



The relative contribution of terrain, land cover, and vegetation structure indices to species distribution models



John W. Wilson^{a,*}, Joseph O. Sexton^b, R. Todd Jobe^c, Nick M. Haddad^a

^a Department of Biology, North Carolina State University, Campus Box 7617, Raleigh, NC 27695, USA

^b Global Land Cover Facility, University of Maryland, Department of Geographical Sciences, 2181 LeFrak Hall, College Park, MD 20782, USA

^c Department of Geography, University of North Carolina, Campus Box 3220, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:

Received 9 February 2013

Received in revised form 1 April 2013

Accepted 20 April 2013

Available online xxxx

Keywords:

Butterflies

Distributions

Habitat assessment

Niche models

Remote sensing

Threatened species

ABSTRACT

Habitat assessments for biodiversity conservation are often complicated by the lack of detailed knowledge of a study species' distribution. As an alternative to resource-intensive field-based methods to obtain such information, remotely sensed products can be utilized in species distribution models to infer a species' distribution and ecological needs. Here we demonstrate how to arbitrate among a variety of remotely sensed predictor variables to estimate the distribution and ecological needs of an endangered butterfly species occurring mainly in inaccessible areas. We classified 19 continuous environmental predictor variables into three conceptually independent predictor classes, terrain, land cover, and vertical vegetation structure, and compared the accuracy of competing Maxent habitat models consisting of different combinations of each class. Each class contributed, though disproportionately, to our most reliable model that considered all 19 variables. We confirm that variables obtained from remote sensors can effectively estimate the distribution and ecological needs of a relatively unknown imperiled species occurring in inaccessible locations. Importantly, increasing the variety of predictor classes through multi-sensor fusion resulted in greater model accuracy than increasing the absolute number of predictor variables.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Considering that habitat loss is a primary driver of species extinctions, detailed habitat assessments are among the most important first steps guiding conservation efforts for imperiled species (Mace and Lande, 1991). Thorough habitat assessments are, however, often complicated by the lack of detailed knowledge of a threatened species' distribution, habitat status, and ecological needs (Anderson and Martinez-Meyer, 2004). Obtaining such information is not trivial. Threatened species are often sparsely distributed, hard to detect, and – due to biotic interactions, historical legacies, and dispersal barriers – not found in all suitable habitat patches (Pulliam, 2000). These qualities make it hard to separate unsuitable habitat from unoccupied suitable habitat (Gu and Swihart, 2004). In addition, time and monetary constraints typically prevent detailed bio-assessments that involve extensive surveys, experiments, and long-term demographic studies. Here we estimate the distribution and ecological needs of a relatively unknown imperiled species occurring in inaccessible locations, and, in doing so, develop methods to evaluate the contribution of a variety of

readily accessible, continuous remotely sensed predictor variables that may be incorporated into species distribution models.

To overcome the challenges associated with imperiled species' habitat assessments, ecologists employ species distribution models (SDMs) to estimate imperiled species' distributions (Elith and Leathwick, 2009). Using spatial data describing distributions and environmental characteristics, SDMs estimate the relationship between the study species' occurrences and the underlying environment. These approximations of the target species' environmental niche are then used to map suitable ecological conditions over an entire study region (Elith and Leathwick, 2009). Because they enable researchers to overcome the challenges associated with resource-intensive bio-assessments, and because of improved model reliability, SDMs have become increasingly popular among ecologists and conservationists (Elith and Leathwick, 2009).

Progress in remote sensing technologies has strongly complemented advances in SDMs. As an alternative to resource-intensive field-based methods, air- and space-borne sensors enable researchers to acquire reliable environmental data at scales relevant to SDMs in a consistent and repeatable way (Gillespie et al., 2008), even from poorly known and inaccessible areas (Raxworthy et al., 2003). Despite their utility, remotely sensed predictor variables remain underutilized in SDMs, possibly because the literature offers little guidance on appropriate datasets (Buermann

* Corresponding author. Tel.: +1 919 272 3522.

E-mail addresses: johnnybirder@gmail.com (J.W. Wilson), jsexton@umd.edu (J.O. Sexton), toddjobe@unc.edu (R. Todd Jobe), nick_haddad@ncsu.edu (N.M. Haddad).

et al., 2008) and interpretation of results obtained from remotely sensed data (Turner et al., 2003). Since the scale at which organisms perceive and interact with their environment is often much smaller than the scale at which many remotely sensed variables are obtained, concerns have also been raised as to whether remotely sensed data can be used to detect environmental variation at scales relevant to SDMs (Bistrat et al., 2011; Laurent et al., 2005).

The accelerating availability of diverse, remotely sensed products has generated questions about which and how many parameters to incorporate into model building. These parameters can be categorized into four conceptually independent remotely sensed predictor classes – terrain, (horizontal) land cover, (vertical) vegetation structure, and climate. Building on a previous effort that only considered land cover variables to track temporal habitat changes (Bartel and Sexton, 2009), we develop SDMs using a range of continuous remotely sensed predictor variables within three of these four remotely sensed predictor classes for an endangered butterfly, the St. Francis' satyr *Neonympha mitchellii francisci*. From these, we developed seven SDMs based on each predictor class independently, and in combination with one another. Four of our SDMs thereby utilized data from more than one sensor simultaneously, termed “multi-sensor fusion” (Hall and Llinas, 1997). Using our SDM results, we compared the performance of each SDM, blocked by data source, in predicting St. Francis' satyr presences. We also evaluated the relative contribution of each predictor variable to St. Francis' satyr distribution. In conducting our investigation, we developed an approach that tests significance of different classes of remotely sensed variables that should be generally applicable to arbitrate among competing models that could include various data inputs.

