

Spatial dependence of predictions from image segmentation: A variogram-based method to determine appropriate scales for producing land-management information

Jason W. Karl ^{a,b,*}, Brian A. Maurer ^b

^a Jornada Experimental Range, U.S.D.A. Agricultural Research Service, Las Cruces, NM, USA

^b Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA

ARTICLE INFO

Article history:

Received 30 July 2009

Received in revised form 18 February 2010

Accepted 19 February 2010

Keywords:

Object-based image analysis

Scale

Variogram

Kriging

Geostatistics

ABSTRACT

A significant challenge in ecological studies has been defining scales of observation that correspond to the relevant ecological scales for organisms or processes of interest. Remote sensing has become commonplace in ecological studies and management, but the default resolution of imagery often used in studies is an arbitrary scale of observation. Segmentation of images into objects has been proposed as an alternative method for scaling remotely-sensed data into units having ecological meaning. However, to date, the selection of image object sets to represent landscape patterns has been largely subjective. Changes in observation scale affect the variance and spatial dependence of measured variables, and may be useful in determining which levels of image segmentation are most appropriate for a given purpose. We used observations of percent bare-ground cover from 346 field sites in a semi-arid shrub-steppe ecosystem of southern Idaho to look at the changes in spatial dependence of regression predictions and residuals for 10 different levels of image segmentation. We found that the segmentation level whose regression predictions had spatial dependence that most closely matched the spatial dependence of the field samples also had the strongest predicted-to-observed correlations. This suggested that for percent bare-ground cover in our study area an appropriate scale could be defined. With the incorporation of a geostatistical interpolator to predict the value of regression residuals at unsampled locations, however, we achieved consistently strong correlations across many segmentation levels. This suggests that if spatial dependence in percent bare ground is accounted for, a range of appropriate scales could be defined. Because the best analysis scale may vary for different ecosystem attributes and many inquiries consider more than one attribute, methods that can perform well across a range of scales and perhaps not at a single, ideal scale are important. More work is needed to develop methods that consider a wider range of ways to segment images into different scales and select sets of scales that perform best for answering specific management questions. The robustness of ecological landscape analyses will increase as methods are devised that remove the subjectivity with which observational scales are defined and selected.

Published by Elsevier B.V.

1. Introduction

Scale is widely recognized as a critical attribute of ecological inquiries that not only defines what patterns and processes can be measured, but also influences observable relationships and governs the inferences that can be made from a set of data (Allen and Starr, 1982; O'Neill et al., 1986b, 1989; Wiens, 1989). In order for data to be useful for management decision-making, it must be collected and analyzed at spatial and temporal scales relevant to processes of interest to managers (O'Neill et al., 1986a) because different patterns

can emerge at different scales for almost any ecosystem (Wiens, 1989).

Scale is a characteristic of a set of observations, and the choice of scale constrains the patterns and processes that are observable (Burnett and Blaschke, 2003). In general terms, scale refers to the grain and extent of observations made in a study area where grain refers to the finest level of spatial and temporal detail observable and extent refers to the maximum area under consideration (Turner et al., 1989). Grain and extent define the upper and lower limits of inference because elements of patterns below the grain cannot be detected and inferences beyond the extent cannot be made without assuming scale-independent uniformity of patterns and processes (Wiens, 1989). Information at scales finer than the observation grain is filtered out and treated as noise, and information at scales larger than the observation extent is also filtered out and becomes context for

* Corresponding author. USDA ARS, Jornada Experimental Range, P.O. Box 80003, MSC 3JER, New Mexico State University, Las Cruces, NM 88003, USA. Tel.: +1 575 646 7015.

E-mail address: jkarl@nmsu.edu (J.W. Karl).

the information retained (Wu, 1999; Burnett and Blaschke, 2003). Thus, observations made with a set grain and extent (i.e., at a specific scale) have the effect of smoothing the dataset to give an interpretable message or signal (Allen and Starr, 1982; Burnett and Blaschke, 2003).

The concept of scale also has ecological interpretations where grain and extent can be defined in terms of how organisms or ecological processes respond to their environment. In this context, grain is the smallest area to which an organism responds to heterogeneity in its environment (Addicott et al., 1987; Kotlair and Wiens, 1990) while extent refers to the coarsest set of environmental patterns to which organisms or ecological processes react (Farina, 1998). From an ecological perspective, scale is relative to an organism or process and is not an inherent property of the environment (Wiens and Milne, 1989). Because natural landscapes are a complex product of many different processes and environmental factors (Peters et al., 2006) that are difficult to characterize when viewed from arbitrarily-selected spatial and temporal scales (Levin, 1992; Burnett and Blaschke, 2003), a significant challenge in ecological studies has been defining scales of observation that match the ecological scales affecting the organisms or processes of interest.

The use of remotely-sensed data from satellite imagery or aerial photography has become commonplace in ecological studies and management. Most studies employing remotely-sensed data have used the de facto scale determined by the image's pixel ground dimensions (i.e., resolution). In the context of the ecological definition of scale discussed above, the default resolution of imagery is an arbitrary scale of observation that may, or may not, correspond to the scale of ecological patterns and processes. Three potential problems may arise when, rather than basing scale selection on ecological processes or organisms of interest, scales are selected in an arbitrary manner or for convenience (Addicott et al., 1987): 1) it becomes difficult to compare results among studies because they each utilize units representing different scales, 2) a given observational unit may not correspond to an ecological patch appropriate for the variable of interest which could lead to misinterpreting relationships between events (see also Wiens, 1989), and 3) because different processes in the same system may occur at different scales, it can be difficult to know the ones to which observed patterns are related.

Scales different from the image resolution can be selected, either from convenience or through an analysis of scaling relationships, by identification of and estimation of sub-pixel features (Fisher, 1997; Foody, 2004) or more commonly by aggregating pixels into regularly shaped units (i.e., grids of square cells). Karl and Maurer (2010) found that correlations between images aggregated by regular, square grid cells and field measurements of semi-arid shrub-steppe ecosystems were unpredictable as the size of the cells increased. Also, as pixels (i.e., grain) become large relative to the extent, analysis results can be influenced by small shifts in the reference grid that defines the pixel boundaries (Jelinski and Wu, 1996).

Segmentation of images into polygons has been proposed as an alternative to pixel-based methods for scaling remotely-sensed data (Burnett and Blaschke, 2003; Wu et al., 2006). Image segmentation works by grouping similar, neighbouring pixels into discrete regions called objects in such a manner that the pixels within an object are more alike than they are to other pixels around the object (Burnett and Blaschke, 2003). Thus, segmentation is a rescaling of the original image data with the intent to produce a set of objects that match conditions on the ground (Woodcock and Harward, 1992). Taken together, a set of image objects that completely tessellate an area into discrete units constitutes one possible observational scale. There are many different methods for aggregating pixels into image objects (see Neubert et al., 2006), and within each method are parameters for controlling the scale that will result. Most significantly, parameters affecting the degree to which information that is filtered from the image—which in turn affects characteristics such as the average size of output objects—have significant effects on how the image information relates to ground observations and what the resulting objects can be used for. Up to some

segmentation level, the information that is lost is expected to be noise that is irrelevant to the analysis objectives. Past some threshold, important information is lost and the accuracy of analysis results may decline (Addink et al., 2007).

