1%) of the newer FIA presences (Table 1). As expected, these sparse species are the most difficult to draw a boundary around the 'zone of primary occupancy'.

The five species classed as broader (Supplementary Table 1) have GSBs with smaller to equal sized areas, as compared to Little's ranges (84–100% of area shared), but captured 95–99% (median 96.8%) of old and 94–99% (median 96.3%) of newer FIA presences (Table 1). These represent species that, based on FIA data, have possibly expanded their range since delineation by Little over 40 years previously. The 73 species with narrower (Supplementary Table 1) GSBs have larger Little's ranges (19–74% of area shared), captured 61–98% (median 90%) of old and 62–97% (median 88%) of newer FIA presences. Species in the no difference class (Supplementary Table 1) have $\geq$75% of the GSB shared with Little's range and capture between 66–99% (median 97%) of old and 62–99% (median 97%) of newer FIA presences (Table 1).

Table 1. Summary of select species with area of Little's range (within the eastern United States) and GSB, and the percentage of area shared by each along with the area beyond the intersection of each range. Additionally, the percentage of all FIA plots reporting presence within the GSB is included. Species are sorted alphabetically by scientific name within each class, and the ranges of values following the groups are indicated. See Table 2 for criteria defining groups and Supplementary Table 1 for the full list of 132 species.

			Area (sq. km)		Percent shared		Percent		Percent FIA	
	Scientific name	Common name	Little	GSB	Little	GSB	Little beyond GSB	GSB beyond Little	1980–1993	1980–2010
Sparse	*Asimina triloba*	pawpaw	1,571,484	157,490	10.0	100	90.0	0	62.1	47.6
	Catalpa speciosa	northern catalpa	45,074	73,229	16.0	9.9	84.0	90.1	63.3	39.7
	Quercus palustris	pin oak	902,159	285,630	30.7	96.6	69.4	3.1	72.0	72.7
	Salix nigra	black willow	3,085,299	758,240	21.2	92.6	77.2	13.9	67.5	54.5
			45,074 – 3,085,299	**27,405 – 758,240**	**9–50**	**10–100**	**50–91**	**0–90**	**52–73**	**39–73**
Broader	*Nyssa biflora*	swamp tupelo	79,722	419,205	100	19.0	0	81.0	95.4	94.5
	Pinus elliottii	slash pine	262,457	447,545	96.6	56.6	3.4	43.4	98.5	98.2
	Pinus taeda	loblolly pine	867,770	1,053,296	98.1	80.8	1.9	19.2	99.2	99.2
	Quercus coccinea	scarlet oak	930,673	930,152	84.9	85.0	15.1	15.0	95.4	95.2
			79,722 – 930,673	**419,205 – 1,053,296**	**84–100**	**19–85**	**0–15**	**15–81**	**95–99**	**94–99**
Narrower	*Chamaecyparis thyoides*	Atlantic white-cedar	138,500	37,230	24.3	90.5	75.7	9.5	71.3	77.4
	Fraxinus pennsylvanica	green ash	3,451,325	2,362,137	68.1	99.8	31.7	0.5	90.7	90.4
	Sassafras albidum	sassafras	2,144,179	1,270,379	58.7	99.0	41.3	1.0	94.0	91.9
	Tsuga canadensis	eastern hemlock	866,790	642,535	72.3	97.6	27.7	2.4	97.7	97.2
	Ulmus americana	American elm	3,773,064	2,450,148	64.7	100	32.7	0.3	97.0	95.2
			138,500 – 3,773,064	**37,230 – 2,450,148**	**19–74**	**86–100**	**26–81**	**0–14**	**61–98**	**62–97**
No difference	*Acer rubrum*	red maple	2,628,588	2,599,880	96.6	97.7	3.4	2.3	99.2	99.2
	Pinus clausa	sand pine	59,096	31,112	32.3	61.4	67.7	38.6	76.6	76.4
	Pinus strobus	eastern white pine	1,021,382	974,543	87.6	91.8	12.4	8.2	97.1	96.6
	Prunus serotina	black cherry	2,869,200	2,609,679	88.5	97.2	11.6	2.6	98.2	98.4
	Quercus alba	white oak	2,436,085	2,402,320	94.6	95.6	5.7	4.1	99.0	99.2
			59,096 – 3,773,064	**31,112 – 2,609,679**	**19–97**	**5–100**	**3–81**	**0–95**	**61– 99**	**62–99**

Table 2. Range classes defined by Little’s range and GSB area, and the portion of FIA. reported presence.