2. Material and methods

2.1. Study species

St Francis' satyr, globally restricted to early-successional wetlands situated on United States Department of Defense lands at Ft. Bragg, NC (35°07'S, 79°08'W, 65,032 ha), is an ideal species for a case study on SDMs utilizing remotely sensed data for a number of reasons. First, the species is listed as Endangered under the United States' Endangered Species Act because of its low population size and limited geographical range. Second, some previously healthy St. Francis' satyr subpopulations are currently in decline as once-suitable habitat transitions toward late-successional stages (Kuefler et al., 2008; Bartel and Sexton, 2009), creating an urgent need to assess the status of suitable habitat to determine the likelihood of population recovery. Third, our study area offers what we believe to be several suitable early-successional wetlands that support St. Francis' satyr's one known host plant, *Carex mitchelliana* sedges, which itself has very limited distributions. Yet, many of these patches remain unoccupied, raising questions about whether we can truly separate unsuitable from unoccupied suitable habitat. Fourth, much of the distribution of St. Francis' satyr falls within the restricted artillery impact zones at Ft. Bragg, where very limited and irregular access complicates efforts to confirm presences of this cryptic species with a short flight period (Kuefler et al., 2008). St. Francis' satyr is thereby representative of many other species whose life history is poorly described, and/or that live in inaccessible areas.

2.2. St. Francis' satyr occurrence

During 2008 we extensively (i.e. daily, during both month-long flight periods, Kuefler et al., 2008) searched for St. Francis' satyr butterflies in all known and accessible colonies ($n = 17$). For each

butterfly observed, we obtained Universal Transverse Mercator (UTM) coordinates using a WAAS-enabled Trimble Nomad 900GL Global Positioning System (GPS) unit (1–3 m accuracy). In total, 138 GPS points were obtained, all within 3 m of butterfly observations to maximize locational accuracy (Graham et al., 2008). Because of the temporary, successional nature of St. Francis' satyr habitat, we based habitat suitability models on locations where we saw St. Francis' satyr during one focal year, 2008 (the year for which we obtained Landsat data, see below).

2.3. Predictor variables

We tested the relative importance of three conceptually independent predictor classes of remotely sensed predictor variables – terrain, (horizontal) land cover, and (vertical) vegetation structure – in explaining St. Francis' satyr distributions (Table 1). We omitted a fourth class consisting of climate measures because such data are usually coarsely scaled (Turner et al., 2003) and thus more appropriate for regional or continental SDMs (Gillespie et al., 2008; Elith and Leathwick, 2009). While some interpolated (e.g. Thornton et al., 1997) and combined (e.g. Herman et al., 1997) climate measures exist, remotely sensed climatic predictor variables are rare, especially for terrestrial surfaces.

Terrain variables, derived from Digital Elevation Models (DEMs) (Li et al., 2005), play an important, though indirect, role in SDMs through their influence on climate (Moore et al., 1990) and vegetation (Franklin, 1995). Five continuous terrain predictor variables were used in this study, which included proxies for moisture (flow accumulation and slope), solar radiation (aspect), and topography (relative slope position and terrain shape, Moore et al., 1990). All terrain variables in this study were derived from the USGS National Elevation Dataset (Gesch et al., 2002), which we obtained at 1/3 arcsec resolution, resampled to 10 m resolution, and processed using tools contained in the ArcGIS Spatial Analyst and TauDEM v. 4.0 (Tarboton, 2009) packages.

Land cover predictor variables, obtained through passive optical multispectral sensors, are used to describe a study area's physiographic and physiognomic characteristics. Most often, land cover

Table 1

Estimates of variable importance of terrain, land cover and vegetation structure variables used to predict St. Francis' satyr presences using the Maxent software package (Phillips and Dudík, 2008).

Remotely sensed class	Variable	Variable importance (%)	Permutation importance (%)
Terrain	Slope	19.4	25.3
	Relative Slope Position	6.3	10
	Terrain Shape	0.5	0.6
	Aspect	0.6	0.2
	Flow Accumulation	0.5	0.2
Land cover	Deciduousness	24.1	41.3
	Summer brightness	9.3	2.3
	Wetness seasonality	1.6	2.2
	Brightness seasonality	1.3	1.2
	Winter greenness	2.0	1.1
	Summer wetness	5.4	0.6
	Winter brightness	0.8	0.4
	Winter wetness	0.9	0.4
	Summer greenness	15.7	0.3
Vegetation structure	Canopy density	2.0	8.5
	Understory density	3.0	2
	Shrub density	1.1	1.6
	Subcanopy density	3.6	1.4
	Midstory density	1.8	0.3

Variable importance is calculated heuristically and thus sensitive to collinearity and the order of variable importance. Permutation importance provides an alternative measure that is calculated from the AUC of the final model, and thus robust to the path of input variables.

variables are converted to categorized thematic land cover-landuse maps before use in SDMs (Laurent et al., 2005; Cord and Rödder, 2011). Recently however, continuous remotely sensed vegetation indices have gained some traction in SDMs (Sexton et al., 2006; Buermann et al., 2008; Cord and Rödder, 2011), though a range of other continuous land cover indices (e.g. for soil and wetness, Kauth and Thomas, 1976; Crist and Cicone, 1984) and land cover measures (e.g. thermal emissivity, Wang et al., 2008) remain largely neglected (although see Bartel and Sexton, 2009). Nine continuous land cover variables were used in this study. To obtain these variables, we georectified one cloud- and snow-free winter (January 2008) and one cloud-free summer (May 2008) Landsat-5 Thematic Mapper (TM) image of the study area (path/row: 16/36; resolution: 30 m). We then converted the six solar-reflective (non-thermal) TM bands (bands 1–5, 7) to radiance (Chander et al., 2007) and estimated surface reflectance using the DOS3 approach (Song et al., 2001). Surface reflectance values were then transformed to Tasseled-Cap “brightness,” “greenness” (Kauth and Thomas, 1976), and “wetness” indices (Crist and Cicone, 1984) in the ArcGIS 10 package. For each Tasseled-Cap index, our SDMs incorporated summer and winter values, and the difference between summer and winter values to characterize seasonality, or “deciduousness.”

