

Wetland features and landscape context predict the risk of wetland habitat loss

KEVIN J. GUTZWILLER^{1,3} AND CURTIS H. FLATHER²

¹Department of Biology, One Bear Place #97388, Baylor University, Waco, Texas 76798 USA

²USDA Forest Service, Rocky Mountain Research Station, 2150 Centre Avenue, Building A, Fort Collins, Colorado 80526 USA

Abstract. Wetlands generally provide significant ecosystem services and function as important harbors of biodiversity. To ensure that these habitats are conserved, an efficient means of identifying wetlands at risk of conversion is needed, especially in the southern United States where the rate of wetland loss has been highest in recent decades. We used multivariate adaptive regression splines to develop a model to predict the risk of wetland habitat loss as a function of wetland features and landscape context. Fates of wetland habitats from 1992 to 1997 were obtained from the National Resources Inventory for the U.S. Forest Service's Southern Region, and land-cover data were obtained from the National Land Cover Data. We randomly selected 70% of our 40 617 observations to build the model ($n = 28\,432$), and randomly divided the remaining 30% of the data into five Test data sets ($n = 2437$ each). The wetland and landscape variables that were important in the model, and their relative contributions to the model's predictive ability (100 = largest, 0 = smallest), were land-cover/land-use of the surrounding landscape (100.0), size and proximity of development patches within 570 m (39.5), land ownership (39.1), road density within 570 m (37.5), percent woody and herbaceous wetland cover within 570 m (27.8), size and proximity of development patches within 5130 m (25.7), percent grasslands/herbaceous plants and pasture/hay cover within 5130 m (21.7), wetland type (21.2), and percent woody and herbaceous wetland cover within 1710 m (16.6). For the five Test data sets, Kappa statistics (0.40, 0.50, 0.52, 0.55, 0.56; $P < 0.0001$), area-under-the-receiver-operating-curve (AUC) statistics (0.78, 0.82, 0.83, 0.83, 0.84; $P < 0.0001$), and percent correct prediction of wetland habitat loss (69.1, 80.4, 81.7, 82.3, 83.1) indicated the model generally had substantial predictive ability across the South. Policy analysts and land-use planners can use the model and associated maps to prioritize at-risk wetlands for protection, evaluate wetland habitat connectivity, predict future conversion of wetland habitat based on projected land-use trends, and assess the effectiveness of wetland conservation programs.

Key words: estuarine and palustrine wetlands; land conservation; landscape context; multivariate adaptive regression splines; National Resources Inventory; policy and planning; risk of wetland habitat conversion; southern United States; spatially explicit predictive model; wetland protection.

INTRODUCTION

Wetlands are crucial habitats for many organisms, yet they continue to be converted to other human land uses. To ensure efficiency of wetland conservation efforts, spatially explicit models that predict the risk of wetland habitat loss are needed to prioritize where limited conservation resources should be applied. We developed a spatially explicit model that predicts the risk of wetland habitat loss throughout the southern United States. We used a modeling approach that considered local and global statistical relations, as well as characteristics of wetland sites and their landscape context. Tests of the model with separate data sets confirmed that it had strong region-wide predictive ability. The model can be used to inform management, planning, and

policy decisions that can reduce wetland habitat loss in the South, and it demonstrates promise for expanding this modeling effort to other regions.

Wetlands often provide valuable ecosystem services in the form of flood control, aquifer recharge, water-quality improvement, and carbon sequestration (Scodari 1997, Mitsch and Gosselink 2007). Furthermore, wetlands contribute prominently to a region's biodiversity because their vegetation and other conditions typically support a diverse biota, many of which are wetland obligates (Zedler and Kercher 2005, Daniels and Cumming 2008). The benefits attributed to wetland habitats are now more important in the United States than they have ever been before because, of the 89.4 million ha of wetland that were present in the conterminous United States during Colonial America, less than 50% remain (Dahl 1990, 2006). Historically, agricultural development was the most significant economic force leading to the conversion of wetland habitats, with up to 87% of wetland losses being

Manuscript received 27 January 2010; revised 21 June 2010; accepted 25 June 2010. Corresponding Editor: C. A. Johnston.

³ Email: kevin_gutzwiller@baylor.edu

attributed to agricultural activities (Frayer et al. 1983). By the early 1990s, agriculture's role as an agent of wetland loss was significantly reduced; agriculture accounted for 20% of wetland conversions while urban and suburban development was responsible for 57% of wetland losses (Brady and Flather 1994).

Much of the reduction in agricultural conversion of wetlands has been attributed to the Wetland Conservation Provisions of the Food Security Act of 1985, commonly called "Swampbuster," which withheld USDA farm program benefits from producers who converted wetlands to grow commodity crops (Williams 2005). However, Swampbuster is not the only federal program affecting wetland conservation. As noted by Mitsch and Gosselink (2007:470), there is no single comprehensive national wetland law in the United States. Rather, there is a collection of statutes (e.g., Clean Water Act, Federal Aid to Wildlife Restoration Act, North American Wetlands Conservation Act), Executive Orders (e.g., Protection of Wetlands, Conservation of Aquatic Systems for Recreational Fisheries), and administrative policies (e.g., no net loss) that guide wetland conservation policy, making for a complicated system of protection and jurisdictional authority.

Nevertheless, wetland conversion continues to occur as a consequence of permitting systems, exemptions, mitigation, and enforcement problems (Hansen 2006). Between 1992 and 1997, just over 204 000 ha of palustrine and estuarine wetlands were lost in the United States; 75% of these losses were attributed to either development (49%) or agriculture (26%; U.S. Department of Agriculture 2000). The greatest loss of wetlands during this period occurred in the South (59% of the national losses), and 78% of the losses in this region were due to development (58%) and agriculture (20%; U.S. Department of Agriculture 2000).

If land-use policy and planning are to be efficient in conserving wetland habitats, it is imperative to identify, in a geographically explicit fashion, those areas where the risks of future wetland habitat conversion are highest. Recent efforts to develop predictive models of wetland conversion have considered physical and socioeconomic variables (Douglas and Johnson 1994) and suites of environmental variables that reflect wetland type, soil and topographic conditions, land-use and land-cover characteristics, and the spatial proximity of roads and cities (Koneff and Royle 2004, Daniels and Cumming 2008). Variation in wetland loss rates at the spatial extent of states in the USA was associated strongly with three variables: land drained (drainage investment), wetland rural acreage, and farmland realty value (Douglas and Johnson 1994). Elevation and a wetness index (derived from slope and location in a watershed) were important predictors of wetland occurrence in Atlantic Coast states (Koneff and Royle 2004). In a large watershed in Costa Rica, slope, elevation, proximity to roads and human settlement, protection status, and geomorphic configuration were

important predictors of wetland loss (Daniels and Cumming 2008).

Here we contribute to this growing body of knowledge about predictors of wetland loss by considering the risk of wetland habitat loss in relation to characteristics of wetlands, of areas immediately adjacent to wetlands, and of landscapes surrounding wetlands. By "risk" we mean the tendency for wetland habitat loss given that a particular inventory point is classified as a wetland. We make this distinction to emphasize that our model does not predict the area of wetland habitat loss (e.g., hectares converted), but rather the likelihood that a particular wetland inventory point will be converted. Development of effective policy will require knowledge about factors associated with wetland habitat conversion in high-risk areas, and a predictive model of the risk of wetland habitat loss is especially necessary for decision making. By focusing on risk we also avoided noted problems associated with wetland area estimation (Government Accounting Office 1998).

Our specific objectives were to (1) develop a predictive model for the risk of wetland habitat loss for the southern United States, where the rate of wetland loss has been highest in recent decades; (2) test the model's predictive performance using separate data sets; (3) produce a predictive map of the risk of wetland habitat loss for the southern United States; (4) construct companion maps of prediction errors to show spatial variation in prediction accuracy; (5) explain how relations embodied in the model can be used to improve understanding about possible drivers of wetland conversion; and (6) indicate how the model and maps can be used to inform policy and land-use planning for conservation of wetland habitats.

METHODS

Study area

Wetland trend analyses based on the National Resources Inventory (NRI; Nusser and Goebel 1997, Nusser et al. 1998) revealed a relatively rapid decline in nonfederal wetlands across a 2.24 million km² region corresponding to the southern United States during the 1990s. This region also supports nearly 48% of the 45 million ha of nonfederal wetland habitats occurring across the conterminous United States (U.S. Department of Agriculture 2000). For these reasons we focused our analysis of wetland habitat conversion across the South.