The choice of parameters used to aggregate pixels into objects, however, is subjective (Wang et al., 2004) and usually involves a trial and error process for deciding which set of objects best represents what the person doing the analysis interprets as meaningful patches on the ground (Burnett and Blaschke, 2003; Feitosa et al., 2006; Navulur, 2007). However, ecologists may be more likely to select intuitive, anthropocentric scales for studying systems rather than the ecological scales at which processes are actually happening (Wiens, 1989).

The optimal scale of segmentation is defined as the set of objects that results in the lowest prediction error of modelled parameters. Feitosa et al. (2006) used a genetic algorithm to iteratively find the combination of segmentation parameters yielding image objects that most closely matched pre-defined patches. Wang et al. (2004) used a multivariate distance measure, Battacharya distance, to determine the segmentation level that gave a mangrove forest classification that most closely matched training data. Addink et al. (2007), using a regression technique, found optimal levels of segmentation for estimating leaf-area index and biomass, and showed that segmentation levels that produced the best results were not the same for each variable.

Changes in scale can also affect the degree of spatial autocorrelation between observations (Cullinan et al., 1997; Fortin and Dale, 2005; Garrigues et al., 2006), but little research has been done on how spatial dependence (i.e., how autocorrelation changes as a function of distance between observations) changes with increasing levels of image segmentation. Geostatistical techniques like semivariance analysis may be useful for understanding how segmentation changes spatial dependence of remotely-sensed data and identifying optimal scales of image segmentation for a given task.

The objective of this study was to determine if, as observational scale increased through successively coarser image segmentation, changes in spatial dependence could be used to identify segmentation levels that most closely match field measurements and could be considered relevant scales for mapping ecosystem features. As an example, we used observations and predictions of percent bare-ground cover, an important attribute for assessing the condition of rangelands (National Research Council, 1994; Pyke et al., 2003), in a semi-arid shrub-steppe ecosystem of southern Idaho. Through comparison of linear regression models and residuals to semivariance calculated from field samples at successively coarser levels of image segmentation, we determined which sets of image objects best represent observed patterns of bare-ground cover. We discuss the benefits and limitations of our approach to defining scales through image segmentation and the implications of scale selection for providing data for land-management planning.

2. Study area

For this study we considered the Bureau of Land Management's (BLM) 97,308-ha Wildhorse grazing allotment in southern Idaho (Fig. 1, 43.028°N, 113.864°W). The majority of the study area is in public ownership with the Bureau of Land Management (BLM) being the largest single land steward—managing approximately 93,317 ha (95.8%) of the study area. Approximately 1305 ha (1.3%) of the study area is in private ownership, and 2843 ha (2.9%) managed by the state of Idaho. The dominant land use in the study area is cattle and sheep grazing.

The defining features of the Wildhorse landscape are expanses of unvegetated black rock created from a series of volcanic eruptions that blanketed this area with basaltic lava flows 15,000 to 2000 years ago (Owen, 2008). Soils in the study area are mostly aridisols with low organic matter and subsurface accumulations of either calcium carbonate or clay (Soil Survey Staff, 2006a,b). The terrain in the

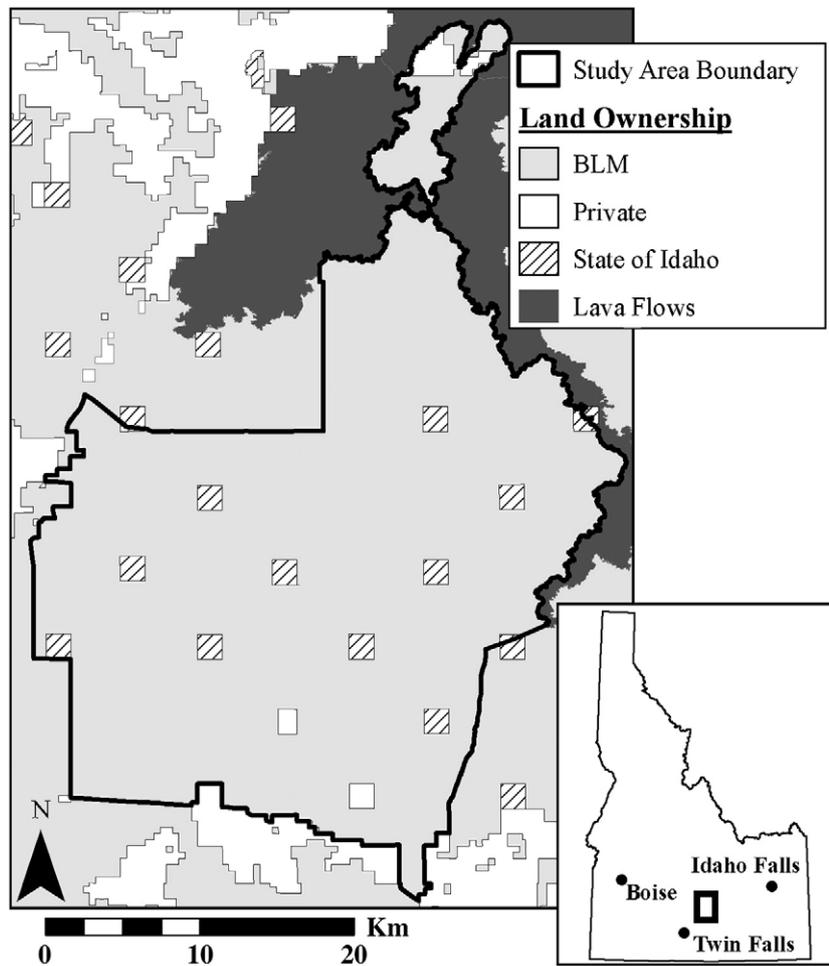


Fig. 1. Location and ownership of the Bureau of Land Management's Wildhorse allotment in southern Idaho.

Wildhorse allotment is flat to gently rolling with elevations ranging from 1272 m to 1557 m. Average annual precipitation ranges from 24.9 cm to 32.6 cm based on the PRISM map of average annual precipitation from 1971 to 2000 (PRISM Group, Oregon State University, <http://www.prismclimate.org>, created June 16, 2006).

Vegetation communities consist of a mosaic of mountain big sagebrush (*Artemisia tridentata* Nutt. ssp. *vaseyana* (Rydb.) Beetle), three-tip sage (*Artemisia tripartita* Rydb.), and Basin big sagebrush (*Artemisia tridentata* Nutt. ssp. *tridentata*) types. Native understory composition is principally bunchgrasses: bluebunch wheatgrass (*Pseudoroegneria spicata* (Pursh) A. Löve), Idaho fescue (*Festuca idahoensis* Elmer), and bottlebrush squirreltail (*Elymus elymoides* (Raf.) Swezey). Cheatgrass (*Bromus tectorum* L.) abundance is variable within the study area, reaching high densities in areas that have frequently burned and sites disturbed by intensive livestock use (e.g., near watering troughs, corrals, and loading areas).

This study area has seen an active fire history with 18 wildfires within the last 20 years, and eight of those being greater than 200 ha. Over the last 20 years, 80% of the Wildhorse allotment has burned. The frequent, large fires in this area have contributed to the spread of cheatgrass and other invasive species in the allotment.