Vegetation structure predictor variables are obtained primarily through active and passive optical remote sensing systems such as Interferometric Synthetic Aperture radar (inSAR) and Light Detection and Ranging (LiDAR) sensors (Lefsky et al., 2002). Though still under-utilized, continuous vegetation structure variables have recently been used more frequently in SDMs (Goetz et al., 2010; Wilsey et al., 2012). Five continuous LiDAR-derived vegetation structure variables were used in this study. To obtain LiDAR data for our study area, we contracted an airborne Optech ALTM 2050 LiDAR System to fly over Ft Bragg in early July 2006 (while deciduous trees were in leaf-on condition) at a height of 4500 ft and a flight speed of approximately 120 knots. This instrument produced a spot spacing of 1 m, horizontal accuracy of 0.3 m, and vertical accuracy of 95% within ± 0.2 m, from which we derived mean vertical density of vegetation in the understory (1–2 m above ground), shrub layer (2–5 m), midstory (5–10 m), subcanopy (10–20 m), and canopy (>20 m). We chose vegetation density as a predictor variable because it best characterizes the relative vertical structure of a habitat, an important aspect of habitat suitability for St. Francis' satyr (Kuefler et al., 2008). LiDAR data were processed at the 10 m resolution using a Maximum-Likelihood model (Moody et al., 2011).

2.4. Species distribution models

To model the distribution of St. Francis' satyr, we used the Maxent software package (Phillips and Dudík, 2008; Elith et al., 2011), as has been previously used to map St. Francis' satyr habitat (Bartel and Sexton, 2009). As a generative model, Maxent compares the environment underlying input presence points against a random sample of background points representing the availability and range of environmental conditions, and produces a raster map ranking each cell in the study area on an index representative of relative habitat suitability. Maxent accounts for interactions among predictor variables (Phillips and Dudík, 2008), and deals with model overfitting by employing the L1-regularization procedure (Hastie et al., 2009; Warren and Seifert, 2011), thus reducing the need to remove correlated predictor variables or preprocessing input before modeling (Phillips and Dudík, 2008; Elith et al., 2011). Originally designed for modeling rare and endangered species distributions, Maxent has gained popularity in part because it is less sensitive to the number of input presence locations (Wisz et al., 2008), and relies on presence-only data, thereby avoiding the

possibility of indeterminacy of habitat vs. non-habitat when faced with issues of imperfect detection (Guisan et al., 2006).

We developed seven Maxent models to estimate habitat suitability of St. Francis' satyr: one model combining all predictor variables from each of the three remotely sensed predictor classes (3 models), one model from each possible pair-wise combination of the three predictor classes (3 models), and one model combining all predictor variables from all three predictor classes (the full model). Developing these seven models allowed us to directly examine the ability of each remotely sensed predictor class, separately and in combination, to predict habitat suitability. Prior to Maxent modeling, all layers were clipped to Ft. Bragg's boundary, to maintain a consistent extent across all SDMs. We also artificially downscaled all land cover predictors, obtained at the 30-m resolution, to 10-m resolution to align predictor variables and avoid losing the fine-scale detail present in LiDAR and DEM-derived variables (at the 10 m resolution). For each Maxent model, we used 10,000 background points and the default settings recommended by Phillips and Dudík (2008) for features and regularization for model training (but see Warren and Seifert, 2011). To estimate errors for model performance evaluation, we implemented Maxent's built-in 10-fold cross-validation on each model to obtain 10 quasi-independent measures of a model's predictive ability (Hastie et al., 2009).

2.5. Model evaluation

We calculated two model performance measures that have been formulated for use in situations where absences are unavailable. First, for each model we calculated the mean area under the receiver operating characteristic curve (AUC) (plotting model sensitivity [errors of commission] against 1-specificity [errors of omission]): a random prediction will result in an AUC value of 0.5 whereas a perfect prediction assumes the maximum possible AUC of 1.0 (Fielding and Bell, 1997), and AUC values >0.75 are suitable for conservation planning (Pearce and Ferrier, 2000, but see Lobo et al., 2008). Second, we calculated mean omission error as the percentage of the withheld 10% test sample of presence points not predicted to fall within suitable habitat. To distinguish suitable from unsuitable habitat along Maxent's continuous habitat suitability index, we used the point on the AUC where model sensitivity is equal to specificity (sensitivity-specificity equality approach, Liu et al., 2005). However, because AUC is sensitive to the L1-regularization term, and omission error is highly sensitive to the amount of habitat predicted to be suitable, it cannot be used to compare model performance directly (Anderson et al., 2003). To allow for direct comparison of model performance, we calculated 'standardized omission error' based on binary habitat maps with the same amount of suitable area, which we set at the mean percentage of suitable area (0.08% of the total study area) predicted across all SDMs.