Data collection

We used data from the NRI to identify individual inventory points (Fig. 1) that were classified as wetland habitat in 1992 and that, by 1997, either remained wetland habitat or were converted to a non-wetland status. We used this period of years because land-cover data available for 1992 made it possible for us to study landscape influences on wetland habitat loss by 1997. The NRI was used to determine the geographic

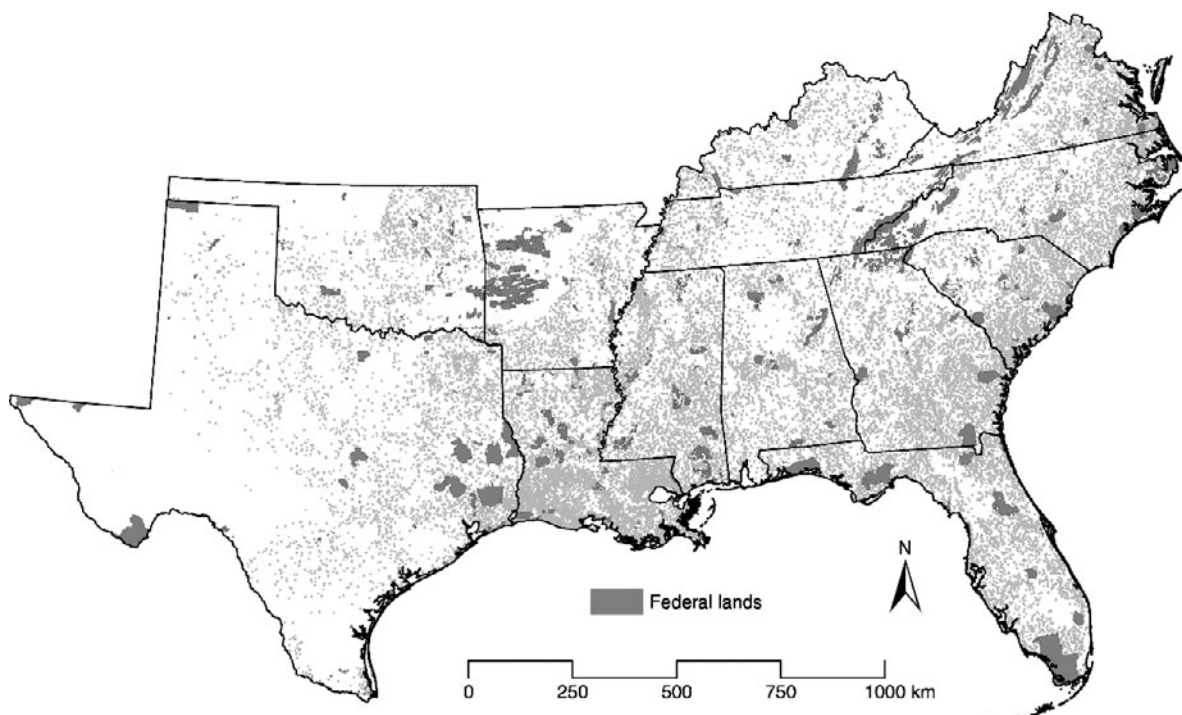


FIG. 1. The distribution of National Resources Inventory (NRI) points (gray dots) that were wetland in 1992 across the Southern Region. Dark gray polygons are federal lands, which were not sampled by the NRI.

locations of wetland habitat for the purpose of determining elevation of each NRI point and the landscape context surrounding each NRI point. Elevation for NRI inventory points was obtained from the U.S. Geological Survey National Elevation Dataset (Gesch et al. 2002), land-use and land-cover data were extracted from the National Land Cover Data (NLCD, Vogelmann et al. 2001), and road data were obtained from the U.S. Bureau of Transportation Statistics as summarized by Watts et al. (2007). Details about resolution, datum, and accuracy for these data sources are provided in Appendix A.

We used the 1992 and 1997 point-based NRI databases to gather data for wetland habitat. Wetland habitat loss (WLOSS), our binary response variable, was an indicator of whether wetland habitat at a NRI sampling point persisted between 1992 and 1997. If wetland habitat did not persist during this period, the response for that point was coded 1 (wetland habitat lost); if wetland habitat did persist, the response for that point was coded 0 (wetland habitat retained).

Wetlands are thought to be influenced by both proximate and landscape-level processes (Daniels and Cumming 2008). We therefore defined two sets of candidate predictors (local and landscape, Table 1) of wetland habitat loss to capture the multi-scale nature of the conversion process. Local variables (inventory-point variables) were derived directly from the NRI, or from the geographic location of the inventory point, and included variables like wetland type (we expected

forested wetlands would be more resistant to conversion than herbaceous wetlands), the broad land-use activities at the point (we expected a wetland habitat would be more prone to conversion if it occurred at a site under intensive land use), and elevation (we expected higher-elevation wetland habitats would be less prone to conversion).

Landscape variables were derived primarily from 30-m land-cover data for the entire United States from the U.S. Geological Survey's NLCD (circa 1992). Although landscape context is thought to influence the likelihood that a particular wetland will be converted, the spatial extent over which a landscape may influence wetland habitat persistence is unknown. We therefore measured landscape variables at three spatial extents that were based on the size of the NRI's primary sample unit (PSU) under the Public Land Survey (one-quarter section [~ 64.7 ha]; Nusser and Goebel 1997:187). Three circular buffers were centered on each NRI point that was classified as a wetland in 1992. The fine-extent buffer (570 m radius) encompassed the PSU; the meso-(intermediate-) extent buffer (1710 m radius) encompassed the PSU and the eight adjacent PSU cells that defined its first-order neighborhood; and the macro-extent buffer (5130 m radius) enclosed the PSU and the 80 adjacent PSU cells that defined its fourth-order neighborhood. To measure landscape variables (Table 1) within each of these nested spatial extents, we used ArcGIS 9.2 software (ESRI 2006) to extract NLCD data for each buffer; custom computer code to calculate

TABLE 1. Local (inventory-point) and landscape variables recorded or measured for 1992 NRI wetland habitat points in the southern United States.

Abbreviation	Description
Inventory-point variables	
BAIL_DIV	Bailey ecoregion division (Bailey 1995) for location of the NRI point (1 = hot continental [division code = 220], 2 = hot continental regime mountains [M220], 3 = subtropical [230], 4 = prairie [250], 5 = tropical/subtropical steppe [310], 6 = savanna [410]).
BROAD_92	Land-cover and land-use category at the NRI point in 1992 (1 = cultivated and non-cultivated cropland, 2 = pastureland, 3 = rangeland, 4 = forestland, 5 = other rural land, 6 = urban, built-up, and rural transportation land, 7 = census water [water bodies > 16.2 ha and perennial streams > 0.2 km wide] and small water areas [water bodies < 16.2 ha and perennial streams < 0.2 km wide]).
COW_92	Cowardin et al. (1979) wetland type for NRI point in 1992 (1 = estuarine, emergent [including open or shallow wetlands with nonpersistent emergents]; 2 = estuarine, scrub-shrub, and forested; 3 = palustrine, emergent [including open or shallow wetlands with nonpersistent emergents]; 4 = palustrine, scrub-shrub, and forested).
EI92	Soil erodibility index (unitless) for the NRI point in 1992 measured as the ratio of potential soil erodibility (based on rainfall and runoff, wind speed, surface soil moisture, susceptibility of the soil to water or wind erosion, and the combined effect of slope length and steepness), and the soil loss tolerance (the maximum annual rate of soil erosion that could occur without causing a decline in long-term productivity).
L_CAP_92	Land capability class at the NRI point in 1992 (0 = not applicable, 1 = low restrictions for agriculture, 2 = moderate restrictions for agriculture, 3 = high restrictions for agriculture).
OWN_92	Land ownership class for the NRI point in 1992 (1 = private, 2 = municipal, 3 = state, county/parish, or tribal/trust, 4 = large water bodies with undetermined ownership).
ELEV	Elevation (m) at the NRI point.
Landscape variables†	
9020D	Density (m/ha) of edge between wetlands (woody and herbaceous) and developed land (residential, commercial, industrial, mined, transportation).
9080D	Density (m/ha) of edge between wetland and agricultural land (orchards, row crops, small grains, fallow, plus grass areas in parks, golf courses, and cemeteries).
AW_20	Area-weighted mean patch size (m ²) for developed land.
AW_80	Area-weighted mean patch size (m ²) for agricultural land.
GI_20	Gravity index (Kline et al. 2001:Eq. 11) for developed land. All patches of developed land within a buffer were used to compute the gravity index (unitless) relative to the NRI wetland point. The center of each patch was used to determine the distance between each patch and the NRI wetland point.
GI_80	Gravity index (Kline et al. 2001:Eq. 11) for agricultural land. All patches of agricultural land within a buffer were used to compute the gravity index (unitless) relative to the NRI wetland point. The center of each patch was used to determine the distance between each patch and the NRI wetland point.
PR1090	Proportion of land area covered by open water, and woody and herbaceous wetlands.
PR4051	Proportion of land area covered by forest and shrubland.
PR7181	Proportion of land area covered by grasslands/herbaceous plants and pasture/hay.
PR20	Proportion of land area covered by developed land.
PR80	Proportion of land area covered by agricultural land.
PR90	Proportion of land area covered by woody and herbaceous wetlands.
RD_DEN	Road density measured as the length of roads in km per square km of area.

Note: Additional details about these variables are provided in Appendix A.

† Each of these variables was measured within the fine (F), meso (M), and macro (L) extents.

GI_20, GI_80, and RD_DEN; and FRAGSTATS 3.3 (Build 3; McGarigal et al. 2002) to compute the remaining variables. We chose landscape variables that we hypothesized would either increase conversion pressure (e.g., the amount, edge density, and proximity of land uses responsible for wetland conversion [agriculture, developed land, and roads]) or resist conversion (e.g., the amount of the landscape remaining in natural vegetation [forest, shrubland, grassland, and wetland]).

