

**Testing a Geostatistical Regression Model as a Tool for the
Management of Soil Organic Matter in a Tropical Watershed**

Kristofer D. Johnson

ESE 502

November 2006

Abstract

Soil Organic Matter (SOM) is an important component of ecosystems and the Earth's Carbon cycle. It is possible that landscapes can be managed so that this desirable soil property can be maximized or protected. There is a current need to predict SOM with statistical and spatial models the distribution of SOM at small scales in forested environments. The current study identifies terrain variables that correlate with SOM data from a small tropical landscape, the Bisley Experimental Watershed in Puerto Rico, and experiments with Geostatistical Regression to predict a high resolution SOM surface. The results show that the Geostatistical Regression model more realistically represents the distribution of SOM at Bisley than Ordinary Kriging. However, the OLS model accomplishes virtually the same results without using information from neighboring point data because spatial dependency is not well represented in this dataset. Although specific values of SOM could not reliably be predicted over the watershed with any of the models, the known distribution of SOM was expressed successfully in the OLS model (i.e. higher SOM values occurred on ridges).

Introduction

Soil Organic Matter (SOM) can be defined as the organic component of soils which is biologically produced and contains the element Carbon (C). Inorganic materials (i.e. minerals) make up the remainder of soil material. SOM is essential to terrestrial ecosystems and benefits plants by holding water, storing nutrients and reducing erosion. SOM also plays an important role in global climate change because the C in SOM exchanges with CO₂ – a greenhouse gas produced from fossil fuel burning. Should CO₂ emissions continue to increase (i.e. “business as usual”) and temperatures rise due to the “greenhouse effect”, a number of impacts are hypothesized. Some of the potential negative impacts may include rising sea levels and increasing occurrence and severity of natural disasters such as drought and hurricanes.

In response to this risk, a number of strategies have been proposed that either limit how much C is released into the atmosphere, or, mitigate the C that is already in the atmosphere. One such strategy, referring to the second of the above options, is to take

advantage of the fact that an accumulation of SOM represents that amount of C which is taken out of the atmosphere and stored in soils (see “Background”). The appeal of this strategy is that it is viewed as a “truly win-win strategy” because the C stored as SOM benefits ecosystems while also providing a natural means to mitigate CO₂ build-up in the atmosphere (Lal 2004). To pursue this option, ecosystems and landscapes must be managed so that SOM storage is maximized. This generally entails either adding organic materials to the soil (e.g. reforestation) or protection of SOM resources that already exist (e.g. abating deforestation).

Because the world’s forests store a significant portion of the world’s C (about 50% of land-based C), forested landscapes are potentially useful for CO₂ mitigation (IPCC 2001). Essential to managing SOM in forested landscapes is an understanding of its spatial distribution. The spatial prediction of SOM is non-trivial because its formation is influenced by many complex factors. Most studies dealing with spatial estimates of SOM are continental or global in scale because SOM is strongly correlated to large scale gradients in climate and geology. However, increasingly important is the question of how to manage SOM on more local to regional scales, such as in a single watershed or parcel of land. At smaller scales, SOM varies greatly over space and with depth in the soil profile and is probably more related to topographic variables and vegetation types. Further, forested landscapes, in particular, provide opportunities to learn about how SOM accumulation can be enhanced and where higher levels of SOM occur. Yet, sparsely represented in the current literature on the prediction of C storage in soils are forested landscapes, and more specifically, tropical forested landscapes.

Interpolated maps are useful tools for managing SOM in landscapes because they identify areas of high and low SOM. Statistical models can also help in the creation of SOM maps. Certain terrain characteristics, for example, are known to be correlated with SOM levels. Variables that measure the distance to the nearest ridge or slope of the soil surface can capture these characteristics and then be used in multiple regression models. These kinds of variables are commonly derived from digital elevation models (DEM) based on contour lines. DEM data is becoming increasingly available and has the advantage of

modeling continuous surfaces that provide information between measured points. Also potentially useful are models that consider spatial structure by utilizing information from neighboring measured points. To wit, “Geostatistical Regression” (GR) is a model that fits data after accounting from both spatial structure as well as secondary information, such as terrain variables.

The current study identifies terrain variables derived from DEM data that explain a portion of the variation in SOM at a small tropical landscape, the Bisley Experimental Watershed in Puerto Rico. These terrain variables are then used as secondary information in the GR model to predict a high resolution SOM surface. The results of the GR model are compared to measured data and two other interpolation models to determine how reliably the GR model can be used to guide the management of SOM in landscapes.

The Bisley Experimental Watershed is a small tropical watershed covering 13 hectares (about 3 city blocks) located in the Caribbean National Forest, Puerto Rico (Figures 1 and 2). Three main vegetation types occur throughout the National Forest at different elevations. Bisley is located in the lower elevation “tabonuco forest” (*Dacryodes excelsa*). Bisley has been the focus on many studies, but is perhaps more frequently associated with the disturbance studies produced as a result of Hurricane Hugo in 1989 (Scatena et al. 1993,1996). The landscape is dissected with many small ephemeral stream channels sometimes separated by only a few meters. Three landforms (ridge, slope and valley) have been determined to characterize the terrain (Scatena and Lugo 1995). SOM was measured at gridpoints spaced every 40 m (85 total observations) and at three depths (0-10cm, 10-35cm, 35-60cm) with a 1-inch coring device (Figure 2).



Figure 1. The location of the Bisley Experimental Watershed in Puerto Rico.

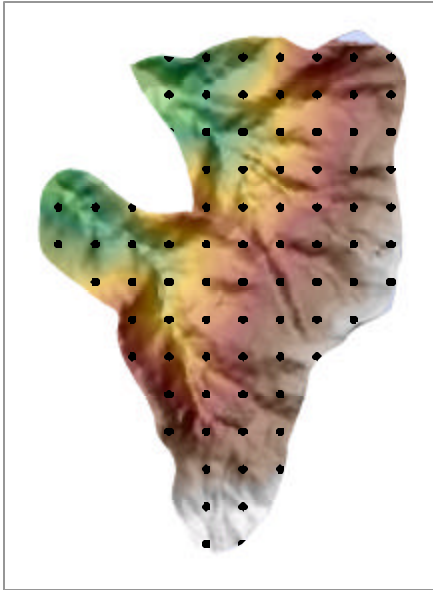
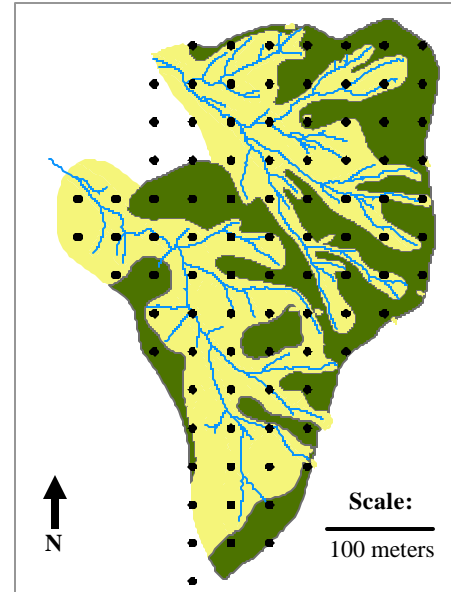


Figure 2 (left). The Bisley Experimental Watershed. This hillshade shows the relatively dissected terrain and elevations are indicated by the color bands. The highest elevations occur in the southernmost area (white). The locations of sampled points are shown by the black dots and are spaced every 40 meters.

