# Using GAP Data in Invasive Plant Ecology and Management

Gary N. Ervin[1]

## Introduction

More than 5,000 species of plants introduced to the United States are estimated to have escaped cultivation and exist in some form of sustained wild populations (Pimentel et al. 2001). The potential costs of non-native species invasions has been estimated in the hundreds of billions of dollars per year for control efforts, environmental impacts, and actual commercial losses (Mullin et al. 2000, Pimentel et al. 2001). However, despite the assertions that control efforts during early stages of establishment are likely to be most successful, the focus of invasive species management remains largely within the realm of removal and eradication of well established species (Mullin et al. 2000). This may be the result of numerous factors, foremost of which is that most widespread species usually are of the most substantial economic interest.

A framework that has been proposed for augmenting programs aimed at controlling economically costly and widespread species is the early detection rapid response (EDRR) system. The concept of a nationwide EDRR system was formulated in 2003 (Westbrooks 2005). The concept of EDRR efforts will become more important over time, as ongoing governmental freezes, reductions, and reallocations of state and Federal funds limit the ability of government agencies to keep pace with the spread of invasive species. One potential means of enhancing the efficacy and efficiency of EDRR efforts is the development of probabilistic habitat models for invasive species of concern. These models could guide the limited search and monitoring efforts that currently are possible based on existing funding allocations. Accurate plant habitat models would facilitate efforts at locating, monitoring, and managing existing populations of invasive species as well as identifying adjacent land areas that are at risk of invasion.

Two major research emphases at Mississippi State University that aim to achieve this integration between invasive species research and management are the Cactus Moth Detection and Monitoring Network <http://www.gri.msstate.edu/research/cmdmn/> and the Invasive Plant Atlas of the Mid-South (IPAMS; <http://www.gri.msstate.edu/ipams/>). The IPAMS was established to record the distribution of native and non-native invasive plants within a five-state region of the south-central United States, and to use those data for developing geospatial habitat expectation models (niche models) to aid in guiding management efforts. The present work is part of that effort, but also demonstrates the utility of data available through the U.S. Geological Survey (USGS) Gap Analysis Program (GAP) in informing the development of such models. The case study presented here deals with a native species, *Baccharis halimifolia* L. (Asteraceae), that is believed to be increasing in its distribution within the southeastern United States (Ervin 2008).

## Methods

### Study Species and Study Area

*Baccharis halimifolia* L. (Asteraceae) is a shrub commonly known as eastern baccharis, silverling, groundsel-bush, or salt-bush. This species is considered native to the Atlantic and Gulf Coast regions of the United States from Texas to Massachusetts (U.S. Department of Agriculture, Natural Resources Conservation Service 2007; Weakley 2007), where it has historically occurred in upland fringes of coastal marshes and back dune habitats (Krischik and Denno 1990). *B. halimifolia* occurs under a wide range of soil and environmental conditions with respect to factors such as soil pH and nutrient concentrations, and has the ability to survive periodic flooding and drought, as well as fire (Westman et al. 1975).

*B. halimifolia* is wind-pollinated and produces large numbers of small, wind-dispersed fruit (0.1 mg dry mass per achene; Krischik and Denno 1990). As many as 1.5 million achenes may be produced per plant, and the highest rates of seed production have been observed from plants growing in open-canopied habitats (Westman et al. 1975, Panetta 1977). Available data suggest that germination of non-buried seeds occurs shortly after dispersal. Seedlings are thought to be able to establish during winter, potentially because dormant neighbors have limited capacity to shade the microhabitats where seedlings must establish (Panetta 1977, 1979b).

---

[1] Department of Biological Sciences, Mississippi State University, Missisippi State, MS.