Statistical analyses

Overview of model construction and evaluation.—We randomly selected observations to create separate data sets for building and testing a predictive model. We used multivariate adaptive regression splines (MARS; Friedman 1991) to relate our binary response variable (WLOSS) to features of wetland sites and the surrounding landscape. Among several candidate models with

comparable mean square error, we identified as our best model the candidate with the fewest number of predictors and the highest prediction accuracy. The potential for effects of spatial autocorrelation on prediction accuracy was addressed using residual interpolation (Koneff and Royle 2004, Miller 2005). Percent correct classification, Kappa statistics, and area-under-the-receiver-operating-curve (AUC) statistics were used to evaluate the predictive performance of the best model. We assessed the relative predictive ability of each predictor, as well as the degree of multicollinearity among predictors. Spatial patterns in the model's predictions and prediction errors were examined through the use of maps. Additional details about MARS and the methods we used to develop the model are in Appendix B.

Allocation of Train and Test data.—Using *k*-fold data partitioning, we used a heuristic to randomly select 70%

of our 40 617 observations to build the model (Train data set; $n = 28\,432$) and 30% of the observations ($n = 12\,185$) to test the model (Fielding and Bell 1997:39–40). To assess variability in the predictive model's performance, we randomly divided the Test data into five separate sets (Test 1, Test 2, . . . , Test 5; $n = 2437$ each).

Model development.—We used software for multivariate adaptive regression splines (MARS2.0, Salford Systems 2001) to develop our model. For predictive modeling and predictive mapping, MARS can provide important advantages (Muñoz and Felicísimo 2004) over standard logistic regression, classification and regression trees, and other similar methods. MARS is a nonparametric multivariate method that simultaneously permits robust assessment of linear and nonlinear influences, simple relations and complex interaction effects, and both local and global statistical relations. MARS derives functions (basis functions) of the original variables that maximize fit and, during a final stage of analysis, retains them in a manner that minimizes the generalized cross-validated mean square error (MSE) for the model. The minimum-MSE model is identified by MARS as its optimal model.

We allowed MARS to fit up to 75 basis functions, and to fit both main (linear) effects and two-way interaction effects. Default settings were used for all other MARS options, except for the assessment of effective degrees of freedom (see Appendix B). Because unequal prevalence of the values of a binary response variable can adversely affect a model's predictive accuracy and stability, we weighted observations to ensure equal prevalence (0.5) of each group (Maggini et al. 2006, Elith and Leathwick 2007, Parviainen et al. 2008).

MARS often provides several candidate models with MSEs that are comparable to that of the optimal model, and the modeler has the option of choosing one of these models for prediction purposes. In our analysis, the model that MARS identified as optimal had 19 basis functions, 13 variables, and $MSE = 0.161$. Because simpler models (those with fewer predictors) tend to be more generalizable (less overfitted), we examined similar candidate models to identify a model that provided prediction performance that approximated that of the MARS optimal model but that had fewer basis functions and variables. In this analysis, our a priori model-selection criteria were parsimony and prediction accuracy. We identified our best prediction model by assessing how well the optimal model and similar candidate models (having 5–18 basis functions, 5–13 variables, and $MSEs = 0.161$ – 0.171) predicted wetland loss for the five Test data sets. For prediction purposes, we identified the optimal binary classification threshold (T) based primarily on the criterion that prediction accuracy for the wetland-loss group should be at least approximately 80%; given that condition, we also wanted a threshold that maximized prediction accuracy for the wetland-retained group and overall (both response groups combined).

We addressed potential effects of spatial autocorrelation on prediction accuracy by using residual interpolation (Koneff and Royle 2004, Miller 2005). This approach can improve prediction accuracy by accounting for spatial patterns in responses that are not captured by the explanatory variables in the model. To implement residual interpolation, we estimated a semivariance function for spatial autocorrelation for our best model's residuals based on the Train data. Then, for each location in a Test data set, we computed an adjusted prediction as the sum of the model's prediction plus the kriged residual (based on the semivariance function). The adjusted prediction reflected the effects of the explanatory variables in our best model, plus any spatial influences represented by the semivariance function.

We used robust estimates of semivariance (SAS Institute 1999, Proc Variogram; Curriero et al. 2002) to fit the semivariance function. This approach assumes stationarity, meaning in the present analysis that there were no substantive broad-scale trends in the residual surface. We confirmed that this assumption was satisfied by examining scatter plots and regressions (SAS Institute 1999, Proc Reg) involving first-, second-, and third-degree polynomial terms for the geographic coordinates of our observations; such terms are often used to adjust for broad-scale spatial trends and to establish stationarity (Legendre 1993). We used ordinary local kriging (SAS Institute, 1999, Proc Krige2d; Schabenberger and Gotway 2005) and the semivariance function to predict residuals for all Test data sets.

Evaluation of predictive performance.—We used SAS software (SAS Institute 1999, Proc Freq) to compute the simple Kappa statistic to determine whether the best model predicted the Test observations better than what would be expected by chance. We used the weighting approach described above (but tailored to the sample sizes of WLOSS categories for each Test data set) to adjust the frequencies of observations before calculating Kappa. This step reduced the chance for adverse effects of disparate sample sizes on Kappa (Fielding and Bell 1997). All predictions of WLOSS were based on the optimal binary threshold (T) for our best model. Values of predicted wetland habitat loss ($PRWLOSS$) $> T$ were classified as predicted wetland-loss observations, and $PRWLOSS$ values $< T$ were classified as predicted wetland-retained observations. To complement Kappa, we computed the AUC statistic, a threshold-independent measure of model predictive ability (Fielding and Bell 1997). We used MedCalc 9.4.2.0 software (MedCalc Software, Mariakerke, Belgium) and SAS Proc Freq (SAS Institute 1999) to compute and compare AUC statistics.

Predictive ability of variables.—We calculated each explanatory variable's relative contribution to the model's predictive ability using importance values estimated by MARS2.0. An importance value indicated the degree to which the fit of the model (based on MSE)

would be degraded if all basis functions involving a variable were dropped from the model (Salford Systems 2001). Importance values ranged from 0% to 100%; higher values indicated greater predictive ability.

Assessment of multicollinearity.—To assess the degree to which one may interpret the independent influence of each variable on the risk of wetland habitat loss, we computed variance inflation factors (VIFs) for the basis functions in the best model using the `vif()` function from the R statistical environment (R Development Core Team 2009). The calculations were based on the Test data sets rather than the Train data because using the data from which the basis functions were originally derived can result in exaggerated estimates of associations among variables (J. H. Friedman, *personal communication*).

Maps of predictions and prediction errors.—We developed a map of the predicted risk of wetland habitat loss (PRWLOSS) for the southern United States. For binary response variables, predicted values from MARS are not constrained to vary from 0 to 1 (Salford Systems 2001). For the map of PRWLOSS, we wanted to use values that would be more common for risk data and hence easier to understand. We therefore plotted a normalized index of PRWLOSS that ranged from 0 to 1, derived from the following formula:

$$\text{PRWLOSS}_{\text{norm}} = \frac{(\text{PRWLOSS} - \text{PRWLOSS}_{\min})}{(\text{PRWLOSS}_{\max} - \text{PRWLOSS}_{\min})}.$$

To map prediction errors, we estimated the absolute error (AE) for each observation (see Kanevski and Maignan 2004:50) in our Test data sets using the following rules, where T is the binary threshold selected to optimize classification accuracy:

- 1) If $\text{PRWLOSS} > T$ and $\text{WLOSS} = 1$, then $\text{AE} = 0$;
- 2) If $\text{PRWLOSS} < T$ and $\text{WLOSS} = 0$, then $\text{AE} = 0$;
- 3) Otherwise, $\text{AE} = |\text{PRWLOSS} - T|$.

The optimal binary threshold was applied to the range of the predicted values, not to the range of the observed values. We assessed model error relative to T rather than WLOSS because we wanted to quantify the distance between PRWLOSS for misclassified observations and the binary decision point defined by T .

To geographically visualize the predicted risk of wetland habitat loss and absolute error of model prediction, we used inverse distance weighting interpolation (Isaaks and Srivastava 1989:257) to generate risk and error maps. Because wetland occurrence across the landscape was rare, we used a relatively coarse search neighborhood, r , of 50 km; maps were generally insensitive to our choice of r over the range of 10–150 km. All of the observations ($n = 40\,617$) in our data set were used to develop the map for the risk of wetland habitat loss; the maps of prediction errors were based on the set of pooled observations from our five Test data

sets ($n = 12\,185$). Federal areas, which are not sampled by the NRI, were masked out for these maps.

Because absolute error does not indicate the direction of the error, we classified incorrectly predicted observations (see rule 3 above) as either errors of omission (failure to predict actual wetland habitat loss) or errors of commission (prediction of wetland habitat loss when no actual loss occurred). An error of omission occurred when $\text{WLOSS} = 1$ and $\text{PRWLOSS} < T$; an error of commission occurred when $\text{WLOSS} = 0$ and $\text{PRWLOSS} > T$. We mapped the locations of NRI wetland points in the Test data sets that were classified as omission and commission errors.