Figure 3 (right). The distribution of tabonuco trees (dark green) and other species (light green) in the Bisley Watershed. The tabonucos also represent ridge areas. The lines represent streams which flow to the northwest.



It was previously observed that SOM content was highest in ridge areas where the soil profiles are deepest and where the forest is dominated by the tabonuco tree (Scatena and Lugo 1995) (Figure 3). The tabonuco tree species is a hurricane-resilient tree that has smaller leaves than the surrounding vegetation. During high winds, the leaves are quickly dropped in order to reduce wind resistance and minimize stress on the tree. Additionally, individual trees are often linked to each other in a network of root grafts so as to avoid being uprooted. The tree and its roots provide physical stability for the soil so that much of the SOM is saved from being transported during heavy rainfall. In contrast, slopes and valleys are dominated by palms and other species that are more easily toppled in strong winds. Further, as water collects at these locations the combination of flooding and high pore pressure in the soils causes a higher occurrence of landslides and treefalls (Scatena and Lugo 1995). Therefore, more SOM preferentially accumulates in stable ridge areas where tabonuco vegetation dominates. This is opposite of what is normally observed for drier areas where SOM tends to be higher in the valleys.

Background

The Earth's temperature has increased by about 0.6°C over the last 100 years. This is likely the largest increase in temperature for any century in the past 1000 years (IPCC 1992). Additionally, the amount of atmospheric "greenhouse gases" has also increased

by about 30% in the last 150 years, presumably due to a combination of the combustion of fossil fuels, deforestation and the agricultural activity (Sparks 2003). The atmosphere warms because the increase of greenhouse gases allows for less of the reflected energy from the Earth's surface, to leave the atmosphere. There are many potential implications to this warming. Sea level may rise and weather patterns could change, possibly leading to increased negative impacts in the health and natural resources sectors (Table 1; UNEP 2005). It should be noted that none of these implications are well understood and there are likely more implications that are not even known.

Table 1. List of Potential Negative Impacts of Global Climate Warming

	Negative Impacts
Health	Weather-related mortality Infectious diseases
Agriculture	Crop yields Irrigation demands
Forest	Forest composition Forest health and productivity
Water Resources	Water supply Water quality
Coastal Areas	Beach erosion Inundation of coastal areas
Species and Natural Habitats	Loss of habitats and species

The most abundant and influential of the “greenhouse gases” is carbon dioxide (CO₂). Like many atmospheric gases, the amount of CO₂ in the atmosphere at any given time is not constant. In fact CO₂ concentrations change daily, or even hourly, as it is converted to and from gaseous, liquid and solid forms of Carbon (C). Like any chemical system, there is a constant flux from one state of C to another in order to achieve equilibrium in Earth's system as a whole. The physical and chemical states of C in Earth's system are categorized into three global “pools”: atmospheric, terrestrial and oceanic. We refer to the movement and recycling of C as it passes from one pool to the next as the “Global C Cycle” (Falkowski 2000; Schlesinger 1997). Figure 4 is a simplified version of global C cycle as we know it.

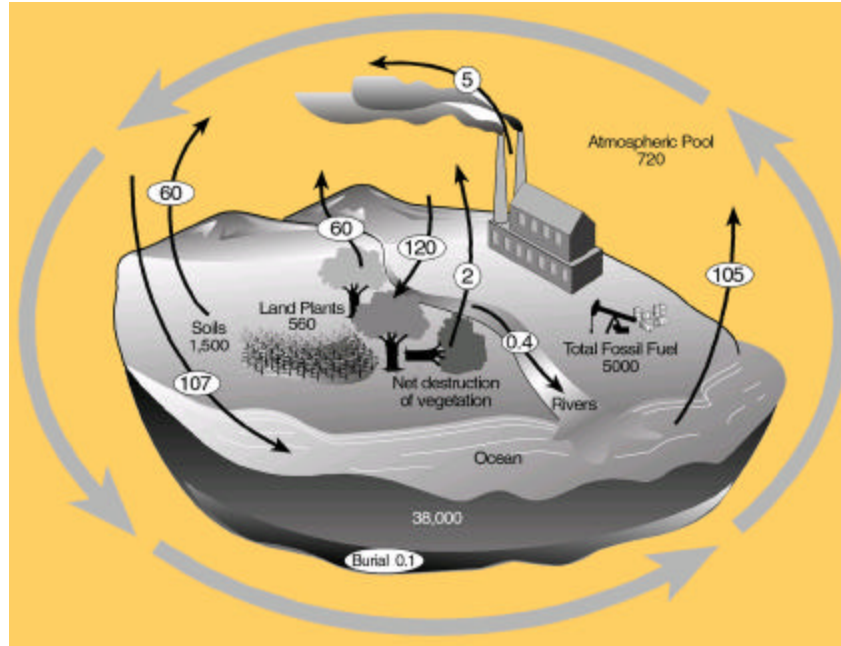


Figure 4. The Global Carbon Cycle. Pool sizes are shown below the pool labels. Circled numbers are the annual fluxes between pools. All values are in units of Petagrams (1Pg = 1×10^{15} grams) (From Kansas State University Research and Extension webpage: www.oznet.ksu.edu/ctec/Carbon/carboncycle.htm)

Because the increase in CO_2 concentrations is thought to be the main culprit of warmer temperatures, most options to mitigate climate change are focused on reducing this C pool. There are at least three ways to reduce atmospheric CO_2 : 1) reduce fossil fuel emissions by using alternative forms of energy, 2) replace fossil fuels with bio-fuels that only add recycled C to the atmosphere and 3) increase the amount of C that is stored in vegetation and soils (Janzen 2004). This study is motivated by the third option, otherwise known as “soil C sequestration”. This option depends on our ability to successfully model the spatial distribution of SOM

SOM and the Terrestrial-Atmospheric C cycle

Most soils can generally be separated into two components, organic matter (SOM) and mineral. Though the SOM component is usually only 0.5 to 5% of the bulk soil, it is essential to most ecosystems. Plants benefit from the physical and chemical properties that SOM and mineral soil together provide such as improving water retention, storing nutrients and reducing erosion (Lal 2004; NRC 2000; Sparks 2003). About half of SOM

is made up of the element C. The term “soil organic C” is often used interchangeably with SOM even though they are not exactly the same.