Little’s range	GSB	Percent FIA	Class
Small	Small	High	No difference
Small	Small	Medium	Sparse
Small	Large	High	Broader
Small	Large	Medium	Sparse
Large	Small	High	Narrower
Large	Small	Medium	Sparse
Large	Large	High	No difference
Large	Large	Medium	Sparse

Discussion

Our method to derive tree species boundaries provides a quick and cost-effective mapped estimate of the primary zone of occupancy for tree species, where the species would have a reasonable probability of being found, as opposed to the more generous definition of range, where the species may be seen occasionally at their fringes of extent. To accomplish this task, we used inputs from three sources: Little’s boundaries published in the 1970s, recent (1980–2010) FIA presence/absence data (approximate locations of plots), and output from the DISTRIB model which statistically predicts suitable habitat based on FIA data and 38 environmental variables (Iverson *et al*. 2008). There are many 20 × 20 km cells that are modeled to have suitable habitat for a species even though there may be no record of the species according to the FIA plot data; indeed, many lesser forested cells, such as in the highly agricultural ‘Corn Belt’ zone (e.g., Iowa, Illinois, and Indiana), do not even have one forested FIA plot. Therefore, by including the vectorized grid cell that represents suitable habitat, the reduced density of FIA data can be partially overcome in the final placement of the boundary. Additionally, our GSB finally represents the primary zone of occupancy and not the full range of occupancy which is represented better with Little’s maps. The GSB also represents an update with the latest data available; with the Little’s boundaries derived from 1960s and early 1970s data, there could be some changes appearing in the intervening period (though we have no concrete evidence for this based on analyses presented here).

For six species with low model reliability and sparse distributions, usually due to too few known presence locations (i.e., rare species), the delineation algorithm produced undesirable results. For such species, more manual editing and agreement among the four authors was needed to more subjectively create the boundary. While FIA data provided the primary source for forest records within the United States, Westfall (2009) examined its accuracy based on the quality assurance program used to validate FIA data. Thirty-eight of the 51 genera for which GSBs were created are included in Westfall’s report, of which 10 have a high accuracy ($\geq$ 75%) of species identification between inventories, 14 are moderate ($\geq$ 50%), and 14 are low ($<$ 50%). Regardless of how well the delineation algorithm or final GSB is perceived to perform, the accuracy of inventory data and model performance of the SDM will primarily drive the delineation of GSBs.

It may seem redundant to begin with a raster output from the SDM, vectorize the grid and then convert it back to a raster, then a polyline, and finally to a polygon; however, the alternative of digitizing would require much more user interaction and the entire process could be biased depending on where the user places line segments and vertices. By converting the original raster to the vectorized grid, the algorithm of counting line segments (to thin out outliers) can be performed. Then with the conversion to a raster, we can define the placement of vertices globally as the centroids of each raster cell, which are then converted into a set of polylines, and which are then manually edited before cleaning up into the final boundary polygon files. Thus, we use some automation where traditional digitizing would require a large amount of user interaction. An advantage of creating polygon boundaries over the line files that were created during the process is the ability to calculate area statistics or perform overlay analyses with other files based on the polygons' geometry and spatial configuration.

While the entire process can usually be completed for one species in under an hour by someone familiar with the data (and with the data all in place), there are disadvantages with this process. These include the manual inspections and editing between automated procedures, generalization along the outer edges where inventory data may be sparse, and determining general rules for defining the primary zone of occupancy versus gaps or islands. Such rules should consider the purpose of the resulting boundary to deal with undesirable artifacts from the delineation algorithm. Additionally, our method requires extensive information (e.g., an extensive inventory data set plus historic boundaries) about the spatial distribution of species. Thus, for other applications, our approach may not produce reliable results; but for locations with adequate inventory data and SDMs (dependent on scale and scope of application), the delineation algorithm and custom tools could be applied to generate a GSB.

Alternative methods could have been used to delineate a polygon boundary from the output of a SDM (e.g., minimum bounding geometry, threshold classification, etc.). However, creating a boundary from the minimum bounding geometry of vertices along the outline of the SDM would need to deal with outliers and uncertainty near the edge. Additionally, classifying the SDM based on a threshold has issues, as discussed by Jiménez-Valverde and Lobo (2007), and, thus, applying a threshold ≥ 1 to DISTRIB's output could have introduced gaps within the range boundary. While outliers and some or most of the uncertainty near the edge may be removed, the threshold method has the potential to allow small outlying areas of presence to remain during the delineation process. Similarly, if FIA data were used to place vertices around the majority of records via clustering, the resulting boundary would be more subjective and similarly weak where FIA plots are few or missing.

