7-27-2001

New approaches for sampling and modeling native and exotic plant species richness

Geneva W. Chong  
*Colorado State University, Fort Collins*

Robin M. Reich  
*Colorado State University, Fort Collins*

Mohammed A. Kalkhan  
*Colorado State University, Fort Collins*

Thomas J. Stohlgren  
*Colorado State University, Fort Collins*

Follow this and additional works at: [https://scholarsarchive.byu.edu/wnan](https://scholarsarchive.byu.edu/wnan)

**Recommended Citation**
Available at: [https://scholarsarchive.byu.edu/wnan/vol61/iss3/9](https://scholarsarchive.byu.edu/wnan/vol61/iss3/9)

This Article is brought to you for free and open access by the Western North American Naturalist Publications at BYU ScholarsArchive. It has been accepted for inclusion in Western North American Naturalist by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Previous research suggests that habitats with high native plant species diversity can be more vulnerable to exotic plant species invasions than less species-rich areas (Stohlgren et al. 1998b, 1999a). To efficiently address threats posed by exotic species to native biodiversity, exotic plant species must be detected early. However, once an invasion has occurred, those populations and surrounding areas must be monitored to provide resource managers with the information needed to contain and control the exotic species.

Because only a small portion of any landscape can be affordably measured (usually <1%), predicting species occurrences or other features over the remainder of the landscape requires accurate multi-scale techniques (Stohlgren et al. 1997d). Most native and exotic plant species, rare habitats, and hot spots of diversity are patchy on most landscapes, and so they are usually missed by single-phase, single-scale transects and small plots (Stohlgren et al. 1998a). In addition, many inventory and monitoring attempts are hampered by unknown sources and amounts of error. For example, inventories based on resource maps must include an assessment of what information would be gained by using a map of higher resolution (smaller minimum mapping unit; Stohlgren et al. 1997b), since many coarse-scale maps fail to recognize rare but important habitats.

This paper describes a sampling and spatial modeling approach that can provide resource managers with a clearer picture of which areas and habitats are vulnerable to invasion by exotic plant species. This information can improve resource management decisions for control of exotic species as well as the inventory and monitoring of native and exotic plant species.

A Multi-species, Multi-scale, Multi-phase Approach

The importance of conserving biological diversity is recognized worldwide, and recognition of the benefits of conserving biodiversity at the ecosystem level, rather than the individual species level (Noss 1983, Agee and Johnson 1988, LaRoe 1993), has resulted in the need to identify areas for protection based on their biodiversity. Stohlgren et al. (1997d) outlined an approach to a landscape-scale...
assessments of plant diversity to complement the National GAP Analysis Program (Scott et al. 1993). The approach recognizes that the resolution of investigation is the 1st source of potential error in identifying important habitats for conservation. Common minimum mapping units (MMUs) of 100 ha, 50 ha, and 2 ha failed to identify rare but important habitats such as aspen stands in the Beaver Meadows area of Rocky Mountain National Park, Colorado (Stohlgren et al. 1997b). Aspen stands are a keystone habitat in this area with unique and rich assemblages of plants (Stohlgren et al. 1997c), birds (T. Mabee personal communication), and butterflies (Simonson 1998). Thus, a key feature of our approach is to sample rare and common habitats with a stratified random sampling design. Unbiased vegetation sampling sites are selected in each stratum (Stohlgren et al. 1997b).

A 2nd source of error in assessing patterns of biodiversity results from single-scale sampling techniques. A comparison of several common sampling techniques demonstrated that small, single-scale plot and linear transect techniques missed many locally rare species, both native and exotic (Stohlgren et al. 1998a). In addition, valid extrapolations to larger areas were impossible. The Modified-Whittaker nested vegetation sampling plot consists of a 20 × 50-m plot that contains ten 1-m² subplots (6 systematically arranged around the inside of the plot perimeter and 4 systematically arranged around the outside of the 100-m² subplot perimeter), two 10-m² subplots (in diagonally opposite corners of the plot), and one 100-m² subplot (in plot center; Stohlgren et al. 1995, 1998a). The multi-scale data allow one to estimate the number of species found in an area larger than the area sampled (Stohlgren et al. 1997c).

