

The spatial legacy of introduction: *Celastrus orbiculatus* in the southern Appalachians, USA

Thomas P. Albright^{1,*}, Dean P. Anderson², Nicholas S. Keuler³, Scott M. Pearson⁴ and Monica G. Turner¹

¹Department of Zoology, University of Wisconsin-Madison, Madison, WI 53706, USA; ²Landcare Research, PO Box 40, Lincoln 7640, New Zealand; ³Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, USA; ⁴Department of Natural Sciences, Mars Hill College, Mars Hill, NC 28754, USA

Summary

1. A variety of abiotic, biotic, human and historic variables related to environmental suitability and propagule pressure determine the distribution of invasive plants in a landscape. Understanding the role of these variables for invasive species is challenging because environmental variables are often correlated, many invaders have broad ecological niches, and invasive distributions are often highly dynamic.

2. We examined the role of environmental variables at multiple spatial scales on the distribution of an invasive vine *Celastrus orbiculatus* (Celastraceae) in the southern Appalachians, USA. Using existing and newly collected occurrence data, we constructed spatial generalized linear mixed models of *C. orbiculatus* occurrence in the whole region as well as the areas nearest (established zone) and farthest (pioneer zone) from the purported introduction locus.

3. For the entire study region, distance from the introduction locus and terrain slope were negatively related to the probability of occurrence of *C. orbiculatus*. The model was highly discriminatory, with an area under the receiver-operating characteristic curve of 0.91.

4. Distance from introduction locus and slope were also important in the models for the pioneer and established zones respectively.

5. The pioneer zone model also indicated a negative relationship with elevation and the established zone model indicated a positive relationship with local topographic position, but these models were less discriminatory.

6. *Synthesis and applications.* Distance from introduction locus was the most powerful variable examined, suggesting an enduring legacy of introduction for the current distribution of *C. orbiculatus*. The varying results in the three zones highlight the importance of accounting for both introduction history and uneven propagule pressure in studies of observed invasive species distributions. The observed relationships with distance and elevation suggest that land managers should prepare for *C. orbiculatus* to expand its distribution in the region over time and with a warming climate.

Key-words: Asiatic bittersweet, Blue Ridge Province, ecological niche, generalized linear mixed models, invasive species, landscape ecology, logistic regression, propagule pressure, species distribution modelling

Introduction

Knowledge of the distribution of invasive plant species and the factors that influence it is important for invasive species

management and control efforts (NRC 2002). However, explaining and predicting the distribution of plants in an exotic range poses several special challenges. First, invasive plants are typically characterized by broad ecological niches, which may be difficult to model (Mitchell, Lancia & Gerwin 2001; Brotons *et al.* 2004). Secondly, scientists and managers typically require information on distribution and ecological niche when the species is in an early phase of invasion (NRC 2002). Especially

*Correspondence author. Department Forest & Wildlife Ecology, 1630 Linden Dr., Madison, WI 53706, USA. E-mail: talbrig1@wisc.edu

in these early phases, spatial distributions are typically far from equilibrium with the environment, resulting in weak or spurious inference of ecological niches (Welk 2004). Instead, distribution and niche inference may be most strongly influenced by the spatial distribution of propagule pressure of a species. Propagule pressure (Lonsdale 1999) may be related to historic factors, such as sites of introduction, human dispersal pathways, and factors related to the reproductive biology and dispersal of the species. Thus, both the actual ecological niche and propagule pressure must be considered in models of incipient invasions. In this study, we examined the roles of various abiotic, biotic, human and historic factors in shaping the distribution of an exotic invasive plant in the southern Appalachians of Tennessee (TN) and North Carolina (NC), USA.

Asiatic bittersweet *Celastrus orbiculatus* Thunb. (Celastraceae) is a woody vine native to East Asia. First introduced to the United States as early as 1736 for horticultural purposes (Miller 2003), it has become naturalized in at least 21 eastern and Midwestern states (Patterson 1974; USDA NRCS 2007). *Celastrus orbiculatus* is especially concentrated immediately south of Asheville, NC, where it has been documented at least as far back as 1895 (McNab & Meeker 1987; Merriam 2003), suggesting a locus of introduction. It is of growing concern as it spreads in the southern Appalachians (Merriam 2003) and possesses several traits that make it both successful as an invader and of acute concern. It has high pollen and seed viability, especially relative to the native congener, *C. scandens* L. (Dreyer, Baird & Fickler 1987). Its coloured capsules have made it attractive for both horticulture and crafts, including a thriving wreath-making trade (Dreyer 1994; NCDA&CS 2007). These uses, coupled with consumption of the fruit by birds and mammals, have resulted in a medium- and long-distance dispersal capability (Dreyer 1994). In full and partial sunlight, it can grow sufficiently to overtop, damage and eventually kill established trees, drawing comparisons to kudzu *Pueraria montana* (Lour.) Merr. (Patterson 1974; Dreyer *et al.* 1987). It can also act as a 'sit-and-wait' invader, persisting under all but the most dense forest canopies until a disturbance brings increased light and rapid growth (Ellsworth, Harrington & Fownes 2004).

The goal of this investigation was to determine the factors influencing the distribution of *C. orbiculatus* in the southern Appalachians so as to improve the ability of managers to focus prevention and mitigation efforts. Because the vine's distribution is still expanding (Merriam 2003), we expected that factors related to propagule pressure integrated over residence time (Wilson *et al.* 2007) would be of primary importance. Specifically, we predicted that locations closer to the introduction locus would be more likely to have *C. orbiculatus*. Because of the role humans play in spread, we expected that road proximity and development would be associated with occurrence. Although the ability of forest to prevent establishment of *C. orbiculatus* is not clear (Ellsworth *et al.* 2004), we predicted that *C. orbiculatus* would tend to occur at the forest's edge. Following a study modelling *C. orbiculatus* occurrence in a 2400-ha watershed (McNab & Loftis 2002), we predicted that *C. orbiculatus* would favour sites that were low relative to neighbouring terrain. We predicted a classic hump-shaped

relationship between elevation and occurrence of *C. orbiculatus* (MacArthur 1972). Additionally, we sought to assess the stability of relationships between occurrence and predictor variables in areas where *C. orbiculatus* was well established and areas where its arrival was more recent. We expected that areas close to the introduction locus would be strongly influenced by factors related to the ecological niche of *C. orbiculatus*, while propagule pressure would be the chief factor in areas distant from the introduction locus.

