Article

Habitat Modeling of Alien Plant Species at Varying Levels of Occupancy

Dawn Lemke [1,2],* and Jennifer A. Brown [1]

[1] Department of Mathematics and Statistics, University of Canterbury, Christchurch 8140, New Zealand; E-Mail: jennifer.brown@canterbury.ac.nz

[2] Department of Biological and Environmental Sciences, Alabama A&M University, Normal, AL 35810, USA

* Author to whom correspondence should be addressed; E-Mail: dawn.lemke@aamu.edu; Tel.: +1-256-372-4562; Fax: +1-256-372-8404.

Abstract: Distribution models of invasive plants are very useful tools for conservation management. There are challenges in modeling expanding populations, especially in a dynamic environment, and when data are limited. In this paper, predictive habitat models were assessed for three invasive plant species, at differing levels of occurrence, using two different habitat modeling techniques: logistic regression and maximum entropy. The influence of disturbance, spatial and temporal heterogeneity, and other landscape characteristics is assessed by creating regional level models based on occurrence records from the USDA Forest Service's Forest Inventory and Analysis database. Logistic regression and maximum entropy models were assessed independently. Ensemble models were developed to combine the predictions of the two analysis approaches to obtain a more robust prediction estimate. All species had strong models with Area Under the receiver operator Curve (AUC) of >0.75. The species with the highest occurrence, *Ligustrum* spp., had the greatest agreement between the models (93%). *Lolium arundinaceum* had the most disagreement between models at 33% and the lowest AUC values. Overall, the strength of integrative modeling in assessing and understanding habitat modeling was demonstrated.

1. Introduction

Invasive species are now a major threat to ecosystems, with the rapid anthropogenic acceleration of species introductions over the last century [1] and the subsequent impact of the species on economies and ecosystems [2]. Invasive species are now recognized as a major component of global environmental change [3–5]. Tools that can accurately assess the impacts of invasive species are becoming essential for identifying areas where management and monitoring efforts should be focused. Species distribution models (SDMs) are one such tool. They are widely used in ecology [6,7] and have broad applications in assessing the relationships between species occurrence, the environment and the impact of ecological change [8]. For invasive species, SDMs are useful for predicting species distributions and ecological niches, and also for assessing potential spread and the suitability of areas that have not yet been invaded. SDMs can be used to assess the impacts of external environmental conditions such as climate change on species distribution [9] and the potential impacts of the species on the landscape [10].

The strength of a SDM is determined, in part, by the correlation of species distribution to input parameters [11] and the number of observation points. Input parameters are often derived from landscape-level digital information and provide a representation of the environmental heterogeneity of the landscape. Typical parameters used in SDM are those that represent climate, habitat diversity, landscape characteristics, habitat patch size and shape, connectivity, regional and local diversity of biota, vegetation structure, and the intensity, frequency and magnitude of disturbance [12–14], all of which vary across spatial and temporal scales [12,15]. Collectively, these factors result in interlaced patterns of species distribution at multiple spatial and temporal scales [16]. Geospatial datasets including remotely sensed data offer significant opportunities for providing information on these characteristics on a larger scale.

There are numerous methods for developing SDMs, many of which have been applied to invasive plants including logistic regression [17,18], fuzzy envelope models [19], genetic algorithms [20], maximum entropy [18,21], and general additive models [22]. These models differ in the underlying assumptions and algorithms, and in their requirement for presence-only species data or for both presence and true absence data. These approaches can be used individually or collectively in an ensemble approach. Ensemble SDMs combine the strengths of several models while limiting the weakness of any one model [23,24] and offer a broad perspective to model results.

In this paper, we illustrate the application of two modeling techniques, logistic regression and maximum entropy, and the ensemble model approach. We discuss the impact of the size of the dataset on the resulting model by comparing the results from three species with different levels of prevalence. We focus on three of the invasive plant species of concern in the Cumberland Plateau and Mountain Region in the United States: privet (*Ligustrum* spp.), tall fescue (*Lolium arundinaceum*) and silktree (*Albizia julibrissin*).

2. Methods

2.1. Study Area

The Cumberland Plateau and Mountain Region (CPMR) extends from northern Alabama, through Tennessee and Kentucky, and into Virginia [25–28] (Figure 1). The region covers 59,000 km^2 and has one of the most diverse woody plant communities in eastern North America [29]. Forest resources and management are a major part of the CPMR economy, particularly in rural communities. Approximately 70% of the land in this area is forested, with over 75% of this comprised of hardwoods [29,30]. Elevations range from 200 to 1200 m [31], with annual rainfall varying from 940 to 1900 mm, and mean minimum winter temperatures of −7 °C to 1.5 °C [32]. Like many of the forests in eastern North America, the native deciduous hardwood forests of the CPMR are characterized by a long history of land-use change driven by agricultural conversion and timber extraction. More recently, urban sprawl and large-scale conversion of land to intensively managed pine plantations have become major contributors to land cover change [33]. McGrath and others [34] found that 14% of native forest cover was lost between 1981 and 2000, predominantly as a result of native forest conversion to pine plantations. Of the 33 invasive species monitored by the United States Forest Service (USFS) [35], 25 of them are found in the CPMR: four trees, seven shrubs, seven vines, five grasses and two forbs.

Figure 1. Study area location map: Cumberland Plateau and Mountain region in the southeastern United States.
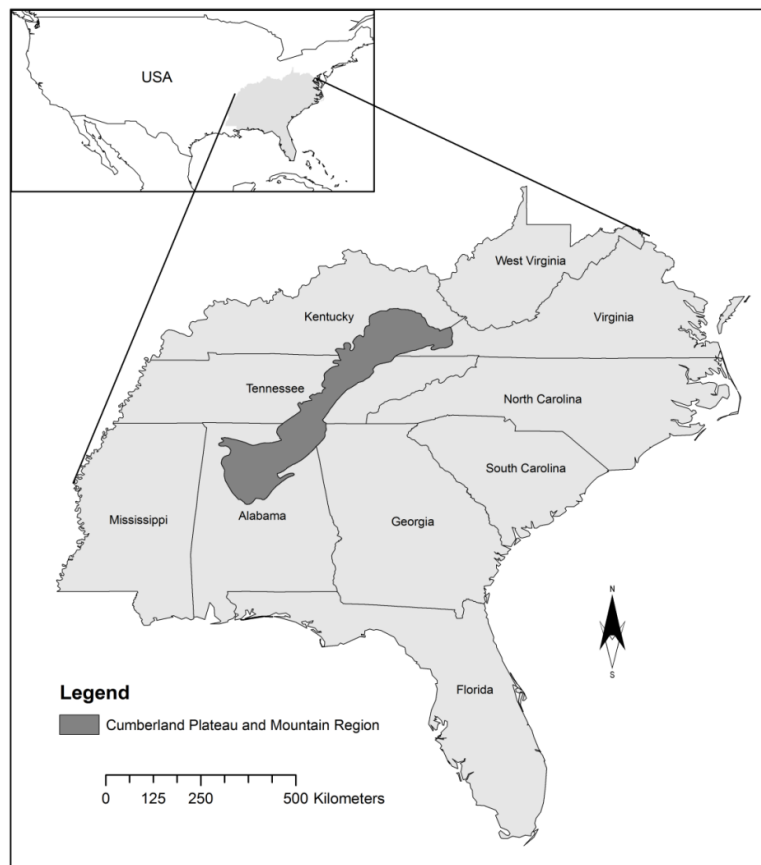
2.2. Species of Interest