Stohlgren et al. (1997c) tested a rapid biodiversity assessment using multi-phase, multi-scale sampling in the Beaver Meadows area mentioned above. Multi-phase sampling refers to using ground-truth plots (Modified-Whittaker), aerial photos, and satellite images to sample a specific characteristic, such as vegetation cover, at overlapping locations (Kalkhan et al. 1995). These multiple layers of data allow assessment of the accuracy of satellite image vegetation classification, and classifications can be improved from multiple layers of data (Kalkhan et al. 1998). The multi-scale Modified-Whittaker vegetation plot sampling design allowed identification of hot spots of biodiversity and a reasonable estimate of the total number of plant species expected to be found in the study area.

**Data Comparability, Analysis, and Synthesis**

In addition to identifying where species of interest and hot spots of diversity occur, multiple threats to native species diversity must be recognized so that appropriate management strategies can be developed. Using comparable sampling methods allows both local and regional analyses and monitoring of species diversity, for example, across management units. Sampling designs and methods must be able to accurately assess the effects of a particular management action or potential resource threat.

For example, data collected using the Modified-Whittaker plot have proven valuable for assessing impacts and outcomes. A grazing study in Rocky Mountain grasslands demonstrated that vegetation composition differences inside and outside grazing exclosures could not be attributed to the effects of grazing alone because of landscape heterogeneity in vegetation distributions that had not been sampled in earlier studies (Stohlgren et al. 1999b). Vegetation sampling in the U.S. central grasslands and Rocky Mountains showed that exotic plant species are invading areas with highest native plant species richness and cover (Stohlgren et al. 1998b, 1999a). Modified-Whittaker plots arranged along transects that cross forest ecotones in Rocky Mountain National Park, Colorado, provided information on understory species richness and species distributions (Stohlgren et al. 2000) and may provide a means to monitor changes in regional climate (Stohlgren et al. 1998c).

In recognition of the strengths of multi-scale sampling, the U.S. Forest Service Forest Health Monitoring Program has modified its single-scale understory vegetation sampling method so that it is comparable to the multi-scale Modified-Whittaker plot (Busing et al. 1999). Grand Staircase–Escalante National Monument, Utah, is using the multi-phase, multi-scale approach to inventory its vascular plant diversity and soil crust development (Stohlgren et al. 1997a). The Smithsonian Institution’s Biodiversity Program has adopted the Modified-Whittaker vegetation sampling
design and successfully used the methods in Peru’s Amazon basin (Stohlgren and Chong 1997). Many other federal and non-federal resource managers are adopting multi-scale approaches to inventory and monitor biodiversity.

Predictive models developed from multi-scale data are an excellent example of data synthesis for resource management (Kalkhan et al. 2000). Modeling small-scale variability in landscape characteristics requires the generation of full-coverage maps depicting characteristics measured at points in the field (Reich and Bravo 1998). While many spatial data sets describing land characteristics have proven reliable for macro-scale ecological monitoring, these relatively coarse-scale data fall short in providing the precision required by more refined ecosystem resource models (Gown et al. 1994). Spatial statistics and geostatistics provide a means of developing spatial models that can be used to correlate coarse-scale geographical data with multi-scale field measurements of biotic and abiotic variables (Kalkhan and Stohlgren 2000).

In summary, we have developed an inventory and monitoring approach where the resulting data are useful for many different applications at various scales. In the remaining portion of the paper, we introduce some preliminary results from our spatial modeling approach as an example of data analysis and synthesis.

**METHODS**

**Field Data**

To demonstrate the model procedures discussed in this paper, we used Modified-Whittaker vegetation data (ninety-four 1000-m² plots) from a 54,000-ha portion of Rocky Mountain National Park, Colorado, USA. Sample points were located based on stratified random sampling in vegetation cover types ranging from wet meadow to alpine tundra (procedure described in Stohlgren et al. 1997c). This data set is used to develop preliminary spatial models to predict species richness (native and exotic) and presence/absence of exotic species in 30 × 30-m cells.