RESULTS

Predictive model

The best model that emerged from among the MARS optimal model and similar candidate models had 10 basis functions, nine variables, $\text{MSE} = 0.164$, and an optimal binary classification threshold (T) of 0.48. Summary statistics for variables involved in the model (Appendix C) indicated that a wide range of environmental conditions were represented in the data. Formulas for the basis functions (BFs) involved in the best model are provided in Table 2. The regression equation for the best model was

$$\begin{aligned} \text{WLOSS} = & 0.774 - (0.229 \times \text{BF2}) + (0.164 \times \text{BF3}) \\ & - (127.824 \times \text{BF8}) + (140.246 \times \text{BF9}) \\ & - (0.574 \times \text{BF29}) + (0.500 \times \text{BF32}) \\ & + (0.714 \times \text{BF42}) + (0.013 \times \text{BF51}) \\ & - (0.133 \times \text{BF55}) - (0.024 \times \text{BF68}). \end{aligned} \quad (1)$$

In Discussion: Understanding statistical relations, we illustrate how one can develop a more intuitive understanding of the basis functions. The wetland and landscape variables that were involved in the model, and their relative contributions to the model's predictive ability as indicated by MARS importance values, were BROAD_92, the land-cover/land-use of the surrounding landscape (100.0); F_GI_20, the size and proximity of development patches within 570 m (39.5); OWN_92, land ownership (39.1); F_RD_DEN, road density within 570 m (37.5); F_PR90, percent woody and herbaceous wetland cover within 570 m (27.8); L_GI_20, the size and proximity of development patches within 5130 m (25.7); L_PR7181, percent grasslands/herbaceous plants and pasture/hay cover within 5130 m (21.7); COW_92, wetland type (21.2); and M_PR90, percent woody and herbaceous wetland cover within 1710 m (16.6).

Multicollinearity among predictors

Across the five Test data sets, VIF values for all of the basis functions (predictors) in the best model were well below 10 (Table 2). VIF values for six of the 10 predictors (BF29, BF32, BF42, BF51, BF55, BF68) were

TABLE 2. Formulas for basis functions associated with the best model for the risk of wetland habitat loss, and variance-inflation-factor statistics for basis functions that were predictors in the model.

Basis function	Formula	Variance inflation factors†	
		Mean	Range
BF2	$\max(0, 1.587 - F_GI_20)‡$	3.0	2.5–4.0
BF3	$(BROAD_92 = 2, 6, \text{ or } 7) \times BF2$	5.2	3.8–6.1
BF8	$\max(0, 0.003 - F_RD_DEN)$	2.7	2.2–3.3
BF9	$(BROAD_92 = 6 \text{ or } 7) \times BF8$	4.2	3.4–5.0
BF12	$\max(0, 12.978 - L_GI_20)$	§	§
BF26	$COW_92 = 2, 3, \text{ or } 4$	§	§
BF27	$BROAD_92 = 6 \text{ or } 7$	§	§
BF28	$BROAD_92 = 1, 2, 3, 4, \text{ or } 5$	§	§
BF29	$(OWN_92 = 1 \text{ or } 2) \times BF27$	1.3	1.1–1.4
BF32	$\max(0, 0.422 - L_PR7181) \times BF27$	2.1	1.8–2.3
BF42	$\max(0, 0.284 - F_PR90) \times BF28$	1.6	1.5–1.7
BF51	$(BROAD_92 = 2, 3, 5, \text{ or } 6) \times BF12$	1.3	1.2–1.3
BF55	$(BROAD_92 = 2, 3, 5, \text{ or } 7) \times BF26$	1.9	1.6–2.1
BF68	$\max(0, 0.515 - M_PR90) \times BF12$	1.9	1.6–2.1

† Statistics for variance inflation factors are for the five Test data sets.

‡ Variable definitions are in Table 1. F, M, and L at the beginning of variable names refer to fine (570 m radius), meso (1710 m radius), and macro (5130 m radius) extents, respectively.

§ Basis functions 12, 26, 27, and 28 were not involved in the best model as predictors (see Eq. 1), but MARS used them to estimate other basis functions that were involved as predictors. Because these four basis functions were not predictors, variance inflation factors were not computed for them.

small (mean = 1.7, range = 1.1–2.3). The other four predictors (BF2, BF3, BF8, BF9) had small to moderate VIF magnitudes (mean = 3.8, range = 2.2–6.1).

Model performance

For the Test data sets, the best model met our threshold criterion that the prediction accuracy for converted wetland habitats (WLOSS = 1) should be at least approximately 80%. In addition, compared to MARS candidate models with more basis functions and variables, the best model had higher prediction accuracies for converted wetland habitats and overall, and compared to candidate models with fewer basis functions and variables, the best model had higher prediction accuracies for retained wetland habitats (WLOSS = 0) and overall. AUCs for the MARS optimal model (19

basis functions, 13 variables) and our best model (10 basis functions, nine variables) did not differ significantly for any of the five Test data sets ($P = 0.141$ – 0.935); for $P = 0.141$, the absolute difference in AUC statistics was only 1.9%.

Performance statistics for the best model were remarkably robust between the Train and Test data sets (Table 3). The means for percent correct classification (of retained wetlands, converted wetlands, and overall) for the Test data sets exhibited little decline (range = 1.1–1.4%) compared to the accuracies of these classifications estimated from the Train data. Kappa statistics for the Test data sets indicated that the best model correctly predicted wetland fates at a rate that far exceeded that expected by chance ($P < 0.0001$), and the difference between mean Kappa for the Test data sets

TABLE 3. Predictive performance statistics for the best model of the risk of wetland habitat loss for the Test data sets and the Train data.

Performance statistics	Test data sets						Train data
	Individual values (rank order)					Mean	
Correct 0s (wetland retained, %)	69.3	70.1	71.1	72.8	73.6	71.4	72.5
Correct 1s (wetland converted, %)	69.1	80.4	81.7	82.3	83.1	79.3	80.7
Correct overall (%)	70.1	74.8	76.2	77.6	78.0	75.3	76.6
Kappa†	0.40	0.50	0.52	0.55	0.56	0.51	0.53
AUC‡	0.78	0.82	0.83	0.83	0.84	0.82	0.85
AUC (adjusted prediction)§	0.70	0.71	0.71	0.73	0.75	0.72	¶

† Kappa > 0 indicates that agreement between observed and predicted values (for wetland loss here) is better than chance agreement, and $0.4 < \text{Kappa} < 0.75$ indicates good agreement (Fleiss 1981).‡ When $0.7 \leq \text{AUC} < 0.8$, discrimination between groups (converted and retained wetlands here) is considered acceptable; when $0.8 \leq \text{AUC} < 0.9$, discrimination is considered excellent (Hosmer and Lemeshow 2000).

§ AUC for adjusted prediction (model prediction + kriged residual).

¶ AUC (adjusted prediction) was not computed for the Train data because assessment of whether residual interpolation improved the prediction accuracy of the model involved comparison of AUCs (for standard vs. adjusted predictions) for Test data sets only.

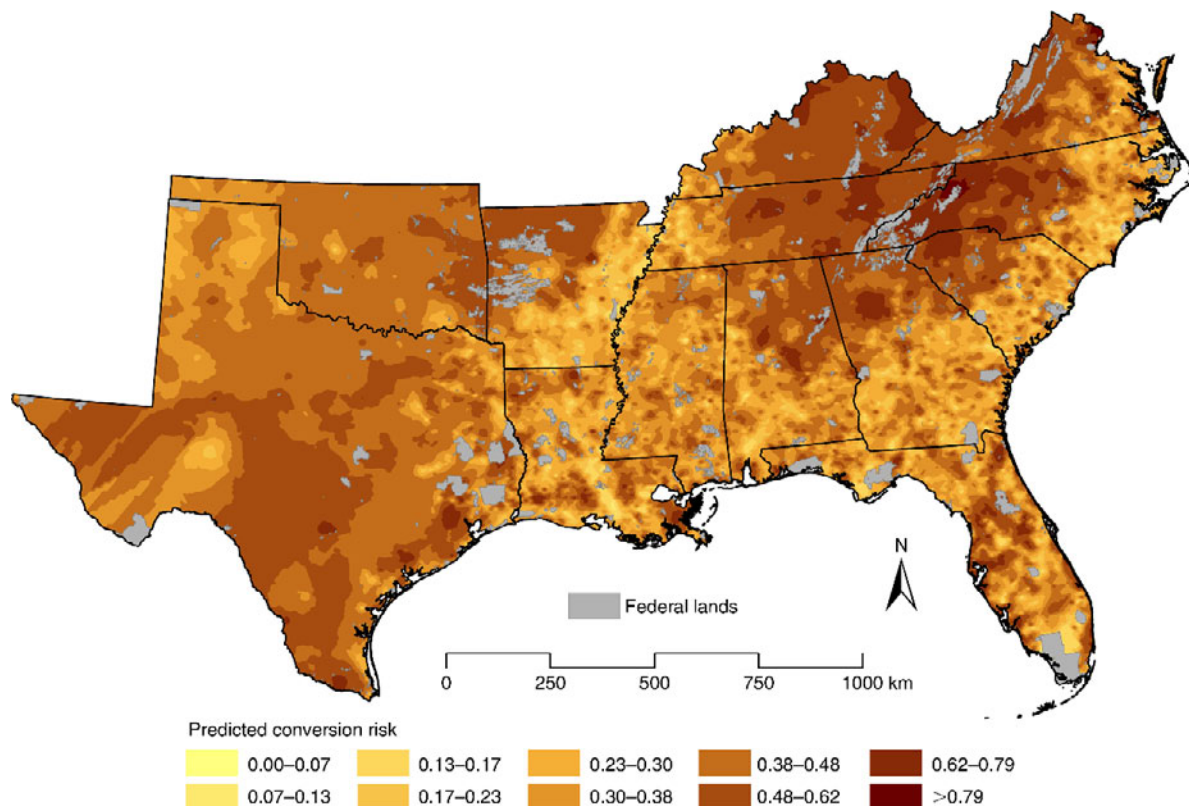


FIG. 2. Predicted risk of wetland habitat loss from 1992 to 1997 for all NRI points that were wetland in 1992 for the southern United States. Gray polygons are federal lands, which were not sampled by the NRI.

and Kappa for the Train data was just 0.02. Threshold-independent AUC statistics also indicated strong predictive ability ($P < 0.0001$), and that the agreement between observed and predicted wetland fate implied by Kappa was insensitive to our selection of the binary threshold. Relative to the AUC for the Train data, mean AUC for the Test data sets declined by only 0.03. Residual interpolation did not improve model predictions. For the Test data sets, mean AUC (adjusted prediction) (based on model predictions + kriged residuals) declined by 0.10 relative to mean AUC (based on model predictions alone), and the differences between AUC (adjusted prediction) and AUC were significant for all Test data sets ($P = <0.001-0.006$).

