ISSN 1387-3547, Volume 12, Number 8

This article was published in the above mentioned Springer issue.
The material, including all portions thereof, is protected by copyright;
all rights are held exclusively by Springer Science + Business Media.
The material is for personal use only;
commercial use is not permitted.
Unauthorized reproduction, transfer and/or use
may be a violation of criminal as well as civil law.

Biol Invasions (2010) 12:2413–2427
DOI 10.1007/s10530-009-9652-8

Author's personal copy

ORIGINAL PAPER

The ecological niche and reciprocal prediction of the disjunct distribution of an invasive species: the example of *Ailanthus altissima*

Thomas P. Albright · Hao Chen · Lijun Chen · Qinfeng Guo

Received: 4 May 2009 / Accepted: 10 November 2009 / Published online: 22 November 2009
© Springer Science+Business Media B.V. 2009

Abstract Knowledge of the ecological niches of invasive species in native and introduced ranges can inform management as well as ecological and evolutionary theory. Here, we identified and compared factors associated with the distribution of an invasive tree, *Ailanthus altissima*, in both its native Chinese and introduced US ranges and predicted potential US distribution. For both ranges separately, we selected suites of the most parsimonious logistic regression models of occurrence based on environmental variables and evaluated these against independent data. We then incorporated information from both ranges in a simple Bayesian model to predict the potential US distribution. Occurrence of *A. altissima* in both ranges exhibited a unimodal response to temperature variables. In China, occurrence had negative relationships with topographic wetness and forest cover and positive relationships with precipitation and agricultural and urban land use. In the US, *A. altissima* was associated with intermediate levels of forest cover and precipitation. The Bayesian model identified 58–80% of 10-arc minute grid cells in the conterminous US as containing suitable areas for *A. altissima*. The best model developed from Chinese data applied to the US matched most areas of observed occurrence but under-predicted occurrence in lower probability areas. This discrepancy is suggestive of a broadening of the ecological niche of *A. altissima* and may be due to such factors as less intense competition, increased potency of allelopathy, and novel genotypes formed from multiple introductions. The Bayesian model suggests that *A. altissima* has the potential to substantially expand its distribution in the US.

Keywords Distribution models · Ecological niche · Generalized linear models · Invasive species · Simple Bayes · Tree-of-heaven

Electronic supplementary material The online version of this article (doi:10.1007/s10530-009-9652-8) contains supplementary material, which is available to authorized users.

T. P. Albright
Department of Zoology, University of Wisconsin-Madison, 430 Lincoln Dr., Madison, WI 53706, USA

H. Chen
School of Remote Sensing Information Engineering, Wuhan University, Wuhan, People's Republic of China

L. Chen
National Geomatics Center of China, Beijing, People's Republic of China

Q. Guo
Southern Research Station, US Forest Service, 200 W.T. Weaver Blvd., Asheville, NC 28804, USA

T. P. Albright (✉)
Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Dr., Madison, WI 53706, USA
e-mail: talbrig1@wisc.edu

Introduction

A critical need exists for knowledge of invasive exotic species in their home ranges (Hierro et al. 2005). Too often, more is known about the distribution, population dynamics, ecosystem-level effects, and other species properties in the introduced range, where species may take on an ecological significance greatly exceeding that in their native range. One key element is a species' ecological niche, which characterizes the range of environmental conditions suitable for growth and reproduction. Linking observed distributions of species in their native ranges to ecological niches offers a framework for predicting species distributions (Peterson 2001). Knowledge of ecological niches and if and how they differ among ranges may offer insights into areas of great theoretical interest. The enemy release hypothesis (ERH) (Keane and Crawley 2002), the evolution of increased competitive ability (EICA) hypothesis (Blossey and Notzold 1995), and the novel weapons hypothesis (Callaway and Aschehoug 2000), in which allelochemicals exuded by an invader are especially potent against nonadapted plants in the invaded range, are all associated with increased abundance of a species in an invaded range. It is also possible that these mechanisms might allow a species to expand into a broader or otherwise different ecological niche. Furthermore, the stochastic nature of invasion and the strong selection forces encountered by an incipient invader might alter its ecological niche (Lee 2002). Finally, because empirical models of ecological niches based on distribution information characterize the realized niche (*sensu* Hutchinson 1957), it is also possible that the biotic interactions are sufficiently different in the novel range that the realized niche might be altered despite a stable fundamental niche. The stability of niches inferred from ecological niche modeling thus requires further investigation in a coordinated manner between native and exotic ranges.

Practical concerns further justify studies involving the native range of invasive species. The ability to explain the factors associated with the distribution of invasive exotic plants and predict their distribution spatially is of great importance for prevention, early warning, and management of invasives (Mack et al. 2000). However, despite advances in ecological understanding, geographic databases, computing, and statistical techniques, it remains surprisingly difficult to predict a species' distribution. Among the challenges are: exotic distributions that are far from equilibrium, hampering efforts to compare niches or predict potential distribution based on exotic range occurrences (Welk 2004); both negative and positive biotic interactions (Mitchell et al. 2006); poor distribution data, particularly in the native ranges (Peterson 2003); sampling bias (Kadmon et al. 2004); and challenges associated with the use of presence only data.

Ailanthus altissima (P. Mill.) Swingle (tree-of-heaven) is a dioecious tree in the Simaroubaceae family native to central China and considered invasive in the US (Miller 2003). Mature females bear an abundant crop of samara that are assisted by wind in their dispersal. In addition, *A. altissima* is capable of vigorous clonal reproduction and resprouting when cut (Kowarik 1995). Intolerant of flooding and shade, it is considered an early successional, gap-obligate species and thrives in highly disturbed sites (Knapp and Canham 2000). It has a high tolerance for a variety of airborne and soil pollutants and is thus associated with reclaimed mines and urban areas (Miller 2003). It exudes a variety of allelopathic compounds, which may assist it in competition with other species (Heisey and Heisey 2003). It has been valued by humans for its medicinal properties (Jin et al. 2006) and ability to grow in harsh conditions. First introduced by gardeners in the 1700s and again during the ca. 1849 California gold rush by Chinese laborers, it is widely considered to be an invasive nuisance species in North America (Miller 2003).

In this study, we identified factors associated with the distribution of *Ailanthus altissima* in both its native Chinese range and exotic range in the US and predicted its potential exotic distribution. We addressed two fundamental questions. First, what factors explain the coarse scale distribution of *A. altissima* in China? Because China appears to encapsulate all the major range boundaries of *A. altissima*, allowing for cold, drought, and competition-driven limits to be expressed, we predicted a hump-shaped relationship between occurrence and both precipitation and temperature-related variables. Because of flood intolerance, we predicted that *A. altissima* will be negatively associated with areas of high topographically induced wetness. Cultural value and use can strongly influence ecological and

biogeographic characteristics of species (Li et al. in review.). We expected that *A. altissima* would be positively associated with human settlement and land-use. Because of its association with open areas and gaps, we predicted that it will be less likely to occur in areas of high forest cover.

For the second question, we asked what factors explain the coarse scale distribution of *A. altissima* in the US, and how do these factors differ from those found in China? To the extent that distributional quasi-equilibrium has been achieved, we expected the relationships with climatic and topographic moisture factors to be similar to those found in China. Because of contrasting forest cover dynamics in the US and China (Houghton et al. 1999; Zhang et al. 2000), we expected the relationship between *A. altissima* and forest cover to differ to a large degree. The association with human settlement and land use should be similar, based on the major role cultural and horticultural link between humans and *A. altissima* in both ranges.