**GIS Data**

The GIS database used to develop the models contained several coverages of independent variables thought to influence variability in species richness and the presence of exotic species. These included a 30-m-resolution Digital Elevation Model (DEM; Department of Interior, U.S. Geological Survey), which was used to create a 30-m grid overlay of percent slope and aspect (GRID, ARC/INFO; ESRI 1997). The database also included 30-m-resolution overlays of Landsat TM bands 1 through 7. The point coverage of the sample data was used to extract point estimates of elevation, slope, aspect, and the digital numbers associated with the 7 Landsat bands (Table 1).

**Geostatistical Analysis**

Multiple regression analysis was first used to explore variation in species richness and presence/absence of exotic species as a function of geographical location, elevation, slope, aspect, and Landsat TM bands 1–7 (Fig. 1). Stepwise regression was used to identify the best linear combination of independent variables.

Residuals of the regression models were computed and used for modeling their semi-
variograms. Model parameters were estimated using weighted least squares (Cressie 1985). We also analyzed residuals for spatial autocorrelation and cross-correlation (Czaplewski and Reich 1993, Reich et al. 1994, Bonham et al. 1995) with geographical variables. Inverse distance sampling was used to define the spatial weights matrix.

Estimates of the residuals were obtained using ordinary kriging. To obtain estimates of species richness (native and exotic) and presence of exotic species, we added regression estimates based on elevation, slope, aspect, etc., and estimated residuals computed using ordinary kriging. Kriging was carried out using the 4 nearest neighbors.

Modified residuals kriging models were cross-validated to assess variability in prediction errors. Cross-validation included deleting a single observation from the data set and predicting the deleted observation using remaining observations in the data set. We repeated this process for all observations in the data set. Summary statistics of estimated values were computed. Accuracies of the kriging models were assessed using the relative mean-squared error suggested by Havesi et al. (1992).

Spatial Integration

The ability to spatially model field data allows integration over any specified geographical region (i.e., point- and plot-level field data, management unit, watershed, region) to obtain a point estimate and associated standard error of prediction. This is accomplished by integrating the 3-dimensional response surface representing the variable of interest over the area of interest and dividing by the area. Since spatially modeled response surfaces can be represented as a grid in ARC/INFO (ESRI 1997), any specified region will contain a finite number (n) of grid cells of uniform size (i.e., 30 × 30 m). Our point estimate of a resource in some bounded region, A, is obtained by summing the point estimates associated with each cell, \( \Phi_i \), and dividing by the number of cells in the bounded region. It is also possible to obtain estimates of variance. Resource managers can

Fig. 1. Flow diagram of statistical procedures.
use this information to determine which areas warrant further field data collection to increase model accuracy.

RESULTS

Regression Models

The regression model developed to describe variability in number of native species includes geographical location, elevation, and Landsat bands 1, 3, 5, 6, and 7 (Table 2). The positive correlation between elevation and number of native species suggests that species richness increases with increasing elevation. The number of native species was higher in northern and western portions of the study area. The significant Landsat TM bands provide information about differences in vegetation and soils throughout the study area and their influence on the richness of native and exotic species (Jensen 1996).

The regression model for number of exotic species includes geographical location, elevation, slope, aspect, and Landsat bands 2, 3, 5, 6, and 7 (Table 3). Exotic species were more prevalent in the southern and eastern portions of the study area and at lower elevations. The positive correlation with slope and the negative correlation with aspect indicate that exotic species are more prevalent on steeper, more northerly exposures. The positive correlation with number of native species indicates that exotic species are invading areas with high native plant species richness. This result agrees with the findings of Stohlgren et al. (1998b, 1999a) and Kalkhan and Stohlgren (2000).

The regression model to predict presence/absence of exotic species is similar to the one developed for number of exotic species (Table 4). The same factors that influence number of exotic species also influence probability of observing an exotic species.

The regression models accounted for 21% and 31% of variability observed in number of native and exotic species, respectively. The model developed to predict presence/absence of exotic species accounted for 38% of observed variability. Residuals of the regression models were positively spatially autocorrelated at the alpha = 0.05 level of significance. No significant cross-correlation was observed between residuals and independent variables used in developing the models. Residuals were approximately normally distributed.

