

Functional Description of Carbon Concentration in a Field

Yue Zhang*, Lucian Wielopolski

Environmental Sciences Department, Brookhaven National Laboratory, Upton, NY 11973

*Department of Applied Math and Statistics, Stony Brook University, Stony Brook, NY 11794

Abstract: The conventional way of mapping carbon distribution in a field is by analyzing discrete soil cores, collected by covering the field either in a random or grid pattern. To optimally deploy the new instrumentation for *in situ* soil carbon analysis based on Inelastic Neutron Scattering (INS), a non-destructive system able to contiguously scan large areas, a better functional description of the carbon concentration in a field is necessary. In particular, its three-dimensional continuous distribution of carbon in a field is required for: a) complete system simulation, and b) assessing the deployment protocol of the new system that would minimize the error propagation. Here, we compare our results from spatial regression with those from ordinary kriging, a widely used method for spatial interpolation of the field data, to illustrate the possibility of fitting three-dimensional carbon functions based on using experimental data from conventional instrument. Such a functional description of carbon concentration should optimize the efficiency of further analyses of soil carbon by the INS system.

1. Introduction

The conventional way of mapping carbon distribution in a field is by analyzing discrete soil cores, collected by covering a field in a grid or random pattern. The cores subsequently are sectioned into pieces of 5 or 10cm in length, homogenized and analyzed for carbon content, thus obtaining a three dimensional (3D) data set of carbon content. More often than not the carbon content in depths (z direction) is projected onto the field's surface at each sample spot (x,y) and the projected carbon content is treated as the response variable associated with coordinates (x,y) in the field, producing a modified 2D data set, notice that bias might be introduced when changes in soil bulk density are ignored. Then, applying ordinary kriging to the 2D data predicts the value of response variable at any un-sampled spot. A new instrument being developed for *in situ* soil carbon analysis based on Inelastic Neutron Scattering (INS) (Wielopolski and Orion, 2000) can

detect carbon concentration in a large field by multiple stationary measurements of large volumes of about 0.3 m³ or by continuously scanning the entire field. The INS system requires a functional description of the carbon concentration in the field for the purpose of: a) simulation of the entire system and b) assessing properly the error propagation. Instead of discrete description of the sampled points we propose to fit a regression model to the original 3D data. This model should optimize the efficiency of further analyses of soil carbon by the INS instrument, and thereby, may afford additional insight into the structure of carbon distribution. We compare our results from spatial regression with those from ordinary kriging (Issaks and Srivastava, 1989, Nielsen and Wendroth, 2003), a widely used method for spatial interpolation, to illustrate the possibility of fitting three-dimensional carbon functions based on using experimental data from conventional instrument.

2. Regression Model

In order to obtain a 3D regression model of carbon concentration and coordinates (x, y), depth (z), we proposed to fit models in two steps.

2.1 Regression Model Step 1: Regression models of carbon concentration and depth (z) at each sample spot (x, y)

Suppose we have n sample spots and the corresponding coordinates are (x_i, y_i) , $i=1, \dots, n$. At each sample spot (x_i, y_i) , we have the carbon concentration (C_c) data at various depth (z). The regression model of carbon and depth is

$$C_c = f(z; \alpha(i), \beta(i), \gamma(i)) + \text{error}, \quad i=1, \dots, n \quad (1)$$

Function f could be exponential function based on the physical characteristics of carbon. Alphas, betas and gammas are the parameters obtained from the regression model.

The errors at sample depth z in regression model (1) are assumed to be normally distributed and independent from each other; ordinary least squares (OLS) methodology is applied to estimate the parameters. Then Moran's I test is performed on the residuals to determine whether autocorrelation exists. If it does, then we use the maximum likelihood method to include terms in the regression model that account for spatial autocorrelation.

2.2 Regression Model Step 2: Regression model of parameters and field coordinates (x, y)

After obtaining the values of alphas, betas and gammas at all the sample spots, we applied a regression technique that incorporates correlated errors to get the model of the above parameters and field coordinates.

$$\alpha(x,y) = g1(x,y) + \text{error} \quad (2)$$

$$\beta(x,y) = g2(x,y) + \text{error} \quad (3)$$

$$\gamma(x,y) = g3(x,y) + \text{error} \quad (4)$$

The errors at all sample spots (x, y) are assumed to be normally distributed and correlated with each other, producing a spatial covariance structure. In this case, OLS is no longer a valid method to estimate the model's parameters. Instead, ML (maximum likelihood) (Upton and Fingleton, 1985) or REML (restricted/residual maximum likelihood) is appropriate.