After addressing fundamental questions about *A. altissima* in both its native and exotic ranges, we then developed models of the predicted distribution of *A. altissima* in the US. More specifically, is it nearing its potential range limits? Are there occurrences in areas of the US that suggest an expansion of the niche from its native range, perhaps due to ERH, EICA, or novel weapons? Inconsistencies between the environmental characteristics of its Chinese and US ranges may provide insights into these and other aspects of the biology of *A. altissima*. Among our motivations for undertaking predictions was to demonstrate the utility of reciprocal modeling of the distribution and ecological niche of invasive species and the incorporation of native range information as priors in a Bayesian model.

Methods

Study areas

The United States and China are two of the largest and most diverse countries in the world, with biomes ranging from tropical forest to tundra (in a contiguous land mass in China). Several factors make them extremely well suited for comparative phytogeographical studies. Strong floristic similarities between East Asia and North America have long been recognized (Gray 1846; Guo 1999). Historically isolated, the recent surge in trade and travel between the regions, which was especially rapid between the US and China, has led to a large intentional and unintentional interchange of species (Yan et al. 2001). In addition to benefiting from their geographic extent and diversity, studies in China and the US are also facilitated by deep botanical traditions and extensive herbarium collections.

Datasets

We used a combination of physical herbarium records, including three of the largest collections of Chinese specimens, and Internet databases in the US and China to develop a geographic database of occurrence records of *A. altissima* and other species native to one region but introduced in the other (Chen et al. 2007; see Acknowledgements). We selected records of spontaneous (i.e., not planted) occurrences that contained coordinates, toponyms, or other location information permitting the assignment of geographic coordinates with a precision of <10 km. The resulting database contained 112 occurrence records in China (Mainland and Taiwan) and 244 in the US (Fig. 1a, b). The number of records included is well above a 20–100-record critical threshold for modeling suggested by Loiselle et al. (2008) and the variety of records included allowed us to sample along the entire climatic gradient of the species' occurrence. Because absence information is rarely available for large areas and would be difficult to ascertain in sampling grains as coarse as used in our study (10 arc minute $\sim$18 km $\times$ 18 km), we instead drew a random sample of 1,000 locations (also known as *pseudoabsences*; Fielding and Bell 1997) in each country in order to characterize the area "available" to the species in each region (Fig. 1c, d).

We selected a variety of variables suitable for characterizing the ecological niche of plant species at coarse scales and answering the questions posed above (Table 1). The Worldclim dataset (Hijmans et al. 2005) is a topographically guided interpolation of a number of climatic variables measured at weather stations around the world. The compound topographic index, an indicator of topographic wetness based on flow accumulation and local terrain slope data, was obtained from the Hydro1 k dataset

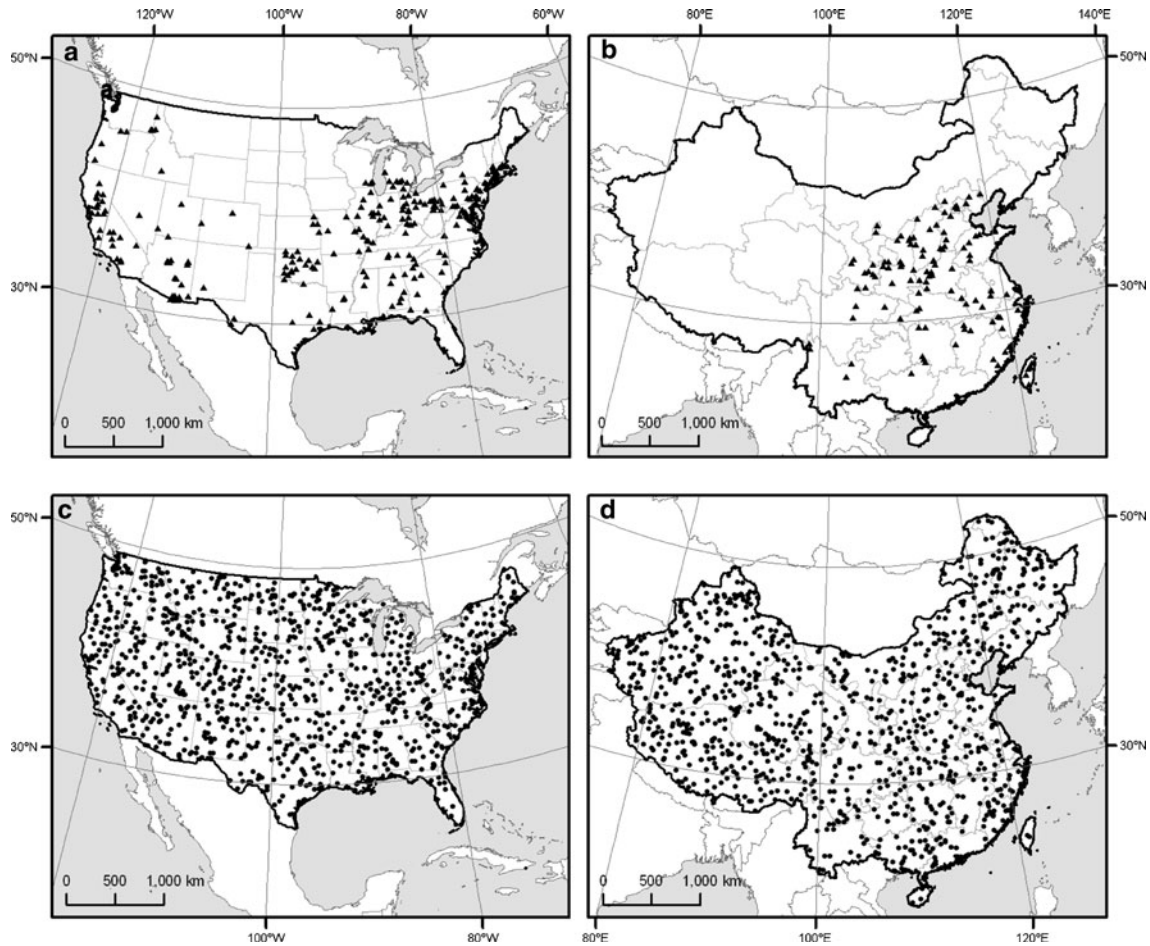

Fig. 1 Locations of occurrence records for *Ailanthus altissima* in the US (**a**) and China (**b**). Randomly placed pseudo-absence locations in the US (**c**) and China (**d**). Figure generated with ArcGIS 9.2

(USGS 1996). Forest cover was obtained from the UN Food and Agriculture Organization/US Geological Survey Forest Resources Assessment dataset, which included estimates of forest cover fraction within 1 km × 1 km pixels (Zhu and Waller 2003). The fraction of agriculture and urban land use in 10 arc minute cells was calculated from a 1 km × 1 km resolution global land cover classification (Loveland et al. 2000). All data were aggregated into 10 arc minute cells by averaging the pixel values of the finer scale cells. While each of these products was available at finer resolutions than 10 arc minutes, we selected this resolution because of our broad-scale focus and concerns about data quality for interpolated climate data over sparsely sampled areas in mountainous terrain (Hijmans et al. 2005). Finally, as necessary, we added an offset to any covariates with ranges that included negative values to ensure that all covariates were positive and thus appropriate for modeling with a quadratic term. Values of these covariates were then extracted using the GIS ERDAS Imagine 9.0 (ERDAS, Inc., Atlanta, GA, USA) by overlaying the occurrence locations and pseudoabsences in both native and exotic ranges.

Analysis

Ideally, several a priori models would be chosen, based on the concept of multiple working hypotheses. However, with ecological niche modeling there is usually insufficient information to identify a small number (e.g., <12) of candidate models in this fashion. Instead, modeling approaches that are heavily driven by data are often employed (Elith et al. 2006).