Predictive maps

The predicted risk of wetland habitat loss (conversion risk, Fig. 2) was higher in and near the Appalachian Mountains, Boston and Ouachita Mountains (in Arkansas and Oklahoma), and in western parts of the study area, including the Hill Country, Trans-Pecos, and Llano Estacado areas of Texas. The risk of wetland habitat loss was typically lower in much of the Coastal Plain, Piedmont, and Mississippi Basin. Throughout the study area, higher predicted risks of wetland habitat loss occurred in and near large urban areas, as illustrated by specific sites in Florida (Fig. 3). Absolute prediction

error was slightly higher in and near the Appalachian Mountains, large parts of southwest Texas, and in and near major cities, whereas absolute error was generally lower elsewhere (Fig. 4). Omission errors ($n = 77$; 20.9% of 369 wetland sites that were converted) were more dense in the lower Mississippi Basin but less dense or absent in and near the Appalachian Mountains and western parts of the study area (Fig. 5a). Commission errors ($n = 3382$; 28.6% of 11 816 wetland sites that were not converted) generally occurred in higher densities in the Coastal Plain, Piedmont, and Mississippi Basin, and in lower densities in the western and northwestern parts of the study area (Fig. 5b).

DISCUSSION

Multicollinearity among predictors

Appropriate interpretation of multiple regression coefficients requires that there be little multicollinearity among predictors. A $VIF \geq 10$ is commonly used as evidence of severe multicollinearity (Neter et al. 1989, Hair et al. 1998), but $VIFs \geq 2$ can indicate multicollinearity that may make interpretation of coefficients difficult (Graham 2003). The magnitudes of VIFs we observed indicated that multicollinearity was weak for most of the predictors. This result indicates that interpretations and inferences about the regression coefficients for most of the predictors

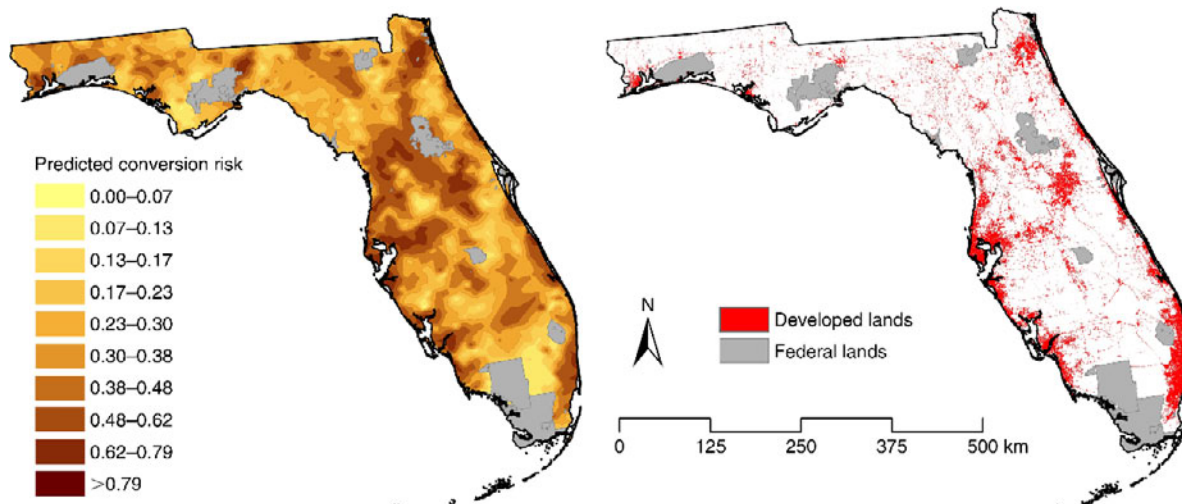


FIG. 3. Spatial correspondence between areas with high predicted risk of wetland habitat loss and urban areas in Florida. Gray polygons are federal lands, which were not sampled by the NRI.

would not be compromised appreciably by correlations among the predictors. Though none of the predictors exhibited multicollinearity that would be considered severe, some caution should be exercised in interpretations and inferences about predictors with moderate VIF magnitudes.

Relative predictive ability of variables

Based on importance values for variables in the best model, variables derived from the NRI inventory points for wetland sites (local predictors) and variables for all three spatial extents (landscape predictors) were important, but their relative predictive ability differed. Variables for the wetland site and for the fine extent generally had greater predictive ability than did variables for the meso and macro extents, and variables for the macro extent had greater predictive ability than did those for the meso extent. These results imply that the risk of wetland habitat loss was most strongly associated with conditions at or within 570 m of a wetland site, less associated with conditions within the macro (5130 m radius) extent, and least associated with conditions within the meso (1710 m radius) extent.

Among the variables that characterized site and fine-extent conditions, the land-cover/land-use type that the wetland point occurred in (BROAD_92) was by far the most important predictor. Formulas for basis functions revealed interaction effects between BROAD_92 and seven other variables (Table 2), indicating that associations between BROAD_92 and wetland habitat loss varied extensively with the values of other site, fine-extent, and macro-extent variables. The complexity of these relations makes it difficult to provide simple explanations for how broad land-use activities affected wetland conversion risk. However, for an example of how to interpret the statistical relations in general, and an interaction effect in particular, see *Understanding*

statistical relations, below. Five of the seven variables involved in interactions with BROAD_92 were site or fine-extent variables, indicating again that the association between wetland habitat loss and environmental conditions was dominated by local or fine-extent conditions.

Of the nine variables involved in the model, the size and proximity of development patches within the fine extent (F_GI_20), road density within the fine extent (F_RD_DEN), and the size and proximity of development patches within the macro extent (L_GI_20), were, respectively, the second-, fourth-, and sixth-most important predictors. Considering coefficient signs (see Eq. 1) and basis-function formulas (Table 2), WLOSS was positively associated with F_GI_20 (in BF2), F_RD_DEN (in BF8), and L_GI_20 (in BF68); however, WLOSS also was negatively associated with F_GI_20 (in BF3), F_RD_DEN (in BF9), and L_GI_20 (in BF51). The positive associations between WLOSS and the three variables are consistent with our observation that areas with high risks of wetland habitat loss occurred in and around urban areas (e.g., Fig. 3). These results also were consistent with earlier research that has documented an increasing incidence of human development as an agent of wetland conversion (Brady and Flather 1994, Dahl 2006, Hansen 2006). Negative associations between WLOSS and the three variables were manifested only for certain categories of BROAD_92 (see basis-function formulas), and this information may be valuable for guiding subsequent research that would explore possible reasons for the negative relations. Importantly, our approach permits a geographically explicit assessment of where conversion risks are predicted to be the highest. Our results also support previous studies that have shown that proximity to human development and accessibility (via roads) are important predictors of the probability of converting

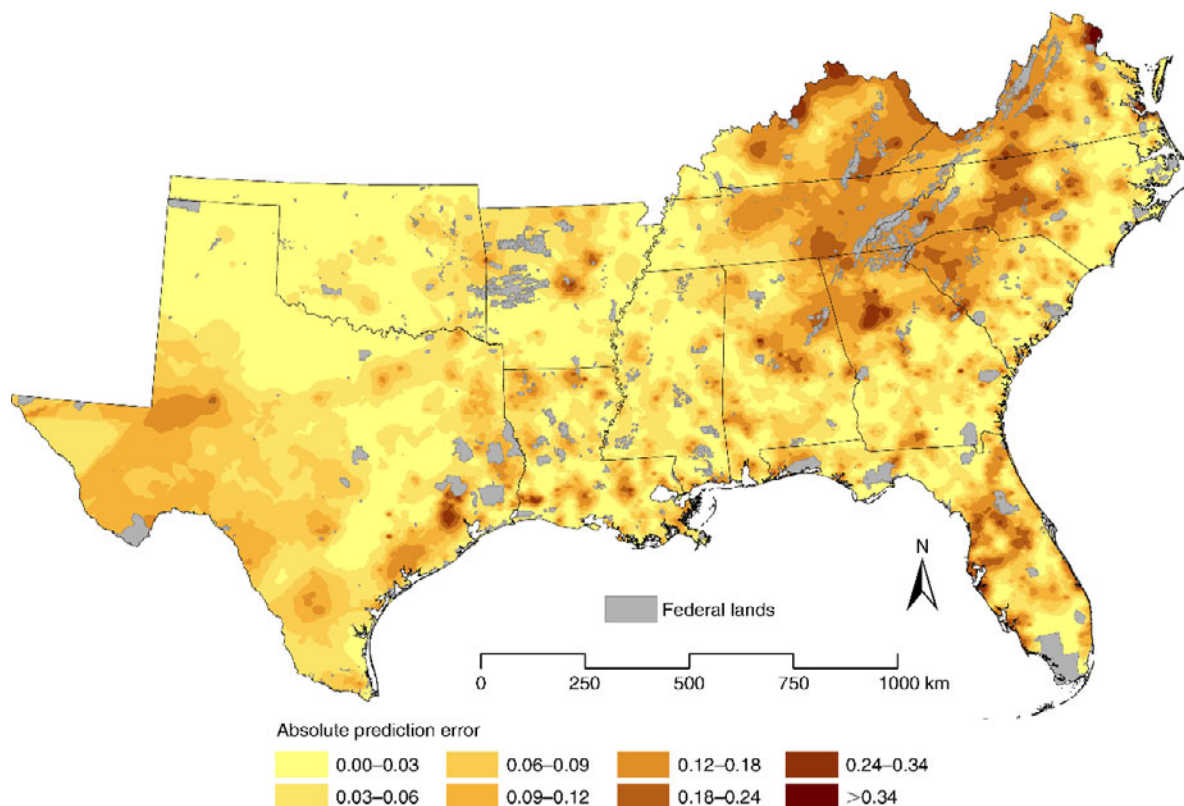


FIG. 4. Absolute errors in predicted risk of wetland habitat loss for the southern United States. Gray polygons are federal lands, which were not sampled by the NRI.

natural vegetation to an intensive land use, whether the focal habitat is forest (Kline et al. 2001) or wetland (Daniels and Cumming 2008).