Despite its historical coastal distribution, *B. halimifolia* is capable of establishing in interior regions of the southeastern United States, particularly in disturbed habitats such as fallow fields and forest edges, as well as arid inland habitats (Krischik and Denno 1990). Areas where *B. halimifolia* has been reported include the interior regions of the coastal plains (Duncan 1954, Krischik and Denno 1990), as well as the Piedmont, Ridge and Valley, Interior Low Plateau, and even in the foothills of the Blue Ridge and Cumberland Plateau (Estes 2004, 2005, Weakley 2007). Duncan (1954) reported that *B. halimifolia* increased its distribution substantially during the first half of the 20th century, and was considered in 1954 to be a weed "of great importance" in Georgia. Outside the United States, *B. halimifolia* has become an invasive weed in Australia, France, Spain, and the Black Sea region of eastern Europe (Westman et al. 1975). The noxious chemistry of its foliage and its preference for disturbed habitats make *B. halimifolia* especially problematic in pasturelands used for cattle production (Kraft and Denno 1982, Boldt 1989, Nesom 2001).

## Data Collection

Data for the present study were collected through roadside surveys of *B. halimifolia* in 17 counties of northeastern Mississippi (Figure 1). This region was expected to provide a gradient of *B. halimifolia* density from the southern extent of the region to the Mississippi-Tennessee border. *B. halimifolia* occurs in high densities in the southern portion of this study area, but has only recently been reported in Tennessee (Estes 2004, 2005). Survey routes consisted primarily of state and federal highways, which provided the most direct means of transecting multiple counties in the study area. Those routes also were readily available as georeferenced data layers. Data layers for the highways were obtained from the Mississippi Automated Resource Information System (MARIS; <http://www.maris.state.ms.us/>). The selected routes were digitized in ArcGIS 9.0 (Environmental Systems Research Institute, Inc.), converted to an ArcGIS shapefile, and transferred to an HP iPAQ HX 2110, running Windows Mobile™ 2003 second edition, version 4.21.1088. Navigation along the routes was performed with the assistance of Farm Works Site Mate version 11.40 (CTN Data Service, Inc.) geographic information system (GIS) software and a Holux compact flash card global positioning system (GPS) unit, model GR-271. Digitized routes were corrected for new highway construction after the surveys by visual inspection within GIS and comparison with *B. halimifolia* locations and an independent land cover data layer (National Land Cover Database [NLCD 2001], Multi-Resolution Land Characteristics Consortium: <www.mrlc.gov> [MRLC 2001]). Data handling within ArcGIS was performed in the Albers map projection (USA Contiguous Albers Equal Area Conic, USGS version) and the 1983 North American Datum geographic coordinate system (NAD 1983). However, data collection in the field was performed in the 1984 World Geodetic System datum (WGS 1984), and data were re-projected to NAD 1983 as necessary within ArcGIS.

The roadside surveys, conducted during November and December 2006 (period of fruit production and dispersal), provided 553 presence points for *B. halimifolia* along the survey route (797 km surveyed). To conduct logistic habitat modeling, absence data are required in addition to presences. In this study, absence data were provided by generating pseudo-absence points within 50 m of the survey routes. This was the distance within which it was estimated that *B. halimifolia* patches could be identified readily during the driving survey, as its dense white clusters of flowers and fruit made reproductive individuals readily visible from a few hundred meters in open landscapes. It is possible that pre-reproductive plants could have been overlooked, although the canopy morphology and leaf color and texture are highly noticeable in the surveyed area, especially within 50 m of the road.

Generation of pseudo-absence points is a well established method of creating statistically valid absence data for ecological modeling when true absence records are not available (see Engler et al. 2004 or Chefaoui and Lobo 2008). For this study, pseudo-absence points were generated by creating, in ArcMap, a buffer of 50 m on each side of the survey route. Each recorded *B. halimifolia* point was buffered by a distance of 200 m, and the area of those point buffers was subtracted from the 100-m-wide route buffer. Five hundred random pseudo-absence points were generated within the remaining route buffer area to represent likely points at which *B. halimifolia* was not present, assuming all patches within 50 m of the survey route were observed and recorded. One-half of the presence and pseudo-absence points were selected to form a training data set, the other half were used in model validation, as described below.