The pathway of atmospheric C to soil C is captured by the terrestrial portion of the global C cycle (Figure 5). The pathway of atmospheric C to soil C begins with plant photosynthesis. Photosynthesis converts CO₂ to plant C in the form of plant carbohydrates. In the next step of this pathway, dead plant C (e.g. leaf, root and trunk material) is decomposed by microorganisms. The decomposition process recycles virtually all of the C in dead plant material (now considered a part of the “soil organic C” pool) to the atmosphere in the form of CO₂, thus completing the cycle. Therefore, the amount of atmospheric C converted to soil C (~60 Pg) and the amount of soil C converted to atmospheric C are approximately equal (Janzen 2004).

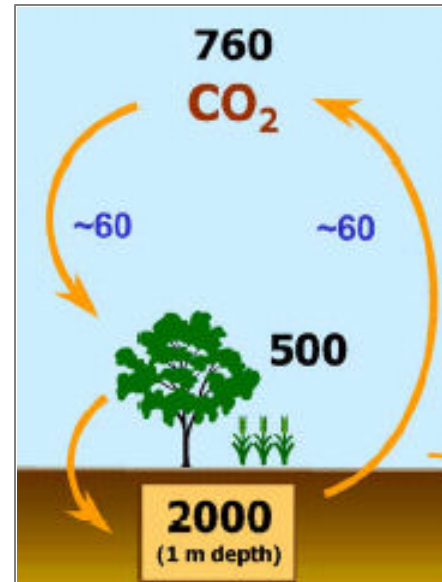


Figure 5. The Terrestrial Carbon Cycle. The large bolded numbers are the pool sizes in units of Pg C of the atmospheric, vegetation and soil pools. The arrows and smaller numbers indicate the exchange paths and rates of C exchange from pool to pool (after Janzen 2004).

It would appear, therefore, that no substantial amount of C would occur in the soil pool. However, note that there is about 2 to 3 times more C in soils than in the atmosphere or in vegetation. This occurs because the rate of decomposition of different organic materials varies. For this reason SOM is further subdivided into more pools differentiated by the length of time the C is held in the soil. For example, easily decomposed organic material may only remain in the soil for less than 10 years, while organic material that has been modified by soil microorganisms may remain for decades or centuries (Sparks 2003). A smaller fraction may still be converted to virtually indecomposable forms (e.g. charcoal produced in fires) (Post and Kwon 2000; Pregitzer 2003; Prentice 2001). Thus, the accumulation of SOM occurs when the balance between the total organic C inputs are greater than the total loss of C due to decomposition over a defined period of time (Prentice 2001; Schlesinger 1997). We can think of this process as giving a dog a turkey

drumstick. The meaty parts are eaten up within minutes while the tougher parts (cartilages) might not be consumed for some hours. Still other organisms such as maggots could finish off all edible parts over the period of weeks until all that is left is the bare bone.

Rates of SOM accumulation not only depend on the length of time, but also on environmental factors (soil moisture, plant chemistry and temperature) which vary over local to global scale (Schlesinger 1997). Therefore, the spatial variation of SOM is complex and difficult to describe. At smaller scales, SOM is also influenced by natural disturbances (e.g. fires), human influences (e.g. forest clearing) and topography (e.g. valley versus ridge landscape position) (Brown and Lugo 1990; Ostertag et al. 2003; Post and Kwan 2000; Silver et al. 2000). To further complicate the issue, SOM varies with soil depth. It is clear that SOM decreases with increasing depth in the soil profile, however it is uncertain that the variation in deeper SOM is controlled by the same factors as variation in SOM closer to the surface (Jobaggy and Jackson 2000). In this study we focus on SOM contained in the top 60 cm of soil. Despite the challenged mentioned above, in many landscapes a portion of the variation in SOM can be explained by one or more environmental factors.

Soil C Sequestration Policy and Feasibility

As policies addressing the mitigation of the “greenhouse effect” have drawn more attention, soil C sequestration practices have become increasingly visible in politics and policy making (Richards 2004; Smith et al. 2000). The conversion of agricultural land to forest represents one of the most promising strategies for C sequestration, especially for marginal or degraded soils (Lal 2004). Similarly, the protection of forests from clearing for agriculture holds C that would otherwise be lost to the atmosphere. Atmospheric CO₂ is estimated to be increasing at a rate of 3.3 PgC/year (IPCC 2001). Lal (2004) estimates that the global potential of C sequestration in soils is about 0.9 ± 0.3 PgC/year, or about one-fourth to one-third the annual increase in CO₂. Though the potential of C sequestration in soils to offset GHG emissions is hopeful, experts predict that soil C sequestration practices will only have significant impacts over 20 to 50 years because the

total amount of C to be sequestered and the duration of storage is finite (Lal 2004; Smith 2004).

The European Union and other countries have committed to the Kyoto Protocol which requires greenhouse gas (GHG) reductions and allows for C sequestration practices to offset GHG emissions (Smith et al. 2000). In the United States, there are no coercive instruments that limit C emissions. However, Richards (2004) showed that voluntary GHG reduction policies introduced in Congress for the years 2000 to 2004 have led to more than 50 pieces of legislation that either directly or indirectly address C sequestration. These policies include tax incentives for C sequestration projects, managing C on federal government lands, and public education and subsidized loans for the protection of forests.

Methods and Data

One of the objectives in this study is to identify factors which explain SOM variation in the top 60 cm of soil in a tropical watershed. The effect of landscape stability on SOM content was modeled by variables derived from a digital elevation model (DEM) produced from a topographic map of the watershed. This approach was chosen because DEM data is becoming increasingly available and provides a continuous surface from which to extrapolate data. Ten meter contour lines were imported into ArcMap and converted into a DEM with a default output cell size of 0.73 meters. Thirteen variables derived from the DEM were analyzed for correlation with SOM. Three of the variables are described here, all of which are significantly correlated with log transformed SOM ($\ln\text{SOM}$) ($p\text{-value} < 0.05$). Flow accumulation ("*flow*"; Figure 6a) is the accumulated weight of flow at a downslope point. This variable was highly skewed, so a natural log transformation was applied. High flow accumulations simulate the distribution of stream channels and valleys as well as the magnitude of water flow over non-stream areas. In general high flow accumulation indicates higher possibility for the physical removal of SOM (Figure 6b).