Associations with topographic conditions (slope, position within a watershed, elevation) have been important in recent predictive models of wetland conversion (Koneff and Royle 2004, Daniels and Cumming 2008), yet neither slope (incorporated into EI92, the soil erodibility index, Table 1) nor elevation were important predictors of wetland habitat loss in the present analysis. It is possible that the proportion of area covered by woody and herbaceous wetland (reflected in basis functions involving F_{PR90} and M_{PR90} in the model) integrated important topographic conditions as well as development costs and flooding risks (see *Improving knowledge about possible causal factors*, below) and was therefore a more robust predictor for our study region than were slope and elevation per se. Recent studies of wetland conversion involved measures of topography that were computed for 10.4-km² grid cells (Koneff and Royle 2004) and 30-m pixels (Daniels and Cumming 2008), whereas in the present study, topographic variables were evaluated at a much finer spatial scale (wetland inventory point). Thus, their comparative unimportance here may also merely reflect these differences in scale.

The patchiness of wetland conversion processes across large landscapes has the potential to reduce the broad-scale predictive accuracy of a model of wetland habitat loss (see Daniels and Cumming 2008). Given the spatial extent of our study area, we thought that socioeconomic drivers of wetland conversion and their interaction with ecological conditions would manifest as regionally conditioned patterns of association, and we were surprised that we found no direct evidence for such relations. An important main effect of $BAIL_DIV$ (variable for ecological region) would have indicated that wetland habitat loss differed among two or more of the Bailey (1995) divisions in the study area. An important two-way interaction effect involving $BAIL_DIV$ would have indicated that the relation between wetland habitat loss and another predictor differed among two or more of the Bailey divisions. The modeling approach we used was robust: it made use of the entire sample, which would tend to provide better statistical power compared to fitting separate regional models, each with smaller sample sizes; our approach also involved a quantitative assessment of predictive importance of interactions involving $BAIL_DIV$. Despite our expectations and the rigor of our analysis, the MARS importance value for $BAIL_DIV$ was 0. This result suggested that, at least at the Bailey division level, whatever differences there may have been in conversion

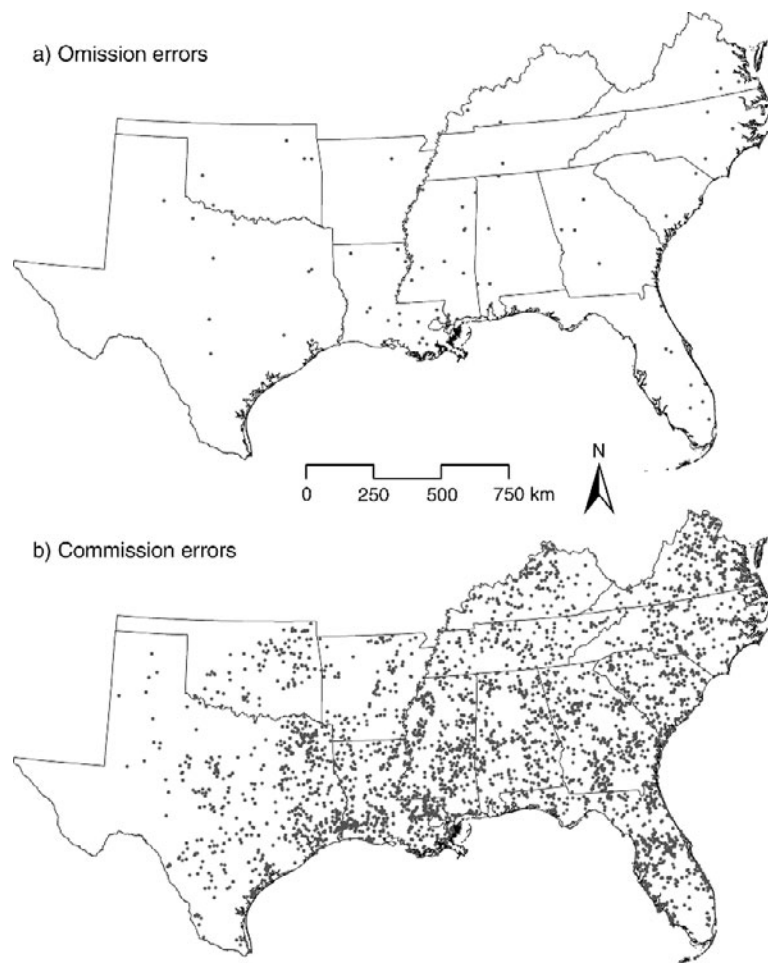


FIG. 5. Geographic distribution (dots) of (a) omission errors (wetland site was converted but predicted to be retained) and (b) commission errors (wetland site was retained but predicted to be converted) in the predicted risk of wetland habitat loss for the southern United States.

processes or associated ecological conditions among regions, these differences were not large enough for this variable to be a strong predictor of wetland habitat loss. An alternative explanation for our results may be that, because MARS considers local (in addition to global) statistical relations, it may have implicitly captured regional variation in drivers of wetland conversion through its specification of basis functions that did not involve BAIL_DIV.

Model performance

The principle of parsimony states that all other things being equal, simpler models should be favored over more complex models (Box and Jenkins 1970). Comparisons of AUCs indicated that the predictive performance of the simpler model we chose (10 basis functions, nine original variables) did not differ significantly from that for the more complex MARS optimal model (19 basis functions, 13 original variables). Correct classification, Kappa, and AUC statistics indicated that our model had real predictive power based on test set

validation across a broad geographic region. And the predictive ability of our model (mean AUC = 0.82, Table 3) compared favorably with that of recent predictive models for wetland conversion (Daniels and Cumming 2008; AUC = 0.81) and grassland conversion (Stephens et al. 2008; AUC = 0.79 and 0.77).

AUC statistics were higher for the unadjusted predictions than they were for the adjusted predictions, indicating that residual interpolation decreased rather than increased prediction accuracy. As noted by Miller (2005), the effectiveness of residual interpolation can be reduced when the proportion of events (typically coded as 1s) for a binary response variable is too low for adequate variogram fitting. In Miller's analyses, event proportions = 0.006–0.011 were generally associated with a loss or no improvement in accuracy, and event proportions > 0.028 were generally associated with accuracy improvement. In our data set, the proportion of events (WLOSS = 1) was 0.032, and our sample sizes were large, so it seems unlikely that the event proportion led to the decrease in accuracy we observed. A more

reasonable explanation for our results may be that any spatial patterns in wetland conversion risk were accounted for by the variables in the model, that the model's residuals contained little if any remaining predictive information derivable from their spatial relations, and that addition of the kriged residuals may have added noise to the predictions. The flexibility of MARS to fit linear, nonlinear, global, and local relations between WLOSS and our set of environmental predictors may have contributed to the absence of prediction improvement via residual interpolation.

Predictive maps

Predicted risks of wetland habitat loss (Fig. 2) were generally greater for highlands (Appalachian region and western parts of the study area) than they were for lowlands (Coastal Plains, Piedmont, Mississippi Basin). Because of their topographic and edaphic characteristics, highlands are likely to be better drained than are lowlands; in addition, the highland areas had fewer wetland points (Fig. 1). Wetlands situated in highlands may therefore be less extensive and more isolated than are wetlands situated in lowlands. Indeed, based on the NLCD for the macro-extent buffer, wetlands in highlands (BAIL_DIV = 1, 2, 4, 5) were smaller (mean wetland patch size = 6.8 ha) and farther apart (mean nearest neighbor distance = 332 m) than were wetlands in lowlands (BAIL_DIV = 3, 6; mean wetland patch size = 21.1 ha; mean nearest neighbor distance = 103 m). Compared to altering a wetland in a generally wet landscape (lowlands), there may be fewer financial costs and risks associated with converting a wetland in a better-drained or dry landscape (see *Improving knowledge about possible causal factors*, below).

Differences in the density of NRI wetland points across the study region may have influenced geographic variation in absolute prediction error. One would expect the magnitude of prediction error for a given area to be inversely proportional to the number of observations involved because smaller sample sizes for an area would result in less information about the wetland features and landscape context that influenced wetland habitat loss in that area. Our results support this idea. In the Coastal Plain, Piedmont, and Mississippi Basin, absolute prediction error was usually relatively low (Fig. 4) and densities of wetland points tended to be higher (Fig. 1), whereas in the Appalachian area, southwest Texas, and in and near major urban areas, prediction error was usually relatively high and the density of wetland points tended to be lower.