To evaluate the ability of each remotely sensed predictor class, separately and in combinations, to predict St. Francis' satyr presences, we compared our model performance measures (as dependent variables) against each of our seven groups of Maxent models (the factors) using two-way ANOVAs. For all significant ANOVA tests, *post hoc* pair-wise comparisons with Tukey's HSD tests were used to further investigate differences in model performance among the seven Maxent models. We also examined each SDM's heuristic estimates of how each predictor variable influenced the final model prediction (Phillips and Dudík, 2008). Lastly, we evaluated the importance of the number of predictor variables in SDM performance using ANCOVA analyses, with AUC values and standardized omission error as dependent variables, number of predictor variables as factors, and number of predictor classes as covariates.

Statistical analyses were conducted using R (R Development Core Team, 2011), and all means are reported ±1 Standard Deviation (SD).

3. Results

Overall, our SDMs performed well in predicting St. Francis' satyr habitat suitability. Mean AUC values across all models (0.95 ± 0.049 ; max possible is 1.0) fell within the range of excellent performance ($AUC > 0.9$, Pearce and Ferrier, 2000; Swets, 1988), while both mean omission rate (0.15 ± 0.06 ; 0.0 and 1.0 is best and worst respectively) and mean standardized omission rate (0.13 ± 0.12) were relatively low. All of our SDMs suggest that suitable St. Francis' satyr habitat is highly restricted, ranging from 0.03% to 0.22% of the total study area (Table 2). SDMs combining all three remotely sensed predictor classes performed best in predicting St. Francis' satyr presences; the full model (together with the terrain-land cover model) had the highest mean AUC, and the lowest mean standardized omission error. Conversely, each of the three one-class SDMs performed, on average, worse than those SDMs combining two and three classes, for both AUC and standardized omission error (Table 2). Despite predicting the greatest proportion of suitable habitat, the SDM based on only vegetation structure captured the least amount of the test sample of presence points, making it the worst performing model.

Comparisons of both AUC values ($F_{6,63} = 131.94$, $P < 0.001$, $r^2 = 0.926$) and standardized omission errors ($F_{6,63} = 63.20$, $P < 0.001$, $r^2 = 0.858$) indicated statistically significant differences in model performance among our seven Maxent models (Table 3). Notably, the SDM based on only terrain predictors performed significantly better than any other single-class model, while SDMs combining terrain variables with other classes performed significantly better than models without terrain variables. Lastly, the number of variables included in an SDM strongly influenced model

performance. After accounting for the number of classes included in each model, models with many variables had significantly higher AUC values than those with few variables (ANCOVA, $F_{1,66} = 7.66$, $P = 0.007$). However, we found no significant differences in standardized omission error values among models with different numbers of variables. Further investigation suggested that predictor class might be more important than absolute number of variables. Although not always statistically significant, the terrain-only SDM (5 predictor variables) was more reliable in predicting St. Francis' satyr presences than the land cover-only SDM (9 predictor variables) and land cover-vegetation structure SDM (14 predictor variables) (Table 2).

The permutation importance values (which are less sensitive to variable order than variable importance values) of our best-performing SDM (the full model) suggest that at least some predictor variables from all three remotely sensed predictor classes contributed to model reliability (Table 1). The most important predictor variables were: a land cover predictor, summer–winter greenness differences (41.3%), followed by two terrain predictors, slope (25.3%) and relative slope position (10%). Canopy density was the fourth-largest contributor (8.5%) to the final model prediction, and the only vegetation structure predictor contributing more than 5% of the overall prediction. The influence of these predictor variables on occurrence reflects St. Francis' satyr's presence in streams and floodplains (from terrain predictors) with semi-deciduous vegetation (from land cover) and an open canopy (from vegetation structure) (Fig. A1).

Competing SDMs resulted in considerable overlap in the extent of predicted suitable St. Francis' satyr habitat (Fig 1). Importantly, areas of suitable habitat predicted from SDMs combining two different predictor classes resembled the intersection of those two predictor classes when considered in isolation (Fig 1). This habitat prediction pattern was also observed when all three predictor classes were modeled together. Despite predicting the least amount of

Table 2

Mean values for each SDM's logistic threshold (specificity-sensitivity equality threshold approach), Area Under Curve (AUC) values, test sample omission error (calculated from a 10% test sample of presence points), amount of suitable habitat, and standardized omission error (0.08% of total area assumed suitable) for St. Francis' satyr using Maxent models. All means reported ±1 SD.

Predictor classes	Number of predictor variables	Logistic threshold	AUC value	Omission error (proportion)	Standardized omission error (proportion)	Suitable habitat (% of total area)
Terrain	5	0.22 ± 0.02	0.97 ± 0.01	0.07 ± 0.08	0.07 ± 0.05	0.07 ± 0.01
Land cover	9	0.21 ± 0.02	0.96 ± 0.01	0.12 ± 0.07	0.17 ± 0.06	0.08 ± 0.01
Vegetation structure	5	0.21 ± 0.03	0.83 ± 0.04	0.27 ± 0.06	0.41 ± 0.08	0.22 ± 0.02
One-class model means	–	0.21 ± 0.02	0.91 ± 0.08	0.18 ± 0.10	0.22 ± 0.16	0.12 ± 0.07
Terrain-Land cover	14	0.14 ± 0.02	0.98 ± 0.01	0.13 ± 0.08	0.04 ± 0.02	0.04 ± 0.00
Terrain-Veg. structure	10	0.16 ± 0.03	0.97 ± 0.01	0.12 ± 0.07	0.10 ± 0.04	0.06 ± 0.01
Land cover-Veg. structure	14	0.19 ± 0.01	0.97 ± 0.02	0.13 ± 0.07	0.08 ± 0.06	0.05 ± 0.01
Two-class model means	–	0.16 ± 0.03	0.97 ± 0.01	0.13 ± 0.07	0.07 ± 0.05	0.05 ± 0.01
Full model (all variables)	19	0.16 ± 0.03	0.98 ± 0.01	0.10 ± 0.06	0.03 ± 0.03	0.03 ± 0.01

Table 3

Difference in mean AUC values (above diagonal; ANOVA: $F_{6,63} = 131.94$, $p < 0.001$, $r^2 = 0.926$), and standardized omission error (suitable habitat fixed at 0.08% of study area, below diagonal; ANOVA: $F_{6,63} = 63.20$, $p < 0.001$, $r^2 = 0.858$) for Maxent models predicting St. Francis satyr occurrences, composed of terrain, land cover, and vegetation structure variables. Significant differences (Tukey's multiple comparisons, 2-tailed, 95%) are indicated with an asterisk. All means, calculated as rows minus columns, reported ±1 SD.