Materials and methods

The study was conducted in the southern Appalachian Mountains, USA, a forested region of high biodiversity, steep environmental gradients and elevation ranging from 250 to 2037 m [see Appendix S1 (Supporting information) for map]. Regional forest cover has increased steadily since clearing occurred in the late 1800s, and housing density and development have increased considerably in the past 30–40 years (Gragson & Bolstad 2006). Recreational and natural resource amenities are important regionally and may be threatened by *C. orbiculatus* and other invaders (McNab & Meeker 1987; NPS 2007).

OCCURRENCE DATA

A database was obtained from the Southern Appalachian Information Node (SAIN) project of the National Biological Information Infrastructure (SAIN 2005). We pooled data from two different sources contained in the database: (1) National Park Service and US Forest Service survey data (NPS/USFS) and (2) Southern Appalachian Volunteer Environmental Monitoring data (SAVEM). Presence and absence data from NPS/USFS were derived from exotic plant surveys in public lands during the period 1994–2004 and included 151 presences in 1869 samples. Data from SAVEM were collected between 2002 and 2003 by trained volunteers along specified survey transects on trails in public forest land. Because volunteers were constantly looking for *C. orbiculatus* along the transect routes, it is possible to infer absence of a species from locations along transects that lack presence observations of that species. Using ArcGIS (ESRI, Redlands, CA, USA), we generated absence points for locations lacking *C. orbiculatus* at c. 100-m intervals along the routes. Because the selection of this interval (and the ensuing number of absence points) influences model intercepts, we included a correction for this in statistical models (below). The resulting SAVEM data set included 18 presences in 1863 samples.

To cover a wider range of areas and conditions, we collected additional field data. During late June to mid-July 2006, we sampled trails and roads in public- and private-forested lands for the absence or presence of *C. orbiculatus*. After identifying several large contiguous areas (hundreds of kilometres each) in the study area that were not sampled in the database data, we used 1 : 24 000–1 : 150 000 maps to identify candidate road or trail corridors, from which we randomly selected. At the beginning of a corridor, a GPS waypoint (Garmin 12XL; Garmin International Inc., Olathe, KS, USA) was recorded. Subsequent sampling plots were established at 250-m intervals until at least four plots were surveyed. Field data collection involved walking along the trail or the roadside, searching for focal species within a 2-m buffer around the trail or forested area at the roadside for a distance of 30 m. For corridors that were forested only on one side, the team surveyed a one-sided 60-m transect in lieu of a two-sided 30-m transect. Every fourth plot was designated for a 30-m off-corridor

sampling transect, perpendicular to the trail or road. Although *C. orbiculatus* was never observed in the 47 off-corridor transects, informal field observations and McNab & Loftis (2002) make clear that *C. orbiculatus* does occur outside of road and trail corridors. The resulting field survey included 41 presences among 348 samples and was combined with the database samples (Appendix S1, Supporting information). Differences in the prevalence of *C. orbiculatus* among the data sets are driven by corresponding differences in distance from the introduction locus, elevation and vegetation composition in the samples making up the respective data sets.

ENVIRONMENTAL DATA

A variety of remotely sensed and GIS data were used to derive variables related to abiotic, biotic and human themes we hypothesized to be important in explaining the distribution of *C. orbiculatus* (Table 1). We explored a range of scales because the scales at which focal variables may be related to occurrence were not known. Elevation and terrain slope were obtained from the National Elevation Database (Gesch *et al.* 2002). To characterize topographic slope position, we used ERDAS Imagine 9.0 (ERDAS, Atlanta, GA, USA) to calculate a terrain index indicating how high or low a pixel is relative to neighbouring pixels (Homer *et al.* 2004) within a moving window of four different sizes (2.3, 4.4, 7.3 and 10.9 ha). Two measures related to insolation were calculated. First, aspect, α (degree), was transformed into 'south-southwestness' (α') (Beers, Dress & Wensel 1966).

$$\alpha' = \sin(\alpha + 22) + 1 \quad \text{eqn 1}$$

Secondly, we calculated an insolation index, s , which accounts for both terrain slope, β , and the transformed aspect (modified from Gustafson, Lietz & Wright 2003).

$$s = \sin((\beta/90) * 180) * \alpha' \quad \text{eqn 2}$$

We assessed both development and forest cover at three scales. Impervious surface and tree canopy percentages were obtained for 30-m (0.1 ha) pixels from the 2001 National Land Cover Dataset (NLCD) (Homer *et al.* 2004). For coarser scales, we used Fragstats 3 (McGarigal 2007) to calculate the fraction of 30-m NLCD pixels in

0.8 and 4.4 ha square windows classified as developed or forest. The density of forest edge was also calculated from 2001 NLCD data using Fragstats. A road distance layer and road density layer within a 200-m radius circle were calculated using Hawth's Analysis Tools for ArcGIS (Beyer 2004). Finally, after designating a point in the Biltmore Forest (35.54°N, 82.53°W) to approximate the centre of the purported introduction locus between Asheville and the Blue Ridge Parkway, we generated a GIS layer of the distance to this locus. This layer, together with the road distance layer were natural log transformed: $\ln(\text{variable} + 1)$. Each GIS layer was projected into Universal Transverse Mercator zone 17, with a 30-m cell size.

Values for environmental variables were obtained by overlaying the coordinates of plots on the environmental GIS layers. Positional accuracy was within one 30-m pixel of known locations. Pooling field survey data with the database data resulted in a combined data set with 210 presences and 3870 absences.

GENERALIZED LINEAR MODELLING

Preliminary analysis included identifying the best variables and model forms in each theme and assessing spatial autocorrelation. Because of the unbalanced nature of our data set (18.4 absences for each presence) and the importance of presences (or, if presences are predominant, absences) in determining the effective degrees of freedom in logistic regression models (Peduzzi *et al.* 1996), we reserved a random 2/3 sample of presences for model development. Memory constraints imposed by subsequent spatial model generation (below) limited development sample size to 2000. Thus, the development data set consisted of 140 presences (inclusion probability, $P_1 = 0.67$) and 1860 absences (inclusion probability, $P_0 = 0.48$) and the validation data set consisted of 70 presences and 2010 absences, randomly selected. When estimating final occurrence probability (below), we added an adjustment to the intercept, β_0 , based on the variable inclusion probabilities for presence and absence samples (Keating & Cherry 2004), such that

$$\beta'_0 = \beta_0 - \ln(P_1/P_0) \quad \text{eqn 3}$$

where β'_0 is the adjusted intercept.