Study species were selected to represent a range of life forms (grass, shrubs, and trees) and occurrence levels (moderate and low percentage of Forest Inventory and Analysis database (FIA) plots occupied) across the CPMR. Privet (the shrub) had moderate occurrence (16% occupied plots), and tall fescue (the grass) and silktree had low occurrence (5% and 2% respectively).

2.2.1. Privet

There are at least eight species of invasive privets (*Ligustrum* spp.) that have been introduced from Asia and Europe into the southern United States as ornamentals [36–38]. The USFS collects information on two species of privet, Chinese privet (*L. sinense*) and European privet (*L. vulgare*) [35]. It can be difficult to distinguish between privet species and instead we have modeled the *Ligustrum* genus as a whole. Privets are the second most abundant invasive plants in the southern region and the most prevalent in the understory of bottomland hardwood forests [39,40]. Chinese privet is the most common species, being present in 20 states ranging from Texas to Massachusetts [36]. All species are still being produced, sold and planted as ornamentals. Privets severely alter natural habitat and critical wetland processes, forming dense stands to the exclusion of most native plants and replacement regeneration. The abundance of specialist birds and the diversity of native plants and bees are dramatically reduced by privet thickets [41,42]. The dense thickets impact forest communities by shading and out-competing many of the native species. Privet can survive in a variety of habitats, including wet or dry areas, but dominates best in mesic forests [39]. Privets produce abundant seeds that are viable for about a year [43], which are predominately spread by birds [44]. Privet also has the ability to increase in density by stem and root sprouts. The fruit produced, however, provides a substantial food source for birds and other wildlife [45].

2.2.2. Tall Fescue

Tall fescue is a grass native to Europe and was first introduced into the United States in the early to mid-1800s. It has been widely planted for turf, forage and erosion control [46]. Tall fescue occurs throughout the continental United States [36] and has been reported as invasive in natural areas [47]. It is still promoted by a variety of agricultural agencies; however, the USFS Southern Region has prohibited the use of endophytically enhanced tall fescue on USFS lands [39]. Tall fescue is a cool season grass that invades native grasslands, savannahs, woodlands and other high-light natural habitats [46]. It spreads mainly through rhizomes and can form extensive colonies that compete with and displace native vegetation. Viable seeds can be dispersed by grazing animals and birds, and remain in the seed bank for extended periods of time [39]. Some varieties of tall fescue have a mutualistic fungal endophyte (*Neotyphodium coenophialum*) that gives them a competitive advantage over some plants, including legumes [48]. As a result, communities dominated by tall fescue are often low in plant species richness [49]. In addition, alkaloids produced by endophyte-infected tall fescue may be toxic to small mammals and of low palatability to ungulates [50]. Tall fescue, which has replaced many acres of native grass, does not supply the type of food and cover that many birds need in order to thrive [51]. The grass supports only a limited number of insects [52], which in turn, are an important

food for both quail and turkey. Grasslands dominated by endophyte-infected tall fescue are expected to support less total herbivore biomass and less predator biomass [51,52]. Tall fescue tolerates nutrient-poor and compacted soils, and grows well in disturbed areas such as highway and railroad right-of-ways. Annual nitrogen inputs are needed to maintain optimal grazing conditions [46]. Tall fescue is adapted to cool, humid climates with moist soils of a pH 5.5 to 7.0 [46]. It will produce top growth when soils are as low as 5 °C and it continues growing into late autumn in the southern United States [46].

2.2.3. Silktree

Silktree is a legume native to south and eastern Asia. It is a small to medium-sized tree that can grow up to 11 m tall. It was introduced to the United States in 1745 and widely planted as an ornamental. Silktree is now found throughout the southern United States along roadsides, beside parking lots bordering power lines and encroaching into forests. Silktree reproduces both vegetative and by seed [39]. The seeds are encased with impermeable seed coats that allow them to remain dormant for many years [53]. Because silktree is sun tolerant, it can grow in a variety of soils and can produce large seed crops and re-sprout when damaged. It is a strong competitor of native trees and shrubs in open areas and forest edges. Dense stands of silktree severely reduce the sunlight and nutrients available for other plants [39]. Silktree can tolerate partial shade but is rarely found in forests with full canopy cover or at higher elevations (above 900 m) where cold hardiness is a limiting factor. However, silktree can become a serious problem along riparian areas where it becomes established along scoured shores and where its seeds are easily transported in water [39]. Although it has been identified as being invasive in forests in the southern United States [39], silktree is still being encouraged as a tree crop species [54]. Ares and others [54] state that in the southern United States, silktree has been considered in agroforestry practices as a forage species for goats and cattle [55,56], and for soil fertility improvement in permaculture systems [57–59]. However, planting of silktree should be evaluated on a site-specific basis because it can become invasive, especially in riparian areas [60]. This mixed message may increase the planting of silktree in the next decade and thus its invasion potential.

2.3. Invasive Plant Occurrence

The USFS, Forest Inventory and Analysis (FIA) program, analyses and reports information on the status, trends and conditions of forests within the United States. It is a periodic survey of all forested land in the United States and has occurred since 1928 [61]. Recent inventories have typically been conducted every 5–7 years in the southeastern states, with approximately 20% of the points assessed every year [35]. In the CPMR there are 2814 FIA sites [35]. An extension of the FIA database focuses on invasive plants, and this database was made available for our study. Data were available for the last completed inventory cycle (2000–2005) and consisted of species absence/presence records.