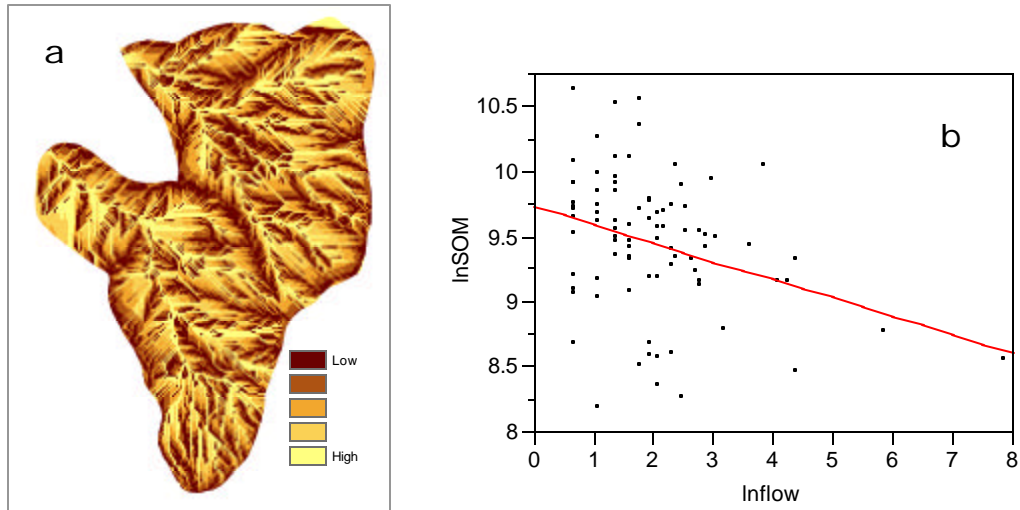


Figure 6. a) Flow accumulation (“flow”). Lighter colors indicate higher flow. The highest flow areas are the stream channels. b) The relationship of lnSOM and *flow* (p-value of linear fit = 0.0021).

Curvature (“*curv*”; Figure 7a) refers to the shape, or relative concavity or convexity of the soil surface. Curvature potentially describes local drainage and erosion processes where, hypothetically, convexity allows for more erosion and therefore less SOM accumulation. A negative curvature value represents a concave surface at that point, while a positive value would indicate a convex shape. In Bisley, SOM is significantly higher when curvature values indicate flatness (-37 to 70) or mild concavity (-221 to -37) and lower when values indicate strong concavity (less than -221) or convexity (greater than 69.8) (Figure 7b).

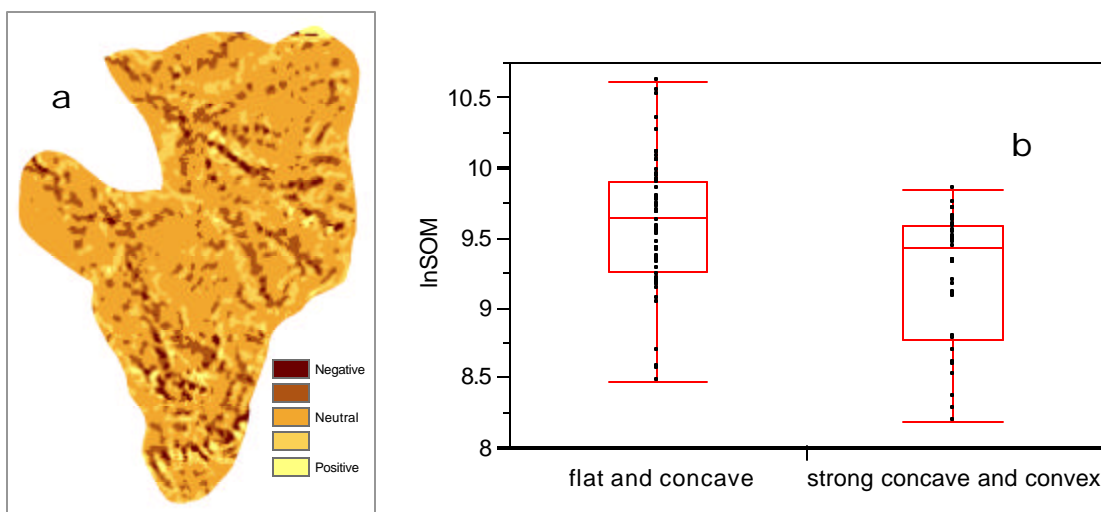


Figure 7. a) Curvature output from digital elevation model. Darker colors are more concave and lighter colors are more convex. “Neutral” color indicates flat areas. b) lnSOM values classified by curvature with statistically significant difference in means (student t-test, p-value < 0.05). The boxes indicate the median, 25th and 75th quartiles. The lines extend to the highest and lowest values.

The distance to stream (“*dstrm*”; Figure 8) was another variable used to represent stable areas versus areas more affected by water disturbance (i.e. ridge and valley areas). In Bisley, areas greater than 5 meters from a stream channel had significantly higher SOM than areas within 5 meters of a stream channel.

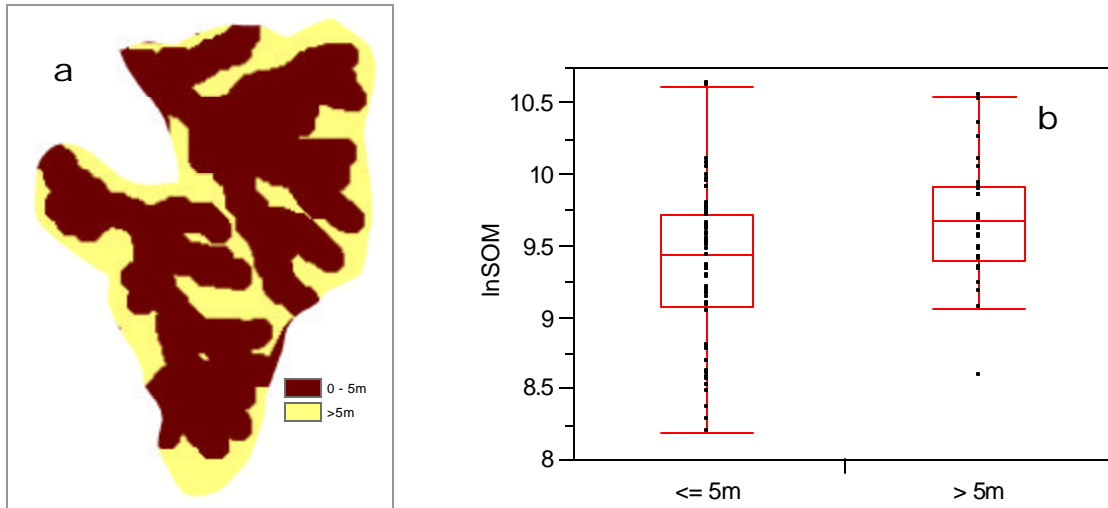


Figure 8. a) Areas in the watershed that are greater than 5 meters and areas that are less than 5m from the nearest stream channel. **b)** lnSOM values classified by *dstrm* with statistically significant difference in means (Tukey-Kramer, *p*-value < 0.05). The boxes indicate the median, 25th and 75th quartiles. The lines extend to the highest and lowest values. The point outside the line is a possible outlier.

