Assessing the Pattern of Error in a Predictive Soil Map

by

Benjamin M. Coakley

A thesis submitted in partial fulfillment of

the requirements for the degree of

Master of Science

(Cartography and GIS)

at the

UNIVERSITY OF WISCONSIN - MADISON

2010

Acknowledgements: I am grateful for the continuing support of my advisor, Dr. Amy Burnicki, and the assistance of Dr. A-Xing Zhu and Dr. Jim Burt.

Abstract: SoLIM is a predictive soil mapping application employing fuzzy logic and expert knowledge to model the relationship between landscape and soil type. The end result of SoLIM's predictive approach is a hardened map of soil types and two confidence statistics describing the uncertainty associated with their prediction. This research valuated SoLIM's performance using a validation dataset consisting of 495 manually-classified soil samples randomly located within a test watershed. The soil types identified manually in each sample point also contained two assessments of measurement uncertainty. The patterns of SoLIM's mispredictions were investigated by comparing the manual and SoLIM soil type predictions: 1) In the spatial domain, using point pattern and geostatistical methods; 2) In the attribute domain; and 3) In relation to SoLIM's internal confidence measures. Because measurement uncertainty was assessed within the validation dataset, the analysis was repeated for different subsets of the validation dataset defined using different levels of measurement uncertainty. No spatial trends were found in the location of errors. Association was found between some input variables used in SoLIM's predictive model and prediction error for some subsets of the validation dataset, but no overall trends were visible. Association was found between SoLIM's internal uncertainty measures and error.

Introduction

Soil mapping is an application which must tolerate high degrees of uncertainty. Traditional soil maps are based on a modeled relationship between the land and soil types. However, collecting exhaustive data on soil types, even for a relatively small area, is enormously expensive. In addition, the underlying model is never made explicit, which makes assessment of the model's validity difficult (Zhu *et al.*, 2001). Predictive soil mapping techniques quantify the relationship between the land and the soil. One effect of explicitly modeling this relationship is that (at least in some approaches) a level of confidence is provided with each prediction. My research objective is to conduct an exploratory analysis of error in a predictive soil map, exploring both the modeled relationship and the confidence rating.

The reference dataset for this research includes explicit measures of uncertainty in the sampling and classification process. This information makes it possible to examine the effects of error in the validation dataset part of the overall accuracy assessment. Subsets of the reference dataset with greater measurement certainty, or less potential error, may be examined separately, and the accuracy compared with the reference dataset as a whole to determine whether the allowance of error in the validation dataset impacts accuracy assessment results .

I. Background

Literature Review

Traditional soil surveys do little to convey error and uncertainty in the data. Soil maps demarcate soil types with crisp boundaries (i.e., as vector polygons), suggesting to the user that the soil characteristics within these boundaries are generally uniform (Heuvelink and Webster, 2001). This approach presents two issues. First, this method of describing soil type does not show any information about variation within a soil zone. Information about this variation may have been collected during the sampling process, but it is not communicated on the map. Instead, the map presents each soil type as a single (or multiple) completely homogeneous zone(s). Second, this method does not show any uncertainty about the location of the boundary (Moore *et al.*, 1993). Soil scientists may have varying degrees of confidence in a given boundary because soil changes gradually over the landscape. This makes the exact location of a boundary between soil types subject to a high degree of uncertainty. For example, a boundary which depends on an unambiguous physical feature (such as a break in slope) may be viewed with more confidence than a boundary which depends on an interpretation or interpolation (Burrough *et al.*, 1997).

These deficiencies affect the usefulness of the crisp boundaries, particularly for fine-scale soil mapping applications. In general, models which use crisp thresholds may be too rigid (Lark and Bolam, 1997). The crisp model "exaggerates the similarity (membership) between an object and the class to which the object is assigned, and ignores the similarities between the object and the other classes" (Zhu, 1997). The

resulting homogenous zones of the map in traditional soil surveys are often not, in reality, similar enough to be useful for modeling applications (Moore *et al.*, 1993).

Fuzzy logic is a useful alternative which can be used to show varying degrees of similarity to many classes (Scull *et al.*, 2003). It can be used for a wide variety of geographic phenomena in cases where there is ambiguity or vagueness in class definitions or spatial interactions (Burrough and McDonnell, 1998). This approach is particularly well-suited for soil mapping, and has been used for this purpose since at least Burrough's work of 1989. A fuzzy classification scheme describes the soil in a particular location in terms of its similarity to many different classes of soil. As soil varies over the landscape, a fuzzy classification scheme can reflect its diminishing similarity to one class and increasing similarity to another, as opposed to drawing a crisp boundary.

While fuzzy logic provides a useful framework for conveying information about soil characteristics, it does little to improve the underlying accuracy of the soil map. Accurate soil mapping has traditionally required extensive soil sampling. Results from these soils samples are then combined with information about landscape features such as topography to establish soil boundaries. This method is based on an assumed relationship between landscape features and soil types. Given that the soil profile is a product of environmental factors, "if the spatial distribution of soil-forming factors is known, soil character may be inferred" (Scull *et al.*, 2003). Moore *et al.*(1993), for example, found significant correlations between measured terrain attributes and soil attributes.

The relationship between landscape features and soil types, however, is complex. Soil is different in different environments; therefore, there is no set list of environmental

variables to consider when defining the relationship (Zhu *et al.*, 2001). Further, the specific interaction between soil types and environmental variables is not fully understood, so relationships which are well-established in one area may not be applicable in another area. The complexity and locally-specific nature of this relationship has meant that soil scientists have had to rely on extensive soil sampling, rather than environmental features, to determine approximate boundaries between soil types.

Predictive soil mapping attempts to improve on traditional soil surveys by describing the relationship between landscape features and soil type specific to a local area. This knowledge is then used to better predict soil classification. There are a wide variety of predictive soil mapping techniques, but they have three common goals: "1) exploit the relationship between environmental variables and soil properties in order to more efficiently collect soil data; 2) produce and present data that better represents soil landscape continuity; and 3) explicitly incorporate expert knowledge in model design" (Scull *et al.*, 2003). Like a traditional soil survey, predictive soil mapping uses knowledge about the relationship between landscape features and soil types to assign soil types to a given location. However, predictive soil mapping techniques quantify these relationships, allowing the assignment of a degree of uncertainty (Scull *et al.*, 2003). This uncertainty measure attempts to capture the confidence level in the prediction.

The SOLIM process

SoLIM – the Soil Land Inference Model – is a predictive soil mapping approach developed at UW-Madison which incorporates expert knowledge developed by soil scientists working in the field regarding the relationship between landscape features and

soil type. SoLIM takes an inferential approach to determining soil similarity at each location. It relies on expert knowledge to describe the relationship between landscape characteristics (such as slope, bedrock geology, and topographic wetness index) and soil type.

SoLIM converts this expert knowledge into a membership function which describes the landscape conditions that are present at each type of soil. The membership functions incorporate degrees of similarity. For example, if the soil scientists describe a particular soil as being present at slopes between 4% and 8%, the membership function will show 100% similarity to the given soil in areas with slope in that range. However, rather than showing 100% similarity at a slope of 4.0% and 0% at a slope of 3.9%, the similarity value will gradually decay to 0% as the slope gets further outside the range. The similarity functions for each of the input variables, representing different attributes of the landscape conditions, are combined in a membership function to give an overall similarity value. Then, using the membership functions and GIS data layers describing the landscape, the map of soil similarity is developed.

In some cases, a single soil type may arise from multiple, dissimilar sets of landscape conditions. For such soils, multiple membership functions may be created and independently evaluated for a single soil type. Also, some soil types arise from conditions that cannot by defined by SoLIM's inferential process. These soil types are not predicted by SoLIM, and locations where they are present are not considered in this analysis.

SoLIM models the variation of soil across the landscape by using a raster model, rather than a vector model. Each pixel of the raster stores the degree of similarity the soil at that location has with each of the soil types being mapped for the area. Although this may seem counterintuitive, the degrees of similarity do not necessarily sum to unity. That is, the similarity values represent the similarity between the conditions at the location and the prototypical conditions where the soil type is found, and not the likelihood that the soil in the location is a particular type. Therefore, as we move across the landscape from one soil type to another, SoLIM can show the gradual diminishing of similarity with one soil type and the gradual increase of similarity with another soil type (Zhu *et al.*, 2001).

As stated previously, SoLIM relies on an explicitly modeled relationship between landscape features and soil types to predict soil type at each location. Because classification depends on this model, misclassification is also likely to depend on the model. Errors in SoLIM predictions may occur under three sets of circumstances: where the input data does not match conditions on the ground (e.g., DEM error); where the knowledge does not accurately describe the relationship between the landscape and the soil; or where the knowledge is not correctly translated to a membership function. The latter two cases will result in a membership function that does not correctly describe the relationship between a landscape feature and the resulting soil type. I hypothesize that this dependency will be visible as a relationship between misprediction and the attributes that are used as inputs to the model. Further, this dependency will be visible in the spatial domain, as a tendency for misprediction to cluster in areas with features that are

incorrectly accounted for by one or more membership functions. Finally, SoLIM's internal uncertainty measures should capture the locations where mispredictions are more likely, visible as a relationship between the confidence assigned to the projection and mispredictions. However, these relationships have yet to be explored in depth.

Therefore, the objective of this research is to determine:

1. whether patterns in misprediction exhibit spatial dependence, i.e., exhibit a pattern in the spatial domain;
2. whether patterns in misprediction are related to the landscape features (e.g., slope, bedrock geology) that are explicitly modeled by SoLIM, i.e., exhibit a pattern in the attribute domain; and
3. whether patterns in misprediction are related to SoLIM's internal assessment of uncertainty in prediction.