Environmental variables hypothesized to be determinants of *B. halimifolia* habitat were soil characteristics and canopy coverage. Soil data included clay content (percent), available water capacity ($cm \cdot cm^{-1}$), bulk density ($g \cdot m^{-3}$), organic matter (percent), pH, cation exchange capacity ($meq \cdot 100g^{-1}$), and permeability ($cm \cdot h^{-1}$). These data were extracted from the USDA NRCS STATSGO data (USDA NRCS 1994). Each variable was represented in the STATSGO database by a high value and a low value for each soil survey mapping unit, and both high and low values were used in the modeling work described here. Canopy cover was obtained from the MRLC National Land Cover Database (NLCD) as 30-m-resolution tree canopy density data (percent cover, to nearest 1 percent). Those data were a USGS Southeastern GAP data product generated initially by the method of Huang et al. (2001).

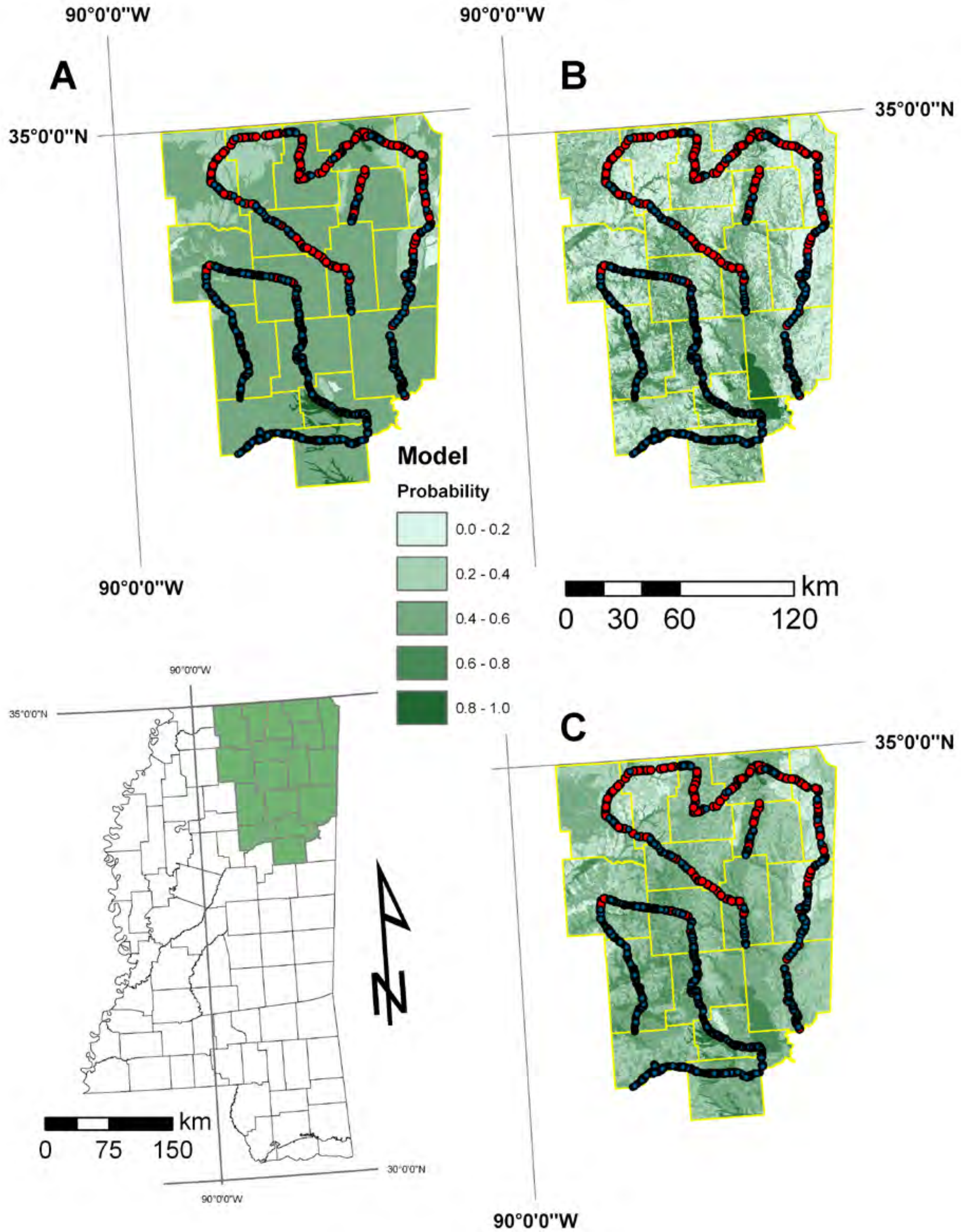

**Figure 1.** Map of study area and depiction of the three best habitat models from these analyses, in terms of probability of occurrence of habitat suitable for *Baccharis halimifolia*, given environmental data across the area. The Mississippi map in the lower-left shows the location of the study area within the state (shaded northeastern counties). Colored points in the habitat probability surfaces (*A*–*C*) indicate presence (blue) and pseudo-absence (red) points for *Baccharis halimifolia*, along the 797 km survey route. Models are (*A*) minimum soil percent organic matter, soil bulk density model; (*B*) canopy, minimum soil percent organic matter, minimum soil percent clay, pH model; (*C*) averaged model.

## Model Development

Presence-absence/distributional models were derived via logistic regression because this approach combines the ability to quantify correlations between predictor and response variables with the utility of incorporating categorical variables, such as the binary presence-absence response variable. The latter feature of logistic regression is attractive for obvious practical reasons but also avoids key statistical issues of ordinary least squares regression, such as failure of dichotomous variables to satisfy assumptions of equal variance, linearity, and normality (Menard 2002).