### Table 2. Regression model used to explain large-scale spatial variability of number of native species (in 30 × 30-m cell) in a 54,000-ha area of Rocky Mountain National Park, Colorado. The x, y-coordinates are in meters (UTM coordinates).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>–454.30</td>
<td>0.030</td>
</tr>
<tr>
<td>X-coordinate</td>
<td>–0.0002</td>
<td>0.0</td>
</tr>
<tr>
<td>Y-coordinate</td>
<td>0.0001</td>
<td>0.015</td>
</tr>
<tr>
<td>Band 1</td>
<td>0.3482</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 3</td>
<td>–0.3936</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 5</td>
<td>0.1072</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 6</td>
<td>0.1452</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 7</td>
<td>–0.1683</td>
<td>0.002</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>0.0058</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*R² = 0.208, standard error = 4.22, n = 940 1-m² plots.

### Table 3. Regression model used to describe large-scale spatial variability in number of exotic species (in 30 × 30-m cell) in a 54,000-ha area of Rocky Mountain National Park, Colorado. The x, y-coordinates are in meters (UTM coordinates).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>83.422</td>
<td>0.065</td>
</tr>
<tr>
<td>X-coordinate</td>
<td>0.00003</td>
<td>0.0</td>
</tr>
<tr>
<td>Y-coordinate</td>
<td>–0.00002</td>
<td>0.039</td>
</tr>
<tr>
<td>Band 2</td>
<td>–0.0966</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 3</td>
<td>0.0560</td>
<td>0.007</td>
</tr>
<tr>
<td>Band 5</td>
<td>0.0386</td>
<td>0.0</td>
</tr>
<tr>
<td>Band 6</td>
<td>–0.0082</td>
<td>0.074</td>
</tr>
<tr>
<td>Band 7</td>
<td>–0.0528</td>
<td>0.0</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>–0.0011</td>
<td>0.0</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>0.0069</td>
<td>0.081</td>
</tr>
<tr>
<td>Aspect</td>
<td>–0.0011</td>
<td>0.066</td>
</tr>
<tr>
<td>Number native</td>
<td>0.0893</td>
<td>0.0</td>
</tr>
</tbody>
</table>

*R² = 0.314, standard error = 0.839, n = 940 1-m² plots.

### Table 4. Regression model used to describe large-scale spatial variability in probability of presence of exotic species (in 30 × 30-m cell) in a 54,000-ha area of Rocky Mountain National Park, Colorado. The x, y-coordinates are in meters (UTM coordinates).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>38.89</td>
<td>0.05</td>
</tr>
<tr>
<td>X-coordinate</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>Y-coordinate</td>
<td>–0.000</td>
<td>0.04</td>
</tr>
<tr>
<td>Band 2</td>
<td>–0.0415</td>
<td>0.001</td>
</tr>
<tr>
<td>Band 3</td>
<td>0.029</td>
<td>0.001</td>
</tr>
<tr>
<td>Band 5</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>Band 7</td>
<td>–0.019</td>
<td>0.001</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>–0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Slope (%)</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td>Aspect</td>
<td>–0.001</td>
<td>0.05</td>
</tr>
<tr>
<td>Number native</td>
<td>0.037</td>
<td>0.001</td>
</tr>
</tbody>
</table>

*R² = 0.383, standard error = 0.377, n = 940 1-m² plots.
Kriging

Model parameter estimates of the semivariograms for the 3 models are given in Table 5. The large range associated with residuals for native species suggests the presence of large-scale spatial continuity in number of native species across the study area. In contrast, the small range associated with exotic species models indicates that exotic species occur in small patches throughout the study area. The large nugget effect relative to the sill for these 2 models also suggests a considerable variation within these patches.

The modified residual kriging model for number of native species had a relative mean-squared error of 8.39 ($R^2 = 0.625$), while, in comparison, the regression model had a relative mean-squared error of 17.81 ($R^2 = 0.208$; Table 6). Kriging the residuals reduced the relative mean-squared error by 53%. The modified residual kriging model for number of exotic species had a relative mean-squared error of 0.501 ($R^2 = 0.506$), which represents a reduction in relative mean-squared error of 29%. Similar mean-squared errors were observed for the probability model of exotic species. Larger errors associated with exotic species models are due primarily to the small-scale spatial heterogeneity associated with the occurrence and density of exotic species. This small-scale spatial heterogeneity makes it difficult to predict spatial variability in presence/absence or number of exotic species at the 1000-m² plot scale.