 Author's personal copy

Table 1 Variables used in modeling of *Ailanthus altissima* occurrence in China and the US

Variable	Short name	Range/units	Native grain
Minimum temperature of the coldest month[a]	mintemp	−38.9 to 17.0°C	10′ from 14,835 stations
Maximum temperature of warmest month[a]	maxtemp	0.54 to 44.8°C	10′ from 14,835 stations
Mean annual temperature[a]	meantemp	−13.8 to 25.5°C	10′ from 24,542 stations
Annual precipitation[a]	precip	0.014 to 4.01 m	10′ from 47,554 stations
Precipitation in the driest month[a]	minprecip	0.000 to 0.19 m	10′ from 47,554 stations
Compound topographic index (CTI)[b]	CTI	Unitless	1 km
Percent forest cover[c]	Forest	0 to 100%	1 km
Agriculture and Urban land uses[d]	AgUrb	0 to 100%	1 km cells into 18 km cells

Sources [a] Worldclim (Hijmans et al. 2005), [b] HYDRO1K (USGS 1996), [c] FAO/USGS (Zhu and Waller 2003), [d] Global Land Cover Characterization (Loveland et al. 2000)

For analyses based on generalized linear models, this may include testing all possible models or some kind of stepwise selection routine. On one hand, these approaches may yield models that describe a dataset very effectively, but on the other hand, they are prone to overfitting, which has particularly deleterious consequences if the model is to be applied outside of the dataset used to generate it (Miller et al. 2004). In this case, we had several hypotheses related to specific covariates, but no specific hypothetical models to evaluate. In order to balance the concerns of overfitting with the risk of failing to capture meaningful relationships, we employed a hierarchical approach to model selection, guided by our research objectives, general principles of broad-scale ecological niche modeling (Austin 2002; Guisan and Thuiller 2005; Peterson 2001), and knowledge of the ecological and physiological characteristics of *A. altissima*. This hierarchical framework allowed us to consider 16 different variables (including square terms plus intercept) and still maintain an events-per-variable ratio >10 (Peduzzi et al. 1996) in all cases by limiting the number considered at any one time.

In this approach, we began by considering model(s) based on variables pertaining to climate (Table S1), which is usually considered to be a first order determinant of plant distributions and of ecological regions (Brown and Lomolino 1998; Omernik 1987). Using generalized linear regression with a logistic link (logistic regression) in R (Hosmer and Lemeshow 2000; R Development Core Team 2006), we examined precipitation-only models and obtained the value of Akaike's Information Criterion (AIC) (Akaike 1974; Burnham and Anderson 2002). Models with lower AIC, a measure of fit and model complexity, minimize Kulback-Leibler (K-L) information loss relative to the full reality conveyed in the data (Burnham and Anderson 2002). Differences between models based on the same dataset (i.e., no missing values, no transformations of the response variable) may be directly compared and ranked by calculating for each model Δ_i, which is an indicator of loss of information relative to the best performing model, $\mathrm{AIC}_{\min}$:

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$$

We also calculated Akaike weights, W_i, which maintain the same model ranking as Δ_i, but provide a more meaningful scaling in that W_i is proportional to the likelihood that a given model is the best considered, (Burnham and Anderson 2002):

$$W_i = \frac{\exp(-1/2\Delta_i)}{\sum_{j=1}^{R} \exp(-1/2\Delta_j)}.$$

Using the same process, we added temperature variables to the best precipitation-only model(s) in order to arrive at the best climate-only model(s). We then evaluated a suite of models composed of the best climate-only model(s) plus several combinations of compound topographic index, forest cover, and agricultural and urban land use (Table S1). The use of quadratic terms was limited in this last stage in order to maintain a sufficient ratio of events to variables (Peduzzi et al. 1996) and avoid overfitting. At each step of this process, if no single model had a value of $W_i > 0.90$, we retained multiple competitive models until this threshold was reached by the sum of individual model weights. In these cases, support for any single model is equivocal and the retention of

multiple models allows them to continue to compete as additional variables are considered. Similarly, if, at the end of the selection process, no single clear favorite emerged, the competing models were considered competitive and we calculated an average model weighted according to the W_i of each candidate (again, including models required for $W > 0.90$):

$$\hat{\bar{\theta}} = \sum_{i=1}^{R} W_i \hat{\theta}_i$$

where $\hat{\bar{\theta}}$ is the average parameter estimate. Our approach acknowledges that there may be support for more than one model instead of adopting a rigid requirement for only one best model. A practical advantage of model averaging is that, by acknowledging model selection uncertainty, parameter estimates can be made more robust due to a reduction of model selection bias (Johnson and Omland 2004).

Resulting models can be used to estimate the relative probability of *A. altissima* occurrence, *w(x)*, (Manly et al. 2002):

$$w(x) = \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots)$$

However, due to the lack of true absence data, these models cannot be used to calculate absolute occurrence probabilities. We mapped model output, applying the models developed in the native range on the environmental covariates of the exotic range as well. We then repeated the entire process to evaluate, select, and apply models based on the exotic range data.

Model evaluation

Verification and validation of models using presence-only data presents an added challenge over situations involving presence/absence or abundance data (Anderson et al. 2003; Boyce et al. 2002). For the finalist models, we performed initial model evaluation by comparing model output to the locations of occurrence in the range they were built upon. Because of the ambiguity of pseudo-absences, it is not possible to ascertain type I (commission) errors, rendering questionable the use of error matrix analysis, Cohen's κ (1960), and area under the receiver operating characteristic curve analysis (Zweig and Campbell 1993). Johnson et al. (2006) proposed a method inspired by Boyce et al. (2002) for evaluating wildlife habitat models based on use-availability data by assessing the degree to which the model output is proportional to probability of use. Used in resource selection functions (Manly et al. 2002), use-availability data are analogous to presence only ecological niche models. Guided by this approach, we binned model output calculated on the pseudo-absences according to deciles of $w(x)$. We then calculated the expected number of occurrences, N_i, in each bin according to: $N_i = N \times w(x_i) / \sum_j w(x_j)$ (modified from Johnson et al. 2006)where N is the number of occurrences in the entire range, $w(x_j)$ is the median relative probability value in bin i. Using deciles and pixels of a uniform size, the area occupied by each bin on a map is equal, so no area normalization is required. Comparison of the relationship between observed and expected occurrences in each bin allows for assessment of proportionality to probability of "use". We repeated this process for the range that was not used to develop the model in order to evaluate its performance in a novel set of conditions.

Prediction of potentially suitable areas

In order to predict potentially suitable areas in the US for *A. altissima*, we used a simple Bayesian approach incorporating informative priors from Chinese data. Using the model structure from the best climate model identified in the US (above), we used parameters estimated from the equivalent model in China to supply prior means and precisions. Using Markov chain Monte Carlo simulations in the software WinBUGS (Spiegelhalter et al. 1999), we generated two chain 2 to 1 million iterations, excluding the first 100,000 for model "burn in." (For WinBUGS code and output, see Appendix 1, on-line). Based on the resulting posterior estimates, we then mapped relative suitability. Without absence information, the intercept term estimated above is not interpretable as it would be in standard logistic regression, obscuring the distinction between "suitable" and "unsuitable" areas. In order to distinguish potentially suitable from unsuitable areas, we produced two maps based on scenarios using thresholds obtained from known presence locations. For the first (liberal) scenario we distinguished suitable from unsuitable by using the lowest prosterior logit value as a cutoff. In the

second (conservative) scenario, we used the top 95% of logit values from presence locations in order to reduce the influence of outlier populations that may not persist or otherwise fail to characterize suitable locations.