From models (2), (3) and (4), alpha, beta and gamma are seen to be functions of coordinates (x, y).

Combining the previous function f in regression model step 1,

$$C_c = f(z; g1(x,y), g2(x,y), g3(x,y)) + \text{error} \quad (5)$$

Thus, carbon is a function of field coordinates x, y and depth z.

3. Data and Results

Due to the lack of 3D carbon concentration data, we illustrate the above two regression model steps with two different data sets.

3.1 Regression model step 1 on Ohio carbon data

We took 54 sample spots of Ohio carbon data (data was provided by Prof. Lal, Ohio State University) and we fitted an exponential function of carbon concentration versus depth thus

$$C_c = \alpha(i) * \exp(-\beta(i) * z) + \gamma(i) + \text{error} \quad (6)$$

$i=1, \dots, 54$

Moran's I test was applied to the residuals of model (6) for every sample spot. There was no autocorrelation significant at level 0.1. Figure 1 shows two extreme examples of fitting. The blue line represents the worst fitting among all 54 sample spots, while the red line is the best. Figure 1

demonstrates that the exponential function in model (6) is appropriate for describing the depth profiles of carbon.

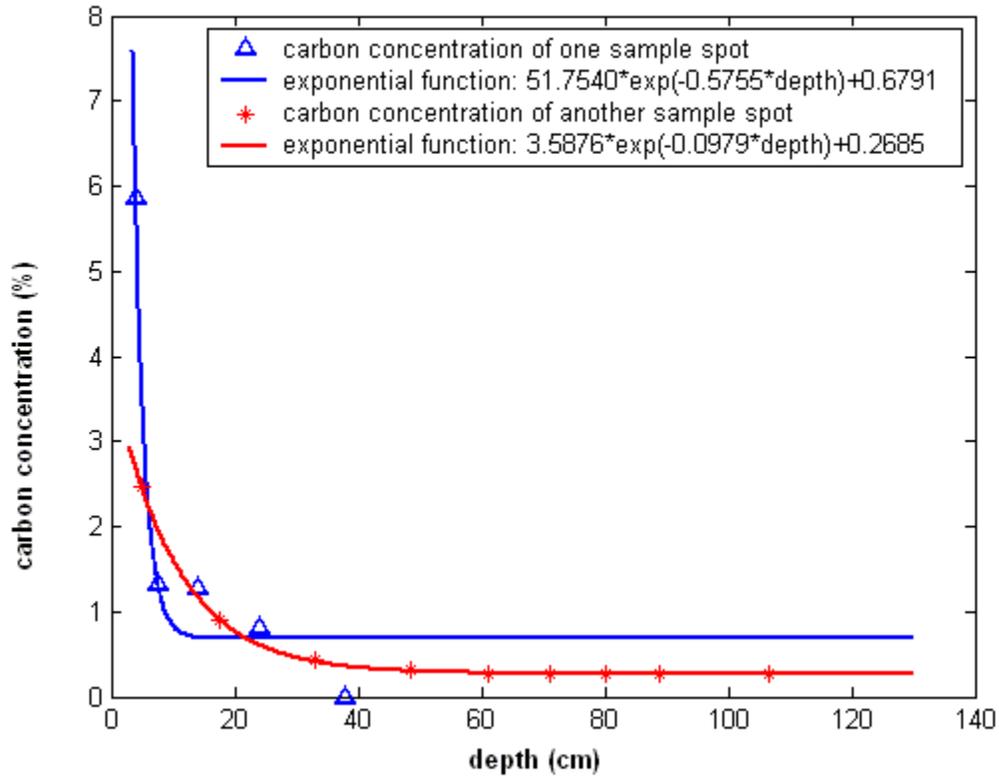


Figure1: Two examples of fitting exponential function of carbon and depth at a given sample spot from Ohio carbon data

Figure 2 shows the correlation between the alphas and betas from these 54 sample spots. Alpha and beta exhibit a strong linear correlation, indicating that the 54 exponential functions of model (6) may have similar properties over the field.