3. Methods

3.1. Field data collection

We used data collected by the Shoshone District BLM (G. Mann and J. Russel, unpublished data) from 468 field observations of percent cover acquired between June 21, 2006 and August 6, 2008. Percent cover of

vegetation and soil surfaces was estimated for each site from a single 15.15 m transect using the line-point intercept method of Keane et al. (2005). The starting location of each transect was recorded with a GPS and differentially corrected.

We excluded 122 of the field observations due to missing data or fires that occurred between when the field measurements were taken and satellite imagery was acquired. Percent bare-ground cover was calculated on the remaining 346 sites. For the purposes of rangeland assessment and monitoring, bare ground is considered land surface not covered by vegetation, rock, or litter (Bedell, 1998; Pellant et al., 2005). Percent bare-ground cover was calculated as the proportion of the transect points where no plant canopy was intercepted and the soil surface was recorded as exposed soil (Herrick et al., 2005). Because regression analyses and kriging methods are sensitive to data normality (Cressie, 1993; Bailey and Gatrell, 1995), we used a square-root transformation to achieve normality of the percent bare-ground measurements.

3.2. Image acquisition and pre-processing

We acquired a Landsat Thematic Mapper (TM) 5 scene from July 11, 2008. We ortho-rectified and geo-registered it to have a nominal ground resolution of 30 m using publically-available 1/3 arc-second digital elevation models and 1-m resolution color aerial photography. We atmospherically corrected the image using Chavez's (1996) dark-object-subtraction method and converted the 8-bit image values to estimated at-sensor reflectance. Because image segmentation is sensitive to correlations between bands (Navulur, 2007), we used Crist's and Kauth's (1986) tasseled-cap transformation to convert the original six TM bands into four nearly orthogonal bands.

We additionally calculated a modified soil-adjusted vegetation index (MSAVI)—an index of vegetation greenness that is designed to be sensitive to changes in vegetation cover in areas with high amounts of exposed soil or rock (Qi et al., 1994; Gilabert et al., 2002). The MSAVI was not used in the segmentation process because it correlates strongly with several of the tasseled-cap bands, but was used in subsequent regression analyses because previous research has shown that vegetation indices correlate well with percent cover of vegetation and bare ground.

3.3. Image segmentation

All image segmentation was done with Definiens Developer 7.0 (<http://www.definiens.com>) using the multi-resolution segmentation algorithm described by Baatz and Schäpe (2000) and Burnett and Blaschke (2003). Multi-resolution segmentation works by merging adjacent pixels in the first iteration and objects in later iterations and evaluating the increase in local heterogeneity. If after merging the local heterogeneity is below a set threshold, the merged objects are retained, otherwise the merge is not kept and a different combination of objects is tried. This process continues until all possible merges below the threshold are made. The multi-resolution segmentation method requires the user to set the threshold value through a unitless scale parameter (p_s). By increasing p_s , the acceptable level of pixel heterogeneity within objects is increased and, on average, segmentation produces larger objects. It is important to note that multi-resolution segmentation does not specify a minimum object size and groups of pixels that are very distinct will be maintained as separate objects while the surrounding objects increase in size. For our analyses, we varied only p_s in multi-resolution segmentation and used default values for all other parameters.

We segmented the tasseled-cap-transformed image for the Wild-horse allotment multiple times to create a set of image segmentations that became progressively coarser in scale. For each successive segmentation run, we incremented p_s by a small amount (Table 1). The image objects for each segmentation level were attributed with the mean and standard deviation of the pixels within the object for all the tasseled-cap bands and the MSAVI and then exported for statistical analysis. For each segmentation level, the output variables were assessed for normality and transformed as necessary using either square-root or log transformations.

3.4. Statistical analysis

All statistical analyses were performed in R version 2.4.2 using the *nlme* (Pinheiro and Bates, 2000) and *gstat* (Pebesma, 2004) packages.

To assess the spatial dependence of the field measurements of percent bare-ground cover, we constructed a semivariogram between

all pairs of observations following Fortin and Dale (2005). We fit an omni-directional, spherical semivariogram model to the percent bare-ground variogram using the ordinary least-squares (OLS) regression method in R's *gstat* package (Pebesma, 2004). The semivariogram model was characterized by its nugget (i.e., variability at distances smaller than the shortest distance between sample points including measurement error), sill (i.e., total observed variation of the variable), and the range (i.e., distance at which two observations could be considered independent) (Fig. 2). The nugget-to-sill ratio (NSR) of a variogram is a commonly used measure of the proportion of the total observed variation that cannot be explained by observed spatial dependence of the variable (Kravchenko, 2003).

We then used a combination of generalized least-squares regression (GLS; Pinheiro and Bates, 2000) and regression kriging (RK; Hengl et al., 2004) to predict percent bare-ground cover from the tasseled-cap band and MSAVI values summarized in the image object polygons at the different segmentation levels.

The first step for each segmentation level was to select the image objects that contained one or more field sample locations and extract them to a new dataset. Coordinate values were also included for the geometric center of each polygon. When more than one sample location fell within an image object polygon, the field-measured percent bare ground for each selected image object was determined by averaging the measurements of all sites within the polygon. We used this method to avoid artificially inflating sample sizes of objects as segmentation levels increased and because it was akin to taking multiple samples within an object.

The second step at each segmentation level was to use GLS regression to establish the relationship between the field measurements of percent bare ground and the image object values and predict percent bare-ground cover. We included all 10 image-band measures (i.e., mean and standard deviation of pixels per object for the four tasseled-cap bands and MSAVI), and the coordinate values in an initial regression model. The only interaction term we considered was between the X and Y coordinate values. Models were created separately for each segmentation level, and we used a backward stepwise procedure to select the most parsimonious model at each level. We used GLS in order to incorporate the spatial covariance between samples (Bailey and Gatrell, 1995). Spatial covariance is incorporated into GLS in an iterative fashion via a semivariogram model found from the residuals of an initial OLS regression model (Hengl et al., 2004). The regression is then rerun with GLS using the semivariogram model to specify covariance of the samples.

The predictions of the GLS regression model at each segmentation level were assessed using leave-one-out cross-validation. In this process, one observation, selected at random, was withheld from the regression and the predicted value for that location compared to the observed value. The omitted point was then replaced and another observation randomly selected. This procedure was repeated 100 times and the results were used to estimate a correlation between predicted and observed measures of percent cover. We employed this

Table 1
Summary information for the ten successively coarser segmentation levels used in this study.