The map of commission errors (Fig. 5b) may do more than simply indicate the location of false positives—it also may indicate wetland habitats that may be predisposed to future conversion. These sites share characteristics with wetland habitats that were converted during the 1992–1997 period, but that for a number of reasons (e.g., regulations, development schedules, economic incentives, and lag effects) were not converted.

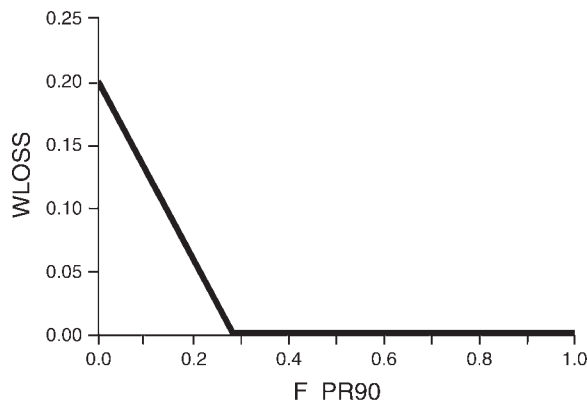


FIG. 6. Contribution to the non-normalized predicted risk of wetland habitat loss (WLOSS) in relation to the proportion of the fine extent covered by woody and herbaceous wetland (F_PR90) when the land cover and land use surrounding the wetland site was cropland, pastureland, rangeland, forestland, or other rural land (when BROAD_92 = 1, 2, 3, 4, or 5).

Whether commission errors do foreshadow future wetland habitat loss can be tested with the next cycle of NRI data. Wetland sites converted post-1997 would be expected to be disproportionately represented among those identified with commission errors in the present study.

Understanding statistical relations

To apply the model, it is important to understand the nature of the basis functions and their contributions to predicting the risk of wetland habitat loss. Many of the statistical associations are complex. Space does not permit us to examine the relations for all of the variables in the model, but we consider relations for one complicated predictor here to illustrate how interpretation of the model's statistical relations can be used to help understand how WLOSS was associated with environmental conditions. The statistical associations do not necessarily reflect causal relations.

WLOSS was positively associated with BF42 (regression coefficient = 0.714). The equation for BF42 was

$$\text{BF42} = \max(0, 0.284 - \text{F_PR90}) \times \text{BF28}$$

where BF28 was an indicator variable:

$$\text{BF28} = \begin{cases} 1 & (\text{if BROAD_92} = 1, 2, 3, 4, \text{ or } 5) \\ 0 & (\text{otherwise}). \end{cases}$$

BF42 therefore represented an interaction involving the proportion of woody and herbaceous wetland within the fine extent (F_PR90) and the broad land-cover/land-use type surrounding the wetland site (BROAD_92). The regression coefficient for BF42 represented the influence of BF42 on WLOSS after the influences of all of the other variables in the predictive model were taken into account.

When BF28 was 1, the contribution to WLOSS at F_PR90 = 0.0 was 0.20 because for F_PR90 = 0.0 and

BF28 = 1, BF42 = 0.284, and the regression coefficient (0.714) multiplied times BF42 (0.284) was 0.20 (Fig. 6). For BF28 = 1, the contribution to WLOSS from F_PR90 declined from 0.20 to 0.00 as F_PR90 increased from 0.00 to 0.284, and for values of F_PR90 > 0.284, the contribution of F_PR90 to WLOSS was 0 (Fig. 6). However, when BF28 was 0 (i.e., when BROAD_92 = 6 or 7), BF42 was 0, and this interaction effect did not contribute to WLOSS. That is, the relation between WLOSS and F_PR90 (Fig. 6) existed only when BF28 = 1. Thus, the influence of F_PR90 depended on BROAD_92, and F_PR90 contributed nonlinearly to WLOSS. The ecological significance of these relations is considered in the next section.

Improving knowledge about possible causal factors

Multicollinearity among predictors was not strong, and lack of improvement in prediction accuracy via residual interpolation indicated there were no missing or misspecified predictors that induced spatial dependence (see Miller 2005). These results provided confidence that the statistical relations embodied in the model can reasonably be used to posit hypotheses about causal agents of wetland habitat loss.

For instance, one explanation for the statistical relations involving WLOSS, BROAD_92, and F_PR90 examined in the previous section may be that, for wetland habitats situated in cropland, pastureland, rangeland, forestland, or other rural land, conversion was not economical once wetland coverage within the fine extent exceeded 28.4%, and conversion became increasingly economical (risk of wetland habitat loss became increasingly higher) as wetland coverage within the fine extent decreased from 28.4% to 0%. If much of the area around a given site was covered by wetlands, conversion may have been too expensive, or the chance of future flooding problems at the converted site may have been too high. At sites surrounded by low amounts of wetland within the fine extent, fewer modifications (draining, filling, levees) may have been necessary for conversion of the site and the adjoining area, or the chance of subsequent flooding at the converted site may have been lower. Thus, F_PR90 may have functioned as a proxy for economic costs and risks, suggesting that economic factors per se deserve explicit examination as possible causal agents.

A series of such interpretations and follow-up analyses for a given variable can be used to refine hypotheses about the factors that actually drive wetland habitat conversion in the South. The 10 basis functions and associated regression coefficients for our model provide a rich source of information for this refinement process.

Potential applications to conservation policy and planning

Because conservation resources are scarce, it is essential to focus on geographic areas where the risks for further wetland habitat loss are greatest. Our model

and associated predictive map of conversion risk can be used to inform the planning of protection efforts by helping to prioritize wetland areas for conservation. Wetland habitats with the highest conservation value and risk of conversion would receive the highest priority for acquisition or some other form of long-term protection, as Stephens et al. (2008) suggested for the conservation of grasslands. However, this approach would need to incorporate information about acquisition or protection cost to optimize conservation benefits from limited funds (Newburn et al. 2005). The maps of prediction errors can be used to incorporate prediction accuracy about wetland habitat loss into the decision process. A focus on areas where prediction errors were relatively low would help to reduce the chance of spending precious resources for minimal conservation benefit.

The risk of wetland habitat conversion can be used in broad-scale evaluations of wetland habitat connectivity. Wetland obligates, including many plants, amphibians, birds, and mammals, may benefit from landscapes in which wetland connectivity is high. A set of wetland sites may be functionally connected by being within the dispersal distance of a species. However, functional connectivity may be eroded over time as wetlands with a high risk of conversion get transformed into some other land use or land cover. Graph-theoretic methods (see Minor and Urban 2007) can be used to identify wetland habitats that are essential for maintaining connectivity while also accounting for differential conversion risk. In this context, the quality of wetland sites for maintaining connectivity would be inversely related to their conversion risk. Such analyses would provide conservationists with fundamental information needed to rank the value of individual wetlands or wetland complexes for maintaining wetland habitat linkages.

The model can be used to aid planning decisions concerning projected development. In our study region, and especially in Florida, urbanization and housing development increased at a greater rate than they did in any other region in the country from the early 1980s into the late 1990s, a pattern that is expected to continue well into the 21st century (Milesi et al. 2003, Alig et al. 2004, Crossett et al. 2004). Such land-use trends can be used to project future development. Our model included three variables that involved development (BROAD_92, F_GI_20, and L_GI_20). Planners can use our model in conjunction with land-development forecasts to anticipate where wetland conversion pressures may be increasing (or decreasing), and then use this information to guide development designs that may lessen wetland conversion pressures and ultimately reduce wetland loss.

Finally, our model can be used to assess the effectiveness of wetland conservation programs. The Wetland Reserve Program is an example of an important land-retirement program that authorizes the U.S. Department of Agriculture to purchase conservation easements to restore and protect wetlands (Williams

2005). But are the wetlands that landowners voluntarily enroll in this program those that face the greatest conversion risk? Characteristics of enrolled wetlands, including their location, can be used to estimate our model predictor variables, enabling prediction of wetland conversion risk for enrolled wetland habitats. If the land-retirement program was actually targeting at-risk wetlands, the average risk for enrolled wetland habitats would be expected to be higher than the average risk for unenrolled wetland habitats. In other words, our model can provide a means to test explicitly for what ecological economists call “additionality”—that conservation should target high-risk areas (e.g., wetland habitats with a high conversion risk) as opposed to areas exposed to limited threats (Naidoo and Ricketts 2006). Although such an exercise does not attain the ideal specified by Scodari (1997), who called for incorporating ecological values of a particular wetland, this analysis would objectively inform evaluations of wetland protection programs from the perspective of conversion risk.

Model improvement

At spatiotemporal resolutions that were relevant to the wetland systems we studied, we did not have access to data for other potentially important predictors of wetland habitat loss, such as zoning and the value of alternative land uses (Newburn et al. 2005), agricultural commodity prices (see Stephens et al. 2008), land prices, financial lending rates, cultural values that affect land-use decisions, or costs of construction equipment, materials, and labor. If such socioeconomic data become broadly accessible in a form that can be tied to specific geographic locales, a potentially important class of predictors that are absent from the present analysis would be available to wetland conservationists. Whenever feasible, we recommend inclusion of such predictors in analyses because they may improve the model's predictive ability and provide new insights about possible drivers of wetland conversion.