Results

The initial stage of the model selection hierarchy for *A. altissima* occurrence in China yielded two competitive models, including linear and quadratic terms, respectively for mean annual precipitation (Table 2). However, when temperature variables were added, one clear favorite climate-only model emerged, which included a linear precipitation term and quadratic terms for minimum and maximum temperatures. Examination of the parameter estimates and graphing the resulting model reveals a positive relationship between occurrence and precipitation and peak in occurrence probability at minimum temperatures of −9°C and maxima at 27°C (Table S2). Of the three non-climatic variables examined in this study, compound topographic index, forest cover, and agricultural and urban land use, each considerably improved upon the best climate-only models in China (Table 3). In concert, these covariates retained importance, with the best model examined including all three. A slightly weaker model included compound topographic index and agricultural and urban land use, but not forest cover (Table S2). Whether added individually to the best climate model or together, *A. altissima* was negatively related to both forest cover and the topographically moist sites identified by high values in the compound topographic index. On the other hand, there was a consistent positive relationship between agricultural and urban land use and *A. altissima*.

As with the Chinese dataset, the initial stage of the model selection hierarchy for *A. altissima* occurrence in the US yielded two competitive models, including

Table 2 China and US precipitation-only and climate model comparisons

Name	Variables	China			US		
		AIC*i*	Δ*i*	*Wi*	AIC*i*	Δ*i*	*Wi*
precip1	precip	**686.4**	**4.2**	**0.111**	**1,158.0**	**1.1**	**0.366**
precip2	precip + precip^2	**682.3**	**0.0**	**0.889**	**1,156.9**	**0.0**	**0.634**
precip3	minprecip	699.2	16.9	0.000	1,180.5	23.6	0.000
precip4	minprecip + minprecip^2	701.1	18.9	0.000	1,182.5	25.6	0.000
clim1.1	[precip1] + meantemp	669.7	134.2	0.000	1,171.5	151.4	0.000
clim1.2	[precip1] + meantemp + meantemp^2	559.9	24.4	0.000	1,044.1	24.0	0.000
clim1.3	[precip1] + mintemp	671.7	136.3	0.000	1,163.6	143.5	0.000
clim1.4	[precip1] + mintemp + mintemp^2	552.6	17.2	0.000	1,079.6	59.5	0.000
clim1.5	[precip1] + maxtemp + mintemp	670.5	135.1	0.000	1,165.5	145.4	0.000
clim1.6	[precip1] + maxtemp + maxtemp^2 + mintemp	646.8	111.4	0.000	1,127.8	107.7	0.000
clim1.7	[precip1] + maxtemp + mintemp + mintemp^2	550.9	15.4	0.000	1,080.7	60.6	0.000
clim1.8	[precip1] + maxtemp + maxtemp^2 + mintemp + mintemp^2	**535.5**	**0.0**	**0.999**	1,062.3	42.2	0.000
clim2.1	[precip2] + meantemp	670.8	135.3	0.000	1,128.7	108.6	0.000
clim2.2	[precip2] + meantemp + meantemp^2	551.5	16.0	0.000	**1,020.7**	**0.6**	**0.425**
clim2.3	[precip2] + mintemp	673.2	137.7	0.000	1,109.0	88.9	0.000
clim2.4	[precip2] + mintemp + mintemp^2	554.6	19.1	0.000	1,034.4	14.3	0.000
clim2.5	[precip2] + maxtemp + mintemp	671.9	136.4	0.000	1,110.6	90.5	0.000
clim2.6	[precip2] + maxtemp + maxtemp^2 + mintemp	648.8	113.4	0.000	1,076.7	56.6	0.000
clim2.7	[precip2] + maxtemp + mintemp + mintemp^2	552.8	17.3	0.000	1,033.4	13.3	0.001
clim2.8	[precip2] + maxtemp + maxtemp^2 + mintemp + mintemp^2	537.2	1.7	*	**1,020.1**	**0.0**	**0.574**

Competitive models are highlighted in bold

* This model was not retained because addition of a parameter increased AIC

Table 3 Full model comparison for US

Name	Variables	AIC*i*	Δ*i*	*Wi*
China				
full1.8.1	[climate1.8] + CTI	532.6	14.5	0.001
full1.8.2	[climate1.8] + CTI +CTI^2	534.5	16.5	0.000
full1.8.3	[climate1.8] + forest	531.2	13.1	0.001
full1.8.4	[climate1.8] + forest + forest^2	530.6	12.6	0.001
full1.8.5	[climate1.8] + AgUrb	530.0	12.0	0.002
full1.8.6	[climate1.8] + AgUrb + AgUrb^2	531.3	13.3	0.001
full1.8.7	[climate1.8] + CTI + forest	523.6	5.6	0.048
full1.8.8	[climate1.8] + CTI + AgUrb	**521.1**	**3.1**	**0.166**
full1.8.9	[climate1.8] + forest + AgUrb	529.5	11.5	0.002
full1.8.10	[climate1.8] + CTI + forest + AgUrb	**518.0**	**0.0**	**0.777**
US				
full2.2.1	[climate2.2] + CTI	1,022.0	14.0	0.001
full2.2.2	[climate2.2] + CTI +CTI^2	1,022.8	14.8	0.001
full2.2.3	[climate2.2] + forest	1,017.5	9.5	0.007
full2.2.4	[climate2.2] + forest + forest^2	**1,012.3**	**4.3**	**0.100**
full2.2.5	[climate2.2] + AgUrb	1,020.2	12.2	0.002
full2.2.6	[climate2.2] + AgUrb + AgUrb^2	1,019.6	11.6	0.003
full2.2.7	[climate2.2] + CTI + forest	1,017.8	9.8	0.006
full2.2.8	[climate2.2] + CTI + AgUrb	1,021.4	13.4	0.001
full2.2.9	[climate2.2] + forest + AgUrb	1,019.5	11.5	0.003
full2.2.10	[climate2.2] + CTI + forest + AgUrb	1,019.5	11.5	0.003
full2.8.1	[climate2.8] + CTI	1,018.2	10.2	0.005
full2.8.2	[climate2.8] + CTI + CTI^2	1,018.9	10.9	0.004
full2.8.3	[climate2.8] + forest	1,022.0	14.0	0.001
full2.8.4	[climate2.8] + forest + forest^2	**1,008.0**	**0.0**	**0.857**
full2.8.5	[climate2.8] + AgUrb	1,022.2	14.2	0.001
full2.8.6	[climate2.8] + AgUrb + AgUrb^2	1,021.2	13.2	0.001
full2.8.7	[climate2.8] + CTI + forest	1,020.1	12.1	0.002
full2.8.8	[climate2.8] + CTI + AgUrb	1,020.0	12.0	0.002
full2.8.9	[climate2.8] + forest + AgUrb	1,023.9	15.9	0.000
full2.8.10	[climate2.8] + CTI + forest + AgUrb	1,021.9	13.9	0.001

Competitive models are highlighted in bold

linear and quadratic terms, respectively for mean annual precipitation (Table 2). The addition of temperature variables resulted in two evenly matched models adding either quadratic mean temperature or quadratic minimum and maximum temperature variables. Calculating a model average from these two shows that peak occurrence was found in locations with approximately 890 mm/year of precipitation, 13°C mean annual temperature, −1°C minimum temperature, and 29°C maximum temperature (Table S2). While the benefits of additional, non-climatic variables in the US were decidedly mixed, there was clear support for adding quadratic forest cover terms to both of the climate finalist models (Table 3). According to the final averaged model, peak probability of *A. altissima* occurred in areas that were 40% forested (Table S2).

Implementing the best native and exotic range models in a GIS resulted in relative probability maps of both Chinese and US ranges (Figs. 2, 3). Comparisons of observed versus expected occurrences in each decile of $w(x)$ show that, in all cases, relative probability is proportional to the expected occurrences (Fig. 4) in the range upon which the model was developed. When evaluated against occurrence data in a different range, model fit was poorer.