Segmentation scale parameter (p_s)	Number of image objects	Median area (ha)	Minimum area (ha)	Maximum area (ha)	Variance of object means (between-object variance)	Mean within object variance
5	10,452	5.77	0.09	491	0.0159	0.004583
10	2660	21.53	0.63	1401	0.018545	0.005676
15	1185	46.7	0.72	1618	0.021825	0.006299
20	694	74.09	2.07	2344	0.026782	0.006743
25	491	117.24	2.97	2129	0.028742	0.007078
30	351	170.29	3.69	2204	0.032919	0.007329
35	267	228.62	7.20	2219	0.035934	0.007505
40	226	282.25	10.44	7640	0.038275	0.00772
45	186	324.92	10.44	3976	0.039849	0.0079
50	162	393.66	10.44	6230	0.04	0.007997

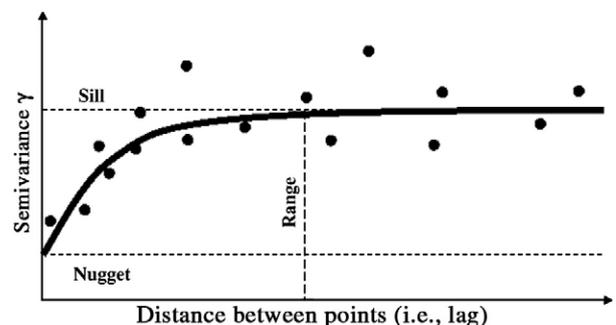


Fig. 2. Example of a sample (empirical) semivariogram (black points) and the variogram model (heavy solid line).

method to have a measure for GLS regression that was comparable to what was available for RK (below).

We hypothesized that segmentation levels yielding the best GLS regression results did so because the image objects minimized spatial dependence between objects. If this were the case, then the predictions of the GLS model for the sample locations would have a similar spatial dependence to the field measurements of percent bare-ground cover and the residuals of the GLS model would show little spatial dependence. To test this, we constructed variograms for the predicted percent bare ground and the residuals of the bare-ground prediction at each segmentation level and compared them to the variogram of the field sample locations by way of their NSR and range.

If substantial spatial dependence exists in regression model residuals, then geostatistical techniques can be employed to improve the predictions of the model. The RK method was developed as a way to exploit spatial autocorrelation in datasets to increase the accuracy of predictions (Odeh and McBratney, 1994; Odeh et al., 1995; Hengl et al., 2004; Karl and Maurer, 2010). Regression kriging uses GLS regression to predict a response variable and then uses the geostatistical technique of simple kriging (Krige, 1966; Goovaerts, 1997; Hengl et al., 2004) to predict the value of the GLS residuals at unsampled locations. The RK predictor for a variable at an unmeasured location, $\hat{z}(s_0)$, is the sum of the GLS regression prediction and the predicted residual:

$$\hat{z}(s_0) = \sum_{k=0}^p \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^n \lambda_i \cdot e(s_i) \quad (1)$$

where the $\hat{\beta}_k$ and $q_k(s_0)$ are the regression coefficient and the value of the k th predictor variable at the unknown location, respectively, the λ_i are the kriging weights determined from the i known locations using a variogram model of the GLS-model residuals, and $e(s_i)$ is the GLS-model residual value at point i (Hengl et al., 2007). All RK predictions and cross-validation were accomplished with the gstat package in R (Pebesma, 2004).

The objective of implementing RK in addition to the GLS modeling was to determine if added predictive capability could be gained through explicitly incorporating spatial dependence into the predictors. At each segmentation level, we used the GLS model and the residuals variogram model developed above to create RK predictions of percent bare-ground cover. We assessed the RK predictions using the same cross-validation procedure described above to derive a correlation between predicted and observed percent bare-ground cover. For each segmentation level, we compared the predicted-vs-observed correlations for the GLS and RK predictions.

4. Results

The empirical variogram constructed for percent bare-ground showed that much of the observed variability in the field measurements could be explained by distance between observations (Fig. 3). The variogram model we fit had spatial dependence of bare-ground cover extending to a range of 14,719 m. The nugget and sill of the variogram model were 0.0036 and 0.0147, respectively, with the NSR being 0.2474—meaning that about one-fourth of the observed variation in percent bare-ground cover could be considered short-range or within plot variability that could not be explained by the model of spatial dependence.

Segmentation of the TM image into successively coarser scales resulted in an exponentially-decreasing number of image objects with scale (Table 1). As p_s increased, the size of the objects also increased. Within any segmentation level, the distribution of object sizes was not normal, and the typical area was best characterized by the median object size (s_{med}). Median object size increased roughly with the square of p_s ($s_{\text{med}} = 0.305 * p_s^{1.85}$, $R^2 = 0.999$). Minimum object area (s_{min}) also increased with the square of p_s ($s_{\text{min}} = 0.0034 * p_s^{2.11}$,

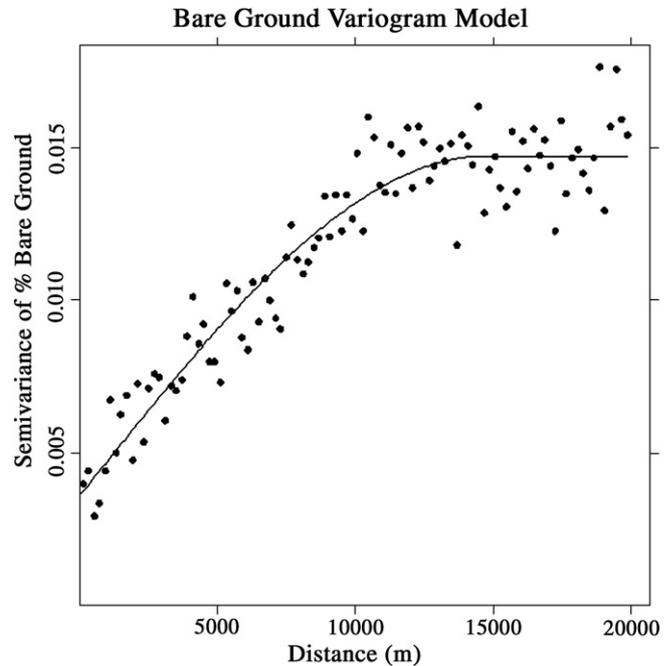


Fig. 3. Empirical semivariogram (dots) and variogram model (line) for percent bare-ground cover in the Wildhorse allotment from field observations. The nugget-to-sill ratio for percent bare ground was 0.2474.

$R^2 = 0.979$) and ranged from a single Landsat pixel at the smallest p_s up to 116 pixels at $p_s = 50$. Maximum area increased with scale parameter until $p_s = 20$ where it remained relatively constant. At $p_s = 40$, maximum object area jumped abruptly.

Cross-validated correlations between the GLS regression predictions and the field measurements at different scales ranged from $R = 0.5553$ to $R = 0.6677$ (Table 2). For the RK predictions, correlations were larger at each segmentation level except $p_s = 50$ and ranged from 0.5792 to 0.7223. There was some evidence that GLS predictions increased slightly in correlation with field measurements as segmentation level increased (Fig. 4, slope = 0.0013, $R^2 = 0.3145$, p -value = 0.0917 for test of zero slope). The RK predictions, however, showed no increasing trend with segmentation level increases (slope = 0.0001, $R^2 = 0.0109$, p -value = 0.7744 for test of zero slope). Correlations reached local maximums at $p_s = 20$ and 35 for GLS and $p_s = 20, 35,$ and 45 for RK predictions.