Table 1. Environmental variables examined in logistic regression modelling of the occurrence of *Celastrus orbiculatus* in the southern Appalachians, USA

Theme	Variable	Short name	Spatial grain (ha)	Sampled range
Elevation	Elevation	Elevation	0.09	263–1917 m
Terrain slope	Slope	Slope	0.09	0–51°
Slope position	Position index (5 × 5)	Position5	2.25	0–100
	Position index (7 × 7)	Position7	4.41	0–100
	Position index (9 × 9)	Position9	7.29	0–100
	Position index (11 × 11)	Position11	10.89	0–100
Insolation	Transformed aspect	Aspect	0.09	0–1000
	Insolation index	Insolation	0.09	–997–997
Forest cover	Canopy cover	Canopy	0.09	0–100%
	Percent forest (3 × 3)	Forest3	0.81	0–100%
	Percent forest (7 × 7)	Forest7	4.41	0–100%
Forest edge	Forest edge density (3 × 3)	Forestedge3	0.81	0–667 m ha ⁻¹
	Forest edge density (7 × 7)	Forestedge7	4.41	0–374 m ha ⁻¹
Development	Impervious surface	Impervious	0.09	0–81%
	Percent developed (3 × 3)	Develop3	0.81	0–100%
	Percent developed (7 × 7)	Develop7	4.41	0–100%
Roads	Road density	Roaddens	13.85	0–106.2 m ha ⁻¹
	Distance to roads	Roaddist	0.09	0–7446 m
Propagule pressure	Distance to introduction locus	Introdist	0.09	2370–156 270 m

First, we performed logistic regression (Hosmer & Lemeshow 2000) using the R command `glm` (R Development Core Team 2006) on each variable separately, including both linear and quadratic terms to account for unimodal (hump-shaped) relationships between the variables and *C. orbiculatus* occurrence. From this pool, we selected the best univariate models within each theme (Table 1) based on Akaike's Information Criterion (AIC) (Akaike 1974). We screened the selected variables for high correlation (Pearson's $r > 0.70$), removing those with weaker AIC and/or high correlations with multiple variables. From the resulting variables in their linear or quadratic forms, we assembled a candidate model superset on which we performed stepwise AIC selection (stepAIC) in the MASS package for R (Venables & Ripley 2002). We then pooled development and validation data, refit the preliminary model, and generated a variogram from deviance residuals using the R package `geoR` (Ribeiro & Diggle 2001) to assess spatial autocorrelation. Pooling the data sets was necessary for this step because of the weak information content of binary response variables in assessment of spatial autocorrelation.

To account for effects of distance relative to the introduction locus, we conducted three parallel analyses, again using the development data set. We developed a 'region-wide' model, which encompassed all development samples. The remaining two analyses were segregated according to a distance threshold 30 km from the introduction locus. This threshold corresponds to an inflection in the relationship between distance and frequency of *C. orbiculatus*, beyond which occurrence was very infrequent. The analysis of development data within 30 km of the locus ($n = 165$, presences = 101), termed 'established', was designed to characterize the ecological niche of *C. orbiculatus* in areas where it has had sufficient colonization opportunities. The third analysis, termed 'pioneer', was outside of this distance, where we hypothesized that only suitable sites experiencing the most propagule pressure would have *C. orbiculatus* ($n = 1835$, presences = 39).

For each analysis region, we began by identifying best logistic regression models using both forwards and backwards stepwise AIC

selection, using AIC_C when required because of small sample size (Bedrick & Tsai 1994). With the best model identified, we constructed a mixed-effects logistic regression model using `glmmPQL` in the MASS package of R (Venables & Ripley 2002), which included a random spatial autocorrelation effect based on findings from variogram analysis. If adding the spatial component resulted in individual covariates losing significance (using a $P < 0.10$ threshold; AIC is not retrievable in mixed-effects logistic models), models were trimmed until remaining covariates met this threshold.

Models were validated by the withheld validation data set using Cohen's κ (Cohen 1960) and receiver-operating characteristic (ROC) plots (Zweig & Campbell 1993). Cohen's κ can be used to assess agreement between two binary data sets in a way that accounts for chance agreement. Values of κ range from -1.0 to 1.0 , with 1.0 indicating perfect agreement and 0.0 indicating no difference from random. In calculating κ , we used a probability threshold of 0.50 to distinguish areas predicted 'present' from those predicted to be 'absent'. Plots of ROC depict the relationship between model sensitivity ($1 - \text{omission error}$) and commission error ($1 - \text{specificity}$) allowing assessment of tradeoffs between error types. We also calculated the area under the ROC curve (AUC), which indicates discriminatory ability irrespective of prediction threshold. The AUC of a random model, which would appear as a diagonal line on an ROC plot, is 0.5 . To evaluate the generality of the model, we also performed validations of the resulting models in each of the three prediction zones.

Results

Univariate logistic regression on the region-wide data set and the resulting AIC differences identified the best variables and model forms in each theme (Table 2). Slope position, development and forest were more informative at the coarser scales (7.3, 4.4 and 4.4 ha respectively), while forest edge was most informative measured in a 0.8 ha window. The resulting

Theme	Variable	Linear sign	Linear delta AIC	Quadratic form	Quadratic delta AIC
Elevation	Elevation	Pos.	-11.72	Hump	-12.48
Terrain slope	Slope	Pos.	-23.63	Hump	-23.87
Slope position	Position5	Pos.	-16.25	Hump	-17.24
Slope position	Position7	Pos.	-22.73	Hump	-26.17
Slope position	Position9	Pos.	-28.75	Hump	-30.31
Slope position	Position11	Pos.	-16.97	Hump	-18.62
Insolation	Aspect	Pos.	-13.01	Hump	-11.16
Insolation	Insolation	Pos.	-23.98	Pos.	-30.53
Forest cover	Canopy	Neg.	0.64	Hump	2.42
Forest cover	Forest3	Neg.	-46.46	Hump	-67.42
Forest cover	Forest7	Neg.	-25.87	Hump	-94.57*
Forest edge	Forestedge3	Pos.	-41.71	Hump	-59.99
Forest edge	Forestedge7	Pos.	-42.73	Hump	-59.29
Development	Impervious	Pos.	-0.88	Pos.	1.06
Development	Develop3	Pos.	-61.15	Hump	-76.91
Development	Develop7	Pos.	-39.75	Hump	-117.67
Roads	Roaddens	Pos.	-5.75	Hump	-36.95
Roads	Roaddist	Neg.	-49.32†	Neg.	-50.47
Introdist	Introdist	Neg.	-393.85		

Table 2. Univariate logistic regression summary

Bold indicates variables and forms retained for further analysis (up to one per theme).