Predictor class	Terrain	Land cover	Veg. structure	Terrain – Land cover	Terrain – Veg. structure	Land cover – Veg. structure	Full model (all variables)
Terrain		0.01 ± 0.02	0.16 ± 0.02**	–0.01 ± 0.02	–0.00 ± 0.02	–0.00 ± 0.02	–0.01 ± 0.02
Land cover	0.09 ± 0.07**		0.14 ± 0.02**	–0.03 ± 0.02*	–0.01 ± 0.02	–0.01 ± 0.02	–0.03 ± 0.02*
Veg. structure	0.33 ± 0.07**	0.24 ± 0.07**		–0.17 ± 0.02**	–0.16 ± 0.02**	–0.16 ± 0.02**	–0.17 ± 0.02**
Terrain – Land cover	–0.04 ± 0.07	–0.13 ± 0.07**	–0.37 ± 0.07**		0.01 ± 0.02	0.01 ± 0.02	–0.00 ± 0.02
Terrain – Veg. structure	0.02 ± 0.07	–0.07 ± 0.07	–0.31 ± 0.07**	0.06 ± 0.07		–0.00 ± 0.02	–0.01 ± 0.02
Land cover – Veg. structure	0.01 ± 0.07	–0.08 ± 0.07*	–0.32 ± 0.07**	0.05 ± 0.07	–0.01 ± 0.07		–0.01 ± 0.02
Full model (all variables)	–0.04 ± 0.07	–0.13 ± 0.07**	–0.37 ± 0.07**	–0.00 ± 0.07	–0.07 ± 0.07	–0.05 ± 0.07	

* <0.05 Significance.