With respect to Little's (1971) ranges, we believe that for many applications (e.g., examining source/sink population dynamics or modeling movement into new habitats), these GSBs might better define the species' range since they exclude outlier populations that may differ from the core distribution, and are based on large amounts of data collected since the time of Little. The work of Little attempts to create a well-established range extent for species based on many sources of data, however, since many herbarium records contain only coarse-level county locations and the resulting distributions were created prior to major advances in GPS and GIS technologies, one can assume that some level of generalization is present in the ranges. Additionally, Little's range maps are aging such that significant changes in presence/absence and actual distributions are possible in the time that has passed (Woodall *et al*. 2009). Little (1971) also used community type boundaries to define edges for individual species, which has the potential to have overestimated the full extent of the range. Furthermore, when recent FIA data is overlaid with Little's range for most of the 132 species, many locations near the edge of the boundary contain few or no survey plots reporting presence. Thus, Little's range maps represent a more broad distribution as compared to the generalized species boundaries presented here. Range maps created by Little represent the distribution of tree species in North America, and still have highly significant scientific values. However, there are instances where the broadening assumptions (e.g., including an entire county based on a single herbarium record) encompassing Little's boundaries are a limiting factor. Such challenges arise predominantly when the distinction between zones of realized occupancy are needed, e.g., for purposes of assessing a species' colonization potential, where a boundary derived from plot level data and modeling results provide a closer link to the SDM inference.

The classification of each GSB into the four categories, albeit subjective, provided a way to describe general trends among species, even though changing any of the thresholds used could place species in another class. Thus, for sparse species, GSBs were generated over an area with FIA plots generally more sparsely dispersed than for other species. The five broader species have GSBs more broadly defined than Little's boundary, and could represent species that have expanded their range since defined by Little or, more likely, presence is now found in places previously thought to be absent due to additional or more extensive sampling. Species with a narrower GSB represent the bulk of the species and have ranges smaller than those defined by Little; these GSBs captured a high percentage of FIA plots, suggesting that the core area can be defined by a smaller area than the full extent as defined by Little. This does not indicate that the species has lost habitat, but rather that the GSB is defining a primary zone of occupancy with a higher density of FIA sampling sites. The no difference GSBs typically share an area similar to that defined by Little's range and also contained a high portion of FIA presence.

Conclusions

In an age when computer-based technologies dominate many aspects of the scientific process, more and more species ranges and distributions are being generated from computer-based models. Though widely needed to assess trends in global biodiversity, large-scale efforts to survey vast tracks of land to delineate species' ranges are not common, and some key efforts are aging (e.g., Little (1971) for North America and Jalas and Suominen (1972) for European vascular plants). In keeping pace with the current practices of using a SDM to define ranges, we offer one way to create a polygon boundary from a grid-based estimate of suitable habitat. Because the resulting boundary maps were created from a combination of FIA presence/absence data, Little's maps, and model outputs of suitable habitat, they provide a reasonable estimate of the primary zone of occupancy within a species' distribution. We believe that the combination of the best characteristics of each of the three input sources provides an overall better product that represents the recent range. It should be noted, however, that these GSBs are based on data and models that are likely to change over time, with additional FIA data and with newer suitable habitat models run at finer resolutions or with different data sets (The latter is underway by the authors.). We believe the GSBs are appropriate for many purposes, among them to provide a basis from which to model migration or assisted migration under climate change. The methods presented here can also be used with other SDM output to delineate ranges of suitable habitat for both animal and plant species. To do so will require recent inventory data for the species and knowledge of its habitat environment.

The generalized species boundaries we describe should not be viewed as replacements for Little's maps or FIA plot data, but rather as supplemental products aimed at the primary zone of a species' distribution, potentially useful to others for various applications.

Our intention is to provide these boundaries for 132 tree species online along with our climate change atlas (http://www.nrs.fs.fed.us/atlas). We also provide details on the methods used to create the GSBs should others wish to generate specific boundaries for their work. Specifics on the protocol are available from the corresponding author.

Acknowledgments

We thank those who have reviewed this manuscript and offered comments. We are greatly indebted to those who have collected FIA data over the years, without which this work would not be possible. We are also grateful for the gigantean effort of Elbert Little as the Chief Dendrologist for the US Forest Service for at least four decades. One of us (Louis R. Iverson) counts a field trip with Dr. Little as a highlight of his career.

References

Brown, J.H., Stevens, G.C., and Kaufman, D.M., 1996. The geographic range: size, shape, boundaries, and internal structures. *Annual Review of Ecological Systems*, 27, 597–623.

Burrough, P.A., and Frank, A.U., eds., 1996. *Geographic objects with indeterminate boundaries*. London: Taylor & Francis.

ESRI (Environmental Systems Resource Institute) 2009.'*ArcMap 9.3*' ESRI: Redlands, CA

Fortin, M.-J., and Drapeau, P., 1995. Delineation of ecological boundaries: comparison of approaches and significance tests. *OIKOS*, 72 (3), 323–332.