III. Data

The study area

As part of the SoLIM validation process, soil samples were collected in a watershed in Dane County, Wisconsin for which SoLIM had previously been used to produce a predictive soil map (Figure 1). For the SoLIM-produced map, similarity functions for 76 different soil types were developed, although only 40 types show up with any frequency. The region under study is part of the Driftless Area, which did not experience glaciations during the most recent glacial period. Generally, the Driftless Area is topographically richer than the rest of Wisconsin, with steep slopes and broad

ridges and valleys. Land use in the study area includes both cultivated areas and forest; there are also large areas of non-inferenceable soils, mostly in valley bottoms. As a result, the polygon defining the study area is highly concave. The total area of predicted soils is approximately 18.95 square miles.

495 locations were randomly defined within the watershed using GIS and then located in the field with survey-grade GPS units (Figure 1). The random locations correspond to the center of SoLIM pixels. The GPS techniques used guaranteed that the sampled locations were within the SoLIM pixel, though not necessarily at the exact center. Core samples were taken at each location, and classified by local soil experts independently of the SoLIM process. These classifications can be directly compared to SoLIM predictions.

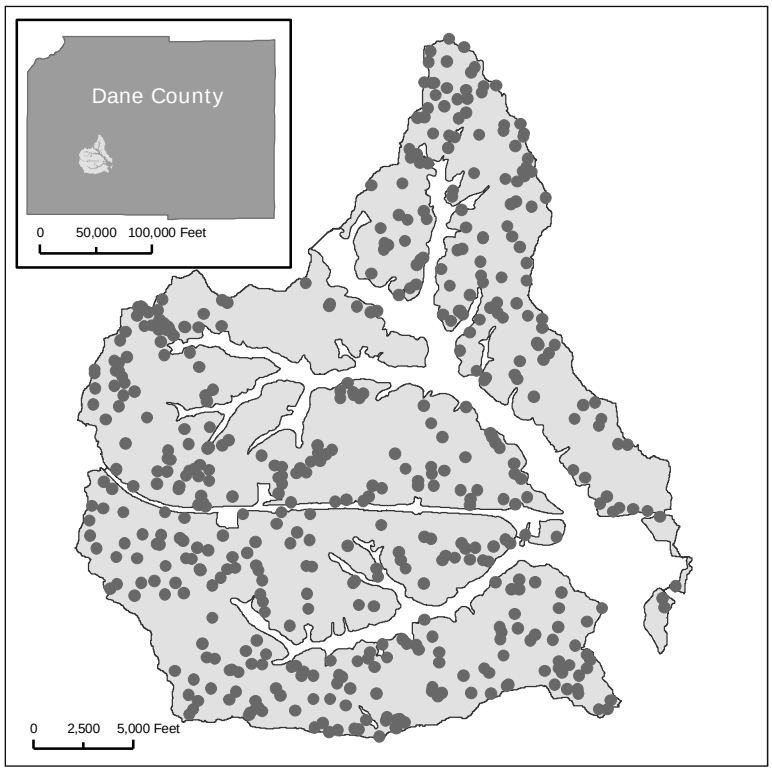

Figure 1: Study area and location of sample points

The purpose of this dataset is to evaluate the performance of SoLIM in a statistically significant way, including the evaluation of internal measures of uncertainty (Zhu *et al.*, 2006). This dataset serves as the validation dataset in this analysis, and SoLIM's predictions will be compared to the field classifications.

Measuring uncertainty

When working with error and uncertainty, it is important to be precise in defining terms. Error represents the difference between the mapped value and the value in the real world, while uncertainty represents the confidence (or lack thereof) in the mapped value. GIS data may contain errors in the spatial, temporal, or thematic values stored, and it is necessary to treat these errors separately (Veregin and Hargitai, 1995). In this analysis, I will examine thematic error, defined as a misprediction; i.e., the soil type assigned in the SoLIM predictive map differed from the soil type determined in the validation data set.

Model measures of uncertainty: Exaggeration and entropy

SoLIM predictions are commonly viewed after a hardening operation, where the soil class representing the most similar soil type is assigned to each raster pixel (Zhu *et al.*, 2001). Two internal measures of uncertainty are calculated for this hardened map. The first, exaggeration, measures the difference between the similarity of the most similar soil type and 1.0. A less-similar soil will have a higher exaggeration statistic, which implies a higher degree of uncertainty (Zhu, 1997). The formula for exaggeration is:

where E_{ij} represents the exaggeration at location (i,j) and $S_{ij}g$ represents the percent match with the most similar soil type at location (i,j) (Zhu, 1997). An exaggeration of 0

means that the conditions at the given location exactly match the conditions for a given soil type, while a value near 1 means that the conditions at the location match the conditions for the most similar soil type at a very low similarity. A value of exactly 1 means that the conditions at the location do not match the conditions for any soil type at all, but this is rare in practice.

The second measure of uncertainty, entropy, measures the similarities with other soil types that were discarded when choosing the most similar soil type. This is a measure of ignorance uncertainty, measuring membership diffusion (Zhu, 1997). A location with several soil types of high similarity will have a higher entropy statistic, implying a higher degree of uncertainty (Goodchild *et al*, 1994; Zhu, 1997).While there are several ways to calculate entropy, SoLIM calculates entropy according to the formula:

where H_{ij} represents the entropy at location (i,j) and $S^{n}_{ij}k$ represents the normalized similarity value for each soil type at location (i,j) (Zhu, 1997). Normalizing similarity values at a location scales them so they sum to 1. An entropy value of 0 means that the conditions at the given location match the conditions for only one soil type, i.e., they do not match the conditions for any other soil type. An entropy value of 1 means that the conditions at a given location match each possible soil type equally. Increasing the number of soil types with non-zero similarity values, or the similarity values associated with each soil type, results in higher entropy values. Taken together, the

exaggeration and entropy measures describe the SoLIM's internal uncertainty about the hardened class assignment at a given location.

Sample measures of uncertainty: Property vs. classification

The validation dataset also contains two measures of uncertainty. First, the soil scientists tasked with collecting the validation dataset gave a confidence rating that the collected sample adequately identifies the properties of the soil at the location. In the event that an obstruction is encountered, for example, all the properties of the soil may not be visible. Second, the scientists gave a confidence rating for their own classification of the soil type. Uncertainties in classification may result from sampled soils which do not exactly match a known soil type, or which match multiple soil types.

Every sampled point, as part of the sampling process, was assigned a property certainty and classification certainty. These measures of certainty were assigned numeric values between 0 and 1. Specifically, perfect certainty is represented by a value of 1.0, intermediate levels of certainty correspond to values 0.7 and 0.4, and an educated guess in absence of real physical evidence is indicated with a certainty score of 0.0. For practical purposes, no point has a value of 0.0 for property certainty. That would indicate that the sample was impossible to take. There are, however, some points that have a classification certainty of 0.0, which means that the assigned soil class is little more than an educated guess.

The sample measures of uncertainty indicate the confidence in the classification of the sampled soils. Higher uncertainty values indicate lower confidence in the classification, and therefore may result in an increased likelihood of an incorrect

classification. Incorrect classifications in the validation dataset will cause false negatives or (less likely) false positives when compared to the predicted soil map. It is necessary, then, to carefully consider how much uncertainty to allow in the validation dataset.

As defined in the data collection process, classification certainty does not depend on property certainty. When assigning a classification certainty, the soil scientists did not consider the uncertainty of the measured soil properties, but instead assumed that the properties were measured perfectly. However, it is not known whether property certainty and classification certainty are independent variables. It is possible that perfect classification certainty requires perfect property certainty. It seems reasonable that, to be completely certain of a classification, one would need to work from a soil sample for which the properties were also known perfectly. However, this relationship is not known.

The possible validation datasets

Based on these measures of uncertainty, I divided the validation dataset into four candidate data sets. The first data set included all the sample points that had both property certainty and classification certainty values equal to 1.0, representing perfect measurement certainty. For these samples, the soil scientists were completely confident they were able to both correctly measure the required soil properties, and to exactly determine the soil class based on these measurements. 66 sample points met these criteria.

The second data set relaxed the requirement for certainty in the sample dataset by including points with a property certainty of 1.0 and a classification certainty greater than or equal to 0.7, resulting in 128 sample points. The goal of this second dataset was to

increase the number of sample points in the validation dataset by allowing one degree of uncertainty in the soil classification process. This increased the likelihood of a misclassification in the validation dataset, but maintained the condition that soil properties were correctly determined. By relaxing the condition for classification certainty to a level that still represents a relatively high degree of confidence in the assigned soil class, the validation dataset was approximately doubled to 128 points. The third data set further expanded the number of sample points in the validation dataset by allowing one degree of uncertainty in either property or classification, but not both. In other words, the third dataset included cases where the property certainty multiplied by the classification certainty was equal to 0.7 or greater. Effectively, one of the certainty values was equal to 1.0 and the other was either 0.7 or 1.0, resulting in 241 points. Because it is not known whether classification and property certainty are independent, this dataset may have the same likelihood of misclassification as the second dataset, or it may have a higher likelihood if classification certainty does depend on property certainty.

For the fourth and final dataset, I chose to not eliminate any points based on measurement uncertainty; i.e., I assumed that the soil class identified was correctly determined in the field based on correctly measured soil properties, regardless of the property certainty and classification certainty values assigned in the sampling process. This represented the complete sample data set of 495 points. While this validation dataset increased the number of sample points available for analysis, it also represented the dataset with the highest potential for incorrect classification in the validation set.