2.4. Landscape Variables

Landscape variables were categorized into six groups: Landsat, anthropogenic, environmental, climate, land use and water. Using ArcGIS [62] and ERDAS [63], all variables were extracted from

available digital information including Landsat imagery, classified land use data, roads, rivers, human population census data and climatic information. All variables were converted to 30 m × 30 m cells across the CPMR [18]. The total number of variables was 41 (Table 1). This initial set was reduced using exploratory data analysis to remove variables that were highly correlated (Pearson's correlation coefficient, *r*). For any two variables that were highly correlated (*r* > 0.8) only one was selected for input into further models. All input variables needed to be able to be displayed on a map. Two variables based on the Normalized Difference Vegetation Index (NVDI), NDVI75 and NDVI90-75, could not be mapped due to inconstancies across Landsat scenes, an artifact of instrumentation, and thus were not suitable for use in further analysis. This left a set of 28 variables (see Table 1).

Table 1. Description of landscape variables categorized into six groups, the resolution of the original data (Res), the citation for other studies that have used the variable, and the original data source. Descriptive statistics are shown for the 28 variables that were used in modeling. (TIGER = Topologically Integrated Geographic Encoding and Referencing, USGS = United States Geological Services, LULC = Land Use Land Cover, NED = National Elevation Dataset, PRISM = Parameter-elevation Regressions on Independent Slopes Model).

	Variable	Variable code	Citation	Res	Source	Mean	SD	Min	Max
Landsat	Disturbance Index for 1975	DI75	[64]	900 m^2	Landsat	9.5	1.3	−11.8	65.5
	Disturbance Index for 1990	DI90	[64]	900 m^2	Landsat	−0.3	1.8	−10.5	38.4
	Disturbance Index for 2000	DI00	[64]	900 m^2	Landsat	0.0	2.0	−9.8	48.1
	Change in Disturbance Index between 1975 and 1990	DI90-75	[64]	900 m^2	Landsat				
	Change in Disturbance Index between 1990 and 2000	DI00-90	[64]	900 m^2	Landsat	0.4	2.2	−40.7	59.9
	NDVI in 1975	NDVI75	[65]	900 m^2	Landsat				
	NDVI 1990	NDVI90	[65]	900 m^2	Landsat	0.57	0.10	−0.94	0.98
	NDVI 2000	NDVI00	[65]	900 m^2	Landsat	0.45	0.15	−0.96	0.99
	Difference in NDVI between 1975 and 1990	NDVI90-75	[65]	900 m^2	Landsat				
	Difference in NDVI between 1990 and 2000	NDVI00-90	[65]	900 m^2	Landsat				

Table 1. *Cont.*

	Variable	Variable code	Citation	Res	Source	Mean	SD	Min	Max
Anthropogenic	Number of people per km² in 2000	CENSUS	[66]	Census block	Census 2000 *TIGER*	24	46	3	2805
	Distance to road	RD_DIST	[66]	900 m²	Census 2000 *TIGER*	397	375	0	3755
	Density of roads within a km² area in 2000	RD_DEN	[66]	900 m²	Census 2000 *TIGER*	1.3	1.0	0	15.6
	Distance to major road	MRD_DIST	[66]	900 m²	Census 2000 *TIGER*	5614	4717	0	26122
	Residential in 2000 or 1990 within a 500 m buffer	RES ALL	[67]	900 m²	USGS LULC				
	Residential presence within a 100 m buffer in 2000	RES100	[67]	900 m²	USGS LULC	0.31	0.49	0	1
	Residential presence within a 500 m buffer in 2000	RES500	[67]	900m²	USGS LULC	0.71	0.49	0	1
Environmental	North	NORTH	[68]	900 m²	USGS NED				
	East	EAST	[68]	900 m²	USGS NED				
	Northness	NORTHNESS	[69]	900 m²	USGS NED	0	0.18	−0.88	0.83
	Eastness	EASTNESS	[69]	900 m²	USGS NED	0	0.19	−0.83	0.84
	Slope	SLOPE	[62]	900 m²	USGS NED	12.9	8.9	0	62.3
	Hillshade	HILL	[62]	900 m²	USGS NED	237	17	59	254
	Curvature	CURV	[62]	900 m²	USGS NED				
	Elevation	DEM	[31]	900 m²	USGS NED	383	168	0	1283
Climate	Average temperature from a 30-year average (1971–2000)	AVET	[32]	900 m²	PRISM				
	Minimum temperature from a 30-year average (1971–2000)	MINT	[32]	900 m²	PRISM	26.5	3.4	19	35
	Maximum temperature from a 30-year average (1971–2000)	MAXT	[32]	900 m²	PRISM				
	Average yearly rainfall from a 30-year average (1971–2000)	RAIN	[32]	900 m²	PRISM	54	5	41	75

Table 1. *Cont.*

	Variable	Variable code	Citation	Res	Source	Mean	SD	Min	Max
Land Cover	Change in forest between 2000 and 1990 within a 100-m buffer	FC100	[67]	900 m^2	USGS LULC				
	Change in forest between 2000 and 1990 within a 500-m buffer	FC500	[67]	900 m^2	USGS LULC	0.12	0.13	−1	0.99
	Proportion of forest in 2000 with in a 100-m buffer	F00 100	[67]	900 m^2	USGS LULC	0.90	0.17	0.03	1
	Proportion of forest in 2000 with in a 500-m buffer	F00 500	[67]	900 m^2	USGS LULC				
	Proportion of farming in 2000 with in a 100-m buffer	FARM100	[67]	900 m^2	USGS LULC				
	Proportion of farming in 2000 with in a 500-m buffer	FARM500	[67]	900 m^2	USGS LULC	0.07	0.13	0	0.98
	Categorical land use in 1990 based on Andersons groupings	LULC90	[67]	900 m^2	USGS LULC		Categorical		
	Categorical land use in 2000 based on Andersons groupings	LULC00	[67]	900 m^2	USGS LULC		Categorical		
Water	Distance from a stream	RIV DIS	[70]	900 m^2	USGS	336	267	0	3288
	Density of streams within a km^2 area	RIV_DEN	[70]	900 m^2	USGS	0.96	0.51	0	6.65
	Occurrence of a wetland or stream within 100 m	WATER100	[67]	900 m^2	USGS LULC	0.05	0.51	0	1
	Occurrence of a wetland or stream within 500 m	WATER500	[67]	900 m^2	USGS LULC	0.30	0.50	0	1

Descriptive statistics (mean, standard deviation (SD), minimum (Min) and maximum (Max)) for the 28 variables were calculated for both the land area covered by the FIA plots and the forested land area in the CPMR. The forested land area in the CPMR was the area depicted by the 2001 National Land Cover Database. This comparison was to determine if FIA data could be extrapolated to the entire forested CPMR (Table 1). The FIA points had a mean that was within one SD of the mean for the forested area of the CPMR for all variables (all but two variables had means within 0.2 SDs). In both cases, the maximum and minimum were very similar, suggesting that although there was some variation in the means, they still represented the full range of the CPMR. Overall, the FIA data are considered to be an adequate representation of the CPMR for this study.