 Author's personal copy

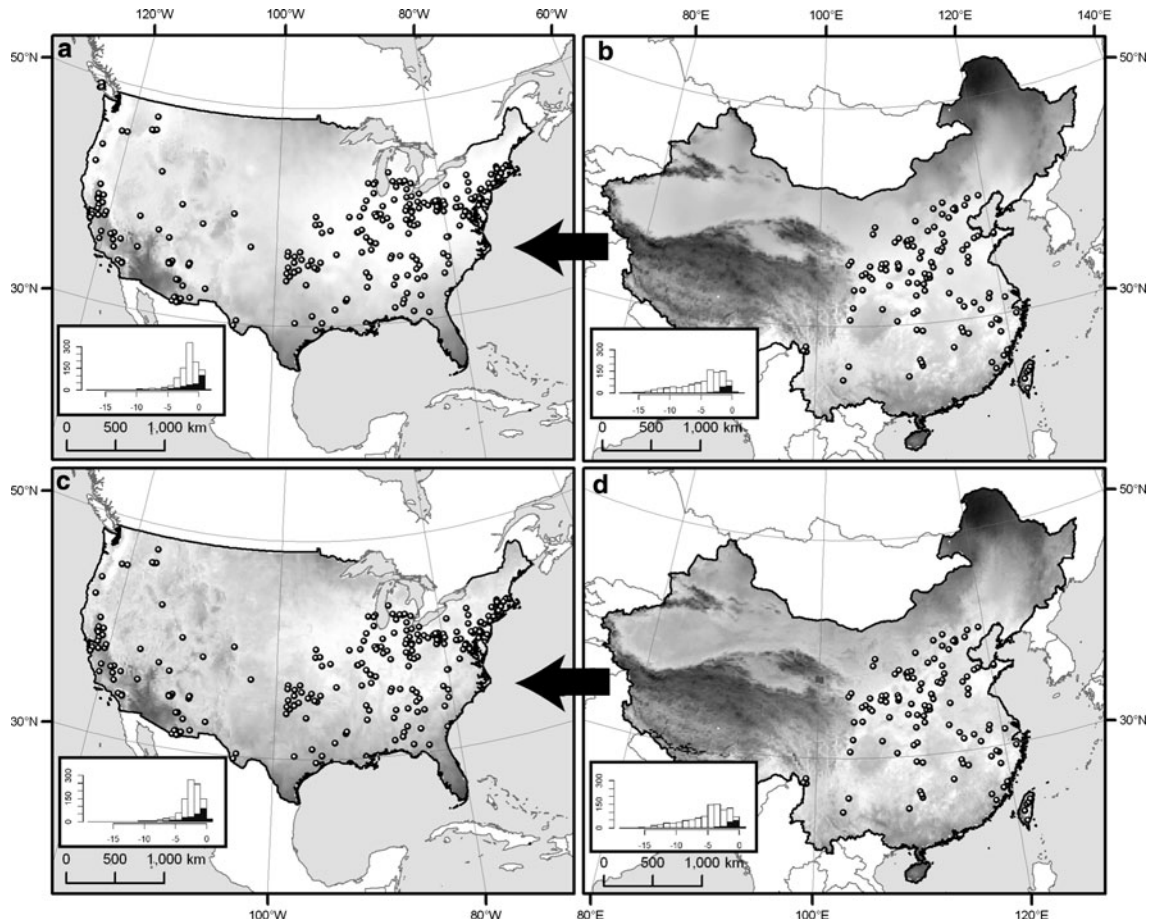

Fig. 2 Relative occurrence probability of *Ailanthus altissima* based on models trained on environmental and occurrence/pseudoabsence data from China for the best climate only model projected onto the US (**a**) and China (**b**) and the best full (climate + other environmental variables) model projected onto the US (**c**) and China (**d**). *Inset* figures are histograms showing distribution of relative occurrence probability among locations with pseudoabsence points (*white*) and occurrence records (*black*). Figure generated with ArcGIS 9.2 and R 2.5.1

Models derived in the US tended to underpredict occurrence for intermediate values of $w(x)$ and overpredict for high values in China (Fig. 4c, d). Both the climate-only and full Chinese models resulted in predictions in the US that were largely in proportion with observations, although the fit was slightly better for the climate-only model. Both the Chinese climate-only and full models underpredicted occurrence at low values of $w(x)$ in the US (Fig. 4a).

Posterior model parameters estimated using a simple Bayesian approach had smaller standard error estimates and were intermediate between the coefficients estimated individually for China and the US using GLMs (Table S3). Prediction based on the posterior parameter estimates was sensitive to choice of threshold for suitability, with 58.3% of the United States predicted to be suitable using the conservative threshold and 80.4% suitable using the liberal threshold (Fig. 5).

Discussion

A broad suite of climate, topography, and land-use-related predictors were most effective in explaining the distribution of *A. altissima* within both its native China and exotic range in the US. In both ranges, *A. ailanthus* is more likely to occur in areas of intermediate temperature conditions (Table S2). While this is not a particularly stunning finding

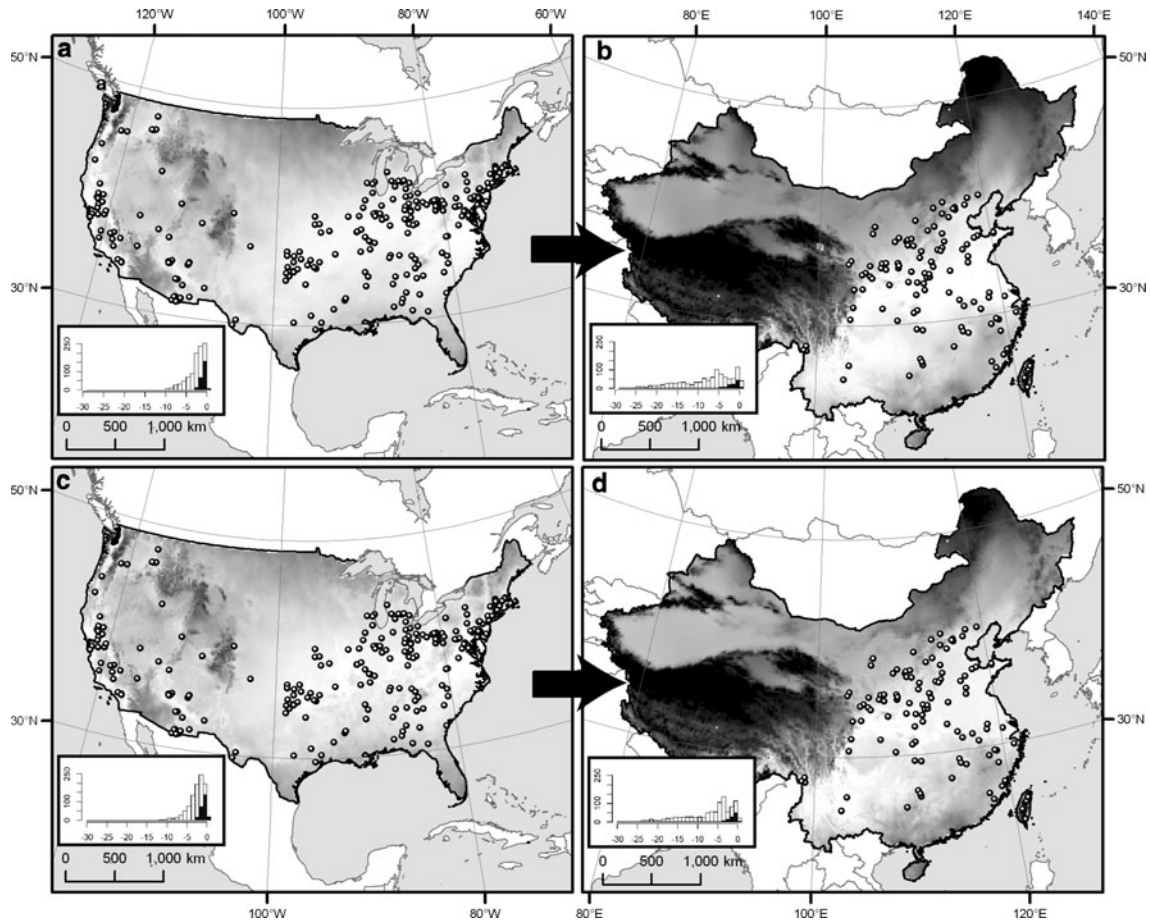

Fig. 3 Relative occurrence probability of *Ailanthus altissima* based on models trained on environmental and occurrence/pseudoabsence data from the US for the best climate-only model projected onto the US (**a**) and China (**b**) and the best full (climate + other environmental variables) model projected onto the US (**c**) and China (**d**). *Inset* figures are histograms showing distribution of relative occurrence probability among locations with pseudoabsence points (*white*) and occurrence records (*black*). Figure generated with ArcGIS 9.2 and R 2.5.1