Logistic regression has been used in successful efforts at modeling species distribution or habitat suitability in several systems. For example, logistic regression was compared with discriminant analysis and artificial neural networks for presence-absence prediction of six bird species on 180 Himalayan streams (Manel et al. 1999). The three methods were found to be equally capable in their overall success at predicting presence or absence. However, logistic regression was described as yielding the most straightforward ability to assess the relative importance of individual environmental variables because of the ability to generate estimated probability of occurrence from regressions on logit-transformed presence-absence data. Manel and colleagues subsequently have used logistic regression in numerous habitat models (e.g., Manel et al. 2000, 2001), as have Buchan and Padilla (2000), Peterson et al. (2003), Underwood et al. (2004), and others.

Derivation and selection of the "best" model(s) of *B. halimifolia* habitat involved two phases. First, the initial set of candidate models to be parameterized consisted only of those containing predictors that were uncorrelated with one another based on simple bivariate correlation analyses (i.e., Pearson correlations). This provided clear indication of the potential explanatory power of each predictor variable independent of other factors. All candidate models then were evaluated by logistic regression and the best models identified by model selection statistics provided through the information-theoretic approach (Burnham and Anderson 2002). Specifically, the log likelihood value from each logistic regression was used to calculate the (corrected) Akaike Information Criterion ($AIC_c$), from which AIC differences were calculated for use in evaluating the resultant models and for determination of the most influential landscape features, in terms of their correlation with presence and absence of *B. halimifolia*. The corrected Akaike Information Criterion was calculated as:

$$AIC_c = -2\times\left(-\frac{n}{2}\log\left(\frac{RSS}{n-(p+1)}\right)\right)+2K+\left(\frac{2K(K+1)}{n-K-1}\right) \quad , \qquad (1)$$

where

$n$ is the number of sample units (presence and absence points)

$RSS$ is the residual sum of squares from each regression model,

$p$ is the number of predictor variables used in the regression model, and

$K$ is the number of parameters estimated in each model (numbers of predictor variables + 2; includes intercept and variance estimate).