2.5. Models

Two modeling techniques were used: binary logistic regression (using a binomial distribution and logit link) [71] and maximum entropy (MaxEnt) [72]. The important difference between the two techniques is that logistic regression uses information on both occurrence and absence to estimate a predictive linear model, whereas MaxEnt uses information from occurrences only [18]. The distribution of each species was modeled, following the methods of Lemke and others [18], using each group of variables (Landsat, anthropogenic, environmental, land use, water and climate) separately (Table 1). These "sub-models" were built using each of the two techniques. Using only variables selected in the final sub-model for each variable group, a final composite model was determined. Logistic regression models were conducted using SAS [73] and MaxEnt models were conducted using a specialized package of MaxEnt [72]. Logistic regression models were derived using a stepwise regression method with Akaike's Information Criterion (AIC) [74] as the selection criterion. MaxEnt models were derived using a manual backward selection method, and variables that had little or no impact on the model were removed. A measure of variable contribution was calculated to identify the key variables determining the occurrence of each species.

The omission rate and Area Under the receiver operator Curve (AUC) were used to assess the reliability and validity of the models. The omission rate is the false negative or the proportion of sites where the species was present but the model predicted absence. To calculate the omission rate, the predicted model values are converted to a binary value (predicted occurrence = 1; predicted absence = 0). The threshold value for this binary conversion was set, for each species, as the value that maximized the sum of the sensitivity and specificity [75]. The AUC provides a single measure of model performance independent of any particular choice of threshold [76].

Rasters were imported from MaxEnt into ArcGIS and the raster calculator was used in creating the logistic regression model. Initial maps with continuous rasters were reclassified into binary rasters based on the cut-off values determined by maximizing the sum of the sensitivity and specificity.

We integrated information from both logistic and MaxEnt using an ensemble approach. While logistic and MaxEnt models may be compared individually to select the best overall model for particular datasets, methods that combine the two models have the potential to reduce the uncertainty associated with any one particular algorithm [23,24]. A number of approaches have been proposed for combining the outputs of individual models for ensemble predictions [23]. Here, we adopt a consensus approach, adding the binary output rasters together to identify areas of agreement and disagreement in the models. Areas of agreement were where both models predicted occurrence or absence, and areas of disagreement were where the predictions of the composite models (the logistic regression or MaxEnt models) differed.

2.6. Data Selection

Models were built for each species, using 70% of the data with the remaining 30% used to test the models (Table 2). For the logistic regression models, the balance between occurrence and absence data points was fixed as 20:80 [77] for the three species, to reduce any effect of having a large binary class imbalance. This was done by under-sampling the absence data points [77].

Table 2. Total number of points, for the occurrence and absence of three species, separated into training and test datasets.

	Training		Test	
	Occurrence	Absence	Occurrence	Absence
Privet	200 (10.4%)	1125 (59.0%)	100 (5.2%)	482 (25.4%)
Tall fescue	65 (3.4%)	1270 (66.6%)	28 (1.5%)	544 (28.5%)
Silktree	31 (1.6%)	1304 (68.4%)	13 (0.7%)	559 (29.3%)

3. Results and Discussion

Of the 42 models run, 41 had better than random predictions (Table 3). All three species had low omission rates and high AUCs. The final composite models were combined to create ensemble models (Figure 2). The species with the strongest agreement was the more prevalent species, privet (93% agreement), while the two low-prevalence species, with the smaller number of occurrence data points, had lower agreement between their composite models (67% agreement for tall fescue and 87% for silktree) [78]. However, despite low prevalence and small datasets, composite models for all three species were acceptable.

Table 3. Threshold (defined as maximum sensitivity plus specificity) and accuracy assessment for the three species (bold denotes strong models with AUC >0.80 and omission rate <0.20) using logistic regression (L) and MaxEnt (M). The variables were grouped into four groups: Landsat, Anthropogenic (Anthro), Environmental (Enviro) and Climate. The composite model is the final, best model.

Species	Model	Group	Threshold	Omission rate		AUC	
				Train	Test	Train	Test
	L	Landsat	0.18	0.07	0.10	0.70	0.66
	M	Landsat	0.47	0.25	0.37	0.74	0.68
	L	Anthro	0.16	0.06	0.08	0.76	0.72
	M	Anthro	0.33	0.09	0.20	0.77	0.70
	L	**Enviro**	**0.16**	**0.04**	**0.05**	**0.84**	**0.82**
	M	**Enviro**	**0.34**	**0.10**	**0.10**	**0.80**	**0.80**
Privet	L	Climate	0.16	0.20	0.25	0.83	0.82
	M	Climate	0.42	0.18	0.30	0.83	0.82
	L	**Land use**	**0.13**	**0.05**	**0.11**	**0.83**	**0.82**
	M	Land use	0.30	0.10	0.13	0.81	0.79
	L	Water	0.17	0.10	0.21	0.66	0.65
	M	Water	0.52	0.51	0.51	0.66	0.67
	L	**Composite**	**0.17**	**0.02**	**0.05**	**0.91**	**0.89**
	M	**Composite**	**0.28**	**0.07**	**0.14**	**0.86**	**0.83**

Table 3. *Cont.*

Species	Model	Group	Threshold	Omission rate		AUC	
				Train	Test	Train	Test
	L	Landsat	0.06	0.32	0.54	0.74	0.65
	M	Landsat	0.45	0.34	0.64	0.76	0.61
	L	Anthro	0.05	0.41	0.40	0.65	0.63
	M	Anthro	0.40	0.26	0.32	0.70	0.67
	L	Enviro	0.05	0.49	0.38	0.60	0.62
	M	Enviro	0.49	0.34	0.50	0.75	0.66
Tall Fescue	L	Climate	0.06	0.28	0.30	0.73	0.70
	M	**Climate**	**0.44**	**0.15**	**0.10**	**0.77**	**0.84**
	L	Land use	0.06	0.36	0.42	0.66	0.59
	M	Land use	0.46	0.24	0.39	0.72	0.60
	L	Water		No Model			
	M	Water	0.47	0.20	0.32	0.61	0.54
	L	Composite	0.05	0.25	0.21	0.78	0.75
	M	Composite	0.42	0.25	0.35	0.82	0.75
	L	Landsat	0.06	0.32	0.54	0.74	0.65
	M	Landsat	0.47	0.29	0.36	0.73	0.73
	L	Anthro	0.02	0.22	0.41	0.75	0.73
	M	Anthro	0.47	0.29	0.21	0.84	0.90
	L	**Enviro**	**0.02**	**0.10**	**0.15**	**0.83**	**0.80**
	M	Enviro	0.30	0.06	0.29	0.82	0.78
Tall Fescue	L	Climate	0.02	0.19	0.17	0.77	0.75
	M	**Climate**	**0.42**	**0.13**	**0.07**	**0.77**	**0.82**
	L	Land use	0.02	0.35	0.37	0.81	0.85
	M	Land use	0.42	0.25	0.43	0.80	0.70
	L	Water	0.02	0.29	0.35	0.74	0.75
	M	Water	0.49	0.32	0.28	0.76	0.76
	L	**Composite**	**0.02**	**0.16**	**0.38**	**0.89**	**0.80**
	M	**Composite**	**0.27**	**0.06**	**0.07**	**0.91**	**0.90**