*No forest cover variable was retained because of high correlation with develop7.

†Roaddist was selected in favour of roaddens because of lower correlation with develop7.

candidate model superset included representative variables from eight of the nine themes plus five quadratic terms (Table 2). A preliminary region-wide model, including slope, develop7 (quadratic) and introdist were identified using step-wise AIC selection. The empirical variogram revealed significant spatial autocorrelation so an exponential variogram model was fitted, resulting in a nugget (semi-variance at zero distance) of 0.07, an apparent sill (limit of semi-variance at long distances) of 0.18, and a range (distance at which sill is reached) of 3900 m (Fig. 1).

REGION-WIDE MODELS

Celastrus orbiculatus occurred in 140/2000 sampling sites in the region-wide development data subset. Region-wide model selection using this data set resulted in retention of the same covariates as the preliminary model, with *C. orbiculatus* negatively related to slope and introdist and having a hump-shaped relationship with develop7 (Table 3). However, both the linear

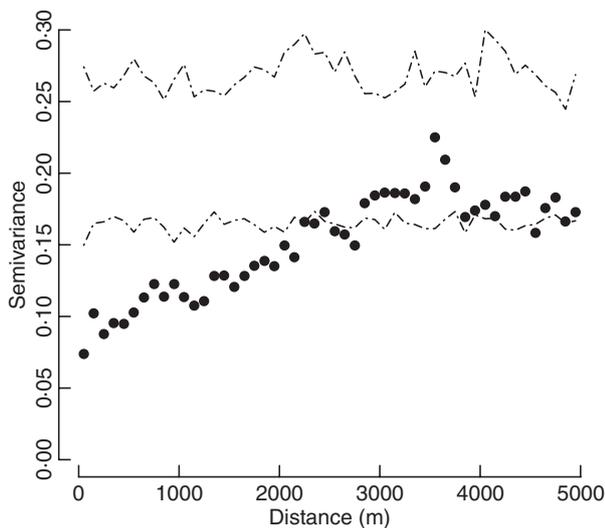


Fig. 1. Empirical semi-variogram from full *Celastrus orbiculatus* distribution model deviance residuals with 95% confidence interval (dashed lines) around a model with no spatial autocorrelation obtained by Monte Carlo resampling.

and quadratic terms for develop7 lost significance when spatial autocorrelation was accounted for, resulting in a final region-wide model retaining only slope and introdist. In this model, *C. orbiculatus* is half as likely to occur in a flat area 20.6 km away from the locus than one just 10 km from the locus. For terrain slope, the odds ratio for each 15° increment is 0.55, such that the odds of occurrence in a non-sloping pixel 20 km from the introduction locus are 1 : 1.22 ($P = 0.45$) but with a 15° slope, the odds are 1 : 2.21 ($P = 0.31$) (Fig. 2). Note that, while probability may be a more intuitive metric and is included where appropriate, we also report odds because, unlike probability, changes in odds are constant across the range of variation of linear covariates. Related to this fact, the odds ratio (Hosmer & Lemeshow 2000) is reported as an indicator of effect size for linear covariates.

ESTABLISHED ZONE MODELS

In the established zone (< 30 km from locus) development data subset, *C. orbiculatus* occurred in 101/140 sampling sites. The best model identified using AIC_C included a negative

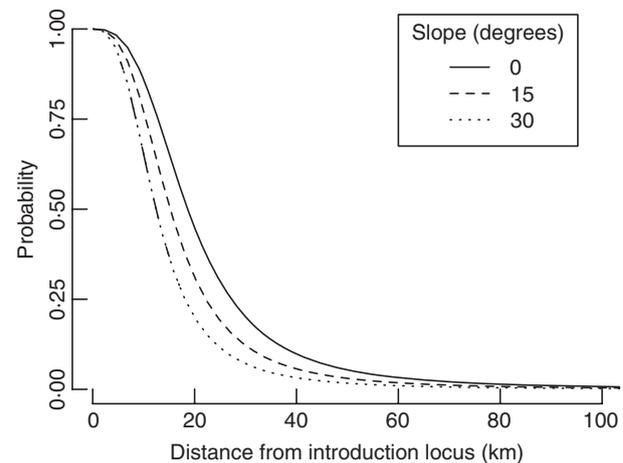


Fig. 2. Variation in probability of *Celastrus orbiculatus* occurrence with distance from an introduction locus south of Asheville, NC and terrain slope estimated from the final region-wide spatial logistic regression model.

Table 3. Model parameter estimates for each of the three zones from logistic regression models, logistic regression models accounting for spatial autocorrelation including the same variables, and reduced (final) spatial logistic models after non-significant variables were removed

Region	Model	Intercept	Slope	Position9	Elevation	Develop7	Develop7 ²	Roaddist	Roaddist ²	Introdist
Region-wide	Aspatial	18.19***	-0.040**			0.076***	-0.001***			-2.902***
Region-wide	Full spatial	18.81***	-0.039*			0.021 ^(NS)	-0.000 ^(NS)			-2.942***
Region-wide	Red. spatial	18.74***	-0.036*							-2.912***
Established	Aspatial	-0.12 ^(NS)	-0.114***	0.031**		0.119***	-0.002***			
Established	Full spatial	0.25 ^(NS)	-0.060**	0.016		0.037 ^(NS)	-0.001 ^(NS)			
Established	Red. spatial	0.39 ^(NS)	-0.055**	0.015						
Pioneer	Aspatial	15.24***		-0.025**	-0.003***			0.753*	-0.102**	-2.289***
Pioneer	Full spatial	36.76 ^(NS)		-0.014 ^(NS)	-0.005 ^(NS)			0.347 ^(NS)	-0.033 ^(NS)	-4.928 ^(NS)
Pioneer	Red. spatial	20.88*			-0.004*					-2.844*

Significance of coefficients is denoted as follows: *** ≤ 0.001 , ** 0.001–0.01, * 0.01–0.05, · 0.05–0.10, ^(NS) > 0.10.

relationship between *C. orbiculatus* and slope, a positive relationship with position9, and a hump-shaped relationship with develop7. Again, develop7 lost all significance in the spatial model, while slope and position9 were retained, albeit with weaker coefficients (Table 3). In the resulting model, the effect size of slope is greater than in the regional model, with an odds ratio of 0.44 for each 15° increase in slope. The effect of varying terrain position is sizeable, with odds increasing from 1.5 : 1 ($P = 0.60$) in low-lying flat areas (position7 = 0) to 6.5 : 1 ($P = 0.87$) along ridge tops (position7 = 100) for an odds ratio of 4.33.