In addition to defining the sample points for the validation dataset, I also needed to define the rule that determined whether the soil class identified in the reference sample point matched the soil class predicted by SoLIM. If the soil class identified by the soil scientists in the field exactly matched the soil class predicted by SoLIM, then agreement existed and the sample point was assigned a value of 0, indicating no error in classification. If the soil classes differed, then a value of 1 was assigned to the sample point, indicating an error in classification. To establish this agreement rule, I used a correlation matrix maintained by soil scientists independently of the SoLIM process (see Appendix A), which described the similarity of each soil type to each other soil type on a scale from 0 to 1. In other words, every soil was assigned a series of values which represent its similarity to every other soil observed in the map. Similarity values were limited to 0, 0.2, 0.4, 0.6, 0.8, and 1, but the most common values were 0.4 and 1.

One possible definition of agreement between the reference and SoLIM soil classes is to define agreement as exact matches; i.e., a soil type is only similar to itself at a 1.0 level. However, lower similarities may still be useful. Soils that match at a numeric value of 0.4, which may seem low, are still very similar and in most cases represent a good enough match for the purposes for which a soil map is used. I chose to establish two agreement rules based on thresholds of 1.0 and 0.4. This resulted in the creation of a series of validation datasets (based on sample uncertainty measures) for each agreement rule.

In all, eight validation data sets, representing four different measurement uncertainty thresholds and two different agreement thresholds, were defined (Table 1).

All analyses were performed on each of these data sets to assess whether allowing for some measurement uncertainty or less than exact matches impacts the pattern of error observed within the hardened SoLIM soil map. In addition, there were some analyses where it was necessary to examine all sample points and only the sample points labeled as errors separately. In those cases, 16 different validation data sets were used.

Table 1: Characteristics of validation datasets

	Match = 1.0	**Match >= 0.4**
Property certainty = 1.0 and Classification certainty = 1.0	Dataset 1	Dataset 2
Property certainty = 1.0 and Classification certainty >= 0.7	Dataset 3	Dataset 4
Property certainty * Classification certainty >= 0.7	Dataset 5	Dataset 6
Property certainty >= 0.0 and Classification certainty >= 0.0	Dataset 7	Dataset 8

IV. Methodology

Examining error in the spatial domain

Point pattern methods

Point pattern methods are used to analyze spatial patterns made up of independent events which occur at point locations in space. In general, attribute values are not considered when looking at point patterns, only the location of the event. These methods are used to compare a given point pattern with Complete Spatial Randomness (CSR), defined as a uniform Poisson process with intensity > 0. If CSR is used as the null model, point pattern methods can be used to determine whether a point process departs from CSR. Departures may be caused by non-uniform intensity over the study area, or by dependence between events. I used point pattern methods to examine both the complete sample datasets, and the subset of the each sample dataset consisting of only errors. To determine whether interaction between error points was meaningful, it was necessary to first be certain that there was no such interaction between the sample points themselves.

Kernel intensity

The first step in point pattern analysis is to look for first-order effects, i.e., a coarse-scale, map-wide trend in error occurrences resulting from a variation in the concentration of errors over the map's extent. Kernel intensity was calculated to assess first-order effects using an isotropic Gaussian kernel. The kernel intensity function examines variations in the concentration of events over the study area by smoothing the number of events that are present over a given distance called the bandwidth. The effect is that rather than just a map of the location of individual event, a surface is calculated illustrating how the concentration of events changes over the map's extent. As with most point-pattern techniques, and particularly with the irregularly-shaped study area examined in this analysis, it is necessary to account for edge effects. The kernel intensity function used in this analysis does so, and is calculated as:

where is the kernel, is the count of nearby points, is the bandwidth and

is the edge correction within the window R (Bailey and Gatrell, 1995). A process which exhibits no first-order variations in intensity, either in distance or direction may be called isotropic (Bailey and Gatrell, 1995).

Nearest neighbor index

To determine whether events interacted within the study area, the nearest neighbor index was calculated to identify whether the classification errors were clustered, dispersed or randomly arranged within the map's extent. The nearest neighbor index, as described by Ebdon (1985), calculates the distance from each event to its nearest neighboring event, and summarizes these distance values by calculating the mean nearest neighbor distance for the whole dataset. The mean is then compared to the expected distance given a random arrangement of events. Mean nearest neighbor distance, $\bar{d}$, was calculated as:

—

where is the distance from each event to its nearest neighboring event and is the number of events. The expected mean nearest neighbor distance for a random arrangement of events, , is:

—

where is the number of points and is the area of the study area. Nearest neighbor index is therefore:

—

and ranges from 0 (where all the sample points are exactly coincident with at least one other sample point) to 2.15 (the maximum possible dispersion), with a random arrangement giving a value of 1.

The standard error of the mean nearest neighbor distance behaves as an ordinary standard error of the mean. As a result, a test statistic constructed from the difference between and is normally distributed, and its significance can be tested.

Ripley's K

The purpose of Ripley's K function is to measure second-order clustering of point events, or the spatial dependence among events. Like nearest neighbor, the K-function examines whether events interact locally and exhibit clustering, dispersion, or are randomly arranged. It differs from the nearest neighbor approach because it takes multiple events into account in its calculation. The nearest neighbor statistic only considers the closest event to a selected event, and therefore examines the smallest scale of spatial dependence. Ripley's K examines multiple events surrounding a selected event, not just the closest, thereby examining spatial dependence over a broader scale. In fact, the K-function is defined over a particular distance *h*, so the scale of the analysis is explicit. Calculating K for a given scale inherently assumes that the pattern is isotropic over the defined scale. K is calculated as:

where A is the area of the study area, d_{ij} is the distance between points i and j of the pattern, and I is an indicator function showing whether d_{ij} is less than h (Bailey and Gatrell 1995). As with all point pattern statistics, it is necessary to consider edge effects; w_{ij} is the edge correction component. For the purposes of this analysis, isotropic edge

correction was used. Isotropic edge correction calculates the portion of the circle defined by *h* that falls outside the study area, and scales the count of events within *h* accordingly.

Calculated values of K are compared to the expected value for a process which exhibits CSR. Under CSR, the expected value for K(h) is πh^2. Greater values indicate clustering, because there are more points than expected near a given point; smaller values indicate dispersion.

When examining the results of the Ripley's K, it is common to examine the transformation of *K* called *L*, rather than *K* itself. *L* is defined as

Under CSR, therefore, L(h) = h. This relationship facilitates interpretation of a graph of L(h) versus h, as departures from the 1-1 line (representing CSR) are easily assessed (Baddely, 2008).

The significance of the observed L-function can be assessed, but this requires the definition of the distribution of L under CSR. It is extremely difficult to formally define the expected distribution of L for an observed point pattern, especially given the added complexity resulting from the inclusion of edge correction. Therefore, to significantly assess the presence of clustering or dispersion in events, a simulation approach was used. This approach arranged events randomly within the study area 19 times and calculated L for each arrangement, then calculates the maximum deviation from CSR. L falls outside the limits established by h + and h - with probability——. The result is a 95% confidence envelope describing the distribution of L under CSR (Baddely,

2008). We can then compare the observed distribution of L to the simulation envelope to assess the significance of clustering or dispersion. A value of L that falls above the simulation envelope indicates significant clustering, while a value of L that falls below the simulation envelope indicates significant dispersion.

Correlograms

Geostatistical methods can also be used to measure spatial autocorrelation in events within the study area. Rather than looking strictly at the distance between events, however, geostatistical methods examine the relationship between the distance between two events and the difference in their attribute values. In this analysis, indicator correlograms were calculated to determine the strength of spatial correlation between errors over a range of distances in the study area. Indicator correlograms are used when the attribute being examined is binary (e.g., correct or incorrect prediction), and measure the spatial autocorrelation between points separated by a given distance, h. Indicator correlograms are calculated as:

where C is the covariance (Bailey and Gatrell, 1995). The covariance measures the tendency for attribute values, i.e. errors and non-errors, to change together. Covariance between two given points and , where *s* represents an indicator variable defining the presence or absence of error, is calculated as:

where E(Y(s)) represents the process, and (s) represents first-order variation in the mean of the process. Plotting indicator correlogram values as a function of distance provides information on the magnitude and range of spatial autocorrelation in the dataset. As with Ripley's K, this calculation requires isotropy over the given scale.

Examining error in the attribute domain

Boxplots

As a first step in examining the relationship between the presence of misprediction and potential explanatory variables, a series of boxplots were constructed. Box plots were only constructed for landscape variables that were continuous; i.e., variables measured as ratio data: elevation, slope, aspect, profile curvature, planform curvature, and topographic wetness index. Boxplots are a quick and convenient way to compare the distribution of attribute values for a variable of interest relative to the categories of a second variable. They are constructed by showing the first quartile, median, and third quartile values as a box. The intra-quartile range is calculated, and values down to (Q1 – 1.5* IQR) and up to (Q3 + 1.5 * IQR) are displayed as tails on the box. Values beyond this range are shown as outliers. Comparing the distribution of values for error versus non-error points gives an indication of a possible relationship between the given landscape variable and the presence or absence of misprediction. If the 'boxes' (or intra-quartile ranges) for each category do not overlap, then there is a strong indication that the landscape variable of interest is related to the presence or absence of error.

T-tests

To evaluate the significance of an association between landscape variables and misprediction, two-sample t-tests were also calculated for the continuous variables. Two-sample t-tests were used to test whether the difference between the means of a given landscape variable for error and non-error points was significant, i.e., whether the difference in means was significantly different from 0.