Franklin, J. and Miller, J.A., 2009. *Mapping species distributions: spatial inference and prediction*, Vol. 338. Cambridge: Cambridge University Press.

Gaston, K.J., 1991. How large is a species' geographic range? *OIKOS*, 61 (3), 434–438.

Gaston, K.J., 2009. Geographic range limits: achieving synthesis. *Proceedings of the Royal Society B*, 276, 1395–1406.

Graham, C.H., and Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, 15, 578–587.

Harrison, P.A., *et al.*, 2006. Modelling climate change impacts on species' distributions at the European scale: implications for conservation policy. *Environmental Science & Policy*, 9, 116–128.

Iverson, L.R., and Prasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, 68, 465–485.

Iverson, L.R., *et al.*, 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management*, 254, 390–406.

Jalas, J. and Suominen, J., 1972. Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 1. Pteridophyta (Psilotaceae to Azollaceae). —The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo, Helsinki. 121 pp.

Jiménez-Valverde, A., and Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, 31 (3). 361–369.

Kirkpatrick, M., and Barton, N.H., 1997. Evolution of a species' range. *The American Naturalist*, 150 (1). 1–23.

Leung, Y., 1987. On the imprecision of boundaries. *Geographical Analysis*, 19 (2), 125–151.

Lister, A., *et al.*, 2005. Strategies for preserving owner privacy in the national information management system of the USDA forest service's forest inventory and analysis unit. *In*: R.E. McRoberts, *et al.*, eds. Proceedings of the fourth annual Forest Inventory and Analysis symposium; 2002 November 19–21; New Orleans, LA. Gen. Tech. Rep. NC-252. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 257 p.

Little, E.L., Jr., 1971. Atlas of United States trees, volume 1, Conifers and important hardwoods: U.S. Department of Agriculture Miscellaneous Publication 1146, 9 p., 200 maps.

MacArthur, R.H., 1972. *Geographical ecology: patterns in the distributions of species*. New York: Harper and Row.

Matthews, S.N., *et al.*, 2011. Modifying climate change habitat models using tree species-specific assessments of model uncertainty and life history-factors. *Forest Ecology and Management*, 262, 1460–1472.

McKenney, D.W., *et al.*, 2007. Potential impacts of climate change on the distribution of North American Trees. *BioScience*, 57 (11). 939–948.

Moore, D.M., Lees, B.G., and Davey, S.M., 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Environmental Management*, 15 (1), 59–71.

Munns, E.N., 1938. *The distribution of important forest trees of the United States*. U.S. Department of Agriculture, Misc. Pub. 287. 176 p.

Ordonez, A. and Williams, J.W., 2013. Climatic and biotic velocities for woody taxa distributions over the last 16 000 years in eastern North America. *Ecology Letters*, 16 (6). 773–781.

Prasad, A.M., *et al.*, 2013. Exploring tree species colonization potentials using a spatially explicit simulation model: implications for four oaks under climate change. *Global Change Biology*, 19 (7), 2196–2208.

Prasad, A.M. , Iverson, L.R., and Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and Random Forests for ecological prediction. *Ecosystems*, 9 (2), 181–199.

Purves, D.W., 2009. The demography of range boundaries versus range cores in eastern US tree species. *Proceedings of the Royal Society B*, 276, 1477–1484.

R Development Core Team (2010). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL, http://www.R-project.org

Rehfeldt, G.E., *et al.*, 2006. Empirical analysis of plant-climate relationships for the western United States. International Journal of Plant Sciences, 167 (6), 1123–1150.

Sangermano, F., and Eastman, J.R., 2012. A GIS framework for the refinement of species geographic ranges. *International Journal of Geographical Information Science*, 26 (1), 39–55.

Smith, W.B. 2006. An overview of inventory and monitoring and the Role of FIA in National Assessments. In: C. Aguirre-Bravo et al., eds. 2006. Monitoring Science and Technology Symposium: Unifying Knowledge for Sustainability in the Western Hemisphere. 2004 September 20–24; Denver, CO. Proceedings RMRS-P-42CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. 223–229.

Sykes, M.T., Prentice, I.C., and Cramer, W., 1996. A bioclimatic model for the potential distributions of North European Tree Species under present and future climates. *Journal of Biogeography*, 23 (2), 203–233.

Westfall, J.A., 2009. *FIA national assessment of data quality for forest health indicators. Gen. Tech. Rep. NRS-53*. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Service, 80.

Woodall, C.W., *et al.*, 2009. An indicator of tree migration in forests of the eastern United States. *Forest Ecology and Management*, 257, 1434–1444.

Woudenberg, S.W., *et al.*, 2010. *The Forest Inventory and Analysis Database: Database description and users manual version 4.0 for Phase 2. Gen. Tech. Rep. RMRSGTR-245*. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 336.