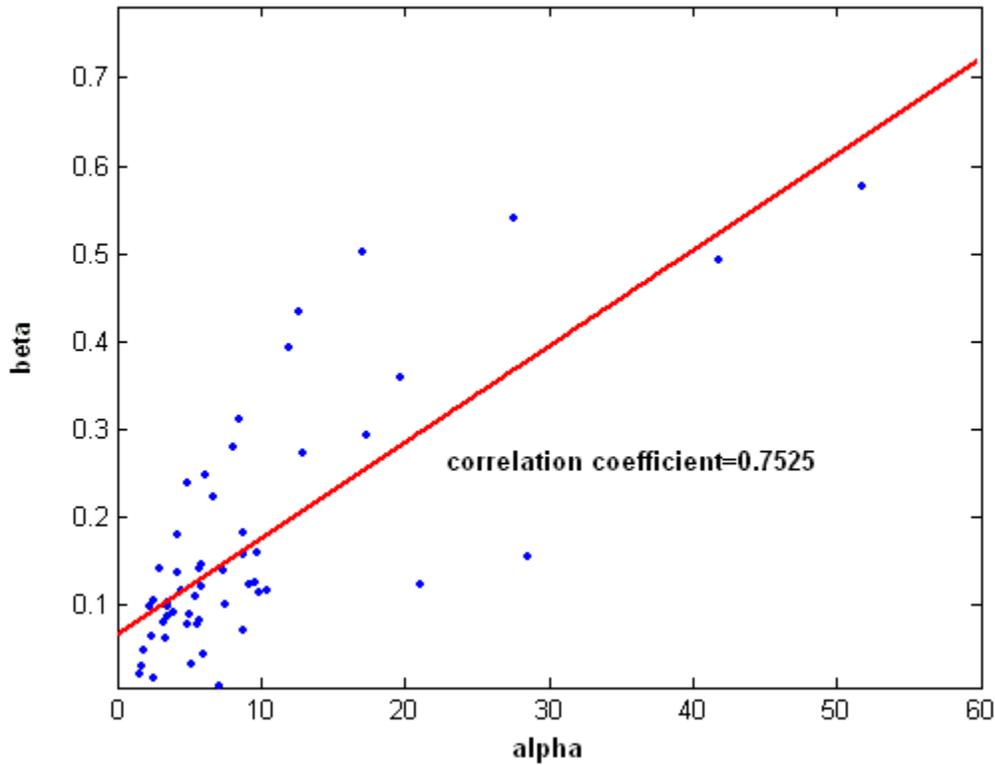


Figure 2: Correlation of alphas and betas from 54 sample spots of Ohio carbon data

3.2 Regression model step 2 on weather station temperature data

To examine the feasibility of regression model step 2, we tested it on 2D weather station temperature data (data is from Upton & Fingleton, 1985). In this data set, the response variable is the temperature recorded at various locations, together with the recorded associated (x, y) coordinates. We fitted a linear regression model of the response variable (temperature) to (x, y) coordinates. The parameters were estimated by REML and the structure of the spatial covariance is assumed to be exponential.

The temperature data shares the following properties with the carbon data:

- (i) The response variable is continuous over the field;
- (ii) The value of response variable is related to spatial locations; and,
- (iii) The response variable may be autocorrelated.

If the linear regression model works well on weather station temperature data, we believe it could be applied to carbon data, thus yielding a three-dimensional functional description of carbon concentration.

The linear regression model of 2D weather station temperature (T) data is

$$T = 18.147 + 6.2248 * x - 4.8853 * y + \text{error} \quad (7)$$

3.3 Ordinary Kriging

Ordinary kriging is a widely used method in spatial interpolation. We briefly introduce the principles of ordinary kriging here. (Isaaks and Srivastava, 1989)

Suppose carbon concentration has been sampled at n locations.

Label the locations s_1, \dots, s_n . The coordinates of s_i are (x_i, y_i) .

The response variable, $V(s_1), \dots, V(s_n)$, has n outcomes v_1, \dots, v_n .

The predicted value at an un-sampled spot s_0 is $\hat{V}(s_0) = \sum_{i=1}^n w_i V(s_i)$, where w_i are weights and $\sum_{i=1}^n w_i = 1$. The prediction error at site s_0 is $R(s_0) = \hat{V}(s_0) - V(s_0)$.

In ordinary kriging, the weights are based on the covariances among points in the sample, and the covariances between the sample points and the point to be predicted. The kriging estimator minimizes the prediction variance $\hat{\sigma}_R^2 = \text{Var}(\hat{V}(s_0) - V(s_0))$. Therefore kriging gives the best linear unbiased estimator.

Comparison between our regression model and ordinary kriging on the 2D weather station temperature data

We compared the estimates from ordinary kriging and our regression model of the 2D weather station temperature data.

Figure 3 shows leave-one-out cross validation of ordinary kriging and regression model for these data.