Chi-square tests

Two of the variables included in SoLIM's predictive model were categorical variables: geology and landscape position. In addition, I also compared the presence or absence of error to the predicted and field-identified soil classes, which are also categorical variables. To test for an association between two categorical variables, e.g.., presence or absence of error and one of the variables listed above, a two-way chi-squared test was used. The chi-square test determined whether the distribution of one variable was related to the distribution of a second by testing whether a significant association occurred between the variables under consideration. The chi-square test requires at least five observations for each category. Therefore, categories with fewer than five observations were eliminated from consideration. In the most extreme case, examining soil type for Dataset 1, all but four categories were eliminated.

Relationship between error and SoLIM uncertainty measures

Boxplots were also constructed to assess a possible relationship between the presence or absence of error and SoLIM's internally-calculated measures of uncertainty,

exaggeration and entropy. The significance of a perceived relationship was assessed by calculating t-tests (see above).

V. Results

Overall characteristics of error by dataset

SoLIM's performance for each validation dataset is show in Table 2. The trend is clear: validation datasets that include points displaying higher levels of measurement uncertainty have lower accuracies for SoLIM predictions. Datasets 1 and 2, which consider points with property and classification certainty values of 1, had the highest overall accuracies, while datasets 7 and 8, which considered all points regardless of measurement certainty, had the lowest. It is also clear that lowering the required similarity for correctness led to higher accuracies for SoLIM predictions; accuracy for dataset 2 is greater than that for dataset 1, accuracy for dataset 4 is greater than that for dataset 3, and so on. It is worth noting that SoLIM performs better for datasets 3 and 4 than for datasets 5 and 6. This implies that property uncertainty and classification uncertainty may not be independent.

Table 2: Overall results by dataset.

	Points	No. correct	Pct. Correct
Dataset 1	66	38	57.6%
Dataset 2	66	53	80.3%
Dataset 3	128	66	51.6%
Dataset 4	128	101	78.9%
Dataset 5	241	101	41.9%
Dataset 6	241	170	70.5%
Dataset 7	495	194	39.2%

Dataset 8	495	336	67.9%

Error in the spatial domain

Overall, sample points exhibit relatively constant density across the study area, and error and non-error points can be found throughout the region (Figure 2). In datasets which permit less measurement uncertainty, and therefore have fewer points available, there is a growing gap in the center of the study area, with fewer error points present. However, errors do not appear to be concentrated in one part of the study area, regardless of the validation dataset considered.

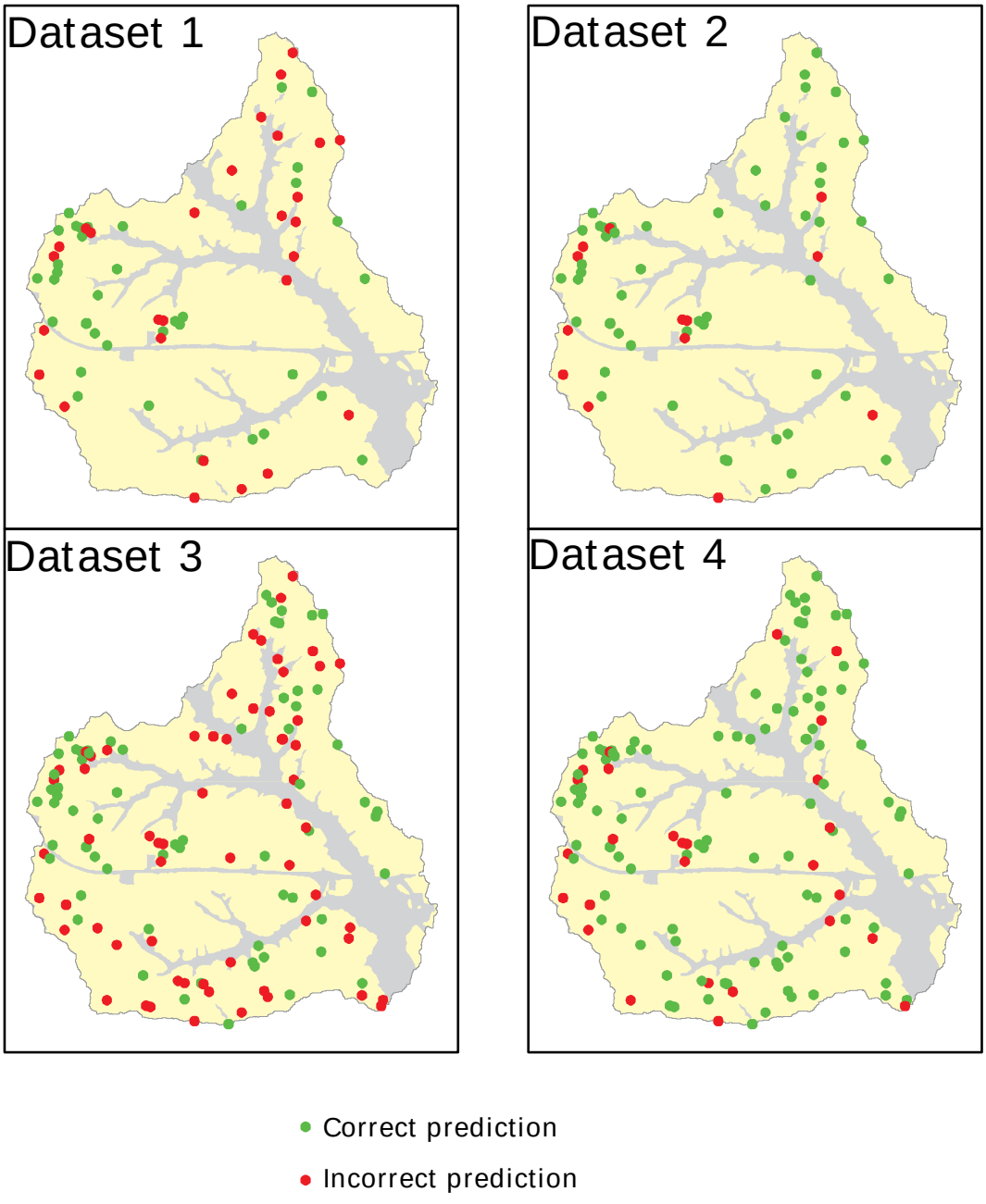

Figure 2: Maps of correct and incorrect predictions by dataset

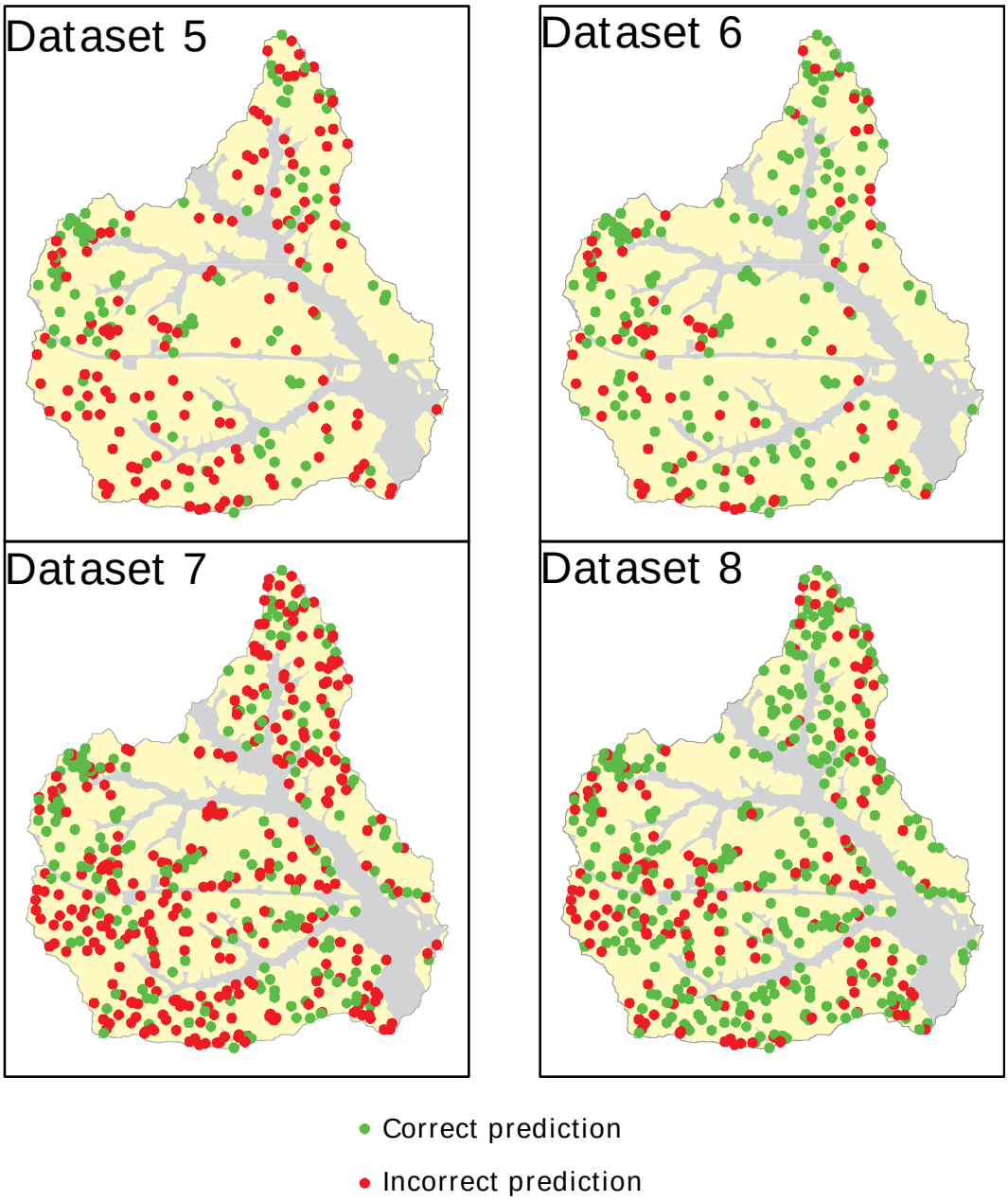

Figure 2 Continued

Kernel Intensity

The kernel intensity function was calculated separately for datasets containing all points and datasets containing only errors (as were all point pattern methods). The resulting maps showed significant variation in the first-order concentration of sample

points and error points over space (Figure 3). This was particularly visible in the concentration of error points. Differences in error intensity were most visible at very coarse scales, with regions at the northern end and southwestern edge having higher concentrations and the central portion having very low concentrations, as previously indicated. The effect is less marked for all sample points, but still visible, with the northern end and northwestern edge showing high concentrations. The results indicate that first-order effects may be present within the study area.