ACKNOWLEDGMENTS

We thank S. J. Brady for background information about the NRI database; M. Knowles for compiling data on candidate predictors and developing predictive maps; M. Hutcheson and F. Kincheloe for recommendations about data file management and computing systems; J. H. Friedman, R. King, A. Kolovos, D. Steinberg, D. S. Tolliver, and D. L. Turner for providing statistical advice; L. S. Baggett and L. S. Porth for assistance with SAS software; and S. J. Brady, A. E. Daniels, C. S. Findlay, S. E. Stephens, and an anonymous referee for reviewing the manuscript. This research was supported in part by funds provided by the Rocky Mountain Research Station, Forest Service, U.S. Department of Agriculture, through Joint Venture Agreements 05-JV-11221611-240 and 07-JV-11221611-291, and by Baylor University.

LITERATURE CITED

Alig, R. J., J. D. Kline, and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. *Landscape and Urban Planning* 69:219–234.

Bailey, R. G. 1995. Description of the ecoregions of the United States. Second edition. Miscellaneous Publication No. 1391

(revised). U.S. Department of Agriculture, Forest Service, Washington, D.C., USA.

Box, G. E. P., and G. M. Jenkins. 1970. Time series analysis: forecasting and control. Holden-Day, London, UK.

Brady, S. J., and C. H. Flather. 1994. Changes in wetlands on nonfederal rural land of the conterminous United States from 1982 to 1987. *Environmental Management* 18:693–705.

Cowardin, L. M., V. Carter, F. C. Golet, and E. T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. FWS/OBS-79/31. U.S. Fish and Wildlife Service, Washington, D.C., USA.

Crossett, K. M., T. J. Culliton, P. C. Wikley, and T. R. Goodspeed. 2004. Population trends along the coastal United States: 1980–2008. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Washington, D.C., USA.

Curriero, F. C., M. E. Hohn, A. M. Liebhold, and S. R. Lele. 2002. A statistical evaluation of non-ergodic variogram estimators. *Environmental and Ecological Statistics* 9:89–110.

Dahl, T. E. 1990. Wetland losses in the United States, 1780s to 1980s. U.S. Department of Interior, Fish and Wildlife Service, Washington, D.C., USA.

Dahl, T. E. 2006. Status and trends of wetlands in the conterminous United States 1998 to 2004. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C., USA.

Daniels, A. E., and G. S. Cumming. 2008. Conversion or conservation? Understanding wetland change in northwest Costa Rica. *Ecological Applications* 18:49–63.

Douglas, A. J., and R. L. Johnson. 1994. Drainage investment and wetland loss: an analysis of the National Resources Inventory data. *Journal of Environmental Management* 40:341–355.

Elith, J., and J. Leathwick. 2007. Predicting species distributions from museum and herbarium records using multi-response models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13:265–275.

ESRI (Environmental Systems Research Institute). 2006. ArcGIS 9.2. ESRI, Redlands, California, USA.

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.

Fleiss, J. L. 1981. Statistical methods for rates and proportions. Second edition. John Wiley, New York, New York, USA.

Freyer, W. E., T. J. Monahan, D. C. Bowden, and F. A. Graybill. 1983. Status and trends of wetlands and deepwater habitats in the conterminous United States, 1950s to 1970s. Colorado State University, Department of Forest and Wood Sciences, Fort Collins, Colorado, USA.

Friedman, J. H. 1991. Multivariate adaptive regression splines. *Annals of Statistics* 19:1–141.

Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. The National Elevation Dataset. *Photogrammetric Engineering and Remote Sensing* 68:5–11.

Government Accounting Office. 1998. Wetlands overview: problems with acreage data persist. GAO/RCED-98-150. Resources, Community, and Economic Development Division, Washington, D.C., USA.

Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84:2809–2815.

Hair, J. F., Jr., R. E. Anderson, R. L. Tatham, and W. C. Black. 1998. Multivariate data analysis. Fifth edition. Prentice Hall, Upper Saddle River, New Jersey, USA.

Hansen, L. 2006. Wetland status and trends. Pages 42–49 in K. Wiebe and N. Gollehon, editors. Agricultural resources and environmental indicators, 2006 edition. Economic Information Bulletin No. 16. U.S. Department of Agriculture, Economic Research Service, Washington, D.C., USA.

- Hosmer, D. W., and S. Lemeshow. 2000. Applied logistic regression. Second edition. John Wiley, New York, New York, USA.
- Isaaks, E. H., and R. M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York, New York, USA.
- Kanevski, M., and M. Maignan. 2004. Analysis and modeling of spatial environmental data. Marcel Dekker, Inc, New York, New York, USA.
- Kline, J. D., A. Moses, and R. J. Alig. 2001. Integrating urbanization into landscape-level ecological assessments. *Ecosystems* 4:3–18.
- Koneff, M. D., and J. A. Royle. 2004. Modeling wetland change along the United States Atlantic Coast. *Ecological Modelling* 177:41–59.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673.
- Maggini, R., A. Lehmann, N. E. Zimmermann, and A. Guisan. 2006. Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* 33:1729–1749.
- McGarigal, K., S. A. Cushman, M. C. Neel, and E. Ene. 2002. FRAGSTATS: spatial pattern analysis program for categorical maps. (<http://www.umass.edu/landeco/research/fragstats/fragstats.html>)
- Milesi, C., C. D. Elvidge, R. R. Nemani, and S. W. Running. 2003. Assessing the impact of urban land development on net primary productivity in the southeastern United States. *Remote Sensing of Environment* 86:401–410.
- Miller, J. 2005. Incorporating spatial dependence in predictive vegetation models: residual interpolation methods. *Professional Geographer* 57:169–184.
- Minor, E. S., and D. L. Urban. 2007. Graph theory as a proxy for spatially explicit population models in conservation planning. *Ecological Applications* 17:1771–1782.
- Mitsch, W. J., and J. G. Gosselink. 2007. Wetlands. Fourth edition. John Wiley and Sons, Hoboken, New Jersey, USA.
- Muñoz, J., and A. M. Felicísimo. 2004. Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15:285–292.
- Naidoo, R., and T. H. Ricketts. 2006. Mapping the economic costs and benefits of conservation. *PLoS Biology* 4:2153–2164.
- Neter, J., W. Wasserman, and M. H. Kutner. 1989. Applied linear regression models. Second edition. Richard D. Irwin, Burr Ridge, Illinois, USA.
- Newburn, D., S. Reed, P. Berck, and A. Merenlender. 2005. Economics and land-use change in prioritizing private land conservation. *Conservation Biology* 19:1411–1420.
- Nusser, S. M., F. J. Breidt, and W. A. Fuller. 1998. Design and estimation for investigating the dynamics of natural resources. *Ecological Applications* 8:234–245.
- Nusser, S. M., and J. J. Goebel. 1997. The National Resources Inventory: a long-term multi-resource monitoring programme. *Environmental and Ecological Statistics* 4:181–204.
- Parviainen, M., M. Luoto, T. Rytteri, and R. K. Heikkinen. 2008. Modelling the occurrence of threatened plant species in taiga landscapes: methodological and ecological perspectives. *Journal of Biogeography* 35:1888–1905.
- R Development Core Team. 2009. R: a language and environment for statistical computing, version 2.10.1. R Foundation for Statistical Computing, Vienna, Austria.
- Salford Systems. 2001. MARS user guide. Salford Systems, San Diego, California, USA.
- SAS Institute. 1999. SAS/STAT user's guide. Version 8. SAS Institute, Cary, North Carolina, USA.
- Schabenberger, O., and C. A. Gotway. 2005. Statistical methods for spatial data analyses. Chapman and Hall/CRC Press, Boca Raton, Florida, USA.
- Scodari, P. F. 1997. Measuring the benefits of federal wetland programs. Environmental Law Institute, Washington, D.C., USA.
- Stephens, S. E., J. A. Walker, D. R. Blunck, A. Jayaraman, D. E. Naugle, J. K. Ringelman, and A. J. Smith. 2008. Predicting risk of habitat conversion in native temperate grasslands. *Conservation Biology* 22:1320–1330.
- U.S. Department of Agriculture. 2000. Summary report: 1997 National Resources Inventory (revised December 2000). Natural Resources Conservation Service, Washington, D.C., USA, and Statistical Laboratory, Iowa State University, Ames, Iowa, USA.
- Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel. 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* 67:650–662.
- Watts, R. D., R. W. Compton, J. H. McCommon, C. L. Rich, S. M. Wright, T. Owens, and D. S. Ouren. 2007. Roadless space of the conterminous United States. *Science* 316:736–738.
- Williams, D. R. 2005. Agricultural programs. Pages 463–499 in K. D. Connolly, S. M. Johnson, and D. R. Williams, editors. Wetlands law and policy: understanding section 404. Section of Environment, Energy, and Resources, American Bar Association, Chicago, Illinois, USA.
- Zedler, J. B., and S. Kercher. 2005. Wetland resources: status, trends, ecosystem services, and restorability. *Annual Review of Environment and Resources* 30:39–74.

APPENDIX A

Details (including resolution, datum, and accuracy) for candidate predictor variables used in the modeling process (*Ecological Archives* A021-044-A1).

APPENDIX B

Information about multivariate adaptive regression splines (MARS) and methods used in model development (*Ecological Archives* A021-044-A2).

APPENDIX C

Summary statistics for response and predictor variables involved in the best predictive model for the risk of wetland habitat loss (*Ecological Archives* A021-044-A3).