PIONEER ZONE MODELS

Celastrus orbiculatus was much less common in the pioneer zone (> 30 km from locus) development subset, occurring in a mere 39/1835 samples. A model for *C. orbiculatus* occurrence indicating a negative relationship with position9, elevation, and introdist, and a hump-shaped relationship with roaddist was selected. After accounting for spatial autocorrelation, however, only elevation and introdist retained significance (Table 3). In this model, each 100-m gain in elevation reduced

the odds of *C. orbiculatus* occurrence by a factor of 0.67. The effect of introdist is similar to that found in the regional model. For areas with the same elevation, the odds of *C. orbiculatus* occurrence are twice as high at 30 km than at 38.3 km from the locus.

MODEL ASSESSMENT

From the spatial models, we created probability maps (e.g. Fig. 3). The κ statistic, plots of ROC, and associated AUC values (Fig. 4) indicate that the region-wide model was the best performer overall ($\kappa = 0.55$, AUC = 0.91) and when evaluated using the established validation data set ($\kappa = 0.24$, AUC = 0.66). It performed weakly on the pioneer validation data set ($\kappa = 0.00$, AUC = 0.60). The established zone model was a weak performer in all zones: established ($\kappa = 0.09$, AUC = 0.58), region-wide ($\kappa = 0.00$, AUC = 0.46) and pioneer ($\kappa = 0.00$, AUC = 0.45). The pioneer zone model was moderately discriminatory in the pioneer zone ($\kappa = 0.27$, AUC = 0.65), a good performer region-wide ($\kappa = 0.46$, AUC = 0.90), but weak in the established zone ($\kappa = 0.10$, AUC = 0.59).

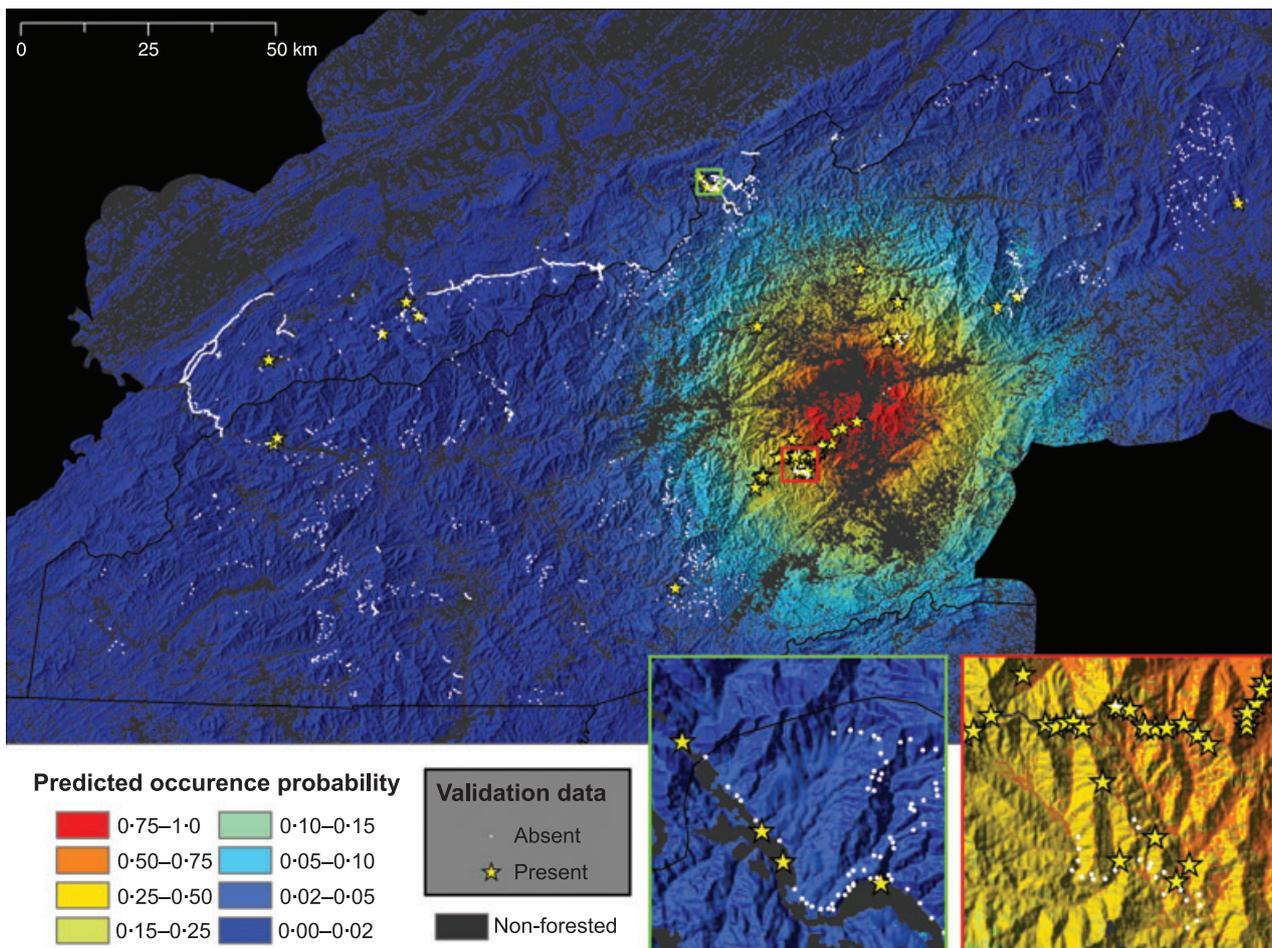


Fig. 3. Predicted occurrence probability for *Celastrus orbiculatus* based on the final region-wide spatial logistic regression model. Magnification of two areas (inset) is also provided.