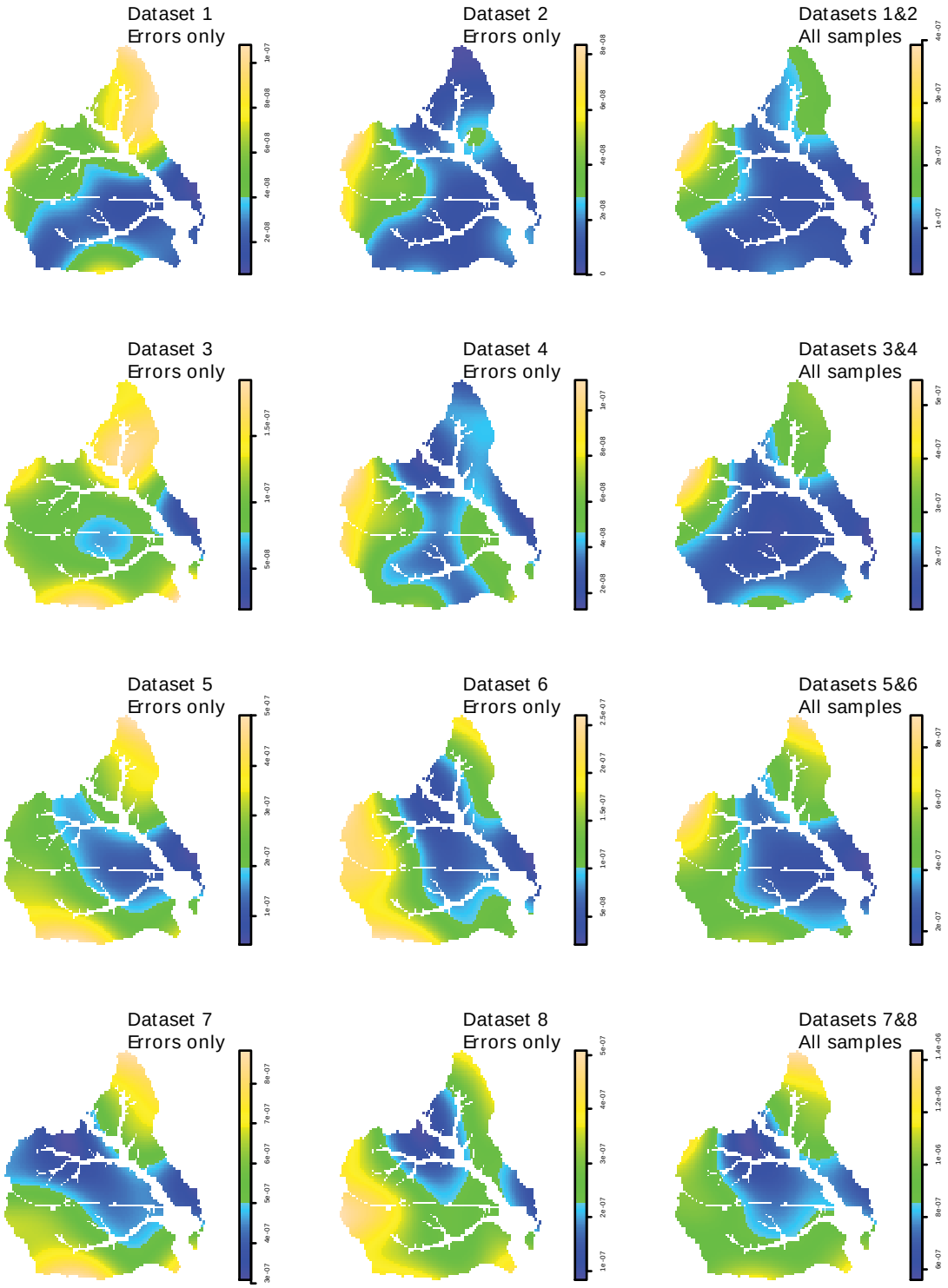

Figure 3: Kernel intensity results by dataset

Nearest neighbor

Nearest neighbor indices, and their associated z-scores, showed no significant evidence of clustering for any of the sample data sets, with only two datasets displaying z-scores greater than 1.5 standard deviations from a random arrangement but having NNI values of ~1 (Table 3).

Table 3: Nearest Neighbor Index by dataset

	NNI (errors)	NNI z-score (errors)	NNI (all points)	NNI z-score (all points)
Dataset 1	0.960377	-0.401509	1.02441	0.379375
Dataset 2	1.178745	1.232927	1.02441	0.379375
Dataset 3	0.992356	-0.11515	0.981219	-0.406494
Dataset 4	1.150804	1.499086	0.981219	-0.406494
Dataset 5	1.024313	0.550344	1.013181	0.39146
Dataset 6	1.010072	0.162355	1.013181	0.39146
Dataset 7	0.964551	-1.176555	0.964514	-1.510384
Dataset 8	1.044961	1.087458	0.964514	-1.510384

This was expected for datasets that include all points (datasets 7 and 8), given that the sample locations were selected randomly, but the trend continued for datasets which eliminate points based on measurement uncertainty. For datasets including error points only, nearest neighbor indices likewise showed no significant evidence of clustering. Although both positive z scores, indicating a trend towards clustering, and negative z scores, indicating a trends towards dispersion, were observed across the various datasets considered, no appreciable pattern based on dataset definition was apparent. All NNI values were close to 1, indicating that any trend towards clustering or dispersion was not noteworthy.

Ripley's *K*

Ripley's *K*, along with the L transformation, was calculated for each validation dataset up to a range of 8000 feet. L showed significant clustering (at 95 percent confidence) between sample points for all datasets at some spatial scales (Figure 4). Clustering is observed beginning at 1,000 feet for datasets 1-6, and is visible all the way to 8000 feet for datasets 1 and 2. This is unexpected for the full validation dataset, as sample points were randomly chosen. While the amount of clustering and the range at which it is significant varies between datasets, there is no overall trend based on measurement error or similarity score.In contrast, L shows no significant tendency towards clustering for most validation datasets composed of error points only. When there is evidence of clustering (strongest in dataset 5 at ranges of 3000 – 8000 feet), the observed clustering was less marked and occurred at greater spatial scales than the clustering of the sample points. Given the observed trend in error intensity (as evidenced by Figure 3), the spatial dependence found over large distances may simply reflect this large-scale pattern.

Spatial autocorrelation

Correlograms did not show any evidence of spatial autocorrelation at any scale (Figure 5). Correlogram values remained near 0 (frequently crossing 0) across all distances for six of the eight validation datasets. The exceptions were Datasets 5 and 7, which showed slight positive correlation between errors and non-errors over small distances (correlation values greater than 0.1). However, the correlogram values quickly approached 0 beyond 2000 feet.