Of the 28 original variables used in developing the models, 15 were ultimately incorporated into at least one of the final composite models, but only seven were used in more than one model (Table 4). Overall, the composite models were dominated by environmental variables (32% of all composite model contributions) and climatic variables (42% of all composite model contributions) with minimum temperature as the single most important variable (40% of all composite model contributions; Table 4). This confirms the validity of matching the ranges of native species with the range of potential invasion, and the approach of integrating elevation, latitude and longitude, as is used to estimate potential invasive distribution [79]. It also suggests that climate change will influence the distribution, and this variation should be integrated into models. Variables in the Landsat and water groups contributed very little to the models, contributing only one variable each to the composite models, and both were at low rates (disturbance index in 2001 at 1%, and water within 500 m at 3%, for all composite model contributions; Table 4). Information on human population, roads and land use (proportion of forest and

proportion of farming) were the most useful anthropogenic variables (Table 4). All of this information is readily available for North America and much of the world, making this level of landscape level modeling very practical.

Figure 2. Spatial representation of the ensemble models combining the logistic regression and MaxEnt composite models (**A**: privet, **B**: silktree, **C**: tall fescue). Areas of high risk and areas of no invasion are where both composite models agree, and areas of moderate invasion are where one composite model predicted invasion and the other did not.
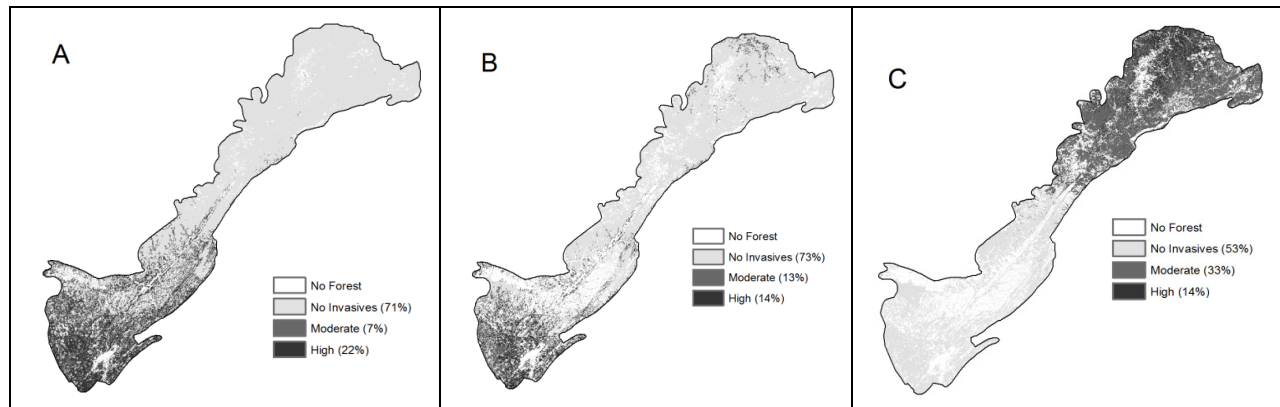
Table 4. Contribution of variables to the final composite models (−, negative; +, positive; ∩ or U for bimodal relationship), dominant variables given in bold (>25%). L = logistic regression; M = MaxEnt.

	Species	Privet		Tall fescue		Silktree	
	Model	L	M	L	M	L	M
Landsat	DI00					(+)6	
Anthropogenic	CENSUS						(+)24
	RD DEN	(+)4	(+)15			(+)**35**	(+)16
	RD DIST	(−)3					
	MRD DIST	(−)3					
	RES100			(−)8			
Environmental	DEM	(−)13	(−)7		(∩)19	(−)**58**	(−)**48**
	NORTHNESS			(−)**30**	(−)7		
	SLOPE		(−)6				
Climatic	MINT	(+)**66**	(+)**55**	(−)**62**	(−)**54**		
	RANN		(U)5		(∩)10		
Land use	F00 100		(−)10				
	FARM500	(+)4			(∩)10		
	LULC90	7					
Water	WATER500					(−)1	(+)12
Proportion forest area invaded		24%	28%	46%	16%	20%	21%

Privet composite models used a range of environmental and anthropogenic variables, with the logistic model having seven variables and MaxEnt having six variables. The logistic model predicted 24% of the forest as having the potential to be invaded, and MaxEnt predicted 28%. Currently, 15% of

FIA plots have privet. Overall, privet was predicted to occur across 22% of the forests by both models. Both composite models were strong, with logistic regression producing a slightly better model. Environmental variables dominated both models, at 73% (MaxEnt) and 79% (logistic regression). Minimum temperature was the single most dominant variable, with higher minimum temperatures having a higher probability of invasion. Both models showed a negative correlation with elevation and a positive correlation with road density, suggesting that privet will be found at lower elevation in areas of higher road density (increased human occupation). The logistic model also suggested privet had a higher chance of occurrence closer to roads and with more farming in the near vicinity. MaxEnt highlighted the trend that the less the forest cover, the more likely the area was to have privet. The logistic model used historical land use as one of the independent variables, associating privet with areas with less forest, more residential land use and more water in 1990. Overall, this suggests that areas of higher human use and disturbance will have more privet. The MaxEnt model also identified slope and rainfall as important, with low slope being more likely to have privet.

For tall fescue, the MaxEnt composite model had the highest AUC (Table 3); however, the MaxEnt model that used only climatic variables had a slightly better omission rate. The logistic regression models had slightly lower validation statistics. Both the MaxEnt and logistic regression composite models were dominated by climatic variables. The MaxEnt composite model showed that tall fescue occurrence was influenced greatly by temperature, elevation, rainfall, farming and aspect. Lower temperature; intermediate levels of farming, rainfall and elevation; and a more southerly aspect were related to a higher occurrence of tall fescue. The logistic regression composite model only used three variables, minimum temperature, aspect and amount of residential land use within 100 m, with low temperature, more southerly slopes and less residential land use having a higher occurrence of tall fescue.