biologically, it does indicate that the sampled climatic conditions are sufficiently broad to capture upper and lower range limits of the species (Austin 2007). Furthermore, in the US, the unimodal response to climatic factors suggests that populations of *A. altissima* exist proximal to moisture- and temperature-related limits to its distribution. Although in western China, the large area of arid, often high-elevation regions of Xizang (Tibet), Qinghai, and Xinjiang may have been highly influential in producing a positive, rather than hump-shaped, relationship overall between occurrence of *A. altissima* and precipitation. Regarding the selection of moisture-related variables, the consistently stronger relationship in both ranges between mean annual, rather than minimum monthly, precipitation and occurrence may be related to the tree's ability to tolerate drought of several months duration (Trifilo et al. 2004; Table 2). The longer-term annual mean may thus be more meaningful to the species' requirements than the amount of rain in the driest month of the year.

Relationships between occurrence and non-climatic variables were stronger in China than in the US (Table 3). *Ailanthus altissima*'s apparent avoidance of areas combining high flow accumulation with low slope (i.e., CTI) offers confirmation of flood intolerance (Knapp and Canham 2000; Table S2). Its positive association with agricultural and urban land use in its native range similarly supports expectations regarding human affinities for this culturally important

 Author's personal copy

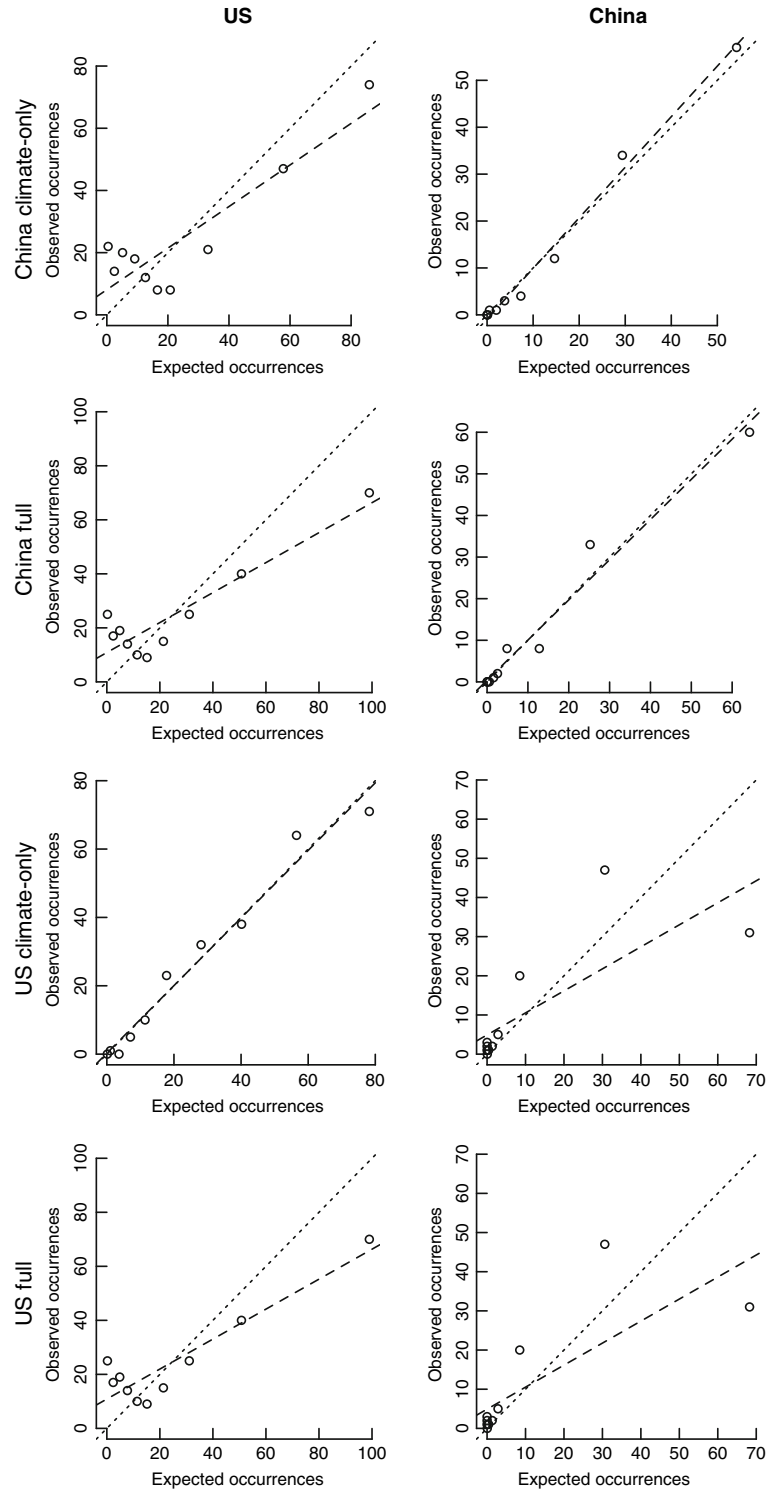

Fig. 4 Comparisons of observed (reference data) versus expected (based on model) numbers of *A. altissima* occurrences in each decile of the prediction range. The *dashed line* is 1:1 and the *dotted line* is a simple linear model fitted to the relationship. The dataset and geographic range on which the models were generated is indicated on the *left* and dataset used for evaluation appears in the *column heading*. Figure generated with R 2.5.1