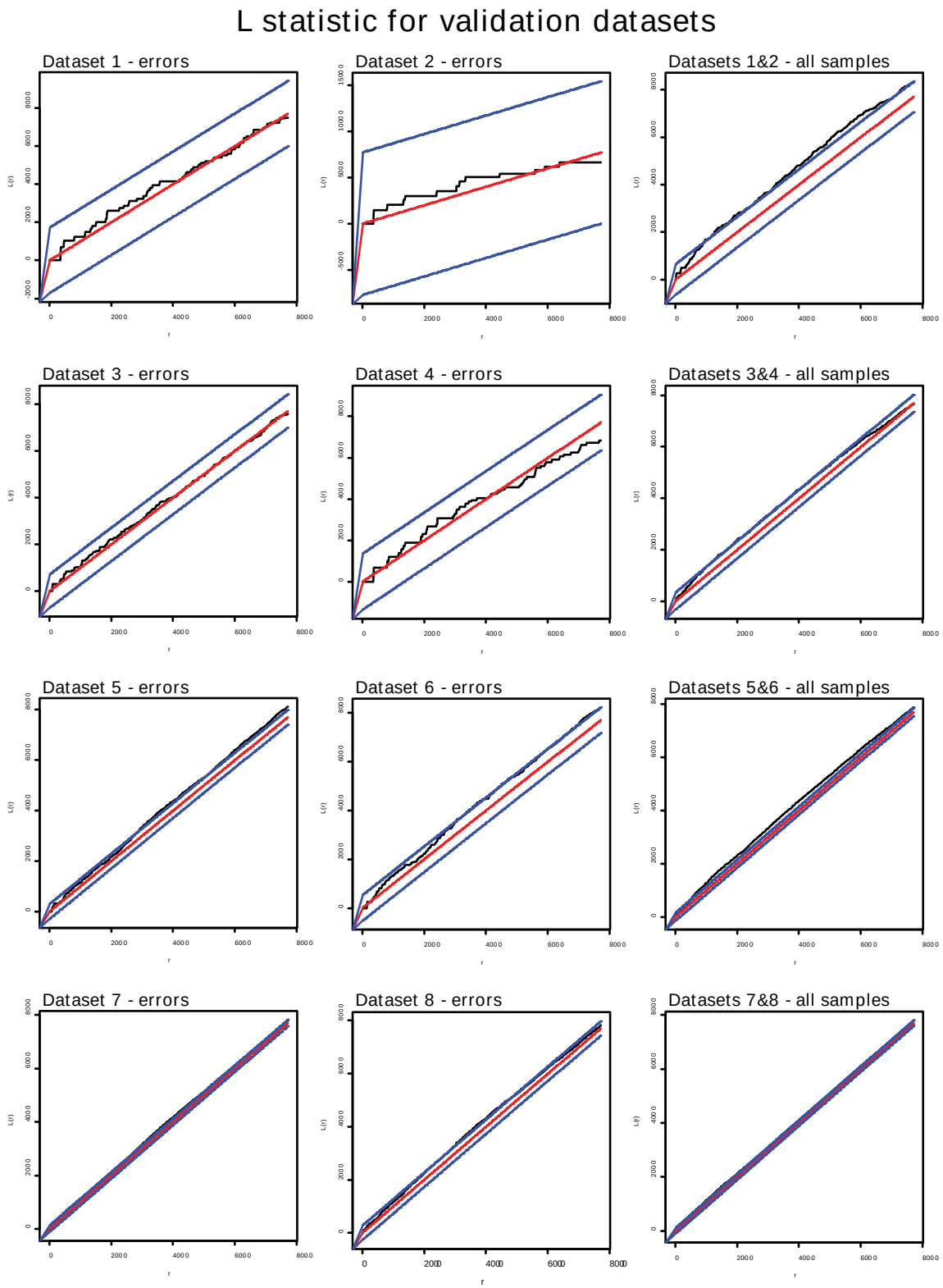

Figure 4: L statistic for validation datasets

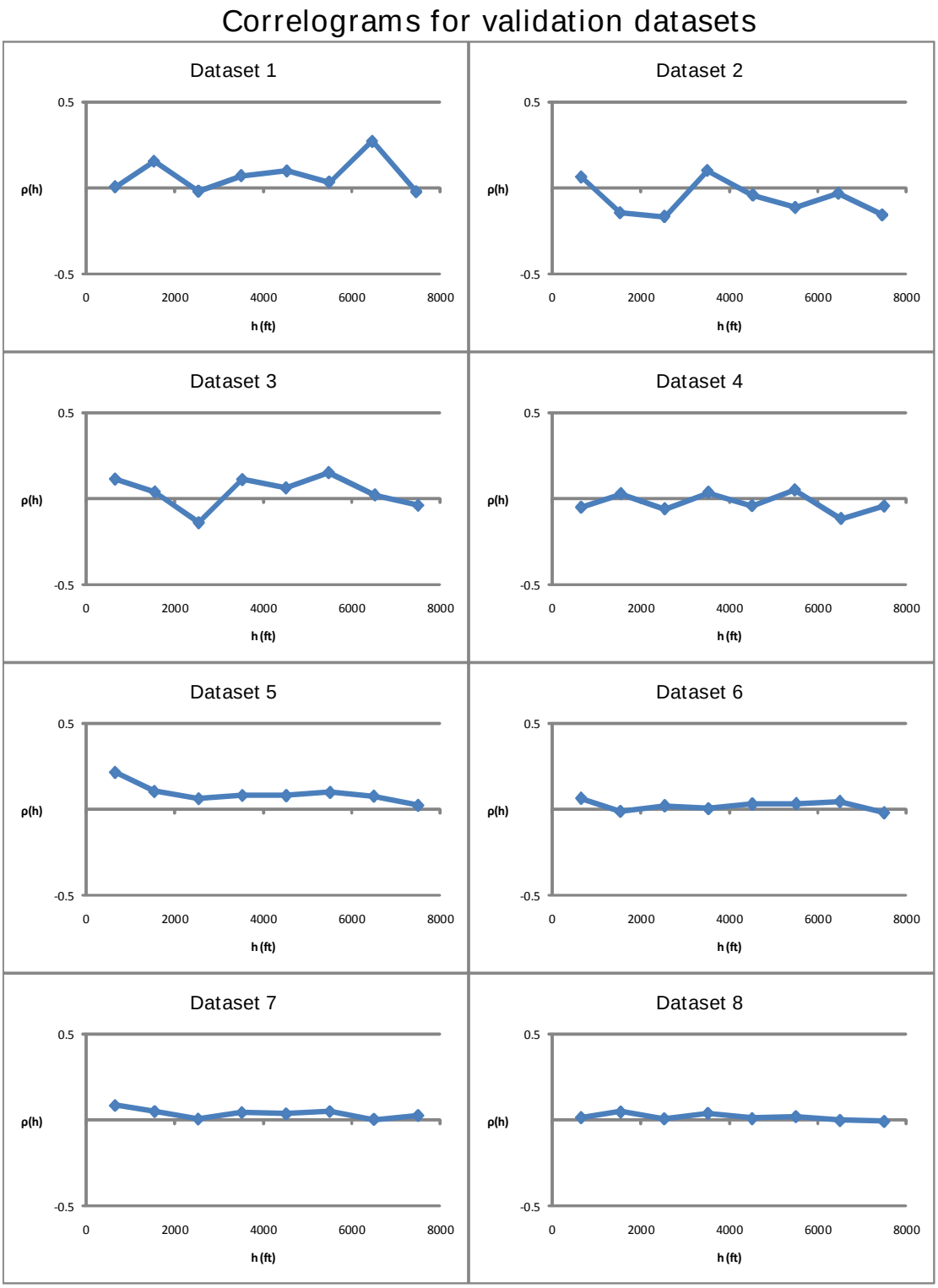

Figure 5: Correlograms by dataset

Error in the attribute domain

Box plots

A total of 48 boxplots were created; each of the six landscape variables were plotted for error vs. non-error points for each validation dataset. The results indicated no visible difference in distribution between error and non-error data points for any of the attributes. Figure 6, chosen as an example, illustrates the distribution of slope values for error and non-error points in Dataset 8. The large degree of overlap observed between error and non-error points was typical for all landscape variables and all datasets examined.

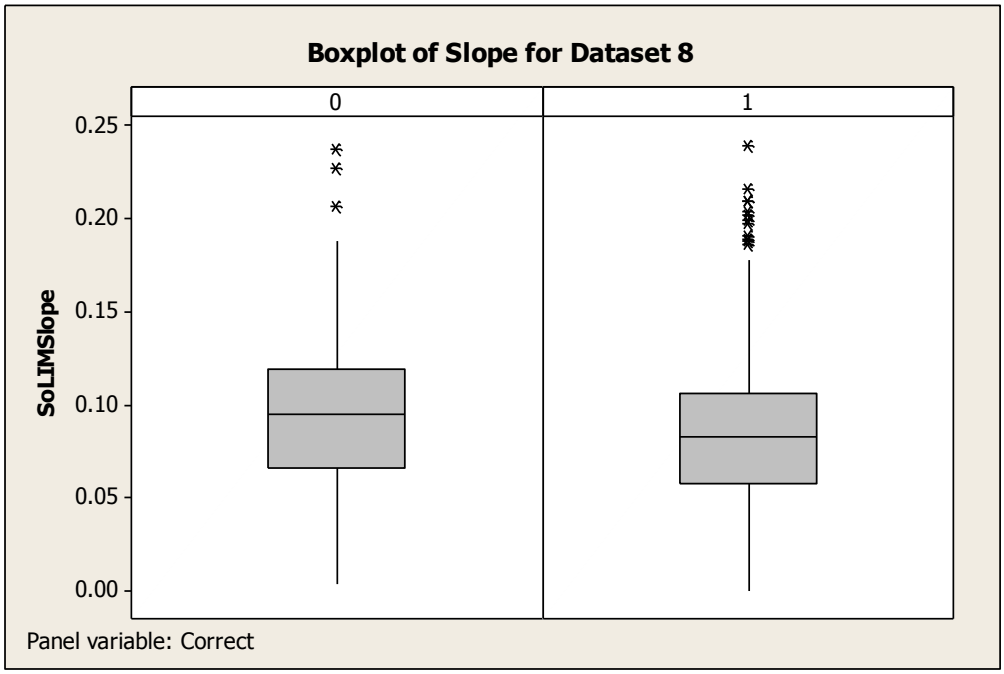

Figure 6: Sample boxplot showing no relationship

T-tests

Two-sample t-tests calculated for the same six variables showed slightly different results (Table 4).