The above variables were used in an Ordinary Least Squares Regression (OLS) where lnSOM was regressed on *flow*, *curv* and *dstrm*. This explained about 25% of the variation of lnSOM in the top 60 cm of soil at Bisley. The relationships between lnSOM and the three terrain variables are shown by equation (1) (note: β_i is the estimated regression coefficient and e is the residual) and the parameter and summary of fit tables (Tables 2 and 3).

$$(1) \quad \ln\text{SOM} = -\beta_1 \text{flow} - \beta_2 \text{curv} + \beta_3 \text{dstrm} + e$$

Table 2. Parameter estimates of multiple regression on lnSOM

Term	Estimate	Std Error	Prob> t
<i>Intercept</i>	9.693737	0.099262	<0.0001
<i>flow</i>	-0.121652	0.041249	0.0042
<i>curv</i>	-0.157437	0.051788	0.0032
<i>dstrm</i>	0.127384	0.052467	0.0174

Table 3. Summary of fit statistics of multiple regression on ln SOM

RSquare	0.273034
RSquare Adj	0.24611
Root Mean Square Error	0.449991
Mean of Response	9.451362
Observations	85

Geostatistical Regression

As noted earlier, Geostatistical Regression (GR) is a method of interpolation which takes advantage of secondary information as well as the observed values from neighboring points (Bailey and Gatrell pp. 189-190; Hengl et al. 2004). Hengl et al. (2004) used GR to predict the concentration of organic matter for a 50 X 50 km square in Croatia from 135 observations. They found that GR performed better than the ordinary kriging method of interpolation (Root Mean Square Error of 53.3% versus 66.5%). In general, GR has not been widely used for the prediction of SOM in the United States or in forested landscapes.

The secondary information component of the GR model includes information that is useful for the prediction of SOM. This information can be obtained from the continuous surfaces of the terrain variables discussed above (Figures 6-8, Figure 17 in Appendix 1). The spatial structure component of the model describes the influence that the known values at neighboring points have on the value at a specific predicted point - estimated from the spherical variogram (Figure 17) (for a discussion of the variogram see Bailey and Gatrell pp.161-166). The mathematical explanation of GR is included in Appendix 1.

Results

Investigating Spatial Dependence

There is apparently a spatial dependence of observed SOM over the whole watershed. When lnSOM is regressed on the X and Y coordinates separately there appears to be some influence of these parameters on lnSOM, though they are not significant (p-value = 0.1245 and 0.1034 respectively). When lnSOM is regressed on the product of these terms, XY, the new parameter is significant (p-value < 0.05) and indicates that SOM

increases from the southwest to northeast portion of the watershed (Figure 9). This relationship will not be modeled in this study since we are more concerned with smaller spatial scales. Also note the sigmoidal pattern with increasing XY. The points at the three “peaks” in Figure 10 are all points that are located on tabonuco/ridge areas. This may reflect the orientation of the major ridges which roughly stretch along the NW-SE axis (i.e. an anisotropic effect). The orientation of the ridges is related to the structure of the underlying geology (Scatena and Lugo 1995). Modeling this effect probably requires a more complex model than considered in this study (see also Ordinary Kriging results in “Results” section).

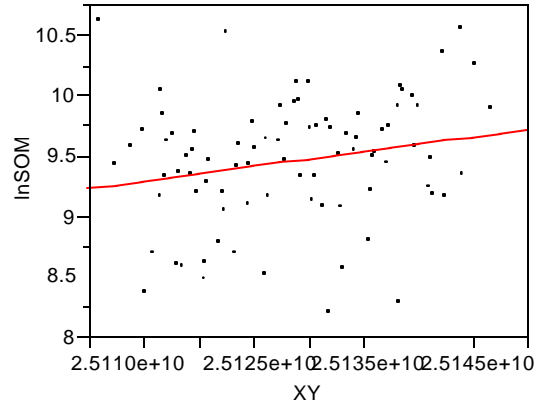


Figure 9. Relationship of lnSOM and the product of the X and Y coordinates in the Bisley Watershed.

Spatial autocorrelation among the measured SOM values were investigated using a spherical variogram. The variogram of the untransformed SOM data revealed a large nugget effect and little spatial structure (bandwidth = 305 meters; Figure 10). This is not surprising since within this small but deeply dissected watershed, the terrain varies substantially between the sampled points. This means that the local scale spatial autocorrelation that is related to the presence of ridges and streams (i.e. areas of stability and disturbance) is probably not being captured by the sampled points. Therefore, for the purposes of estimating small scale (20-30 meters) patterns of SOM in a landscape such as Bisley, more intensive sampling is required between the points.

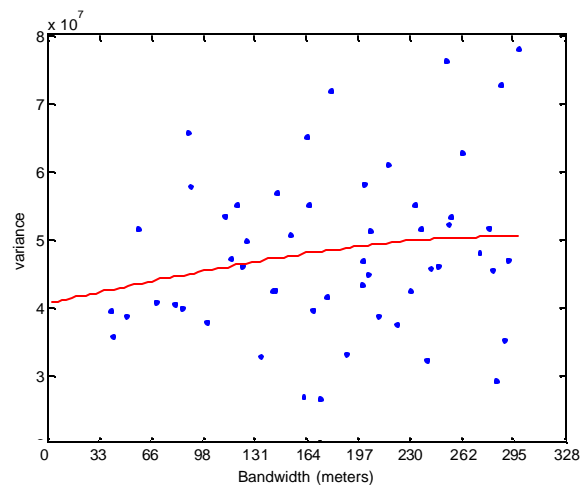


Figure 10. Variogram of the SOM and coordinate data in the Bisley Watershed.

The variogram model for the log transformed SOM values (Figure 11, bandwidth = 305 meters; a bandwidth of 160 meters did not improve the fit) suggests that spatial dependency exists at scales smaller than the range (<103.5 meters). However, the substantial scatter about the fitted line makes this interpretation suspect. To check the spatial dependency, the residuals of the OLS model were fit to a variogram and compared to the original variogram (Figure 12). There is no spatial dependence after accounting for the variation in SOM with the three terrain variables. Note that the existing OLS model probably already accounts for the spatial dependency at this scale. This may be due to the presence of the “distance to stream” variable. Recall that major ridges are identified by the distribution of tabonuco trees (Figure 3). Major ridge to stream transitions occur within a range of about 100 meters so that variation at this scale is represented in the dataset.

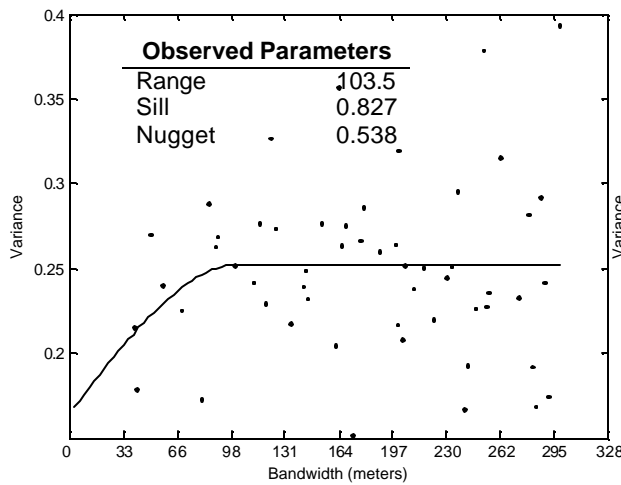


Figure 11. Variogram and parameters of observed lnSOM.