The silktree was the only species to integrate a high portion of anthropogenic variables into the composite models (Table 4). The MaxEnt composite model predicted 21% of the area to have probable occurrence of silktree, and showed its occurrence to be influenced by elevation, population density, road density and water bodies. The variables lower elevation, higher population and road density, and nearby water bodies were related to a higher occurrence of silktree. The composite logistic model also utilized a number of anthropogenic variables. The logistic model was dominated by elevation but road density also had a major role in the model. The logistic composite model was the only composite model to use a Landsat variable. The logistic model also suggested that low elevation and high road density are important contributors to silktree occurrence, with higher disturbance in the landscape also being important.

4. Conclusions

Remote sensing has been identified as an emerging tool for biodiversity science and conservation [80]. However, in this work, the introduction of remotely sensed medium resolution (30 m) data had little value in the overall model development. Only one of the composite models, the logistic regression model for silktree, used any Landsat variables. The silktree model used the Landsat disturbance index for 2001 but this only had a 5% contribution to the model. Given the time put into developing the Landsat variables, we would suggest that for future work, this information adds little value to the predictive ability of models and is probably unnecessary at a landscape scale. The large size of the

study area (59,000 km^2) made it impractical to use remotely sensed data at a finer resolution due to the computer processing power required for analysis. Exploring different abstraction resolutions, as suggested by Sester [81], would be a worthwhile study, possibly on a smaller scale, to identify an optimal resolution.

The use of the two different modeling approaches, logistic regression and MaxEnt, strengthens the validity of the results. The inclusion in the models of similar variables with the same direction of relationships gives confidence to any inference about the importance of these variables. In examining all the composite models, there was only one variable that had a different relationship between the two types of modeling: water in the tall fescue composite models. In this model, water had a positive relationship with MaxEnt (12% contribution) but a small weak relationship in the logistic regression (1% contribution to the model).

The ensemble approach and mapping the agreement and disagreement of composite models within each species showed privet to have a very strong agreement (93%), silktree a moderate agreement (87%) and tall fescue a limited agreement (67%). This is a reflection of the model strength, the number of occurrence points and the applicability of the independent variables in predicting the species of interest. Tall fescue had the lowest agreement of the three species, even though it was not the species with the smallest number of occurrence points. There may be a number of reasons for this; for example, only forested landscapes were modeled rather than grasslands. Other reasons could be the suitability of the independent variables or the scale of the independent variables. Independent variables were used at a 30 m × 30 m resolution and habitat characteristics that function at a smaller scale may be driving the distribution of tall fescue.

Models such as those developed by this research can be used as tools for landscape management, forest stand assessment or long-term forest monitoring programs. We recommend the use of an ensemble modeling approach to combine different models. One of the greatest benefits of large-scale GIS models is that they can outline the main characteristics of species distribution areas and be used to predict environmental favorability in regions where their distribution is less documented [82]. They can also be integrated into forest management decision support systems [83] and assist in developing long-term management plans.

Acknowledgments

Conflict of Interest

The authors declare no conflict of interest.

References

1. Hulme, P.E.; Pyšek, P.; Nentwig, W.; Vila, M. Will threat of biological invasions unite the European Union? *Science* **2009**, *324*, 40–41.
2. Vilà, M.; Basnou, C.; Pyšek, P.; Josefsson, M.; Genovesi, P.; Gollasch, S.; Nentwig, W.; Olenin, S.; Roques, A.; Roy, D.; *et al.* How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Front. Ecol. Environ.* **2010**, *8*, 135–144.
3. Mainka, S.A.; Howard, G.W. Climate change and invasive species: Double jeopardy. *Integr. Zool.* **2010**, *5*, 102–111.
4. Ricciardi, A. Are modern biological invasions an unprecedented form of global change? *Conserv. Biol.* **2007**, *21*, 329–336.
5. Vitousek, P.M.; D'Antonio, C.M.; Loope, L.L.; Rejmanek, M.; Westbrooks, R. Introduced species: A significant component of human-caused global change. *N. Z. J. Ecol.* **1997**, *21*, 1–16.
6. Guisan, A.; Thuiller, W. Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* **2005**, *8*, 993–1009.
7. Smolik, M.G.; Dullinger, S.; Essl, F.; Kleinbauer, I.; Leitner, M.; Peterseil, J.; Stadler, L.M.; Vogl, G. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *J. Biogeogr.* **2010**, *37*, 411–422.
8. Guisan, A.; Lehmann, A.; Ferrier, S.; Austin, M.; Overton, J.M.C.; Aspinall, R.; Hastie, T. Making better biogeographical predictions of species' distributions. *J. Appl. Ecol.* **2006**, *43*, 386–392.
9. Kearney, M.R.; Wintle, B.A.; Porter, W.P. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conserv. Lett.* **2010**, *3*, 203–213.
10. Ficetola, G.F.; Thuiller, W.; Miaud, C. Prediction and validation of the potential global distribution of a problematic non-native invasive species; the American bullfrog. *Divers. Distrib.* **2002**, *8*, 49–56.
11. Hoffman, J.D.; Aguilar-Amuchastegui, N.; Tyre, A.J. Use of simulated data from a process-based habitat model to evaluate methods for predicting species occurrence. *Ecography* **2010**, *33*, 656–666.
12. Kumar, S.; Stohlgren, T.J.; Chong, G.W. Spatial heterogeneity influences native and alien plant species richness. *Ecology* **2006**, *87*, 3186–3199.
13. Stohlgren, T.J.; Binkley, D.; Chong, G.W.; Kalkhan, M.A.; Schell, L.D.; Bull, K.A.; Otsuki, Y.; Newman, G.; Bashkin, M.; Son, Y. Exotic plant species invade hot spots of native plant diversity. *Ecol. Monogr.* **1999**, *69*, 25–46.
14. With, K.A.; Crist, T.O. Critical thresholds in species' responses to landscape structure. *Ecology* **1995**, *76*, 2446–2459.
15. Pickett, S.T.A.; Cadenasso, M.L. Landscape ecology: Spatial heterogeneity in ecological systems. *Science* **1995**, *269*, 331–334.
16. Wagner, H.H.; Fortin, M.J. Spatial analysis of landscapes: Concepts and statistics. *Ecology* **2005**, *86*, 1975–1987.
17. Collingham, Y.C.; Wadsworth, R.A.; Huntley, B.; Hulme, P.E. Predicting the spatial distribution of non-indigenous riparian weeds: Issues of spatial scale and extent. *J. Appl. Ecol.* **2000**, *37*, 13–27.