Table 4: T-test results by model input and dataset

Dataset 1 Variable	DF	T-value	P-value
Elevation	54	-0.95	0.348
Slope	60	-0.32	0.753
Aspect	60	-1.99	0.052
Wetness	63	-0.91	0.367
Profile	63	-0.64	0.522
Planform	58	-0.62	0.541
Dataset 2 Variable	DF	T-value	P-value
Elevation	17	0.57	0.578
Slope	18	0.83	0.419
Aspect	**21**	**-2.39**	**0.027**
Wetness	17	0.14	0.891
Profile	19	0.56	0.582
Planform	57	-1.08	0.283
Dataset 3 Variable	DF	T-value	P-value
Elevation	**125**	**-2.08**	**0.039**
Slope	125	-0.47	0.637
Aspect	**125**	**-2.35**	**0.020**
Wetness	124	-0.29	0.772
Profile	120	-1.92	0.057
Planform	104	0.36	0.717
Dataset 4 Variable	DF	T-value	P-value
Elevation	39	0.05	0.963
Slope	46	0.42	0.676
Aspect	45	-1.69	0.098
Wetness	37	0.97	0.336
Profile	57	-0.51	0.609
Planform	111	-0.34	0.736
Dataset 5 Variable	DF	T-value	P-value
Elevation	211	-1.45	0.149
Slope	219	-0.52	0.602

Aspect	**210**	**-2.82**	**0.005**
Wetness	188	-0.56	0.575
Profile	188	-1.24	0.217
Planform	177	0.57	0.567

Dataset 6			
Variable	DF	T-value	P-value
Elevation	144	1.72	0.088
Slope	133	0.79	0.432
Aspect	136	-0.62	0.534
Wetness	127	0.48	0.631
Profile	139	0.47	0.636
Planform	148	0.35	0.727

Dataset 7			
Variable	DF	T-value	P-value
Elevation	398	-0.83	0.408
Slope	426	-0.29	0.771
Aspect	**410**	**-2.13**	**0.034**
Wetness	363	-0.15	0.880
Profile	392	0.11	0.909
Planform	425	0.78	0.435

Dataset 8			
Variable	DF	T-value	P-value
Elevation	**325**	**3.31**	**0.001**
Slope	**301**	**2.69**	**0.008**
Aspect	307	-0.9	0.366
Wetness	331	-0.32	0.752
Profile	365	0.78	0.438
Planform	268	0.34	0.732

Tests exhibiting a significant association between the landscape variable in question and error (at the 95% confidence level) are shown in bold. Topographic wetness index, profile curvature, and planform curvature showed no significant association with error for any of the eight datasets. Slope showed significant association with error for dataset 8, and elevation for datasets 3 and 8. Aspect showed significant association with error for datasets 2, 3, 5, and 7. Four of the six significant test results were observed for validation

datasets that required exact matches between the reference and SoLIM-identified soil classes when defining agreement.

Chi-squared

Chi-square test results varied greatly between validation datasets (Table 5).

Table 5: Two factor chi-square results by categorical variable and dataset.

Dataset 1 Variable	DF	Chi-sq	P-Value
Slope position	3	6.685	0.083
Geology	4	3.484	0.480
Predicted soil	6	11.103	0.085
Sampled soil	3	3.262	0.353

Dataset 2 Variable	DF	Chi-sq	P-Value
Slope position	3	6.09	0.107
Geology	4	6.578	0.160
Predicted soil	**3**	**8.553**	**0.036**
Sampled soil	2	1.685	0.431

Dataset 3 Variable	DF	Chi-sq	P-Value
Slope position	4	2.094	0.718
Geology	5	7.577	0.181
Predicted soil	**10**	**32.913**	**0.000**
Sampled soil	7	13.159	0.068

Dataset 4 Variable	DF	Chi-sq	P-Value
Slope position	4	6.325	0.176
Geology	**5**	**11.469**	**0.043**
Predicted soil	**10**	**29.282**	**0.001**
Sampled soil	7	4.678	0.699

Variable	DF	Chi-sq	P-Value
Slope position	4	2.145	0.709
Geology	7	13.6	0.059
Predicted soil	**14**	**42.904**	**0.000**
Sampled soil	**14**	**45.281**	**0.000**

Dataset 6

Variable	DF	Chi-sq	P-Value
Slope position	4	1.056	0.901
Geology	**7**	**17.835**	**0.013**
Predicted soil	**15**	**46.076**	**0.000**
Sampled soil	15	17.847	0.271

Dataset 7

Variable	DF	Chi-sq	P-Value
Slope position	4	3.773	0.438
Geology	**7**	**29.941**	**0.000**
Predicted soil	**19**	**75.655**	**0.000**
Sampled soil	**22**	**122.852**	**0.000**

Dataset 8

Variable	DF	Chi-sq	P-Value
Slope position	4	1.266	0.867
Geology	**7**	**44.064**	**0.000**
Predicted soil	**19**	**99.132**	**0.000**
Sampled soil	**22**	**57.548**	**0.000**

Tests with significant associations between error and the categorical variable of interest (at the 95% confidence level) are shown in bold. The soil class predicted by SoLIM had the strongest relationship to the presence or absence of error, with seven of eight validation datasets exhibiting significant associations. Significant associations were also observed between the field-identified soil class and error / non-error for datasets 5, 7 and 8 – datasets that contained points with greater measurement uncertainty. Geology exhibited a significant association with the presence or absence of error for validation datasets 4, 6, 7 and 8 – again, datasets with greater measurement uncertainty.

Relationship to SoLIM's uncertainty measures

Boxplots

A set of 16 boxplots were created for the eight validation datasets based on SoLIM's two internal measures of uncertainty. Again, no visible difference in the distribution of

exaggeration or entropy values was observed between error and non-error points for nearly all the validation datasets. The one exception corresponded to Dataset 1, the dataset with the lowest measurement error (Figure 7).

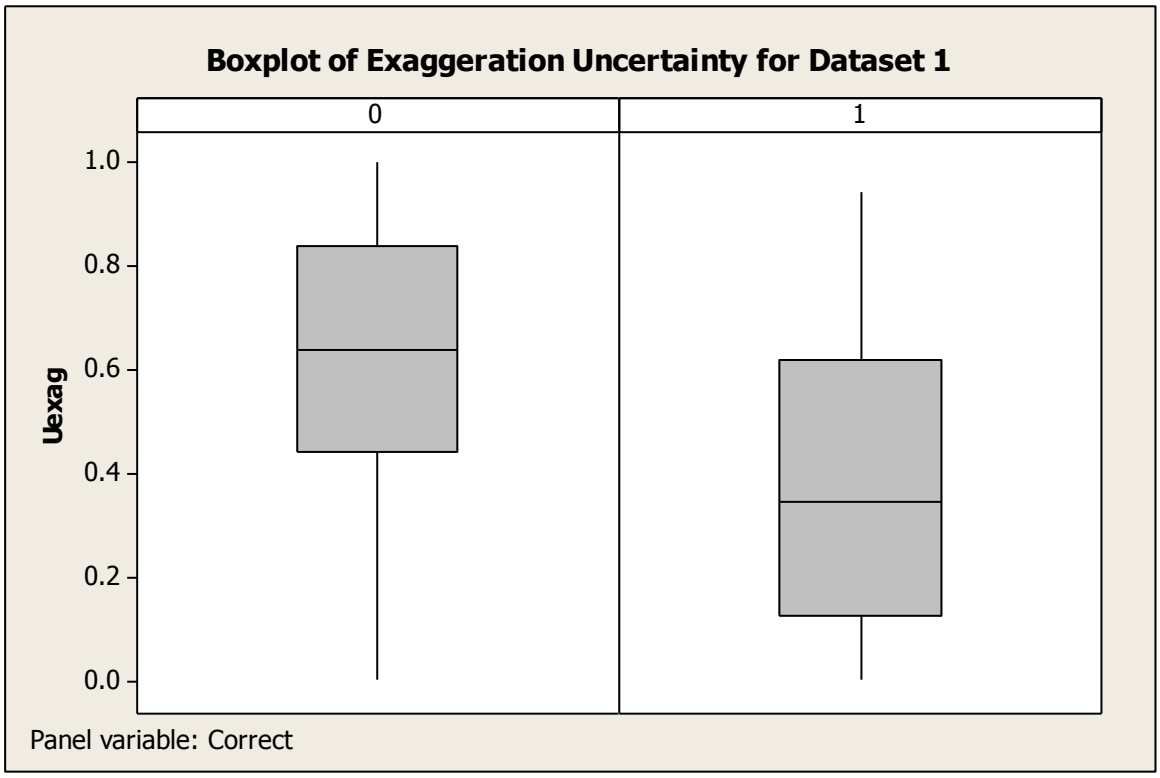

Figure 7: Boxplot for exaggeration uncertainty by error for Dataset 1

There was a visible difference in the distribution of exaggeration uncertainty when comparing error and non-error data points, with lower exaggeration uncertainty values corresponding to correct predictions (i.e., non-error points).

T-tests

There was significant association between the internal measures of uncertainty and error for half of the validation datasets (Table 6).

Table 6: Two-factor t-test results by uncertainty measure and dataset

Dataset 1			
Variable	DF	T-value	P-value
Exag	**59**	**3.57**	**0.001**
Ent	**54**	**2.09**	**0.042**

Dataset 2			
Variable	DF	T-value	P-value
Exag	18	1.33	0.202
Ent	**17**	**3.79**	**0.001**

Dataset 3			
Variable	DF	T-value	P-value
Exag	**125**	**2.76**	**0.007**
Ent	125	1.41	0.162

Dataset 4			
Variable	DF	T-value	P-value
Exag	38	0.43	0.667
Ent	36	1.75	0.088

Dataset 5			
Variable	DF	T-value	P-value
Exag	**206**	**3.35**	**0.001**
Ent	214	1.39	0.166

Dataset 6			
Variable	DF	T-value	P-value
Exag	132	0.62	0.534
Ent	**116**	**2.6**	**0.011**

Variable	DF	T-value	P-value
Exag	**404**	**3.98**	**0.000**
Ent	**419**	**2.82**	**0.005**

Dataset 8			
Variable	DF	T-value	P-value
Exag	309	-0.13	0.900
Ent	**290**	**4.05**	**0.000**

Again, significant associations are shown in bold. An association between exaggeration uncertainty and error was seen in datasets 1, 3, 5, and 7 – all the datasets which require a match threshold of 1.0 for predictions to be considered correct.