** <0.01 Significance.

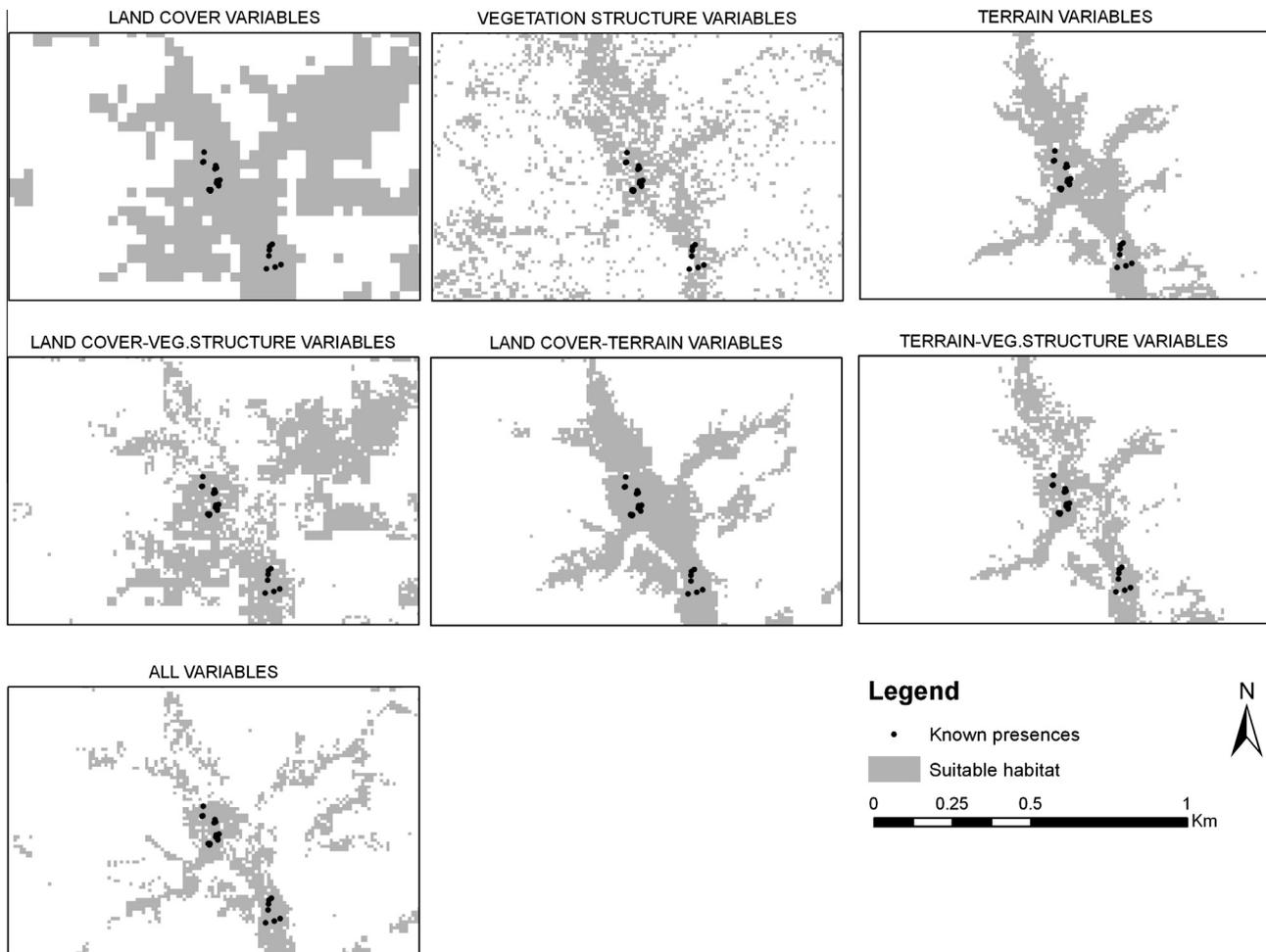


Fig. 1. Distribution of suitable habitat along an example drainage on Ft. Bragg, NC, to illustrate the differences in predictions between terrain, land cover, and vegetation structure variables in predicting St. Francis' satyr presences.

suitable habitat (Table 2), the SDM combining all predictor classes predicted suitable habitat towards the northeast; these areas were not predicted by the terrain-based SDM, and therefore a reflection of the land cover class' strong influence on the final prediction (Fig 1).

We found little correlation between our predictor variables (Table B1). Terrain variables provided the greatest amount of unique information, with only one (Pearson's correlation value of 0.54 for relative slope position vs. terrain shape) of 99 pair-wise comparisons providing a correlation >0.3 . Among the 85 pair-wise combinations of each vegetation structure and other variables, four combinations provided correlations >0.3 but <0.5 , while one comparison yielded a correlation greater than 0.5 (density of understorey vs. shrub layer: 0.59). Among the 90 pair-wise combinations of each land cover and other variables, eight correlations fell between 0.3 and 0.5, twelve correlations between 0.5 and 0.7, and six correlations between 0.7 and 1. However, apart from the six correlations between 0.3 and 0.5 (comprising comparisons between subcanopy density and each summer and winter land cover index), all values higher than 0.3 were from comparisons among different land cover indices.

4. Discussion

Our estimates of SDM reliability indicate that we were successful in exclusively using continuous remotely sensed predictor variables to model a rare species' habitat suitability. By comparing

model reliability of SDMs consisting of conceptually independent predictor classes, we also found that all remotely sensed classes contributed, though disproportionately, to final model predictions. Moreover, the variety of variables included in each SDM contributed more to model reliability than the number of variables. Our results, admittedly based on a single study case at a local scale, suggest that increasing the variety of continuous remotely sensed variables leads to significant improvements in SDM reliability. Further research is required to confirm whether our approach, which is highly flexible, is applicable to other species, spatial scales, modeling algorithms, and predictor variables.

Our study confirms that current remote sensors are able to provide environmental predictor variables relevant to SDMs (Laurent et al., 2005; Bradley and Fleishman, 2008; Bistrat et al., 2011). Despite challenges in interpreting remote sensing-based SDM output (Bradley and Fleishman, 2008; Cord and Rödder, 2011), remotely sensed datasets provide unbiased, high-resolution environmental data over larger areas through logistically and economically more efficient means than traditional field-based methods (Gillespie et al., 2008). Continuous remotely sensed products retain information along environmental gradients (Bradley and Fleishman, 2008; McGarigal et al., 2009; Cord and Rödder, 2011) and thus reduce classification and interpretation errors common in categorized predictor variables (Laurent et al., 2005; McGarigal et al., 2009). This is particularly beneficial when modeling species distributions at ecological transitional zones or in heterogeneous environments that are difficult to characterize using categorical or interpolated predictor variables (Álvarez-Martínez et al., 2010).

The availability of an up-to-date time-series of synoptic products (Turner et al., 2003) allows researchers to track temporal changes in habitat availability (Bartel and Sexton, 2009; Tuanmu et al., 2011), assess catastrophic and human-driven environmental changes (Tuanmu et al., 2011), as well as account for inter-annual and seasonal changes in the environment (this study, Cord and Rödder, 2011; Tuanmu et al., 2011). Furthermore, the consistency and near global coverage of many remotely sensed products may allow for greater model transferability over space, a major challenge of SDMs (Randin et al., 2006).

The choice of predictor variables significantly influences SDM predictions (Peterson and Nakazawa, 2008). We found that utilizing a greater variety of remotely sensed predictor variables produced more accurate models than solely increasing the number of variables, perhaps because more predictor classes capture a greater amount of environmental variation, including complementary attributes not detected in single-sensor models (Swatantran et al., 2012). Our results support this, particularly our maps showing how areas of suitable habitat predicted from multi-sensor fusion resembled the intersection of those predicted for each class when considered in isolation. The availability of multiple remotely sensed datasets allows modelers to simultaneously use climate, terrain, land cover, and vegetation structure classes with great ease over larger spatial scales (Gillespie et al., 2008), which in turn allows researchers to detect more subtle variations in a species' habitat than obtained from single-class models (Swatantran et al., 2012). It would be interesting to know how downscaled models would respond to the inclusion of remotely sensed climate predictors at broader scales, assuming that climatic predictors correlate with terrain predictors at local scales (Moore et al., 1990; Tarboton, 2009), and land cover predictors at coarse scales (Zimmermann et al., 2007).