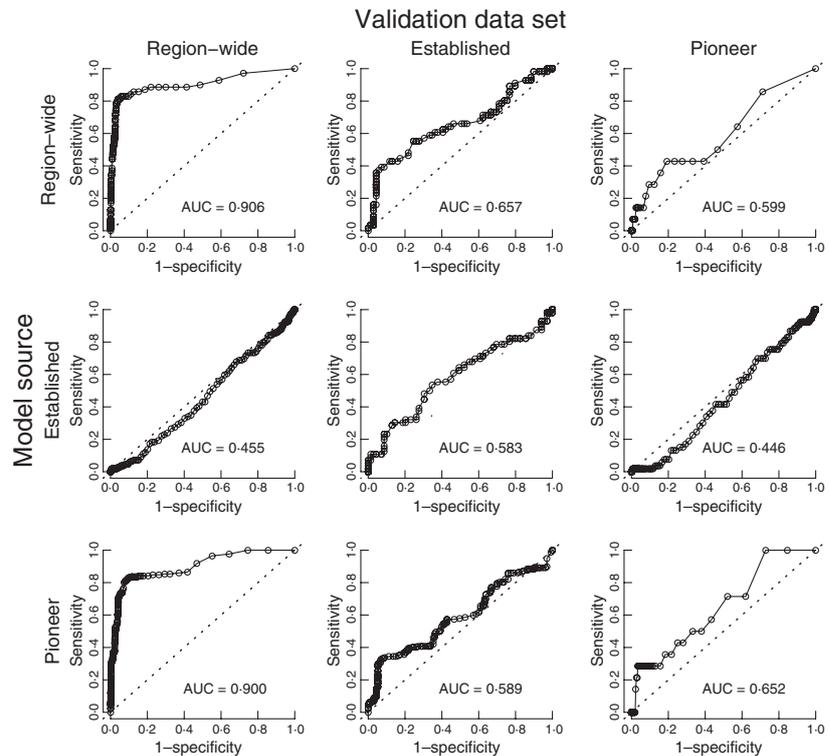


Fig. 4. Receiver operating characteristic plots for the final spatial logistic regression models in the three modelling zones (rows) evaluated against validation data from each of the zones (columns). The dashed 1 : 1 line represents a random prediction.

Discussion

VARIABLES INFLUENCING *CELASTRUS ORBICULATUS* DISTRIBUTION

Univariate and multivariate model results confirm the importance of the locus of introduction in explaining the current distribution of *C. orbiculatus* in the southern Appalachians. Although occurrence of an organism will always be negatively related to distance from the centre of its distribution if a sufficiently large area is considered, in this case, the species also occurs well outside the bounds of the limited study area. The introduction locus probably experienced multiple introductions in proximity to Asheville, which may have resulted in enhanced adaptive potential (Lavergne & Molofsky 2007). Although the importance of distance from source populations is supported by theory (Skellam 1951), as well as observational (Alston & Richardson 2006) and modelling studies (Higgins, Richardson & Cowling 2000), the effect is not always as strong or as enduring as found here (but see Foxcroft *et al.* 2004). The topography of the southern Appalachians fragments the region at a scale that may impede natural dispersal vectors, allowing only minimal (most likely human-linked) nascent invasion foci to develop on the periphery (Moody & Mack 1988).

Elevation and local topography play important roles in the distribution of *C. orbiculatus*, but not always in a manner consistent with our predictions. *Celastrus orbiculatus* was more likely to appear at lower altitudes in the pioneer model. We did not detect a declining probability of occurrence at lower elevations as expected, perhaps because these elevations were inade-

quately sampled. Similarly, only *c.* 1% of samples were obtained at elevations > 1500 m, which may have been inadequate for identifying declining occurrence in the other models. On the other hand, movement up to higher elevations, which tend to be heavily forested and remote, may be underway, given the known occurrence of *C. orbiculatus* in relatively cold portions of New England (Mehrhoff *et al.* 2003). The position9 variable was positively related to occurrence in the established region, suggesting that *C. orbiculatus* favours locally high locations. However, this relationship is equivocal, not being found significant in the region-wide or pioneer multivariate spatial models. Interpretation of the negative relationship observed between *C. orbiculatus* occurrence and slope in the region-wide and established models is challenging because of the important correlations between slope and other variables. For example, human development, roads, and agriculture are often limited to gently sloping terrain. Furthermore, we noticed that heavily sloping areas were often associated with *Rhododendron* spp. and *Kalmia latifolia* L. thickets, which are characterized by low plant diversity and a lack of *C. orbiculatus*.

Univariate and non-spatial multivariate models suggest that other variables may play roles in shaping the distribution of *C. orbiculatus*. The occurrence of *C. orbiculatus* exhibited a hump-shaped relationship with roaddist, although this variable was retained only in the pioneer logistic model (Table 3). Note that the estimated coefficients indicate that occurrence probability was maximized at a mere 39 m from roads, with a gradual decline in occurrence at distances beyond this peak. Because this distance is approaching the level of spatial uncertainty in our data, it is likely that the decline is the most meaningful aspect of this relationship. Non-spatial multivariate

models for the region-wide and established zones did not include the road distance variable, but instead included positive relationships between development and *C. orbiculatus* occurrence. Both roads and development are probably associated with increased propagule pressure (Lonsdale 1999; Gelbard & Belnap 2003; McKinney 2004), confirming our prediction and underlining its importance in explaining the current distribution of *C. orbiculatus*. Although forest cover was not included in any modelling, the lack of *C. orbiculatus* observed in off-corridor samples appears to confirm the prediction that interior forest areas would be less favoured. Similarly, univariate models suggest that *C. orbiculatus* is more likely to appear in areas with intermediate or high forest edge (Table 2). This variable may not have been selected in final models because of the inclusion of road distance and development variables with which it was highly correlated ($r = 0.63$ for both). In no case were the best multivariate models able to eliminate spatial autocorrelation in residuals, which we attribute to endogenous spatial dependence associated with short distance dispersal.

COMPARISON OF MODELLING REGIONS

Given the large contrast in *C. orbiculatus* occurrence between areas that are close to the introduction locus and those that are far away, it is not surprising that the region-wide model, which was the only one to span the full range of distances, was the most discriminatory. Nonetheless, even when evaluated within the established and pioneer zones, the model had discriminatory value, suggesting that relationships between occurrence and both slope and distance were robust throughout the entire region.