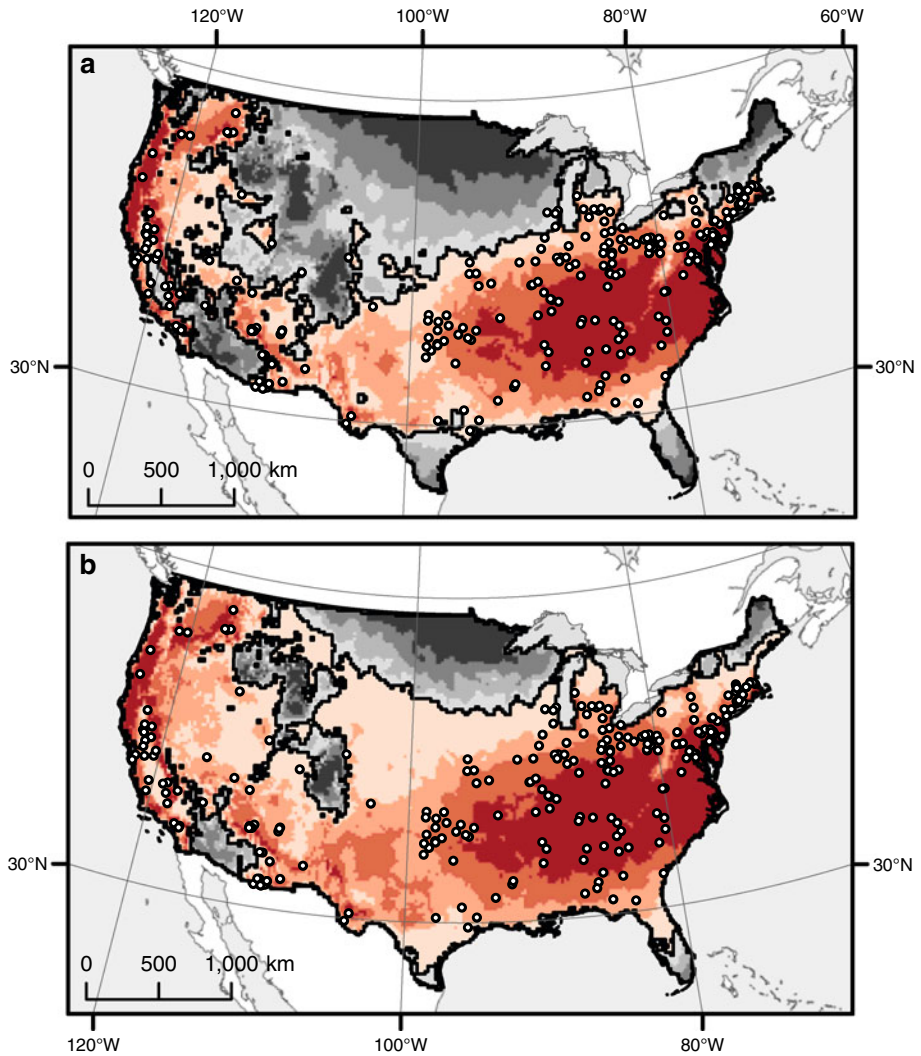

Fig. 5 Relative environmental suitability for *Ailanthus altissima* estimated by a simple Bayesian model based on distribution information from the US and prior parameters derived from China. In the maps, *point symbols* represent occurrence records and suitable area (*inside bold line*) are defined as areas with relative probability (modeled logit values using posterior parameter estimates) equal to or greater than the minimum value among all US occurrence records (**a**) or among the top 95% of US occurrence records (**b**). Both suitable and unsuitable areas (*outside bold line*) are divided into quartiles of relative probability. Figure generated with ArcGIS 9.2

species in China. The apparent negative relationship with forest cover is also consistent with expectations of scarcity in the most heavily forested areas. While there was a strong preference for areas with intermediate forest cover in the US, the relationships of *A. altissima* to both CTI and agricultural and urban land use weakened in the US, relative to China. The explanation for this is not entirely clear, but may be related to insufficient introduction opportunities for such relationships to become apparent in a novel range.

Predicted relative likelihood in China is highly proportional to observed occurrences in the best climate and averaged full models (Fig. 4). In both, there are virtually no occurrences in the low probability areas of either map (Fig. 2b, d). The addition of the non-climatic covariates appears to have only minor effects on the predicted map, other than

 Author's personal copy

increasing the spatial heterogeneity due to their finer scale of variation. There are few large tracts of predicted suitable habitat that lack occurrence records of *A. altissima*. Two exceptions are in the mountainous region of Northern Yunnan and Southern Sichuan provinces and an isolated area (relative to the rest of China) on the south slope of the Himalayas. Possession of this latter area is disputed and it is currently administered by India. Due to the remoteness of both of these regions, they may be poorly represented in herbarium collections.

Projections of the best climate-only and full *A. altissima* occurrence probability models developed in China onto the US yield maps that are in general agreement with occurrence records (Fig. 2a, c). One notable exception is the Central Valley of California, an area of intensive agriculture and irrigation. This area is also near a locus of what were likely multiple and genetically diverse introductions of *A. altissima* associated with Chinese miners and laborers in the ca. 1849 gold rush (Greer and Aldrich 2005). The genetic diversity and potential for novel hybrids (Lavergne and Molofsky 2007), facilitated by irrigation, may allow numerous populations of *A. altissima* to persist in an area that would otherwise appear to be too arid. At the other extreme, the Pacific Northwest is an area of very high predicted occurrence probability. Yet, this area has very few records of occurrence. Two explanations are worth considering. One is that this region simply has not been subjected to enough propagule pressure of *A. altissima* and, given sufficient time, will begin to see more occurrences. An alternative explanation is that model performance is simply poor in this area, which has a Marine West Coast climate that has no analog in China (Williams et al. 2001). Another deviation involves a number of occurrences near the Gulf Coast of the Southeastern US. This warm and moist region is not predicted to be highly suitable for *A. altissima*. Southern range limits are often determined only indirectly by temperature, with competition and predation being more proximal limiting factors (MacArthur 1972). This apparent protrusion of the range southward is suggestive of an explanation based on ERH, EICA, or novel weapons. Comparing the best climate-only and full prediction maps, the main difference appear to be the increased detail of the model that considers non-climatic factors. In particular the Mississippi alluvial plain (or "delta") is a flat area with high flow accumulation that sees a reduction in probability going from the climate-only to the full model, which includes CTI.

Naturally, the climate-only and full models developed using the US data set are superior to the projected models based on Chinese data at the task of identifying occurrences in the US (Figs. 2a, c, 3a, c, 4). The projected maps succeed in capturing the Central Valley and also have much lower probabilities in the high rainfall areas of the Pacific Northwest (owing to the inclusion of the quadratic precipitation term). Projecting the US-based relative probability models onto China results in a map that appears generally successful in identifying the native range of *A. altissima*, with the primary exception being a failure to identify upland area of Taiwan as suitable (Fig. 3b, d).

Some advantages of Bayesian frameworks for ecological studies have been well articulated, including their analytic flexibility (e.g., accounting for hierarchies in data) and transparency (e.g., quantification of errors) (Clark 2005). We argue that informative priors in Bayesian models can provide a simple and currently underexploited enhancement useful for ecology and especially invasion biology. Many studies related to the distribution or ecological niche of invasive species focus on either the native (Peterson 2003) or invaded ranges (Albright et al. 2009). Informative Bayesian priors offer a way of incorporating information from both ranges, each being weighted by the amount of support provided by the respective datasets. In our case, we have used a simple Bayesian framework to predict the ultimate distribution of *Ailanthus altissima* in the US by using priors from the native range and "updating" this with data from the invaded range. The resulting prediction (Fig. 4) appears reasonable but is perhaps most useful in ensuring that dispersal limitation does not unduly limit the predicted range of invader, which is a risk with models built solely on the invaded range. Furthermore, unlike models built solely on the native range, the occurrences in novel environmental space (e.g., California) also influence the prediction. Because the model predicts a future distribution, we cannot validate it with the present dataset. However, future studies through simulations or field data, may investigate this aspect.

This study has illustrated the value of reciprocal modeling of the distribution and ecological niche of an invasive species. It has also shown that characterization of the broad scale ecological niche of a species in its native range can produce a model that successfully predicts exotic range distribution, although areas lacking exotic occurrences may also be included. In this case, the model includes temperature and moisture variables as well as topographic moisture, forest cover, and agricultural and urban land use. Climate variables appear to be slightly more robust than non-climatic terrain and landscape composition variables when making predictions in novel ranges. The resulting posterior model obtained from a simple Bayes approach offers a reasonable estimate of areas of potential distribution for *A. altissima* assuming current environmental conditions. How changes in climatic means and extreme events (Easterling et al. 2000) will affect the distribution of *A. altissima* and other exotic species is an emerging question. The model selection framework is appropriate for ecological niche modeling when inference about specific relationships between covariates and a species is desired. While often neglected, some type of evaluation of model quality is crucial. The observed versus expected comparison employed in this study is simple to calculate and should be considered in future efforts. Studies involving other species are needed in order to evaluate the generality of these findings.