Nugget-to-sill ratios for GLS predictions ranged from 0.1000 to 0.5435 (Table 2). This meant that depending on the segmentation level, anywhere between 10% and approximately half of the variation in predicted percent bare-ground cover could not be attributed to spatial dependence. For the GLS residuals, NSR ranged from 0.000 to 0.5451, and, in general, as NSR for the residuals decreased the GLS-model correlations increased ($R^2 = 0.6927$). Ranges of the variogram models at different segmentation levels were from 15,189 m to 22,041 m for the predicted values and 3035 m to 12,507 m for the prediction residuals (Table 2). Ranges for the residual variogram models were always shorter than those of the predicted values and tended to decrease with increasing segmentation level as more of the spatial dependence in bare-ground cover was accounted for by coarser image segmentation. Segmentation levels whose residual variograms had the longest ranges also tended to have the largest differences between the GLS regression and RK predictions of bare-ground cover ($R^2 = 0.5502$).

At fine levels of segmentation (e.g., $p_s = 5$), the form of spatial dependence for the GLS residuals more closely matched the field empirical variogram than did the spatial dependence from the GLS predictions (Fig. 5). As segmentation level increased to $p_s = 20$, the variograms for the predicted values became more similar in form to

Table 2

Variogram characteristics of the predicted and residual values and cross-validation results for the regression analysis at each segmentation level. The empirical variogram for percent bare ground from the field observations had the following properties: nugget = 0.0036, sill = 0.0147, nugget-to-sill ratio (NSR) = 0.2474, range = 14,719 m.

Segmentation scale parameter (p_s)	Predicted				Residual				Cross-validated correlation	
	Nugget	Sill	NSR	Range (m)	Nugget	Sill	NSR	Range (m)	GLS	RK
5	0.0005	0.0033	0.1551	21,436	0.0049	0.0101	0.4783	12,507	0.5829	0.6779
10	0.0006	0.0028	0.2186	22,041	0.0052	0.0096	0.5451	10,297	0.5553	0.6459
15	0.0007	0.0057	0.1182	16,221	0.0047	0.0087	0.5436	4556	0.6073	0.6450
20	0.0006	0.0042	0.1394	15,189	0.0011	0.0067	0.1666	3213	0.6229	0.6911
25	0.0018	0.0046	0.4003	21,594	0.0023	0.0063	0.3667	5277	0.6200	0.6801
30	0.0012	0.0053	0.2249	18,422	0.0013	0.0053	0.2379	3954	0.6287	0.6568
35	0.0016	0.0061	0.2634	15,985	0.0000	0.0048	0.0000	3035	0.6677	0.6859
40	0.0023	0.0059	0.3850	19,833	0.0008	0.0051	0.1544	4178	0.6650	0.6799
45	0.0050	0.0092	0.5435	18,500	0.0000	0.0041	0.0000	5247	0.6461	0.7223
50	0.0005	0.0050	0.1000	18,500	0.0013	0.0044	0.2938	3461	0.5874	0.5792

the field-derived variogram, and the spatial dependence of the regression residuals decreased (i.e., range decreased, Fig. 5). Beyond $p_s = 20$, the variogram range of GLS predictions values increased again until $p_s = 35$ where the predicted-values variogram again closely matched the field-derived variogram. When measured by the predicted-values variogram range and NSR, the segmentation levels that most closely matched the field-derived bare-ground cover variogram (i.e., $p_s = 20$ and 35) corresponded to the locally maximum correlations for the GLS-only predictions (Fig. 6). This also held for the RK results with the exception of $p_s = 45$ which had the strongest predicted-to-observed value correlation but whose spatial dependence of prediction values was very different from the field-derived variogram (Fig. 6).

When the correlation between predicted and observed bare-ground cover was plotted against the difference in range between the predicted-values variogram and the field-derived variogram for the GLS regression predictions there was some evidence for a general decrease in the strength of the correlation as the difference in ranges increased (Fig. 7, $R^2 = 0.2188$, p -value = 0.1728 for test of zero slope). For RK, however, there was no trend in correlation with changes in difference between the variogram ranges ($R^2 = 0.0050$, p -value = 0.8463 for test of zero slope). These results are consistent with results above that segmentation levels producing predictions with spatial dependence matching that observed in the field yielded the best predictions when geostatistical techniques were not employed. The use of geostatistical techniques, however, could achieve similar results regardless of scale because they accounted for spatial dependence not captured by the regression prediction.

5. Discussion

When considering only GLS regression, the scales where spatial dependence of prediction most closely matched that of the field measurements in terms of NSR and range performed the best. The likely reason for this lies in the fact that image segmentation is a non-arbitrary aggregation of pixels into units that have implicit meaning with reference to the landscape being studied (Hay and Marceau, 2004). By grouping together similar adjacent pixels into objects in a manner that minimizes variance within each object (Baatz and Schäpe, 2000; Burnett and Blaschke, 2003), image segmentation preserves the variance between the objects. When the variance between objects is maximized, the objects are better able to model the large-scale variation in the observations. Hence there is less unexplained variation in the residuals. Scales that match the field variogram perform better because the objects at that scale are defined in such a way that the variance between the objects is similar to the variance of the field measurements not only in magnitude, but also in covariance between samples. By contrast, aggregation of pixels by regular grids of arbitrary shapes tends to decrease variance as scale coarsens (Wiens, 1989). This not only obscures the spatial dependence of a variable but also decreases the chance of defining a set of arbitrary units with similar variance as the field measurements.

Aggregating spatial data into non-overlapping areal units can change the patterns present in the original set of observations—a phenomenon known as the modifiable areal unit problem (MAUP, Openshaw and Taylor, 1981; Openshaw, 1984; Dark and Bram, 2007). Both the number of units used to tessellate a landscape and the configuration of those units have been demonstrated to affect the magnitude and the variance spatial information (Dark and Bram, 2007, see also Svancara et al., 2002). Our results suggest that another possible effect of the MAUP is changes in spatial dependence of landscape patterns. Burnett and Blaschke (2003) suggested that while image segmentation does not solve the MAUP, it can minimize its effects because data are aggregated with respect to patterns present in the original image. Our results support this claim, and offer an additional technique for evaluating (and potentially minimizing) MAUP effects for a specific objective, namely the comparison of the spatial dependence of segmentation-based predictions to that of the original field measurements.

The results from our GLS regressions support Addink et al.'s (2007) observation that a specific, most appropriate scale can be defined with regard to a particular ecosystem attribute. Addink et al. also demonstrated that the most appropriate scale may not be the same for all variables. However, many inquiries involve more than a single attribute, and using regression-based techniques alone, it would be important to discover for each variable being considered the scale that would provide the most accurate results. But our results also demonstrated that good predictions were possible across a range of scales when the spatial autocorrelation between objects was accounted for (i.e., in the case of bare-ground cover, RK produced strong correlations over most scales

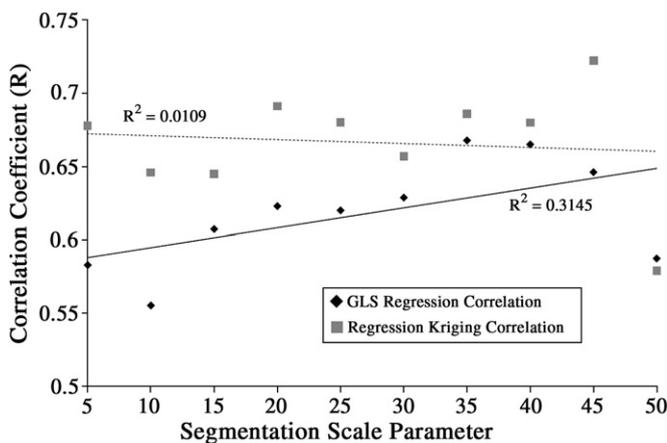


Fig. 4. Changes in correlation with segmentation scale between predictions of percent bare ground from generalized least-squares (GLS) regression and regression kriging (RK). The R^2 values reported on the graph are for the fit of the points to the trend line.