There are several explanations for the weakness of the model developed in the established zone. The established zone encompasses a narrower range of conditions than the entire region, which may have weakened inference. Furthermore, with an apparently broad ecological niche, propagule pressure may be a much more important factor in explaining distribution of *C. orbiculatus* at this scale. Thus, excluding a major proxy for propagule pressure (distance to introduction locus) may have simply resulted in a model reflective of the generalist nature of the species. Finally, model selection based on data in the established zone was challenged by having a mere 39 absence observations. Peduzzi *et al.* (1996) recommended a minimum of 10 events (in this case, absences) per variable, although there is evidence that high overall sample size (presences plus absences) can compensate for this to a degree (Reineking & Schroder 2006). Because overall sample size was only 140, the small number of absences undoubtedly resulted in a low-information data set and weakened inference.

In the pioneer zone, *C. orbiculatus* was exceedingly rare. In fact, the maximum predicted probability of occurrence among either validation or development data in this zone was only 0.17. Nonetheless, the pioneer zone model provided some discrimination ability in all zones. Owing to the inclusion of the introduction distance variable, the model was actually most discriminatory when evaluated against the region-wide valida-

tion data set (Fig. 4). As with the established zone, model selection using the pioneer zone data set was probably hampered by a paucity of events (in this case, 39 presences), although this was tempered by a large overall sample size ($n = 1835$).

CONCLUSIONS AND MANAGEMENT RECOMMENDATIONS

This study affirms several lessons for species distribution modelling studies. We offer a cautionary note about the sensitivity of modelling studies to input data quality and composition. Among key questions researchers should ask are whether a species is approaching distributional equilibrium, whether only a portion of the geographic and environmental space suitable to the species is being included, and whether sample size and the number of events in presence/absence data sets are sufficient for making meaningful inference. It is also important to consider spatial autocorrelation. This study revealed that several relevant covariates identified by model selection on non-spatial models were probably elevated in importance because of pseudoreplication associated with spatial autocorrelation (Lennon 2000). Although we argue that it is important to address this issue through the use of autoregressive functions, spatial random effects and other techniques (Dormann *et al.* 2007), researchers should be aware that these techniques themselves involve considerable assumptions (e.g. stationarity) and should be interpreted with as much caution as conventional logistic regression models (Hawkins *et al.* 2007). For this reason, we have chosen to include results from both approaches.

This study also offers insights relevant to the management of *C. orbiculatus* and other exotic invasive species. The role of propagule pressure, as indicated by the effect of distance to both the introduction locus and roads, is paramount and this study adds to evidence of its general importance in governing the distribution of exotic species (Foxcroft *et al.* 2004; McKinney 2004; Lambrinos 2006). While cold currently appears to limit the distribution of *C. orbiculatus*, climate change may render high-elevation sites increasingly vulnerable to this and other plant invaders, as extreme cold temperatures become less frequent (Walther 2004). As time passes and *C. orbiculatus* becomes entrenched on the landscape of the southern Appalachians, the observed roles of variables related to propagule pressure should diminish. In this scenario, *C. orbiculatus* emanates from the established zone along road corridors and aided by long distance dispersal events (Higgins & Richardson 1999), eventually occupying areas that meet the broad requirements of its ecological niche. Finally, hitherto unmentioned is the looming forest upheaval that will occur as large tracts of hemlock (*Tsuga* spp.) in the southern Appalachians succumb to hemlock woolly adelgid *Adelges tsugae* Annand (Kincaid 2007). This may present a significant opportunity to *C. orbiculatus* and other invaders, as appears to be the case in New England (Small, Small & Dreyer 2005). Although there is considerable uncertainty in predicting the spread of invasive species, it is clear that *C. orbiculatus* will remain an important, if unwanted, participant in the ecology of the forests of the southern Appalachians.

Acknowledgements

This research was supported by NSF Grant DEB 9632854 and DEB-0218001 to the Coweeta LTER program and United States Geological Survey grant (144-ME50) and made possible by the helpful contributions of many people and organizations. P. Nabors of the NBII assisted in access and interpretation of database data. H. McNab provided addition background. C. Brunton, D. Flores, J. Nix and T. Kuhman all participated in summer 2006 field sampling, which augmented the efforts of numerous anonymous naturalists. V. Radeloff, J. Lennon, D. Waller, J. Zedler, J. Zhu, D. Gorchov and two anonymous reviewers all provided valuable comments.