Acknowledgments Research funded by United States Geological Survey grant (144-ME50). Specimen data was accessed from the following herbaria and databases: MO, ARIZ, ASU, AUA, calflora, CM, COLO, CRISIS, FTG, ILLS, INVADERS, IPANE, KUN, LL, LSU, MIL, MISS, MO, MOR, MSC, NY, OKL, OS, OSC, PH, PUL, SAT, SJSU, TAES, TAMU, TEX, UNA, USNM, WIS, JEPS, WTU, UCR, MONTU, F, NAU, UMO, SUWS, UWPL, WSP, UA, USGB, CSUC, QUE, UNB, UBC, NSPM, ALTA, SASK, PE, KUN. The assistance from these institutions is greatly appreciated. We especially thank curators and staff at Missouri Botanical Gardens, The Field Museum, and The Institute of Botany of the Chinese Academy of Sciences for exceptional assistance during our visits there. TPA thanks N. Keuler for statistical consultation, M. Turner and V. Radeloff for comments and support in completing this work, and Z. Zhu for facilitating this collaboration. QG was also supported by National Science Foundation grant (DEB-0640058) and USDA. Comments by two anonymous reviewers improved this manuscript. Author contributions: TPA, QFG, and LJC conceived the study; TPA, QFG, LJC, and CH collected the data; TPA, LJC, and CH analyzed the data; and TPA led the writing.

References

Akaike H (1974) New look at statistical-model identification. IEEE Trans Automat Contr AC 19:716–723

Albright TP, Anderson DP, Keuler NS, Pearson SM, Turner MG (2009) The spatial legacy of introduction: *Celastrus orbiculatus* in the southern Appalachians. J Appl Ecol (in press)

Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol Modell 162:211–232

Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol Modell 157:101–118

Austin M (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol Modell 200:1–19

Blossey B, Notzold R (1995) Evolution of increased competitive ability in invasive nonindigenous plants—a hypothesis. J Ecol 83:887–889

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecol Modell 157:281–300

Brown JH, Lomolino MV (1998) Biogeography. Sinauer Associates, Sunderland, p 691

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York, p 488

Callaway RM, Aschehoug ET (2000) Invasive plants versus their new and old neighbors: a mechanism for exotic invasion. Science 290:521–523

Chen H, Chen LJ, Albright TP (2007) Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: a case of ragweed (*Ambrosia artemisiifolia* L.) distribution in China. Chin Sci Bull 52:1223–1230

Clark JS (2005) Why environmental scientists are becoming Bayesians. Ecol Lett 8:2–14

Cohen JE (1960) A coefficient of agreement for nominal scales. Educ Psychol Measur 20:37–46

Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO (2000) Climate extremes: observations, modeling, and impacts. Science 289:2068–2074

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Gray A (1846) Analogy between the flora of Japan and that of the United States. Am J Sci Arts II 2:135–136

Greer G, Aldrich PR (2005) Genetics and biochemical variation of US Ailanthus altissima populations: a preliminary discussion of a research plan. 16th US Department of

Agriculture interagency research forum on gypsy moth and other invasive species. USDA Forest Service, Northeastern Research Station, Annapolis

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8:993–1009

Guo Q (1999) Ecological comparisons between Eastern Asia and North America: historical and geographical perspectives. J Biogeogr 26:199–206

Heisey RM, Heisey TK (2003) Herbicidal effects under field conditions of *Ailanthus altissima* bark extract, which contains ailanthone. Plant Soil 256:85–99

Hierro JL, Maron JL, Callaway RM (2005) A biogeographical approach to plant invasions: the importance of studying exotics in their introduced and native range. J Ecol 93:5–15

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Hosmer DW, Lemeshow S (2000) Applied logistic regression. Wiley, New York, p 373

Houghton RA, Hackler JL, Lawrence KT (1999) The US carbon budget: contributions from land-use change. Science 285:574–578

Hutchinson GE (1957) Concluding remarks. Population studies: animal ecology and demography. Cold Spring Harb Symp Quant Biol 22:415–427

Jin MH, Yook J, Lee E, Lin CX, Quan Z, Son KH, Bae KH, Kim HP, Kang SS, Chang HW (2006) Anti-inflammatory activity of *Ailanthus altissima* in ovalbumin-induced lung inflammation. Biol Pharm Bull 29:884–888

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends Ecol Evol 19:101–108

Johnson CJ, Nielsen SE, Merrill EH, McDonald TL, Boyce MS (2006) Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. J Wildl Manage 70:347–357

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl 14:401–413

Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. Trends Ecol Evol 17:164–170

Knapp LB, Canham CD (2000) Invasion of an old-growth forest in New York by *Ailanthus altissima*: sapling growth and recruitment in canopy gaps. J Torrey Bot Soc 127:307–315

Kowarik I (1995) Clonal growth in *Ailanthus altissima* on a natural site in West Virginia. J Veg Sci 6:853–856

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. Proc Natl Acad Sci USA 104:3883–3888

Lee CE (2002) Evolutionary genetics of invasive species. Trends Ecol Evol 17:386–391

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, Lohmann LG, Montiel OM (2008) Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J Biogeogr 35:105–116

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu Z, Yang L, Merchant JW (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int J Remote Sens 21:1303–1330

MacArthur RH (1972) Geographical ECOLOGY. Harper & Row Publishers, New York, p 269

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecol Appl 10:689–710

Manly BF, MacDonald LL, McDonals TL, Thomas DL, Erickson WP (2002) Resource selection by animals: statistical design and analysis for field studies. Kluwer, Dordrecht, p 221

Miller JH (2003) Nonnative invasive plants of southern forests: a field guide for identification and control. General Technical Report, p 93. USDA Forest Service, Southern Research Station, Asheville, NC

Miller JR, Turner MG, Smithwick EAH, Dent CL, Stanley EH (2004) Spatial extrapolation: the science of predicting ecological patterns and processes. Bioscience 54:310–320

Mitchell CE, Agrawal AA, Bever JD, Gilbert GS, Hufbauer RA, Klironomos JN, Maron JL, Morris WF, Parker IM, Power AG, Seabloom EW, Torchin ME, Vazquez DP (2006) Biotic interactions and plant invasions. Ecol Lett 9:726–740

Omernik JM (1987) Ecoregions of the conterminous United States. Ann Assoc Am Geogr 77:118–125

Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 49:1373–1379

Peterson AT (2001) Predicting species' geographic distributions based on ecological niche modeling. Condor 103:599–605

Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. Q Rev Biol 78:419–433

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Spiegelhalter D, Thomas A, Best N (1999) WinBUGS version 1.2 user manual. MRC Biostatistics Unit, Cambridge

Trifilo P, Raimondo F, Nardini A, Lo Gullo MA, Salleo S (2004) Drought resistance of *Ailanthus altissima*: root hydraulics and water relations. Tree Physiol 24:107–114

USGS (1996) HYDRO1k elevation derivative database. Retrieved from http://lpdaac.usgs.gov/gtopo30/hydro/index.asp on 22 February 2004

Welk E (2004) Constraints in range predictions of invasive plant species due to non-equilibrium distribution patterns: Purple loosestrife (*Lythrum salicaria*) in North America. Ecol Modell 179:551–567

Williams JW, Shuman BN, Webb T (2001) Dissimilarity analyses of late-quaternary vegetation and climate in eastern North America. Ecology 82:3346–3362

Yan X, Zhenyu L, Gregg WP, Dianmo L (2001) Invasive species in China—an overview. Biodivers Conserv 10:1317–1341

Zhang PC, Shao GF, Zhao G, Le Master DC, Parker GR, Dunning JB, Li QL (2000) Ecology—China's forest policy for the 21st century. Science 288:2135–2136

Zhu ZL, Waller E (2003) Global forest cover mapping for the United Nations Food and Agriculture Organization Forest Resources Assessment 2000 program. For Sci 49:369–380

Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots—a fundamental evaluation tool in clinical medicine. Clin Chem 39:561–577