After the initial round of model and environmental factor screening, four additional models were added which included combinations of the most influential, but correlated, environmental factors. Those models were added to permit evaluation of the strength of models with specific suites of soil variables, along with canopy cover. Thus, the tiered approach to developing habitat models incorporated (1) logistic regression to test sets of candidate hypotheses for suites of factors influencing *B. halimifolia* occurrence, and (2) quantitative determination of the model(s) that best represented landscape features associated with distribution of this species, based on the presence-absence data set and *a priori* determined environmental variables thought to be of importance.

During model development and validation, model adequacy also was assessed with a suite of criteria including simple success rate (percent of test points correctly classified) as well as more complex evaluation metrics (Table 1). The $AIC_C$ mentioned above is one such metric that provided information on the relative strength of models, including the efficiency of prediction, as AIC calculation "penalizes" models with larger numbers of variables. Once models were selected for implementation from among those developed with the training data, they were validated with the second half of the data, using the same set of assessment metrics. Manel et al. (2001) evaluated seven assessment metrics for presence-absence models and found that Cohen's kappa performed best, in part because it was less influenced by prevalence of presence or absence points (i.e., the relative proportion of presence or absence points within the data set). Cohen's kappa essentially represents the proportion of all possible cases predicted correctly as present or absent by a model after accounting for the effects of chance. Because it is a standardized value, kappa can be used to compare models that include different suites of predictor variables. The True Skill Statistic is suggested to improve upon kappa by having a smaller degree of correlation with prevalence of occurrences versus absences in the data set (Allouche et al. 2006).

**Table 1.** Model assessment metrics used in this study.

[**Description**: After Δ $AIC_c$, based on detail provided by Fielding and Bell (1997) and Allouche et al. (2006). **Abbreviation**: Δ, delta; AICc, Akaike Information Criterion]

| Assessment metric | Description |
|---|---|
| Δ $AIC_c$ | Difference in $AIC_c$ between an individual model and the model with the lowest $AIC_c$ value |
| Success rate | Proportion of points correctly classified as presence or absence |
| Sensitivity | Proportion of actual presences correctly predicted |
| Specificity | Proportion of actual absences correctly predicted |
| Positive predictive power | Proportion of predicted presences that represent actual presence points |
| Negative predictive power | Proportion of predicted absences that represent actual absence points |
| Cohen's kappa | "Proportion of specific agreement;" effectively corrects overall accuracy by accuracy expected by chance; see references for details |
| True skill statistic | (Sensitivity + specificity) - 1 |

## Results

The three best models, based on the Akaike Information Criterion, and as evaluated within the training data set, all included canopy cover as one of the environmental variables correlated with presence of *B. halimifolia* (Table 2). The three models were similar in terms of their performance based on a suite of standard model assessment criteria (overall prediction success, sensitivity, negative predictive power, and Cohen's kappa). However, another set of three models lacking the canopy cover variable performed well when evaluated by specificity, positive predictive power and the True Skill Statistic (TSS). Based on these results, all six of these models were carried into the validation phase of model evaluation.

In validation assessment, the set of six predictive habitat models performed similarly when evaluated against the training data set (Table 2). The models including canopy cover performed better in overall prediction success, sensitivity, negative predictive power, and Cohen's kappa, whereas those without canopy cover exhibited slightly better values for specificity, positive predictive power and TSS. Because of this ambiguity, an additional model was evaluated; that model was based on the average of the logit-transformed predicted probability occurrence determined by the top model in each of the two subsets of models with or without canopy cover (Table 2, bottom row). This model's performance appeared to be influenced heavily by the model that included canopy cover, while values for specificity, positive predictive power, and TSS were intermediate between the two models being averaged.

The best model in each subset of validated models (Table 2, highlighted rows), along with the averaged model are depicted in Figure 1*A–C*. The model that lacked canopy cover (Figure 1*A*) predicted a greater than 50 percent chance of *B. halimifolia* presence for most of the study area. The other two models yielded a more heterogeneous habitat probability surface when projected across the study area in GIS (Figure 1*B,C*). Both models, however, included a large region of high predicted probability of occurrence in the southeastern portion of the study area. As discussed below, that appears to have been an artifact of the resolution of soils data that were used in these analyses.