References

- Akaike, H. (1974) New look at statistical-model identification. *IEEE Transactions on Automatic Control*, **AC19**, 716–723.
- Alston, K.P. & Richardson, D.M. (2006) The roles of habitat features, disturbance, and distance from putative source populations in structuring alien plant invasions at the urban/wildland interface on the Cape Peninsula, South Africa. *Biological Conservation*, **132**, 183–198.
- Bedrick, E.J. & Tsai, C.L. (1994) Model selection for multivariate regression in small samples. *Biometrics*, **50**, 226–231.
- Beers, T.W., Dress, P.E. & Wensel, L.C. (1966) Aspect transformation in site productivity research. *Journal of Forestry*, **64**, 691–692.
- Beyer, H.L. (2004) *Hawth's Analysis Tools for ArcGIS*. <http://www.spatial-ecology.com/htools> (accessed 16 November 2005).
- Brotons, L., Thuiller, W., Araujo, M.B. & Hirzel, A.H. (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **76**, 378–382.
- Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Dreyer, G.D. (1994) *Element Stewardship Abstract for Celastrus orbiculata*. The Nature Conservancy, Davis, CA.
- Dreyer, G.D., Baird, L.M. & Fickler, C. (1987) *Celastrus-scandens* and *Celastrus-orbiculatus* – comparisons of reproductive potential between a native and an introduced woody vine. *Bulletin of the Torrey Botanical Club*, **114**, 260–264.
- Ellsworth, J.W., Harrington, R.A. & Fownes, J.H. (2004) Survival, growth and gas exchange of *Celastrus orbiculatus* seedlings in sun and shade. *American Midland Naturalist*, **151**, 233–240.
- Foxcroft, L.C., Rouget, M., Richardson, D.M. & Mac Fadyen, S. (2004) Reconstructing 50 years of *Opuntia stricta* invasion in the Kruger National Park, South Africa: environmental determinants and propagule pressure. *Diversity and Distributions*, **10**, 427–437.
- Gelbard, J.L. & Belnap, J. (2003) Roads as conduits for exotic plant invasions in a semiarid landscape. *Conservation Biology*, **17**, 420–432.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M. & Tyler, D. (2002) The national elevation dataset. *Photogrammetric Engineering and Remote Sensing*, **68**, 5–11.
- Gragson, T.L. & Bolstad, P.V. (2006) Land use legacies and the future of southern Appalachia. *Society & Natural Resources*, **19**, 175–190.
- Gustafson, E.J., Lietz, S.M. & Wright, J.L. (2003) Predicting the spatial distribution of aspen growth potential in the upper Great Lakes region. *Forest Science*, **49**, 499–508.
- Hawkins, B.A., Diniz, J.A.F., Bini, L.M., De Marco, P. & Blackburn, T.M. (2007) Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. *Ecography*, **30**, 375–384.
- Higgins, S.I. & Richardson, D.M. (1999) Predicting plant migration rates in a changing world: the role of long-distance dispersal. *American Naturalist*, **153**, 464–475.
- Higgins, S.I., Richardson, D.M. & Cowling, R.M. (2000) Using a dynamic landscape model for planning the management of alien plant invasions. *Ecological Applications*, **10**, 1833–1848.
- Homer, C., Huang, C.Q., Yang, L.M., Wylie, B. & Coan, M. (2004) Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing*, **70**, 829–840.
- Hosmer, D.W. & Lemeshow, S. (2000) *Applied Logistic Regression*, 2nd edn. Wiley, New York.
- Keating, K.A. & Cherry, S. (2004) Use and interpretation of logistic regression in habitat selection studies. *Journal of Wildlife Management*, **68**, 774–789.
- Kincaid, J.A. (2007) Compositional and environmental characteristics of *Tsuga canadensis* (L.) Carr. forests in the southern Appalachian Mountains, USA. *Journal of the Torrey Botanical Society*, **134**, 479–488.
- Lambrinos, J.G. (2006) Spatially variable propagule pressure and herbivory influence invasion of chaparral shrubland by an exotic grass. *Oecologia*, **147**, 327–334.
- Lavergne, S. & Molofsky, J. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3883–3888.
- Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.
- Lonsdale, W.M. (1999) Global patterns of plant invasions and the concept of invasibility. *Ecology*, **80**, 1522–1536.
- MacArthur, R.H. (1972) *Geographical Ecology*. Harper & Row Publishers, New York.
- McGarigal, K. (2007) *Fragstats Documentation*. University of Massachusetts, Amherst, MA.
- McKinney, M.L. (2004) Citizens as propagules for exotic plants: measurement and management implications. *Weed technology*, **18**, 1480–1483.
- McNab, W.H. & Loftis, D.L. (2002) Probability of occurrence and habitat features for oriental bittersweet in an oak forest in the southern Appalachian mountains, USA. *Forest Ecology and Management*, **155**, 45–54.
- McNab, W.H. & Meeker, M. (1987) Oriental bittersweet: a growing threat to hardwood silviculture in the Appalachians. *Northern Journal of Applied Forestry*, **4**, 174–177.
- Mehrhoff, L.J., Silander Jr, J.A., Leicht, S.A. & Mosher, E. (2003) *IPANE: Invasive Plant Atlas of New England*. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT.
- Merriam, R.W. (2003) The abundance, distribution and edge associations of six non-indigenous, harmful plants across North Carolina. *Journal of the Torrey Botanical Society*, **130**, 283–291.
- Miller, J.H. (2003) Nonnative invasive plants of southern forests: a field guide for identification and control. *General Technical Report*. USDA Forest Service, Southern Research Station, Asheville, NC.
- Mitchell, M.S., Lancia, R.A. & Gerwin, J.A. (2001) Using landscape-level data to predict the distribution of birds on a managed forest: Effects of scale. *Ecological Applications*, **11**, 1692–1708.
- Moody, M.E. & Mack, R.N. (1988) Controlling the spread of plant invasions – the importance of nascent foci. *Journal of Applied Ecology*, **25**, 1009–1021.
- NCDA&CS (2007) *New Pest Alerts*. North Carolina Department of Agriculture and Consumer Services, Raleigh, NC <http://www.ncagr.gov/plantindustry/plant/entomology/alert.htm#OR> accessed 29 July 2009.
- NPS (2007) *Great Smoky Mountains: Nature and Science*. National Park Service, Gatlinburg, TN, <http://www.nps.gov/grsm/naturescience/non-natives.htm> accessed 29 July 2009.
- NRC (2002) *Predicting Invasions of Nonindigenous Plants and Plant Pests*. National Academy of Sciences, Washington, DC.
- Patterson, D.T. (1974) *The ecology of oriental bittersweet, Celastrus orbiculatus, a weedy introduced ornamental vine*. PhD thesis, Duke University, Durham, NC, USA.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, **49**, 1373–1379.
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*, 2.4.0 edn. R Foundation for Statistical Computing, Vienna.
- Reineking, B. & Schroder, B. (2006) Constrain to perform: regularization of habitat models. *Ecological Modelling*, **193**, 675–690.
- Ribeiro Jr, P.J. & Diggle, P.J. (2001) geoR: a package for geostatistical analysis. *R-NEWS*, **1**, 15–18.
- SAIN (2005) *Southern Appalachian Information Node*. National Biological Information Infrastructure, Knoxville, TN.
- Skellam, J.G. (1951) Random dispersal in theoretical populations. *Biometrika*, **38**, 196–208.
- Small, M.J., Small, C.J. & Dreyer, G.D. (2005) Changes in a hemlock-dominated forest following woolly adelgid infestation in southern New England. *Journal of the Torrey Botanical Society*, **132**, 458–470.
- USDA NRCS (2007) *The PLANTS Database*. National Plant Data Center, Baton Rouge, LA.
- Venables, W.N. & Ripley, B.D. (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York.

- Walther, G.R. (2004) Plants in a warmer world. *Perspectives in Plant Ecology Evolution and Systematics*, **6**, 169–185.
- Welk, E. (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: purple loosestrife (*Lythrum salicaria*) in North America. *Ecological Modelling*, **179**, 551–567.
- Wilson, J.R.U., Richardson, D.M., Rouget, M., Proches, S., Amis, M.A., Henderson, L. & Thuiller, W. (2007) Residence time and potential range: crucial considerations in modelling plant invasions. *Diversity and Distributions*, **13**, 11–22.
- Zweig, M.H. & Campbell, G. (1993) Receiver-operating characteristic (ROC) plots – a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561–577.

Received 3 April 2009; accepted 4 August 2009
Handling Editor: Jack Lennon

Supporting Information

Additional supporting information may be found in the online version of this article.

Appendix S1. Map of study area and presence/absence data.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.