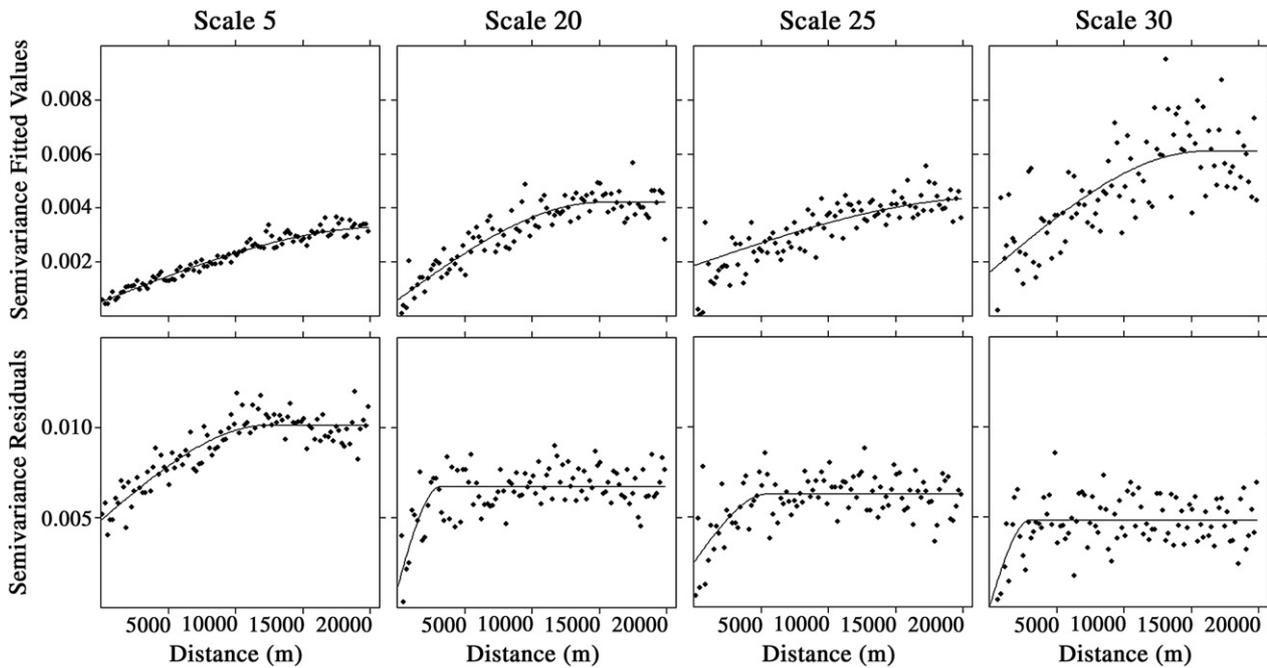


Fig. 5. Empirical semivariograms for percent bare-ground predictions and residuals from the generalized least-squares (GLS) regression models at four different levels of image segmentation. Semivariograms for predicted values became more similar to the field-derived spatial dependence up to segmentation level (p_s) of 20.

considered). This finding is important because it allows a single scale to be used to effectively consider the distribution of multiple attributes. Our results support Wiens (1989); see also Wu and Li. (2006) concept of scaling domains (i.e., ranges of scales within which the relationship between variance and scale is consistent due to a similar set of underlying processes that control landscape patterns). Karl and Maurer (2010) found that scaling through image segmentation was a better method than square pixel-aggregation techniques for characterizing scaling domains and thresholds from remotely-sensed data and field observations.

For making RK predictions, the scales whose GLS-residual semivariograms had low NSR and long ranges showed the largest gain in correlation to field observations over GLS. This suggested that the potential to improve overall predictions by predicting residual values at unsampled locations could be assessed by considering the NSR and range of a regression-residuals variogram. Low NSR and long ranges usually indicate that higher accuracies can be achieved in making

predictions to unsampled locations (Kravchenko, 2003). Using kriging and inverse-distance weighting (another spatial interpolator) on soil properties, Kravchenko (2003) found that variables with strong spatial structure could be mapped more accurately than those that had weak spatial structure regardless of the variance of the soil property.

The strong correlation between the RK prediction and field-measured bare-ground cover at $p_s=45$ did not follow the patterns observed for RK at other scales. The likely reason for this result has to do with how the Landsat image was segmented. First, we used the original Landsat pixels to create image objects at each scale rather than deriving coarser scales through merging of objects in the next finest scale. This was because merging objects at one scale to produce a coarser scale enforces a patch hierarchy on the study area that may not be appropriate, and segmentation that does not enforce strict hierarchies can show ephemeral objects that appear at one scale and are later replaced by different sets of objects (Hay et al., 2003). Segmentation up to $p_s=40$ grouped pixels similarly at each scale, steadily produced a set of image objects that gradually grew larger by,

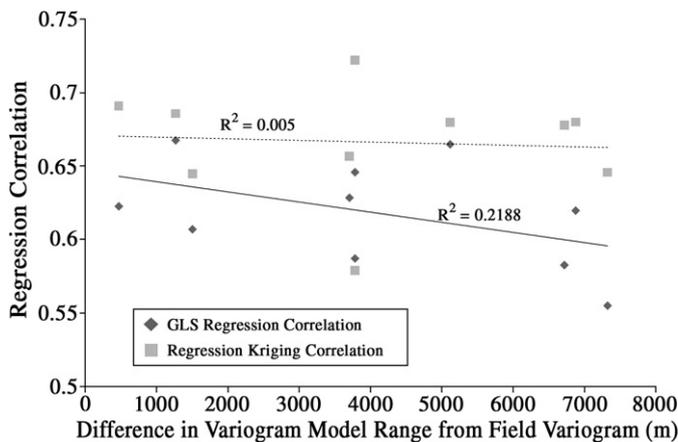


Fig. 6. Plot of variogram range versus nugget-to-sill ratio for the field measurements of bare-ground cover and generalized least-squares regression predictions for the ten segmentation levels considered. Segmentation levels closest to the field-observation variogram parameters had the strongest correlation between predicted and observed values for percent bare ground.