**Table 2.** Assessment and validation criteria used in selecting the “best” logistic models to represent suitable habitat for *Baccharis halimifolia* across the survey area.

[Numbers in bold are the highest value for each criterion in each phase of model assessment and validation. The Averaged Model is an average of the two best models from the validation process (highlighted rows); its assessment criteria were highest or equal to the highest for three of the assessment criteria (bold font). **Abbreviations**: Ca, canopy (USGS GAP); OM, minimum soil percent organic matter; Cl, minimum soil percent clay; pH, maximum soil pH; OMx, maximum soil percent organic matter; BD, soil bulk density, in gram per cubic meter]

| Environmental variables | Δ AICc | Success rate | Sensitivity | Specificity | Positive predictive power | Negative predictive power | Cohen’s kappa | True skill statistic |
|---|---|---|---|---|---|---|---|---|
| Internal assessment | | | | | | | | |
| Ca, OM, Cl | **0.0** | 0.64 | **0.62** | 0.70 | 0.83 | 0.44 | 0.27 | 0.32 |
| Ca, OM, Cl, pH | 1.2 | **0.65** | **0.62** | 0.70 | 0.83 | 0.44 | **0.28** | 0.32 |
| Ca, Cl, OMx | 2.6 | 0.64 | **0.62** | 0.67 | 0.80 | **0.46** | 0.26 | 0.29 |
| OM | 39.8 | 0.61 | 0.58 | **0.79** | **0.95** | 0.23 | 0.18 | **0.37** |
| OM, pH | 40.3 | 0.61 | 0.58 | **0.79** | **0.95** | 0.23 | 0.18 | **0.37** |
| OM, BD | 41.8 | 0.61 | 0.58 | 0.78 | 0.94 | 0.24 | 0.19 | 0.36 |
| Validation | | | | | | | | |
| Ca, OM, Cl | na | 0.61 | 0.58 | 0.81 | **0.95** | 0.24 | 0.19 | 0.39 |
| ***Ca, OM, Cl, pH*** | na | **0.64** | **0.61** | 0.70 | 0.83 | 0.42 | **0.26** | 0.31 |
| Ca, Cl, OMx | na | 0.63 | **0.61** | 0.67 | 0.81 | **0.44** | 0.25 | 0.29 |
| OM | na | 0.61 | 0.58 | 0.81 | **0.95** | 0.24 | 0.19 | 0.39 |
| OM, pH | na | 0.61 | 0.58 | 0.81 | **0.95** | 0.24 | 0.19 | 0.39 |
| ***OM, BD*** | na | 0.62 | 0.58 | **0.82** | **0.95** | 0.25 | 0.21 | **0.40** |
| ***Averaged Model*** | na | **0.64** | **0.62** | 0.72 | 0.86 | 0.41 | **0.27** | 0.34 |

## Discussion

Natural history information on *B. halimifolia* suggests that canopy cover would be informative in terms of predicting areas of suitable habitat. *B. halimifolia* is wind dispersed, requiring open habitats for long-distance dispersal, and published reports indicate that seeds germinate and establish more readily and that plants produce more viable seed in open canopied habitats than in shaded areas (Westman et al. 1975; Panetta 1977, 1979a). Further, observations before and during the road surveys suggested *Baccharis* is highly correlated with disturbed habitats, another correlate of low canopy coverage (Ervin 2008). Because of these factors, it is not surprising that the canopy cover data provided through the USGS GAP Project were present in the better models resulting from these analyses.