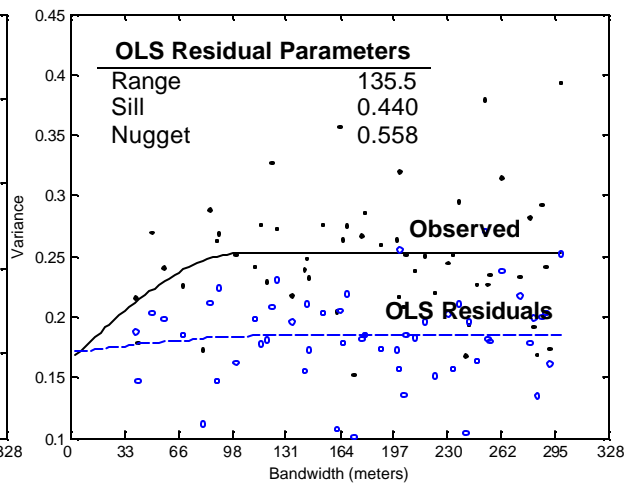


Figure 12. Variogram of observed lnSOM (solid line) and the residuals of the OLS regression (dashed line).

Because of these results, we can already conclude that using information from neighboring points is not particularly helpful for modeling the small scale distribution of SOM in Bisley using this dataset. The reasons for this are: 1) The observed values are located every 40 meters, which cannot account for the variation of SOM values that likely occur in between these locations, 2) the OLS model seems to already account for

spatial dependency among the observed variables. Nonetheless, since this is partly a study about the usefulness of the GR, we will proceed with the study using the GR results and compare the m to the Ordinary Kriging and the OLS models.

Interpolated Map Results

Three surfaces for lnSOM were created with Ordinary Kriging (OK), OLS and GR models. The colors in all the following figures represent the same intervals for lnSOM where darker colors are higher SOM values. The OK model (see Bailey and Gatrell, pp. 191-192) shows lnSOM variation in the watershed at larger scales (Figure 13). We see that SOM is generally higher in the middle and northeast portions of the watershed and roughly reflects the distribution of the ridge/tabonuco areas (Figure 3). While this information would be useful for larger scale studies (e.g. the entire tabonuco forest in the Caribbean National Forest), it is less helpful for the purposes of managing SOM at the landform scale in Bisley. For example, the details of the map does not show small scale patterns of SOM for making decisions about how much, or what portion, of the watershed to protect for SOM storage.

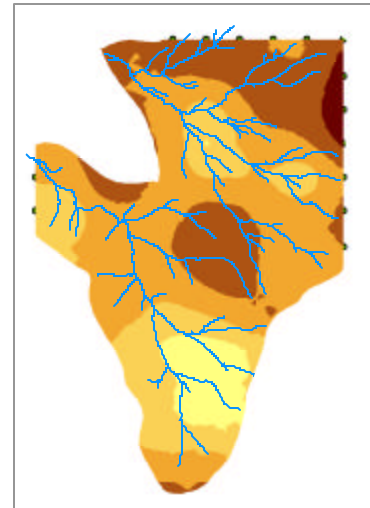


Figure 13. The Ordinary Kriging Model. Darker colors indicate higher lnSOM.

For the OLS and GR models, prediction estimates were calculated for points spaced every 2 meters (in MATLAB). Then these points were interpolated with a spline (in ArcMap). In contrast to OK, the OLS surface has much improved resolution because in this model we took advantage of the continuous surfaces of our DEM derived variables (Figure 14). Recall that it was our objective to model the effect of landscape stability as represented by the ridge/tabonuco areas. The figure below clearly highlights these areas as higher in SOM and the stream areas as lower in SOM. Thus, accounting for topography in this model accomplished a better sense of the distribution of SOM than OK. The fit of the Observed lnSOM on the Predicted lnSOM had an adjusted R^2 of 0.13 and is a measure of performance for the OLS model. This low correlation suggests that

this model still needs more work if it is to be more reliably used for quantitative SOM estimates.

The GR model basically shows the same surface for SOM distribution in this landscape (Figure 15). As already noted, information from neighboring points was not very helpful in the prediction of SOM at Bisley at this scale.

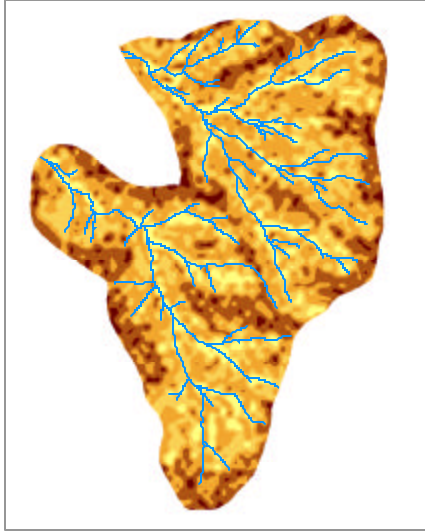


Figure 14. OLS Model. Darker colors indicate higher lnSOM

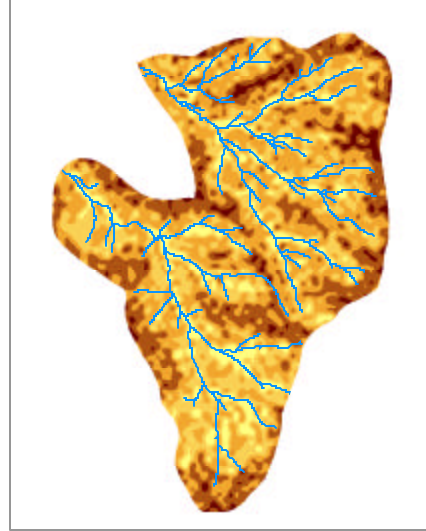


Figure 15. GR Model. Darker colors indicate higher lnSOM

Uncertainty

A map of the standard errors shows that the highest errors occur in the stream channels while the ridges have the lowest errors (Figure 16) (see Equation 4 in Appendix). This may be due to the flow accumulation and distance to stream variables. For example, flow accumulation is probably overcompensating for low SOM values in these areas because there is likely a threshold in this landscape at which high flow accumulation does not influence SOM content as much as would be expected in the model. Notice also that the highest errors are concentrated near the mouths of the streams, where flow accumulation is highest.

Table 4. The observed SOM for each landform represented in the watershed and prediction intervals for selected points. Units are in Megagrams SOM per hectare area.