18. Lemke, D.; Hulme, P.E.; Brown, J.A.; Tadesse, W. Distribution modelling of Japanese honeysuckle (*Lonicera japonica*) invasion in the Cumberland Plateau and Mountain Region, USA. *For. Ecol. Manag.* **2011**, *262*, 139–149.

19. Robertson, M.P.; Villet, M.H.; Palmer, A.R. A fuzzy classification technique for predicting species' distributions: Application using invasive alien plants and indigenous insects. *Divers. Distrib.* **2004**, *10*, 461–474.

20. Underwood, E.C.; Klinger, R.; Moore, P.E. Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. *Divers. Distrib.* **2004**, *10*, 447–459.

21. Hoffman, J.D.; Narumalani, S.; Mishra, D.R.; Merani, P.; Wilson, R.G. Predicting potential occurrence and spread of invasive plant species along the North Platte River, Nebraska. *Invasive Plant Sci. Manag.* **2008**, *1*, 359–367.

22. Dullinger, S.; Kleinbauer, I.; Peterseil, J.; Smolik, M.; Essl, F. Niche based distribution modelling of an invasive alien plant: Effects of population status, propagule pressure and invasion history. *Biol. Invasions* **2009**, *11*, 2401–2414.

23. Araujo, M.B.; New, M. Ensemble forecasting of species distributions. *Trends in Ecol. Evol.* **2007**, *22*, 42–47.

24. Stohlgren, T.J.; Ma, P.; Kumar, S.; Rocca, M.; Morisette, J.T.; Jarnevich, C.S.; Benson, N. Ensemble habitat mapping of invasive plant species. *Risk Anal.* **2010**, *30*, 224–235.

25. Smalley, G.W. *Classification and Evaluation of Forest Sites on the Southern Cumberland Plateau*; General Technical Report Southern-23; U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station: New Orleans, LA, USA, 1979.

26. Smalley, G.W. *Classification and Evaluation of Forest Sites on the Mid-Cumberland Plateau*; General Technical Report Southern-38; U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station: New Orleans, LA, USA, 1982.

27. Smalley, G.W. *Classification and Evaluation of Forest Sites in the Cumberland Mountains*; General Technical Report Southern-50; U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station: New Orleans, LA, USA, 1984.

28. Smalley, G.W. *Classification and Evaluation of Forest Sites on the Northern Cumberland Plateau*; General Technical Report Southern-60; U.S. Department of Agriculture, Forest Service, Southern Forest Experiment Station: New Orleans, LA, USA, 1986.

29. Ricketts, T.H.; Dinerstein, E.; Olson, D.M.; Loucks, C.J.; Eichbaum, W. *Terrestrial Ecoregions of North America: A Conservation Assessment*; Island Press: Washington, DC, USA, 1999.

30. Homer, C.; Huang, C.; Yang, L.; Wylie, B. Development of a 2001 national land cover database for the United States. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 829–840.

31. Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. The national elevation dataset. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–11.

32. PRISM Group. 30-year average (1971–2000) PRISM data. Oregon State University: Corvallis, OR, USA; Available online: http://www.prismclimate.org (accessed on 9 December 2007).

33. Wear, D.N.; Greis, J.G. Southern forest resource assessment: Summary of findings. *J. For.* **2002**, *100*, 6–14.

34. McGrath, D.A.; Evans, J.P.; Smith, C.K.; Haskell, D.G.; Pelkey, N.W.; Gottfried, R.R.; Brockett, C.D.; Lane, M.D.; Williams, E.D. Mapping land-use change and monitoring the impacts of hardwood-to-pine conversion on the Southern Cumberland Plateau in Tennessee. *Earth Interact.* **2004**, *8*, 1–24.

35. U.S. Department of Agriculture, Forest Service. *The Forest Inventory and Analysis Database: Database Description and Users' Guide*, Version 3.0; USDA FS: Washington, DC, USA, 2007.

36. U.S. Department of Agriculture (USDA) Plants. National Plants Database. USDA: National Plant Data Team, Greensboro, NC, USA, 2011. Available online: http://plants.usda.gov (accessed on 10 April 2011).

37. Dirr, M.A. *Manual of Woody Landscape Plants: Their Identification, Ornamental Characteristics, Culture, Propagation and Uses*; Stipes Publishing: Champaign, IL, USA, 1998.

38. Maddox, V.; Byrd, J.; Serviss, B. Identification and control of invasive privets (*Ligustrum spp.*) in the middle southern United States. *Invasive Plant Sci. Manag.* **2010**, *3*, 482–488.

39. Miller, J.H.; Chambliss, E.B.; Loewenstein, N.J. *A Field Guide for the Identification of Invasive Plants in Southern Forests*; General Technical Report Southern Research Station-119; U.S. Department of Agriculture, Forest Service, Southern Research Station: Asheville, NC, USA, 2010.

40. Merriam, R.W.; Feil, E. The potential impact of an introduced shrub on native plants diversity and forest regeneration. *Biol. Invasions* **2002**, *4*, 369–373.

41. Wilcox, J.; Beck, C.W. Effects of *Ligustrum sinense* Lour. (Chinese privet) on abundance and diversity of songbirds and native plants in a southeastern nature preserve. *Southeast. Nat.* **2007**, *6*, 535–550.

42. Hanula, J.L.; Horn, S.; Taylor, J.W. Chinese privet (*Ligustrum sinense*) removal and its effect on native plant communities of riparian forests. *Invasive Plant Sci. Manag.* **2009**, *2*, 292–300.

43. Shelton, M.G.; Cain, M.D. Potential carry-over of seeds from 11 common shrub and vine competitors of loblolly and shortleaf pines. *Can. J. For. Res.* **2002**, *32*, 412–419.

44. Greenberg, C.H.; Walter, S.T. Fleshy fruit removal and nutritional composition of winter-fruiting plants: A comparison of non-native invasive and native species. *Nat. Areas J.* **2010**, *30*, 312–321.

45. Stromayer, K.; Warren, R.J.; Johnson, A.S.; Hale, P.E.; Rogers, C.L.; Tucker, C.L. Chinese privet and the feeding ecology of white-tailed deer: The role of an exotic plant. *J. Wildl. Manag.* **1998**, *62*, 1321–1329.

46. Hannaway, D.; Fransen, S.; Cropper, J.; Teel, M.; Chaney, M.; Griggs, T.; Halse, R.; Hart, J.; Cheeke, P.; Hansen D. *Tall Fescue (Festuca arundinacea* Schreb). A Pacific Northwest Extension Publication PWN 504; Oregon State University: Corvallis, OR, USA, 1999.