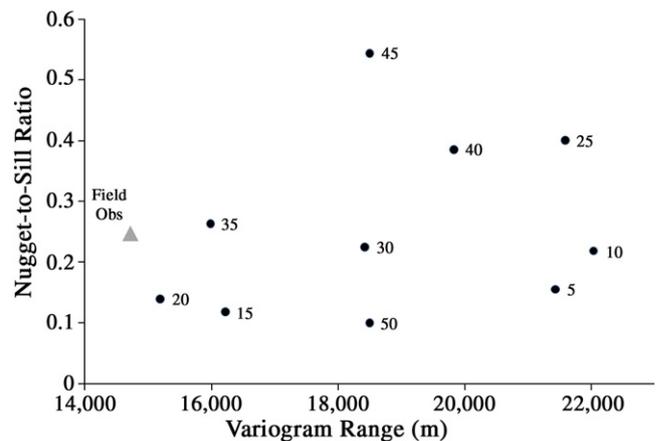


Fig. 7. Correlation between predicted and observed percent bare-ground cover for 10 segmentation levels plotted against the difference between the range of the variogram model of the predictions and the range of the variogram model derived from the field observations. The R^2 values reported on the graph are for fit of the points to the trend lines.

in effect, dissolving boundaries between adjacent objects. At $p_s = 45$, while most of the small objects had the same boundaries as at finer scales, the largest image objects had different boundaries and configurations (i.e., the largest objects at $p_s = 45$ were smaller and simpler than at $p_s = 40$). This was evident in Table 1 where the maximum object size for $p_s = 45$ was much smaller than for $p_s = 35$ or 40. Because this event happened at one of our coarsest scales, and because the number of image objects at scales coarser than $p_s = 50$ gave too few objects for reliable regression or variogram estimates, we could not tell if the strong correlation at $p_s = 45$ was a unique event or signified a scaling transition. Owing to the fact that the jump in correlation appeared to have come from a reconfiguration of the object boundaries, the strong correlation at $p_s = 45$ suggests that other ways of combining pixels into meaningful objects (i.e., by exploring other parameter options in addition to p_s) may provide better results than what we reported here. Feitosa et al. (2006) used a genetic algorithm to vary three multi-resolution segmentation parameters to find the combination that best matched pre-defined patches—a similar approach could be adapted to identifying scales that maximize regression-based prediction accuracy.

The success of the technique we have outlined in this paper for identifying appropriate levels of image segmentation for making predictions of ecosystem attributes depends on two factors. First, there must be enough spatial dependence in the attribute of interest to detect and use in conjunction with available remote-sensing data. Percent bare-ground cover had a well-defined spatial dependence in the Wildhorse allotment, making it a good candidate for this study. Precision of the field measurements is an important consideration when assessing spatial dependence because imprecise measurements can inflate the variogram nugget (Webster and Oliver, 2007) and decrease the observed spatial dependence. Second, enough field samples must be collected to accurately characterize the spatial dependence that is present, and the samples must be collected in the right manner. Webster and Oliver (1991) found through simulations that variograms created using fewer than 50 samples did a poor job of estimating actual spatial dependence. They recommended at least 100 samples to construct a reliable variogram. In addition, the distribution of the sample locations is important. If samples are taken on a regular grid, the shortest lag distance possible will be the grid spacing and it will be impossible to evaluate short-range spatial dependence if samples are spaced far apart. One strategy when selecting sample locations for use in geostatistics is the use of multi-level or nested sampling (Webster and Oliver, 2007) to achieve a variety of distances between samples.

6. Conclusion

The fact that the ability to predict ecosystem attributes from remotely-sensed data is closely tied to the scale of analysis is not new. It is a product of the well-known relationship between measurements of an attribute's variance and the scale of observation that occurs because the attribute is non-randomly distributed across a landscape (Wiens, 1989; Horne and Schneider, 1995; Fuhlendorf and Smeins, 1999; Wu et al., 2006). The contribution of our results is that the spatial dependence of predictions also changes with scale and this information can be used to select specific scales that maximize predictive ability. However, our results also support the idea that methods which consider the spatial variance of an ecosystem variable, make accurate predictions possible over a range of appropriate scales and make the need to identify a single, best scale less critical.

The process of collapsing information on landscape patterns into discrete units is an attempt to reduce the complexity of ecological systems so that the relationships between patterns and processes can be understood (Burnett and Blaschke, 2003). Scaling theory holds that there may be many appropriate ways for aggregating landscape data depending on the objective, and in our results we found evidence of

this. Image segmentation shows much promise for scaling image data for ecological analyses, but more work is needed to develop methods that consider a wide range of different ways to segment images into coarser scales and select sets of scales that perform best for answering specific management questions. The robustness of ecological analysis of landscapes will increase as methods are developed to remove the subjectivity with which these discrete units are drawn.

Acknowledgments

J. Russel and G. Mann of the Bureau of Land Management provided us with the field data from the Wildhorse allotment. J. Qi, S. Riley, G. Roloff, and R. Unnasch provided insightful discussion during this study and invaluable review of this manuscript. This work was supported by the Idaho Chapter of The Nature Conservancy, the M. J. Murdock Charitable Trust, the Lava Lake Foundation for Science and Conservation, and The Nature Conservancy's Rodney Johnson and Katherine Ordway Science Endowment.

References

- Addicott, J.F., Aho, J.M., Antolin, M.F., Padilla, D.K., Richardson, J.S., Soluk, D.A., 1987. Ecological neighborhoods: scaling environmental patterns. *Oikos* 49, 340–346.
- Addink, E.A., de Jong, S.M., Pebesma, E.J., 2007. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing* 73, 905–912.
- Allen, T.F.H., Starr, T.B., 1982. *Hierarchy: Perspectives for Ecological Complexity*. University of Chicago Press, Chicago, IL.
- Baatz, M., Schöpe, A., 2000. Multiresolution segmentation — an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informationsverarbeitung XII*. Wichmann-Verlag, Heidelberg, pp. 12–23.
- Bailey, T.C., Gatrell, A.C., 1995. *Interactive Spatial Data Analysis*. Addison-Wesley.
- Bedell T. E. 1998. *Glossary of terms used in range management*. 4th Edition, Society for Range Management. Direct Press, Denver, Colorado.
- Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modeling* 168, 233–249.
- Chavez Jr., P.S., 1996. Image-based atmospheric corrections — revisited and improved. *Photogrammetric Engineering and Remote Sensing* 62, 1025–1036.
- Cressie, N., 1993. *Statistics for Spatial Data*, revised edition. Wiley and Sons, New York, NY.
- Crist, E.P., Kauth, R.J., 1986. The tasseled cap de-mystified. *Photogrammetric Engineering and Remote Sensing* 52, 81–86.
- Cullinan, V.I., Simmons, M.A., Thomas, J.M., 1997. A Bayesian test of hierarchy theory: scaling up variability in plant cover from field to remotely sensed data. *Landscape Ecology* 12, 273–285.
- Dark, S.J., Bram, D., 2007. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography* 31, 471–479.
- Farina, A., 1998. *Principles and Methods in Landscape Ecology*. Chapman & Hall, London, U.K.
- Feitosa, R.Q., Costa, G.A.O.P., Cazes, T.B., Feijo, B., 2006. A Genetic Approach for the Automatic Adaptation of Segmentation Parameters. Salzburg University, Salzburg, Austria.
- Fisher, P., 1997. The pixel: a snare and a delusion. *International Journal of Remote Sensing* 18, 679–685.
- Foody, G.M., 2004. Sub-pixel methods in remote sensing. In: de Jong, S.M., van der Meer, F.D. (Eds.), *Remote Sensing Image Analysis: Including the Spatial Domain*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 71–92.
- Fortin, M.J., Dale, M., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press, Cambridge, UK.
- Fuhlendorf, S.D., Smeins, F.E., 1999. Scaling effects of grazing in a semi-arid grassland. *Journal of Vegetation Science* 10, 731–738.
- Garrigues, S., Allard, D., Baret, F., Weiss, M., 2006. Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sensing of the Environment* 103, 81–96.
- Gilbert, M.A., Gonzalez-Piqueras, J., Garcia-Haro, F.J., Melia, J., 2002. A generalized soil-adjusted vegetation index. *Remote Sensing of the Environment* 82, 303–310.
- Goovaerts, P., 1997. *Geostatistics for Natural Resource Evaluation*. Oxford University Press, New York, NY.
- Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *Journal of Photogrammetry and Remote Sensing* 57, 327–345.
- Hay, G.J., Marceau, D.J., 2004. Multiscale object-specific analysis (MOSA): an integrative approach for multiscale landscape analysis. In: de Jong, S.M., van der Meer, F.D. (Eds.), *Remote Sensing and Digital Image Analysis: including the spatial domain*. Kluwer Academic Publishers, Dordrecht, Germany.
- Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. *Computers & Geosciences* 33, 1301–1315.
- Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120, 75–93.