Association between entropy uncertainty and error was seen in datasets 1, 2, 6, 7, and 8, though it is unclear what these datasets have in common.

VI. Discussion

Error by dataset

Two trends are evident in the results. First, as is expected, SoLIM performs much better if we count predictions of similar soil types as correct. That is, if instead of requiring an exact match to count as a correct prediction, only a similar match is required, SoLIM has many more correct predictions. This effect is visible at a similar magnitude across all the data sets, with gains in overall accuracy ranging from 22.7% to 28.7%.

The second trend illustrates that SoLIM performs better on data sets with higher property and classification certainty. That is, in cases where the soil scientists were more certain of the soil type they were classifying, SoLIM is more likely to make correct predictions. The magnitude varies, but the trend is very clear.

There are two potential explanations for this second trend. One is that measurement uncertainty in some cases led to misclassification in the field. This misclassification in the validation dataset then led to an inaccurate match between the field-identified soil and the SoLIM-identified soil, resulting in an error. This would suggest that the SoLIM-identified soil was correct, while the field-identified soil was wrong. Another possible explanation is that these soils are genuinely ambiguous, that both SoLIM and a trained soil scientist have difficulty distinguishing between two soils. If this is the case, SoLIM's performance would reflect actual difficulty in distinguishing

between soil types because of inherent similarities between different soils. This explanation is reinforced by the association between SoLIM's entropy measure and error for some datasets. This explanation, however, is somewhat undermined by of our ability to look at similar soils for the analysis rather than just exact matches.

Error in the spatial domain

Ripley's *K* showed some evidence of clustering of sample points, and weaker evidence of clustering of error points, at ranges of thousands of feet in some of the data sets. It is likely that this clustering is caused by first-order variations in error or sample point intensity. Kernel intensity results show a first-order effect, and these first-order variations appear to explain this apparent second-order spatial structure.

Second, all of the spatial tools used (nearest-neighbor index, the k-function, and the correlograms) show weak evidence, at best, of clustering in the errors in any of the datasets. However, this conclusion may not mean that there is no spatial structure to error whatsoever, just that the sampling procedure used to create the validation dataset was not dense enough to show a potential structure. The density of sample points, particularly in the data sets with higher measurement certainty, was very low. The average distance from a sample point to its nearest neighboring sample point for datasets 1 and 2 was nearly 1500 feet. In reality, two points that are 1500 feet apart are not likely to be in the same position on the landscape or share similar values in the other measured landscape variables. It is therefore difficult to draw the conclusion from the spatial tools used for this analysis that there is absolutely no correlation between position in the

landscape and error. Our sample points are most likely too sparse to reveal a spatial structure to error.

Error in the attribute domain

SoLIM uses an explicit relationship between landscape features and predicted soil type, and it is reasonable to assume that a relationship between landscape features and error, or incorrect predictions of the soil type, would exist. I see very little evidence that this is true. There is no consistent relationship between most of the input variables to the SoLIM model and incorrect predictions. One exception is geology, but even that relationship becomes weaker in datasets with higher measurement certainty. In places where SoLIM is doing well, the relationship between geology and error is weaker. The other exception is aspect, which is difficult to explain.

There is significant association between error and predicted soil for all but one of the validation datasets. This relationship can be taken a step further. There are some cases where a single soil type may arise from more than one set of landscape conditions. The model codes these as different soils. In some cases, the chi-squared analysis shows significant differences in error between different membership functions for the same soil type. For example, there are separate entries for “Dubuque” and “Dubuque Medium”, representing two different membership functions, but the same observable soil type. The model performs significantly better for “Dubuque Medium” than for “Dubuque”. This indicates that in addition to there being an association between error and predicted soil type, there is an association between error and the different membership functions for soil

types. This strongly implies that error in SoLIM predictions is not random, but is associated with the membership functions for particular soil types.

On the other hand, there is no association between sampled soil type and error for the more selective data sets. This indicates that the errors are essentially at the model level rather than based on physical attributes found in the field. That is, there are no patterns to the error when we describe them based on the sampled reality; there are only patterns when we describe error based on the modeled membership functions.

The fact that strong association is visible between soil type and error may help to explain why we see some association between geology and error, given that certain soil types are associated with specific geologies. If a specific soil type with a strong association to a particular geology was accurately (or inaccurately) classified, then this relationship may be strong enough to produce similar patterns for geology; i.e., geology demonstrates a relationship with error. This may help explain the association between aspect and error as well. Although aspect is measured as ratio data, it is likely to be treated as a categorical variable when developing membership functions. (That is, a given soil is likely to be described as present on "east-facing slopes", not on "slopes with aspects between 25 and 70 degrees".) Again, if more accurately or less accurately classified soils are associated with specific aspect ranges, this may result in a relationship between aspect and error.

Relationship to SoLIM's uncertainty measures

Exaggeration uncertainty showed an association with error for all of the datasets which require a 1.0 match for a prediction to be considered correct. Exaggeration

uncertainty measures only the difference between the similarity for the most likely soil class and 1.0. Therefore, it is successful only when the requirement is that the predicted class matches exactly; when similar soils also count as a correct match, the relationship is not visible. Entropy uncertainty showed an association with error for five of the eight datasets, though the pattern there is less clear. Still, entropy uncertainty successfully identifies incorrect points under a variety of circumstances. Note also that both exaggeration and entropy show association with error for the datasets with the highest measurement certainty, indicating that SoLIM's internal measures of prediction assessment are successful in identifying regions of the map prone to higher rates of misprediction. The results for SoLIM's internal measures of uncertainty confirm that both exaggeration and entropy are useful measures of SoLIM's performance.

VII. Conclusion

The results of this analysis failed to find evidence to support the primary hypotheses. Classification errors did not demonstrate any spatial structure, most likely resulting from the sampling density used to produce the validation dataset. 500 samples for an 18 square mile area may not be sufficient to show spatial patterns in the error. It may be necessary to take multiple samples on the same land form or slope position before a spatial pattern in error occurs. Even at the highest density (i.e., the most inclusive data set), the sample dataset does not have this level of detail.

Error in the attribute domain similarly shows surprisingly little. In particular, there is no association between most of the model inputs and error. On the other hand,

there is strong association between predicted soil type and error. The strongest conclusion that we can draw is that errors appear to be associated with individual soil types, or more specifically with the membership function for an individual soil type. Errors are not strongly associated with individual inputs to the model across soil types. As a result, it may be possible to improve the performance of the model as a whole by focusing on a few soil types rather than changing the entire SoLIM approach to classifying soil. Without performing extensive sampling and analysis, however, it is impossible to say which membership functions require improvement.

The use of multiple validation datasets shows two interesting trends in the internal uncertainty measures of the predictive model. First, the two uncertainty measures perform differently depending on the threshold for correctness. Exaggeration uncertainty is only associated with error in the case that predictions must match exactly to count as correct, whereas entropy uncertainty is associated with error in a wider variety of cases. Second, both uncertainty measures are associated with error for the datasets with the highest measurement certainty (dataset 1 for exaggeration, datasets 1 and 2 for entropy). However, this association does not necessarily remain in datasets with lower measurement certainty. This clearly demonstrates that accuracy assessment results are impacted by uncertainty within the validation dataset, with a potential for slightly different conclusions based on the dataset they selected. Researchers should use caution if they are aware of potential misclassifications in their reference dataset.

References

Baddely, A. (2008). Analysing Spatial Point Patterns in R. Clayton South, Australia: CSIRO.

Bailey, T., & Gatrell, A. (1995). Interactive Spatial Data Analysis. Harlow Essex, England : Longman Scientific & Technical.

Burrough, P. (1989). Fuzzy Mathematical Methods for Soil Survey and Land Evaluation. *Journal of Soil Science*, *40*(3), 477-492.

Burrough, P., vanGaans, P., & Hootsmans, R. (1997). Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma*, *77*(2-4), 115-135.

Burrough, P. A., & McDonnell, R. (1998). *Principles of Geographical Information Systems*, Spatial information systems. (p. 333). New York: Oxford University Press.

Ebdon, D. (1985). Statistics in Geography. New York: B. Blackwell.

Goodchild, M.F., Chin-Chang, L., and Leung, Y. (1994). Visualizing Fuzzy Maps. *Visualization in Geographic Information Systems* (H.M. Hearnshaw and D.J. Unwin, editors). New York: John Wiley & Sons.

Heuvelink, G. B. M., & Webster, R. (2001). Modelling soil variation: past, present, and future. *Geoderma*, *100*(3-4), 269-301.

Lark, R., & Bolam, H. (1997). Uncertainty in prediction and interpretation of spatially variable data on soils. *Geoderma*, *77*(2-4), 263-282.

Moore, I., Gessler, P., Nielsen, G., & Peterson, G. (1993). Soil Attribute Prediction Using Terrain Analysis. *Soil Science Society Of America Journal*, *57*(2), 443-452.

Scull, P., Franklin, J., Chadwick, O., & McArthur, D. (2003). Predictive soil mapping: a review. *Progress in Physical Geography*, *27*(2), 171-197.

Veregin, H., & Hargitai, P. (1995). An evaluation matrix for geographical data qualty. In *Elements of spatial data quality* (pp. 167-188). Oxford, UK: Elsevier Science.

Zhu, A. (1997). Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering and Remote Sensing*, *63*(10), 1195-1202.

Zhu, A., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, *65*(5), 1463-1472.

Zhu, A., Stiles, C. A., Burt, J. E., & Fehrenbacher, D. J. (2006). Assessment of SoLIM Predictions of Soil Series and Depth to the P-enriched Rountree Formation over a Landscape Continuum. Unpublished research proposal.