To overcome potential confusion when comparing between increasing the variety of remotely sensed predictor variables and the absolute amount of variables, we categorized remotely sensed variables into four conceptually independent classes, of which we used three. Even so, much variability exists within different classes, complicating the highly subjective task of increasing predictor variable variety. Such decisions are important, because naively maximizing the amount of SDM variables may compromise SDM building through misleading results, while there is also a danger of over-fitting (Elith and Leathwick, 2009; Tuanmu et al., 2011). Two main schools of thought address the issue of model predictors (Elith and Leathwick, 2009; Warren and Seifert, 2011). The first school argues that the choice of variables should be made *a priori*, based on existing knowledge of ecophysiological and biophysical processes driving a species' distribution. The second school argues for maximizing the number of variables, and allowing the model to identify those that are important. Our results support aspects from both schools: maximizing the number of variables did not necessarily improve model predictions. However, the inclusion of environmental variables aimed at maximizing the types of environmental variation also improved our understanding of a relatively unknown species' relationship with its environment; such information may otherwise have been lost if those variables were omitted *a priori*. The difference between these two schools reflects subtle differences in intention (Elith and Leathwick, 2009). While the first school focuses on examining a species' response to a specific set of variables, the second school focuses more on prediction, especially for species for which little knowledge exist. So, in essence, the first approach avoids overfitting by placing greater emphasis on fundamental, invariant niche relationships, whereas the latter approach merely seeks empirical patterns.

While terrain and land cover variables are expected to strongly influence SDMs at local scales (Pearson and Dawson, 2003), we were surprised by the weak contribution of vegetation structure

to our model predictions. In fact, our full model hardly improved upon models that did not include (expensive) vegetation structure variables. Vegetation structure is generally important in SDMs (Kattwinkel et al., 2009). We also expected that vegetation structure would contribute to habitat suitability for St. Francis' satyr, as suitable habitat – early successional wetlands – is distinct from the surrounding landscape (Kuefler et al., 2008). We propose three explanations. First, and most likely, correlations among vegetation structure and other variables may reduce the predictive power of variables relating to vegetation structure (Swatantran et al., 2012). Though we detected little to no correlation among vegetation structure and other variables, it is possible that terrain and land cover variables together already captured similar aspects of the environment (see Bolstad et al., 1998; Hill and Thompson, 2005), leading to undetected correlations among variables. If correlations between vegetation structure and the combination of other sources of remotely sensed variables exist, then habitat suitability can be detected with high accuracy using freely available remotely sensed data (i.e. terrain and land cover variables) without the need to obtain expensive and sparsely distributed vegetation structure variables (Swatantran et al., 2012). Second, suitable St. Francis' satyr habitat changes rapidly (Kuefler et al., 2008); consequently there may be a mismatch between our vegetation structure variables, collected 2 years before our St. Francis' satyr presences, and St. Francis' satyr habitat we attempted to model. Third, vegetation structure alone may not play a prominent role in habitat suitability; Wilsey et al. (2012) found that vegetation structure improved SDMs of black-capped vireos, *Vireo atricapilla*, only when used in combination with other variables and in the absence of other vegetation variables.

In light of this, we show that continuous remotely sensed predictor variables, which offer many notable advantages over categorical variables, can be used exclusively and effectively in SDMs to estimate a species' distribution, habitat status, and ecological needs. While we have used only a small subset of potentially useful remotely sensed variables, a great number exist. Incorporating a wider range of remote sensing predictor classes into models can significantly improve model reliability, and detect highly subtle variations in a species' habitat needs. The application of continuous remotely sensed variables shows high promise to obtain direct, accurate, relatively inexpensive, and logistically feasible habitat assessments, even for little known species in remote and inaccessible terrain.

Acknowledgements

We thank the Ft. Bragg Endangered Species Branch and the Strategic Environmental Research and Development Program (SERDP) for financial support, Brian Ball and Erich Hoffman for facilitating our work, and Lesley Starke for comments on earlier drafts of this manuscript. JWW was supported on a NASA Earth and Space Science Fellowship.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.biocon.2013.04.021>.

References

- Álvarez-Martínez, J.M., Stoorvogel, J.J., Suárez-Seoane, S., de Luis Calabuig, E., 2010. Uncertainty analysis as a tool for refining land dynamics modelling on changing landscapes: a case study in a Spanish NaturalPark. *Landscape Ecol.* 25, 1385–1404.