Landform and Observations	Observed Mean SOM	Prediction Interval
Valley n=9	Max: 199 Min: 52 Mean: 111	<i>At Point "V"</i> Upper: 470 Lower: 8 Predicted: 63
Slope n=62	Max: 407 Min: 36 Mean: 133	<i>At Point "S"</i> Upper: 188 Lower: 29 Predicted: 73
Ridge n=14	Max: 377 Min: 96 Mean: 215	<i>At Point "R"</i> Upper: 490 Lower: 75 Predicted: 192

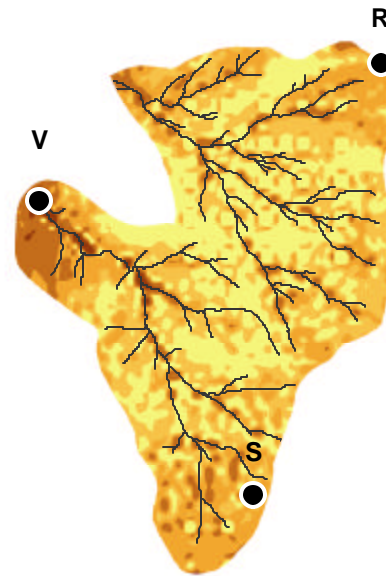


Figure 16. Standard Error map of the GR model. Darker colors indicate higher standard errors. Points labeled V (valley), R (ridge), and S (slope) were chosen to calculate prediction intervals.

To investigate how well the model captures the range of values possible for the three landform types, SOM values were chosen from ridge, slope and valley locations. Prediction intervals were calculated and back-transformed from lnSOM (Table 4). The values of the maximum, minimum and mean observed SOM values also appear for comparison. Note that the predicted values fall in between the observed values for each landform type. In all cases, the lower prediction bound was lower than the lowest observed SOM value. Also, in all cases except slope, the upper prediction bound was higher than the highest observed SOM value.

To compare how well each model performed, the maximum, minimum and mean values of the observed data for the watershed were compared to the range of values for each model (after back-transformation of lnSOM; Table 5). The OK model was able to capture the mean of the watershed. This is not surprising since it does not use information besides observed values and distances in its prediction. The other two models were similar in their ranges and means of SOM values and were somewhat

successful in capturing the mean. All the models fared less well in capturing the extreme values of SOM over the watershed.

Table 5. The range of predicted SOM (Megagrams SOM per hectare) for the Ordinary Kriging (OK), Ordinary Least Squares (OLS) and Geostatistical Regression (GR) models and the observed SOM of the watershed.

	Observed	OK	OLS	GR
Max	407	229	198	256
Min	36	86	39	38
Mean	144	143	120	120

Discussion and Future Work

The lack of spatial structure at this sampling scale confirms that SOM is highly variable, especially in this landscape which dissected by many small streams and ridges.

Unfortunately, any kriging model is contingent upon the observation that things closer together are more alike than things farther apart. In order to successfully use the GR model in this landscape the spatial dependency at smaller scales must be estimated.

There is no getting around using the need to gather more data at smaller scales.

However, this need not be as overwhelming a task as it seems. The whole watershed does not need to be sampled. One option is to choose 6 to 9 transects which representatively sample the three landform types (ridge, slope and valley). The samples should be spaced about 2 meters apart and sampled by the exact same method as the rest of the grid points. Similarly, a nested sampling pattern could be used. In this case, samples are taken at the cardinal points of an existing gridpoint at 2, 5, 10 and 20 meters from the center. This approach could also be done for each landform although it would be more intensive than the first option.

The GR is best applied in situations where spatial dependency at multiple scales is known and good correlation exists between the explanatory variables and the observed values.

Therefore, another potential problem with our regression models is that our original OLS equation (which is also used in the GLS estimation) still only explains about 25% of the variation in SOM. Two additional steps could help improve the OLS model. First, modeling the data after it has been stratified by soil depth or landform type will result in separate regressions which may explain more of the variation in each case. Secondly,

nearly all the data used (both dependent and independent variables) are highly skewed – even after log transformation. It has been suggested by Hengl et al. (2004) that logit transformations may be a better way to manage this type of dataset.

When compared to the observed data, none of the three models tested did a satisfactory job of predicting SOM distribution in Bisley. This is no doubt a reflection of the observations mentioned above. Nonetheless, there is a better way to evaluate and compare each model's performance that can be used in future analysis. This entails setting aside a set of validation points that are not used in the regression analysis. From the predicted and validation values can be calculated the Adjusted-R² and mean prediction error (Hengl et al. 2004).

In sum, the application of the GR model was limited by the lack of spatial dependency in the data and therefore could not produce a reliable map for the purpose of SOM management. Although further analysis is needed, this study represents experimentation with the GR model which may have good potential as a tool for future SOM management as well as for other soil management situations. It should be noted that SOM is probably one of the most (if not *the* most) variable soil property in any given landscape. Therefore, practice with this variable should lend experience to future studies on the spatial patterns and analysis of other soil nutrients.

References

- Bailey, T. C., and A. C. Gatrell, 1995, *Interactive Spatial Data Analysis*: Harlow, England, Prentice Hall, 413 p.
- Brown, S., and A. E. Lugo, 1990, Effects of forest clearing and succession on the carbon and nitrogen content of soils in Puerto Rico and US Virgin Islands: *Plant and Soil*, v. 124, p. 53-64.
- Class Notes. Smith, Tony E., ESE 502 Spatial Data Analysis Spring 2006.
- Falkowski, P., R. J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hogberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffen, 2000, The global carbon cycle: a test of our knowledge of earth as a system: *Science*, v. 290, p. 291-296.
- IPCC, 2001, *Climate change: the scientific basis*: Cambridge, UK, Cambridge University Press.
- Janzen, H. H., 2004, Carbon cycling in earth systems - a soil science perspective: *Agriculture, Ecosystems and Environment*, v. 104, p. 399-417.
- Jobbagy, E. G., and R. B. Jackson, 2000, The vertical distribution of soil organic carbon and its relation to climate and vegetation: *Ecological Applications*, v. 10, p. 423-436.
- Lal, R., 2004, Soil carbon sequestration to mitigate climate change: *Geoderma*, v. 123, p. 1-22.
- N. R. C., 2000, *Ecological Indicators for the Nation*, National Academy Press, p. 64-115.
- Ostertag, R., F. N. Scatena, and W. L. Silver, 2003, Forest floor decomposition following hurricane litter inputs in several Puerto Rican forests: *Ecosystems*, v. 6, p. 261-273.
- Post, W. M., and K. C. Kwan, 2000, Soil carbon sequestration and land-use change: processes and potential: *Global Change Biology*, v. 6, p. 317-327.
- Pregitzer, K. S., 2003, Chapter 6: Carbon Cycling in Forest Ecosystems with an Emphasis on Belowground Processes, in J. M. Kimble, L. S. Heath, R. A. Birdsey, and R. Lal, eds., *The Potential of U.S. Forest Soils to Sequester Carbon and Mitigate the Greenhouse Effect*: Boca Raton, CRC Press, p. 93-108.
- Prentice, I. C., 2001, *The Carbon Cycle and Atmospheric Carbon Dioxide*, *Climate Change 2001: The Scientific Basis IPCC*: Cambridge, UK, Cambridge University Press, p. 183-237.