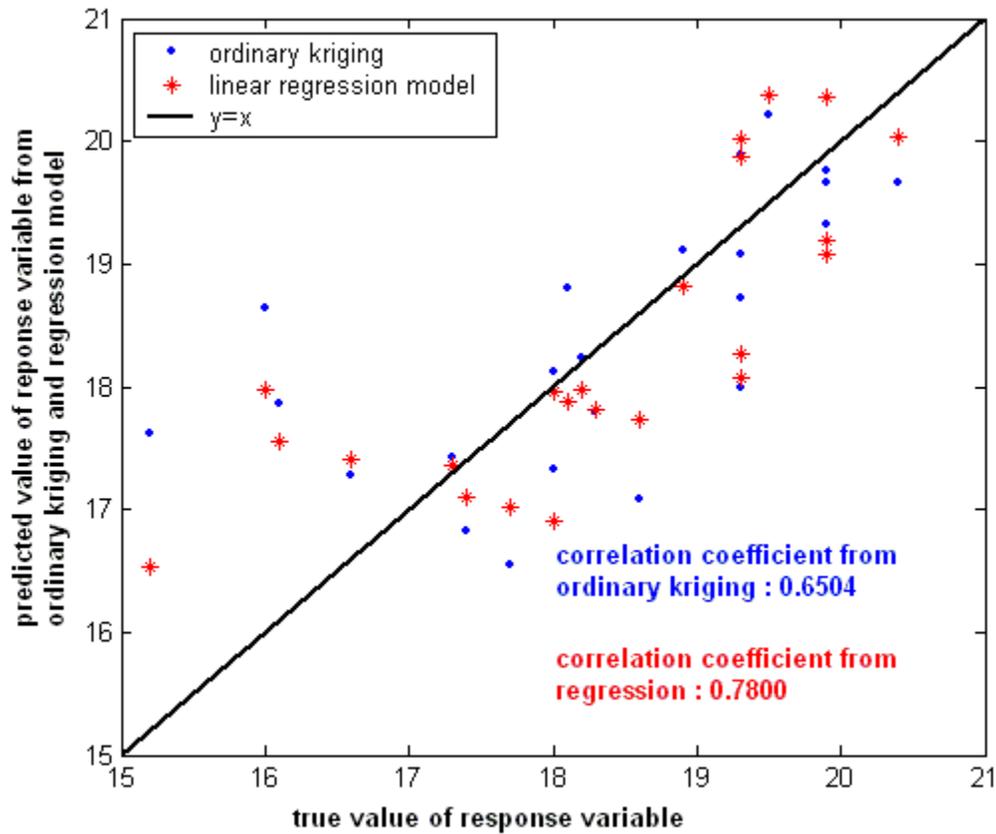


Figure 3: Leave-one-out cross validation of ordinary kriging and regression model for the 2D weather station temperature data.

Allen's predicted residual sum of squares (PRESS) of ordinary kriging and the regression model for the weather station data are

$$\frac{1}{n} \text{PRESS} = \frac{1}{n} \sum_{i=1}^n (\hat{V}(x_i, y_i) - V(x_i, y_i))^2$$

$$\frac{1}{n} \text{PRESS of ordinary kriging} = 1.1094$$

$$\frac{1}{n} \text{PRESS of regression} = 0.7383$$

From the PRESS values of ordinary kriging and the regression model, we see that the performance of the latter on the temperature data is better than that of ordinary kriging. Furthermore, we randomly selected 50 un-sampled spots within the sampling region of weather station temperature data, and compared the predicted value of the response variable (temperature) from ordinary kriging and the regression model. Figure 4 shows the close correlation between the predicted values from the two methods.