One observation from this exercise that was not expected, however, was the synergy provided by combining the USDA NRCS STATSGO data with canopy cover. The STATSGO data are very coarse-grained, with a mean sample unit of about 640 acres (259 ha). This is in large part explains the homogeneous nature of the predictive surface provided when the model included only soil data (Figure 1*A*). This coarseness also probably contributed to the large, homogeneous high probability region in the southeastern region of Figure 1*B* and 1*C* (although an inadequate number of sample points in that area might also be responsible). Nevertheless, when the coarse-grained soil data were combined with the more detailed and fairly accurate canopy coverage data, the result was a much more realistically heterogeneous probability surface and slightly better performance in assessment criteria (Table 2). Averaging of the two top models further improved upon this performance. Additional improvement likely will be possible with the incorporation of the SSURGO data, which are based on an average soil survey unit of about 64 acres (26 ha). Those data sets were not available for all the study area at the time these models were developed. These data are now available for all the surveyed counties in Mississippi.

This work suggests that the data provided by the USGS GAP project generally can be quite beneficial to efforts at modeling potential habitat for invasive plant species. This project has promise for contributing substantially to early detection-rapid response (EDRR) efforts as part of state and regional invasive species management plans. If accurate predictive models for key invasive plants can be developed, those models will be valuable tools to advise survey efforts contributing to EDRR programs. In addition to guiding detection and monitoring programs, these models can be used to assess the relative importance of different habitat variables, and that information may inform management efforts. In the present work, for example, canopy coverage seemed to be the most influential single variable. This correlates with the known biology of the target species and suggests that programs aimed at managing spread of *B. halimifolia* should focus on open canopied habitats, particularly newly opened areas which would be prone to colonization and development of satellite populations. In light of the biology of this species, the results also suggest that canopy-free land cover, such as transportation corridors and utility rights-of-way, might contribute to spread of *B. halimifolia*. All these pieces of information could prove very useful in monitoring or managing to reduce the spread of *B. halimifolia*, and the approach used here could serve as a template for studying and advising management of other invasive plant species.

## Acknowledgments

The work presented here was supported by funding from the U.S. Geological Survey Biological Resources Discipline (04HQAG0135), U.S. Department of Agriculture (2007-55320-17847 and 2008-35320-18679), and the Mississippi Computational Biology Consortium (NSF EPSCoR #EPS-0556308). Assistance in data collection was provided by T. Ervin and G.L. Ervin. Chris Holly provided a very helpful critique of an earlier version of this paper, as did two anonymous reviewers.

## References Cited

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa, and the true skill statistic. *Journal of Applied Ecology* 43: 1223-1232.

Boldt, P.E. 1989. Biology and host specificity of *Trirhabda bacharidis* (Coleoptera: Chrysomelidae) on *Baccharis* (Asteraceae: Asterae). *Environmental Entomology* 18: 78-84.

Buchan, L.A.J., and D.K. Padilla. 2000. Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. *Ecological Applications* 10: 1442-1455.

Burnham, K.P., and D.R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Springer, New York.

Chefaoui, R.M., and J.M. Lobo. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological modelling* 210: 478–486.

Duncan, W.H. 1954. More and more weeds in Georgia. *Bulletin of the Georgia Academy of Science* 12:99-103.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41: 263–274.

Ervin, G.N. 2008. Distribution, habitat characteristics, and new county-level records of *Baccharis halimifolia* L. on a portion of its present US range boundary. *Southeastern Naturalist*, accepted.

Estes, D. 2004. Noteworthy collections: middle Tennessee. *Castanea* 69: 69-74.

Estes, D. 2005. The vascular flora of Giles County, Tennessee. *Sida* 21: 2343-2388.

Fielding, A.H., and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.

Huang, C., L. Yang, B. Wylie, and C. Homer. 2001. A strategy for estimating tree canopy density using LANDSAT 7 ETM+ and high resolution images over large areas. *Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry*, Denver, Colorado, 5-7 November, 2001. Available from http://www.mrlc.gov/mrlc2k_nlcd.asp

Kraft, S.K., and R.F. Denno. 1982. Feeding responses of adapted and non-adapted insects to the defensive properties of *Baccharis halimifolia* L. (Compositae). *Oecologia* 52: 156-163.