Richards, K. R., 2004, A brief overview of carbon sequestration economics and policy: *Environmental Management*, v. 33, p. 545-558.

Scatena, F.N., W. Silver, T. Siccama, M.J. Sanchez, 1993, Biomass and nutrient content of the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico, before and after Hurricane Hugo, 1989: *Biotropica*, v.25(1), p. 5-27.

Scatena, F.N. and A. Lugo, 1995, Geomorphology, disturbance, and the soil and vegetation of two subtropical wet stepland watersheds of Puerto Rico: *Geomorphology*, v. 13, p. 199-213.

Scatena, F.N., S. Moya, C. Estrada, J.D. Chinea, 1996, The first five years in the reorganization of aboveground biomass and nutrient use following Hurricane Hugo in the Bisley Experimental Watersheds, Luquillo Experimental Forest, Puerto Rico: *Biotropica*, v. 28, p. 424-440.

Schlesinger, W. H., 1997, *Biogeochemistry: An Analysis of Global Change*: San Diego, Academic Press, 588 p.

Silver, W. L., R. Ostertag, and A. E. Lugo, 2000, The potential for carbon sequestration through reforestation of abandoned tropical agricultural and pasture lands: *Restoration Ecology*, v. 8, p. 394-407.

Smith, P., 2004, Soils as carbon sinks: the global context: *Soil Use and Management*, v. 20, p. 212-218.

Smith, P., D. S. Powlson, J. U. Smith, P. Falloon, and K. Coleman, 2000, Meeting Europe's climate change commitments: quantitative estimates of the potential for carbon mitigation by agriculture: *Global Change Biology*, v. 6, p. 525-539.

Sparks, D. L., 2003, *Environmental Soil Chemistry*: San Diego, Academic Press.

Appendix 1 – Geostatistical Regression

The Geostatistical Regression model is mathematically expressed (in matrix notation) as:

$$(2) \quad \hat{z}_{(s_0)} = \mathbf{x}_{(s_0)}' \hat{\boldsymbol{\beta}}_{\text{gls}} + \hat{\mathbf{c}}_{(s_0)}' \hat{\mathbf{C}}_s^{-1} (\mathbf{y} - \mathbf{x} \hat{\boldsymbol{\beta}}_{\text{gls}})$$

where $\hat{z}_{(s_0)}$ is the predicted value (e.g. predicted SOM) at one location, s_0 , in the watershed (Figure 17). The matrix of all observed explanatory variable values (e.g. *flow*, *curv*, and *dstrm*) is denoted \mathbf{x} and the vectors of explanatory variables at location s_0 are denoted $\mathbf{x}_{(s_0)}$. The vector \mathbf{y} represents all the sampled observations (e.g. the measured SOM values for the watershed). The estimated matrices $\hat{\mathbf{c}}_{(s_0)}$ and $\hat{\mathbf{C}}_s$ refer to the covariances obtained from the spherical variogram parameters of the regression residuals after calculating $\hat{\boldsymbol{\beta}}_{\text{gls}}$ (see below). Specifically, $\hat{\mathbf{c}}_{(s_0)}$, is the estimated covariance between all the predicted point and each of the sampled locations and $\hat{\mathbf{C}}_s$ is the estimated covariance of all pairs of sampled locations.

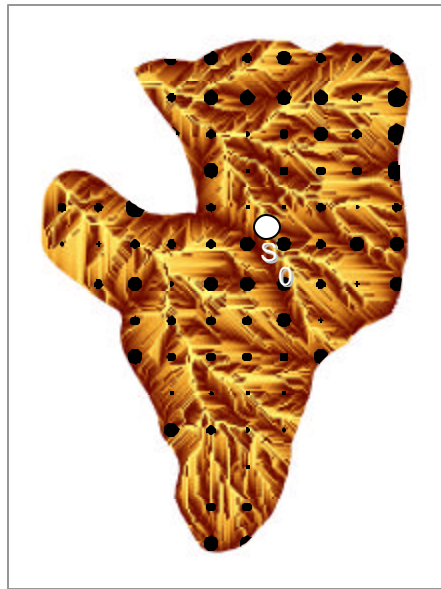


Figure 17. The prediction of SOM at point s_0 (white dot) is estimated by the values of neighboring observed points (black dots; larger dots are higher in SOM). Additionally s_0 is estimated from secondary information extracted from a continuous surface. To illustrate secondary information, the variable *flow* is shown at right and is represented by the background colors.

The trend coefficient, $\hat{\boldsymbol{\beta}}_{\text{gls}}$, is estimated by Generalized Least Squares (GLS) regression which accounts for spatial autocorrelation:

$$(3) \quad \hat{\boldsymbol{\beta}}_{\text{gls}} = (\mathbf{x}' \hat{\mathbf{C}}_s^{-1} \mathbf{x})^{-1} \mathbf{x}' \hat{\mathbf{C}}_s^{-1} \mathbf{y}$$

It is possible to calculate the $\hat{\mathbf{C}}_s$ component of this equation, and therefore the solution for $\hat{\mathbf{B}}_{\text{gls}}$, from the parameters of the spherical variogram of the OLS residuals. However, it is common to first use the OLS residuals to obtain GLS coefficients. Then the GLS coefficients are used to calculate a new set of residuals and variogram parameters. This process can be iterated until the GLS and OLS trend estimates converge. After convergence, the final covariance parameters ($\hat{\mathbf{c}}_{(s_0)}$ and $\hat{\mathbf{C}}_s$) and the final $\hat{\mathbf{B}}_{\text{gls}}$ derived from (3) are used in the GR model (2) (Bailey and Gatrell, pp 189-190; Hengl et al. 2004). The above steps were operationalized in MATLAB (for further explanation see “Class Notes; Continuous Data Analysis, pg. 72-76”).

Standard Error Estimate

The standard error matrix for each of the predicted points of the GR model includes the kriging variance of the residuals and the error variance related to $\hat{\mathbf{B}}_{\text{gls}}$:

$$(4) \quad \hat{\mathbf{s}}_\varepsilon = \left[\hat{\mathbf{s}}^2 - \hat{\mathbf{c}}_{(s_0)}' \hat{\mathbf{C}}_s^{-1} \hat{\mathbf{c}}_{(s_0)} + (\mathbf{x}_{(s_0)} - \mathbf{x}' \hat{\mathbf{C}}_s^{-1} \hat{\mathbf{c}}_{(s_0)})' (\mathbf{x}' \hat{\mathbf{C}}_s^{-1} \mathbf{x})^{-1} (\mathbf{x}_{(s_0)} - \mathbf{x}' \hat{\mathbf{C}}_s^{-1} \hat{\mathbf{c}}_{(s_0)}) \right]^{\frac{1}{2}}$$

where $\hat{\mathbf{s}}^2$ is the variance matrix of all predicted points (Hengl et al 2004) (for further explanation see “Class Notes; Continuous Data Analysis, pg. 72-76”).