47. Fleming, C.A.; Wofford, B.E. The vascular flora of Fall Creek Falls State Park, Van Buren and Bledsoe Counties, Tennessee. *Castanea* **2004**, *69*, 164–184.

48. Pedersen, J.F.; Lacefield, G.D.; Ball, D.M. A review of the agronomic characteristics of endophyte-free and endophyte-infected tall fescue. *Appl. Agric. Res.* **1990**, *3*, 188–194.

49. Spyreas, G.; Gibson, D.J.; Middleton, B.A. Effects of endophyte infection in tall fescue (*Festuca arundinacea*, Poaceae) on community diversity. *Int. J. Plant Sci.* **2001**, *162*, 1237–1245.

50. Clay, K.; Schardl, C. *Evolutionary Origins and Ecological Consequences of Endophyte Symbiosis with Grasses*; University Chicago Press: Chicago, IL, USA, 2002.

51. Schardl, C.L.; Leuchtmann, A.; Spiering, M.J. Symbioses of grasses with seedborne fungal endophytes. *Annu. Rev. Plant Biol.* **2004**, *55*, 315–340.

52. Rudgers, J.; Clay, K. An invasive plant-fungal mutualism reduces arthropod diversity. *Ecol. Lett.* **2008**, *11*, 831–840.

53. Creager, R.A. Seed Germination, physical and chemical control of catclaw mimosa (*Mimosa pigra var. pigra*). *Weed Technol.* **1992**, *6*, 884–891.

54. Ares, A.; Burner, D.M.; Brauer, D.K. Soil phosphorus and water effects on growth, nutrient and carbohydrate concentrations, d13C, and nodulation of silktree (*Albizia julibrissin* Durz.) on a highly weathered soil. *Agrofor. Syst.* **2009**, *76*, 317–325.

55. Addlestone, B.J.; Mueller, J.P.; Luginbuhl, P.M. The establishment and early growth of three leguminous tree species for use in silvopastoral systems in the southern USA. *Agrofor. Syst.* **1998**, *44*, 253–265.

56. Bransby, D.I.; Sladden, S.E.; Aiken, G.E. Silktree as a Forage Plant: A Preliminary Evaluation. In *Proceedings of the Forage Grassland Conference*, Georgetown, TX, USA, 5 April 1992; American Forage Grass Council; Volume 1, pp. 28–31.

57. Matta-Machado, R.P.; Jordan, C.F. Nutrient dynamics during the first three years of an alley cropping agroecosystem in southern USA. *Agrofor. Syst.* **1995**, *30*, 351–362.

58. Rhoades, C.C.; Nissen, T.M.; Kettler, J.S. Soil nitrogen dynamics in alley cropping and no-till systems on ultisols of the Georgia Piedmont, USA. *Agrofor. Syst.* **1997**, *39*, 31–44.

59. Jordan, C.F. Organic farming and agroforestry: Alley cropping for mulch production for organic farms in southern United States. *Agrofor. Syst.* **2004**, *61*, 79–90.

60. Loewenstein, N.J.; Loewenstein, E.F. Alien plants in the understory of riparian forests across a land use gradient in the Southeast. *Urban Ecosyst.* **2005**, *8*, 79–91.

61. Birdsey, R.A.; Schreuder, H.T. *An Overview of Forest Inventory and Analysis Estimation Procedures in the Eastern United States–with an Emphasis on the Components of Change*; USDA Forest Service, Rocky Mountain Forest and Range Experiment Station: Fort Collins, CO, USA, 1992.

62. Environmental Systems Research Institute (ESRI). *ArcGIS*; Environmental Systems Research Institute: Redlands, CA, USA, 2009.

63. *Earth Resources Data Analysis System* (*ERDAS IMAGINE 9.2.*); Intergraph Corporation: Norcross, GA, USA, 2008.

64. Healey, S.P.; Cohen, W.B.; Yang, Z.Q.; Krankina, O.N. Comparison of tasseled cap-based Landsat data structures for use in forest disturbance detection. *Remote Sens. Environ.* **2005**, *97*, 301−310.

65. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150.

66. United States Bureau of the Census (USBOC). *Tiger Files*; USBOC, Geography Division: Washington, DC, USA, 2000.

67. Anderson, J.R.; Hardy, E.E.; Roach, J.T.; Witmer, R.E. A Land Use and Land Cover Classification System for Use with Remote Sensor Data. In *United States Geological Survey Professional Paper 964*; United States Government Printing Office: Washington, DC, USA, 1976.

68. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM *vs*. CCA spatial modeling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122.

69. Piedallu, C.; Gegout, J. Efficient assessment of topographic solar radiation to improve plant distribution models. *Agric. For. Meteorol.* **2008**, *148*, 1696–1706.

70. Simley, J.D.; Carswell, W.J., Jr. *The National Map—Hydrography*; U.S. Geological Survey Fact Sheet 2009-3054; U.S. Geological Survey: Reston, VA, USA, 2009.

71. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*; Wiley Interscience: New York, NY, USA, 2000.

72. Phillips, S.; Anderson, R.; Schapire R. Maximum entropy modelling of species geographic distributions. *Ecol. Model.* **2006**, *190*, 231–259.

73. *SAS*, Version 9.2; SAS Institute: Cary, FL, USA, 2009.

74. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.

75. Manel, S.; Williams, H.C.; Ormerod, S.J. Evaluating presence-absence models in ecology: The need to account for prevalence. *J. Appl. Ecol.* **2002**, *38*, 921–931.

76. Lobo, J.M.; Jiménez-Valverde, A.; Real. R. AUC: A misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* **2008**, *17*, 145–151.

77. Oommen, T.; Baise, L.G.; Vogel, R.M. Sampling bias and class imbalance in maximum-likelihood logistic regression. *Math. Geosci.* **2010**, *43*, 99–120.

78. Wisz, M.S.; Hijmanss, R.J.; Peterson, A.T.; Graham, C.H.; Guisan, A. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* **2008**, *14*, 763–773.

79. Peterson, A.T. Predicting the geography of species' invasions via ecological niche modelling. *Q. Rev. Biol.* **2003**, *78*, 419–433.

80. Turner, W.; Spector, S.; Gardiner, N.; Fladeland, M.; Sterling, E.; Steininger, M. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* **2003**, *18*, 306–314.

81. Sester, M. Optimization approaches for generalization and data abstraction. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 871–897.

82. Barbosa, A.M.; Real, R.; Vargas, J.M. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecol. Model.* **2009**, *220*, 747–754.

83. Ducheyne, E.I.; De Wulf, R.R.; De Baets, B. A spatial approach to forest-management optimization: Linking GIS and multiple objective genetic algorithms. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 917–928.