- Anderson, R.P., Martinez-Meyer, E., 2004. Modeling species' geographic distributions for conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biol. Conserv.* 116, 167–179.
- Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Model.* 162, 211–232.
- Bartel, R.A., Sexton, J.O., 2009. Monitoring habitat dynamics for rare and endangered species using satellite images and niche-based models. *Ecography* 32, 1–9.
- Bistrat, S.A., White, M.A., Beard, K.H., Cutler, D.R., 2011. Predicting the distribution potential of an invasive frog using remotely sensed data in Hawaii. *Divers. Distrib.* 18, 648–660.
- Bolstad, P.V., Swank, W., Vose, J., 1998. Predicting southern Appalachian overstory vegetation with digital terrain data. *Landscape Ecol.* 13, 271–283.
- Bradley, A.B., Fleishman, E., 2008. Can remote sensing of land cover improve species distribution modeling? *J. Biogeogr.* 35, 1158–1159.
- Buermann, W., Saatchi, S., Smith, T.B., Zutta, B.R., Chaves, J.A., Milá, B., Graham, C.H., 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *J. Biogeogr.* 35, 1160–1176.
- Chander, G., Markham, B.L., Barsi, J.A., 2007. Revised Landsat-5 Thematic Mapper radiometric calibration. *IEEE Geosci. Remote Sens. Lett.* 4, 490–494.
- Cord, A., Rödder, D., 2011. Inclusion of habitat availability in species distribution models through multi-temporal remote-sensing data? *Ecol. Appl.* 21, 3285–3298.
- Crist, E.C., Cicone, R.C., 1984. Application of the Tasseled Cap concept to simulated Thematic Mapper data. *Photogramm. Eng. Rem. Sens.* 50, 343–352.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanations and prediction across space and time. *Annu. Rev. Ecol. Syst.* 40, 677–697.
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* 17, 43–57.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49.
- Franklin, J., 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geog.* 19, 474–499.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., Tyler, D., 2002. The national elevation dataset. *Photogramm. Eng. Rem. Sens.* 68, 5–11.
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modelling biodiversity from space. *Prog. Phys. Geog.* 32, 203–221.
- Goetz, S.J., Steinberg, D., Betts, M.G., Holmes, R.T., Doran, P.J., Dubayah, R., Hofton, M., 2010. Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology* 91, 1569–1576.
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., NCEAS Predicting Species Distributions Working Group, 2008. The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239–247.
- Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrences on wildlife-habitat models. *Biol. Conserv.* 116, 195–203.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E., 2006. Using niche-based models to improve the sampling of rare species. *Conserv. Biol.* 20, 501–511.
- Hall, D.L., Llinas, J., 1997. An introduction to multisensory data fusion. *Proc. IEEE* 85, 6–23.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inferences, and Predictions*. Springer-Verlag, New York.
- Herman, A., Kumar, V.B., Arkin, P.A., Kousky, J.V., 1997. Objective determined 10-day African rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.* 18, 2147–2159.
- Hill, R.A., Thompson, A.G., 2005. Mapping woodland species composition and structure using airborne spectral and LiDAR data. *Int. J. Remote Sens.* 26, 3763–3779.
- Kattwinkel, M., Strauss, B., Biedermann, R., Kleyer, M., 2009. Modelling multi-species responses to landscape dynamics: mosaic cycles support urban biodiversity. *Landscape Ecol.* 24, 929–941.
- Kauth, R.J., Thomas, G.S., 1976. The Tasseled Cap – A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In: *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*. Purdue University, West Lafayette, Paper 59.
- Kuefler, D., Haddad, N.M., Hall, S., Budgens, B., Bartel, B., Hoffman, E., 2008. Population structure and habitat use of the endangered Saint Francis satyr butterfly, *Neonympha mitchellii francisci*. *Am. Midl. Nat.* 159, 298–320.
- Laurent, E.J., Shi, H., Gatzliolis, D., LeBouton, J.P., Walters, M.B., Liu, J., 2005. Using the spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns. *Remote Sens. Environ.* 97, 249–262.
- Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. *Bioscience* 52, 19–30.
- Li, Z., Qing, Z., Gold, C., 2005. *Digital Terrain Modeling: Principles and Methodology*. CRC Press, Boca Raton.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28, 385–393.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17, 145–151.
- Mace, G.M., Lande, R., 1991. Assessing extinction threats: toward a reevaluation of IUCN threatened species categories. *Conserv. Biol.* 5, 148–157.
- McGarigal, K., Tagil, S., Cushman, S.A., 2009. Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecol.* 24, 433–450.
- Moody, A., Haddad, N., Morris, W.F., Walters, J., 2011. Habitat connectivity for multiple rare, threatened, and endangered species on and around military installations. Final report. Strategic Environmental Research and Development Program, Alexandria VA.
- Moore, I.D., Grayson, R.B., Ladson, A.R., 1990. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* 5, 3–30.
- Pearce, J., Ferrier, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecol. Model.* 128, 127–147.
- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimatic envelope models useful? *Global Ecol. Biogeogr.* 12, 361–371.
- Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modeling: an example with *Solenopsis invicta* and *Solenopsis richteri*. *Global Ecol. Biogeogr.* 17, 135–144.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31, 161–175.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- R Development Core Team, 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? *J. Biogeogr.* 33, 1689–1703.
- Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., Peterson, A.T., 2003. Predicting distributions of unknown reptile species in Madagascar. *Nature* 426, 837–841.
- Sexton, J.O., Ramsey, R.D., Bartos, D.L., 2006. Habitone analysis of quaking aspen in the Utah Book Cliffs: effects of site water demand and conifer cover. *Ecol. Model.* 198, 301–311.
- Song, C., Woodcock, C.E., Seto, K.C., Lenney, M.P., Macomber, S.A., 2001. Classification and change detection using Landsat TM data: when and how to correct for atmospheric effects? *Remote Sens. Environ.* 75, 230–244.
- Swatantran, A., Dubayah, R., Goetz, S., Hofton, M., Betts, M.G., Sun, M., Simard, M., Holmes, R., 2012. Mapping migratory bird prevalence using remote sensing data fusion. *PLoS ONE* 7, e28992.
- Swets, K.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Tarboton, D.G., 2009. TauDEM (Terrain Analysis Using Digital Elevation Models) Package. <<http://hydrology.neng.usu.edu/taudem/>>.
- Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorology variables over large regions of complex terrain. *J. Hydrol.* 190, 214–251.
- Tuanmu, M.-N., Viña, A., Roloff, G.J., Liu, W., Ouyang, Z., Zhang, H., Liu, J., 2011. Temporal transferability of wildlife habitat models: implications for habitat monitoring. *J. Biogeogr.* 38, 1510–1523.
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314.
- Wang, W., Liang, S., Meyers, T., 2008. Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sens. Environ.* 112, 623–635.
- Warren, D.L., Seifert, S.N., 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecol. Appl.* 21, 335–342.
- Wilsey, C.B., Lawler, J.J., Cimprich, D.A., 2012. Performance of habitat suitability models for the endangered black-capped vireo build with remotely sensed data. *Remote Sens. Environ.* 119, 35–42.
- Wis, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., The NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763–773.
- Zimmermann, N.E., Edwards Jr., T.C., Moison, G.G., Frenscino, T.S., Blackard, J.A., 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *J. Appl. Ecol.* 44, 1057–1067.