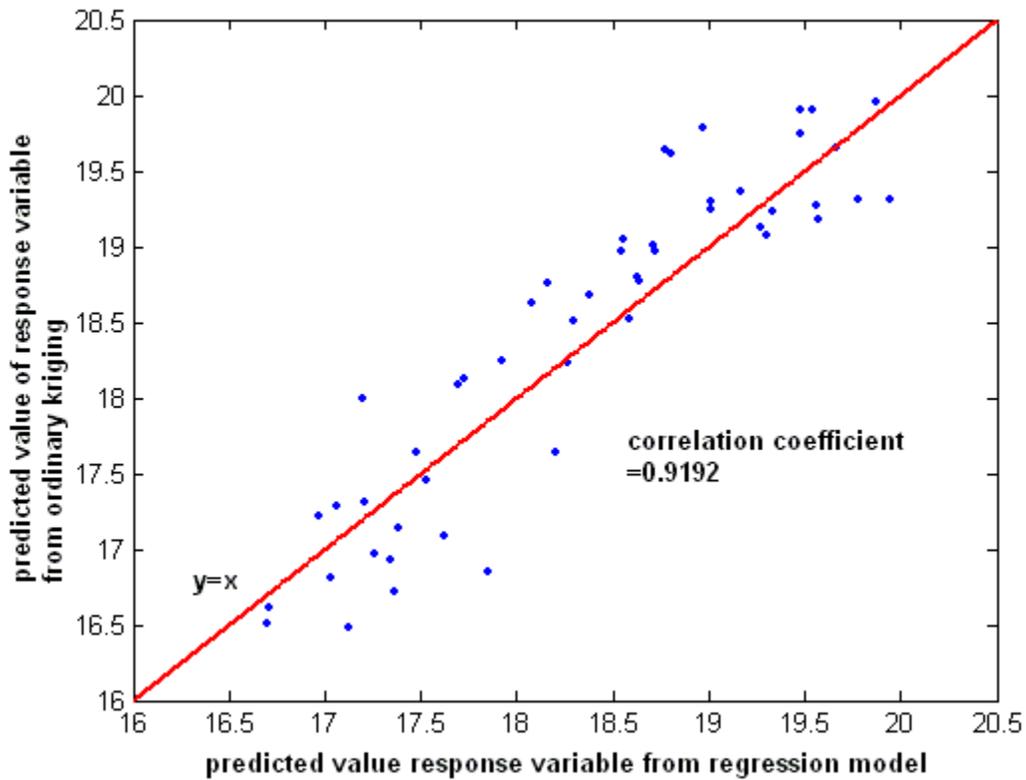


Figure 4: Correlation between predicted value of temperature from ordinary kriging and linear regression model from 50 randomly selected un-sampled spots.

Regression model: $\text{Temperature} = 18.147 + 6.2248 * x - 4.8853 * y + \text{error}$

The predicted values from ordinary kriging and the regression model show strong linear correlation, indicating they give similar results for the weather station temperature data.

4. Conclusions

4.1 Comparisons between the regression method and the conventional method of describing carbon distribution

- (i) The regression model describes carbon profiles at continuous depth, while, in contrast the conventional method uses only projections of carbon concentration on to the field's surface.
- (ii) The regression model provides a three-dimensional structure of carbon concentration while conventional method offers only a two-dimensional contour map.
- (iii) The regression model is simple provided that there is a systematic structure of the response variable over the field, which is true for carbon concentration.
- (iv) The three-dimensional function obtained by the regression model could be used in conjunction with a new instrument for *in situ* soil carbon analysis based on Inelastic Neutron Scattering (INS). The new system is non-destructive and can scan large areas contiguously, thus necessitating a functional description of carbon concentration in the field.
- (v) Ordinary kriging, which is used to derive a contour map of carbon in the conventional method, is more sensitive to local variation than is the regression model.

4.2 Regression model of Ohio carbon data and weather station temperature data

Carbon concentration is approximately an exponential function of depth at a given sample spot in Ohio carbon data.

The linear regression model performs better than ordinary kriging when there is a systematic structure of the response variable over the field, as in the weather station temperature data. Since carbon concentration also has a similar systematic structure as temperature, we believe linear regression model could be applied to carbon data, thus offering a three-dimensional functional description of carbon.

5. Future work

We will apply the proposed two-step regression model to real three-dimensional (x,y,z) carbon concentration data, derive the error estimate of the proposed regression model and then compare this estimate to that from ordinary kriging. We will develop the application of the proposed

regression model to improve variogram calculations. Furthermore, we will incorporate the proposed regression model into Monte Carlo simulation for the INS system.

Acknowledgements

We thank Dr. Avril Woodhead for reviewing this paper and Prof. Lal from Ohio State University for providing Ohio carbon data. This manuscript has been authored by Brookhaven Science Associates, LLC under Contract No. DE-AC02-98CH10886 with the US Department of Energy. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges, a world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes.

Reference:

- Isaaks, E.H., and Srivastava, R.M., 1989. An introduction to applied geostatistics, Oxford University Press.
- Nielsen D.R., and Wendroth O., 2003. Spatial and Temporal Statistics Sampling Field Soils and their Vegetation, CATENA VERLAG GMBH.
- Upton, G., and Fingleton, B., 1985. Spatial Data Analysis by Example, Wiley.
- Wielopolski, L., Orion, I., Hendrey, G., and Rogers, H., 2000. Soil carbon measurements using inelastic neutron scattering, IEEE Trans Nucl.Sciences, 47, 914-917.