Krischik, V.A. and R.F. Denno. 1990. Differences in environmental response between the sexes of the dioecious shrub, *Baccharis halimifolia* (Compositae). *Oecologia* 83:176-181.

Manel, S., J.M. Dias, S.T. Buckton, and S.J. Ormerod. 1999. Alternative methods for predicting species distribution: An illustration with Himalayan river birds. *Journal of Applied Ecology* 36: 734-747.

Manel, S., S.T. Buckton, and S.J. Ormerod. 2000. Testing large-scale hypotheses using surveys: The effects of land use on the habitats, invertebrates, and birds of Himalayan rivers. *Journal of Applied Ecology* 37: 756-770.

Manel, S., H.C. Williams, and S.J. Ormerod. 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology* 38: 921-931.

Menard, S. 2002. *Applied logistic regression analysis*, 2nd ed. Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-106. Sage Publications, Thousand Oaks, CA.

Mullin, B.H., L.W.J. Anderson, J.M. DiTomaso, R.E. Eplee, and K.D. Getsinger. 2000. Invasive plant species. Council for Agricultural Science and Technology Issue Paper, Number 13, February 2000, 18 p.

(MRLC) Multi-Resolution Land Characteristics Consortium. 2001. National Land Cover Dataset 2001. U.S. Geological Survey, U.S. Department of the Interior. Available from http://www.mrlc.gov/.

Nesom, G. 2001. Groundsel tree – *Baccharis halimifolia* L. USDA Natural Resources Conservation Service Plant Guide. Available from http://plants.usda.gov/factsheet/pdf/fs_baha.pdf

Panetta, F.D. 1977. The effects of shade upon seedling growth in groundsel bush (*Baccharis halimifolia* L.). *Australian Journal of Agricultural Research* 28: 681-690.

Panetta, F.D. 1979a. The effects of vegetation development upon achene production in the woody weed, groundsel bush (*Baccharis halimifolia* L.). *Australian Journal of Agricultural Research* 30: 1053-1065.

Panetta, F. D. 1979b. Germination and seed survival in the woody weed, groundsel bush (*Baccharis halimifolia* L.). *Australian Journal of Agricultural Research* 30: 1067-1077.

Peterson, A.T., M. Papes, and D.A. Kluza. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* 51: 863-868.

Pimentel, D., S. McNair, J. Janecka, J. Wightman, C. Simmonds, C. O'Connell, E. Wong, L. Russel, J. Zern, T. Aquino, T. Tsomondo. 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems and Environment* 84:1–20.

Underwood, E.C., R. Klinger, and P.E. Moore. 2004. Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. *Diversity and Distributions* 10: 447-459.

(USDA NRCS) United States Department of Agriculture, Natural Resources Conservation Service. 1994. State Soil Geographic (STATSGO) Database for MS. Available from http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/index.html

(USDA NRCS) United States Department of Agriculture, Natural Resources Conservation Service. 2007. The PLANTS Database. National Plant Data Center, Baton Rouge, LA. Available from http://plants.usda.gov

Weakley, A.S. 2007. *Flora of the Carolinas, Virginia, and Georgia and Surrounding Areas*. UNC Herbarium, North Carolina Botanical Garden, University of North Carolina at Chapel Hill. Available from http://www.herbarium.unc.edu/flora.htm

Westbrooks, R. 2005. Progress on the development of a national early detection and rapid response system for invasive plants in the United States. Abstracts of the Second New England Invasive Plant Summit, September 16-17, 2005, Framingham, MA. Available from http://invasives.eeb.uconn.edu/ipane/summit05/summit05home.htm

Westman, W.E., F.D. Panetta, and T.D. Stanley. 1975. Ecological studies on reproduction and establishment of the woody weed, groundsel bush (*Baccharis halimifolia* L.: Asteraceae). *Australian Journal of Agricultural Research* 26:855-870.