DISCUSSION

We have outlined a comprehensive approach to sampling and modeling native and exotic plant species for natural resources management. Our approach provides an alternative to individual-based reaction-diffusion and spatially explicit simulation models and their assumptions and limitations (see Higgins et al. 1996). For example, the use of full-coverage, fine-scale variables (e.g., Landsat TM data with a 30 × 30-m resolution) is a valuable addition to spatial modeling and addresses problems relating to lack of empirical data and inappropriate scales that affect previously mentioned types of models. Also, our approach is based on current species locations, and so no direct assumptions are made about dispersal or autecology (see Kot et al. 1996)

The multi-phase, multi-scale sampling and modeling methods are easily modified for application across management units and even biomes and taxonomic groups (e.g., birds and butterflies as well as plants). The sampling methods are efficient and accurate in the field, and the development of automated data management and analysis tools will facilitate data use for local management as well as basic research and synthesis.

Data collected for a rapid assessment of plant diversity patterns were immediately useful for modeling native and exotic plant distributions across the landscape (Kalkhan and Stohlgren 2000). This information could be used by resource managers to set priorities and quickly target hot spots of exotic plant diversity for control efforts. Alternatively, or simultaneously, they might target areas of recent invasion where control efforts are relatively less expensive. Likewise, corridors of invasion, such as roads and riparian zones, might be targeted for control (Greenberg et al. 1997, Stohlgren et al. 1998b).

Additional variables, such as soil characteristics, will enhance the models’ predictive capabilities (Kalkhan and Stohlgren 2000). The combination of spatial statistics and stepwise multiple regressions greatly increased the predictive capabilities of our models for estimating the numbers of native and exotic species and the probability of encountering an exotic species. One of the strengths of this modeling approach is the ability to develop maps of “uncertainty” based on subsampling the data with Monte-Carlo simulations (Kalkhan et al. 2001).

### Table 5. Parameter estimates of the semivariograms used to describe spatial continuity in the residuals.

<table>
<thead>
<tr>
<th>Regression model</th>
<th>Nugget</th>
<th>Sill</th>
<th>Range</th>
<th>Semivariogram model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>15.94</td>
<td>1450</td>
<td>32704728</td>
<td>exponential</td>
</tr>
<tr>
<td>Exotic</td>
<td>0.763</td>
<td>1.117</td>
<td>151.4</td>
<td>Gaussian</td>
</tr>
<tr>
<td>Probably exotic</td>
<td>0.081</td>
<td>0.187</td>
<td>215.1</td>
<td>Gaussian</td>
</tr>
</tbody>
</table>
This provides land managers with a spatial representation of the confidence of the model and completeness of plot data. The multi-phase sampling approach (i.e., data from ground-truth plots, air photos, and Landsat TM images) provides additional ways to assess vegetation classification accuracy and determine where more ground-truth plots are needed (Kalkhan et al. 1998).

Similar models can be developed for individual species in more restricted areas (with a greater density of sample points) to better understand their ecology (where they are able to occur) and patterns of spread. We are developing spatial models for many common invasive plant species in Rocky Mountain National Park to better understand the effects of grazing, natural and prescribed fire, and rapid climate change on invasive plant species.

ACKNOWLEDGMENTS

Many excellent taxonomists and field technicians, cited in individual papers, made this work possible. Funding was provided by the Biological Resources Division of the U.S. Geological Survey. The staffs of Rocky Mountain National Park, the Midcontinent Ecological Science Center, and the Natural Resource Ecology Laboratory provided invaluable support. Michelle Lee, Kelly Rimar, and an anonymous reviewer provided helpful comments on earlier drafts of the manuscript. To all we are grateful.

LITERATURE CITED


REICH, R.M., R.L. CZAPLEWSKI, AND WA. RECHTOLD. 1994. Spatial cross-correlation in growth of undis-


Received 15 March 2000
Accepted 24 October 2000