- Herrick, J.E., Van Zee, J.W., Havstad, K.M., Burkett, L.M., Whitford, W.G., 2005. Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems. USDA-ARS Jornada Experimental Range, Las Cruces, New Mexico.
- Horne, J.K., Schneider, D.C., 1995. Spatial variance in ecology. *Oikos* 74, 18–26.
- Jelinski, D.E., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology* 11, 129–140.
- Karl, J.W., Maurer, B.A., 2010. Multivariate correlations between imagery and field measurements across scales: comparing pixel aggregation and image segmentation. *Landscape Ecology* 25, 591–605.
- Keane, R., Lutes, D., Key, C., and Benson, N. FIREMON - sampling methods documents. 2005. Moscow, ID, Fire Research and Management Exchange System, National Biological Information Infrastructure.
- Kotlair, N.B., Wiens, J.A., 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos* 59, 253–260.
- Kravchenko, A.N., 2003. Influence of spatial structure on accuracy of interpolation methods. *Soil Science Society of America Journal* 67, 1564–1571.
- Krige, D.G., 1966. Two-dimensional weighted moving average trend surfaces for ore-evaluation. *Journal of the South Africa Institute of Mining and Metallurgy* 66, 13–38.
- Levin, S.A., 1992. The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.
- National Research Council, 1994. Rangeland Health: New Methods to Classify, Inventory, and Monitor Rangelands. National Academy Press, Washington, D.C.
- Navulur, K., 2007. Multi-spectral Image Analysis using the Object-oriented Paradigm. CRC Press, Taylor and Francis Group, LLC, Boca Raton, FL.
- Neubert, M., Herold, H., Meinel, G., 2006. Evaluation of remote sensing image segmentation quality – further results and concepts. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Salzburg, Austria.
- O'Neill, R.V., Deangelis, D.L., Waide, J.B., Allen, T.F.H., 1986a. A Hierarchical Concept of Ecosystems. Princeton University Press, Princeton, New Jersey.
- O'Neill, R.V., Johnson, A.R., King, A.W., 1989. A hierarchical framework for the analysis of scale. *Landscape Ecology* 3, 193–205.
- O'Neill, R.V., Johnson, A.R., King, R.V., 1986b. A Hierarchical Concept of Ecosystems. Princeton University Press, Princeton, New Jersey.
- Oppenshaw, S., 1984. Ecological fallacies and the analysis of areal census data. *Environment and Planning A* 16, 17–31.
- Oppenshaw, S., Taylor, P.J., 1981. The modifiable areal unit problem. In: Wrigley, N., Bennett, R.J. (Eds.), *Quantitative Geography: A British View*. Routledge & Kegan Paul Ltd., London, pp. 60–69.
- Odeh, I., McBratney, A.B., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma* 63, 197–214.
- Odeh, I., McBratney, A.B., Chittleborough, D., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression kriging. *Geoderma* 67, 215–226.
- Owen, D. E. *Geology of Craters of the Moon*. 2008. Arco, ID, Craters of the Moon National Monument, National Park Service.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30, 683–691.
- Pellant, M., Shaver, P., Pyke, D. A., and Herrick, J. E. Interpreting indicators of rangeland health, version 4. Technical Reference 1734-6. 2005. Denver, CO, U.S. Department of the Interior, Bureau of Land Management, National Science and Technology Center.
- Peters, D.P.C., Yao, J., Huenneke, L.F., et al., 2006. A framework and methods for simplifying complex landscapes to reduce uncertainty in predictions. In: Wu, J., Jones, K.B., Loucks, O.L. (Eds.), *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer, Dordrecht, The Netherlands, pp. 131–146.
- Pinheiro, J.C., Bates, D.M., 2000. *Mixed-effects Models in S and S-PLUS*. Springer-Verlag, New York.
- Pyke D. A., J. E. Herrick, P. Shaver, and M. Pellant. 2003. What is the standard for rangeland health assessments? Pages 764–766 Durban, South Africa.
- Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. *Remote Sensing of the Environment* 30, 1311–1319.
- Soil Survey Staff. U.S. general soil map (STATSGO2) for Idaho. USDA Natural Resource Conservation Service. 7-5-2006a. 3-3-2009a.
- Soil Survey Staff, 2006b. *Keys to Soil Taxonomy* 10th ed. USDA Natural Resource Conservation Service, Washington, D.C.
- Svancara, L.K., Garton, E.O., Chang, K., Scott, J.M., Zager, P., Gratson, M., 2002. The inherit aggravation of aggregation: an example with elk aerial survey data. *Journal of Wildlife Management* 66, 776–787.
- Turner, M.G., O'Neill, R.V., Gardner, R.H., Milne, B.T., 1989. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology* 3, 153–162.
- Wang, L., Sousa, W.P., Gong, P., 2004. Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing* 25, 5655–5668.
- Webster, R., Oliver, M.A., 1991. Sample adequately to estimate variograms of soil properties. *European Journal of Soil Science* 43, 177–192.
- Webster, R., Oliver, M.A., 2007. *Geostatistics for Environmental Scientists*. John Wiley & Sons Inc., Hoboken, NJ.
- Wiens, J.A., 1989. Spatial scaling in ecology. *Functional Ecology* 3, 385–397.
- Wiens, J.A., Milne, B.T., 1989. Scaling of 'landscapes' in landscape ecology, or, landscape ecology from a beetle's perspective. *Landscape Ecology* 3, 87–96.
- Woodcock, C., Harward, V.J., 1992. Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing* 13, 3167–3187.
- Wu, J., 1999. Hierarchy and scaling: extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing* 25, 367–380.
- Wu, J., Jones, B.K., Li, H., Loucks, O.L., 2006. *Scaling and Uncertainty Analysis in Ecology*. Springer, Dordrecht, The Netherlands.
- Wu, J., Li, H., 2006. Concepts of scale and scaling. In: Wu, J., Jones, K.B., Li, H., Loucks, O.L. (Eds.), *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer, Dordrecht, the Netherlands, pp. 3–15.