Can Topographical and Yield Data Substantially Improve Total Soil Carbon Mapping by Regression Kriging?

A. N. Kravchenko* and G. P. Robertson

ABSTRACT

Accurate mapping of total soil C on the field scale is essential for evaluating efforts to sequester soil C and for providing individual producers with information on C sequestration potentials of their fields. Data on easily measured secondary variables that are strongly related to soil C are believed to be helpful in improving mapping accuracy. The objective of this study was to assess improvement in mapping accuracy due to dense topographical and long-term yield monitoring information. Approximately 1200 total C measurements at the 0- to 5-cm depth along with topographical and 7-yr crop grain yield data were collected at twelve 60- by 60-m plots at the Long-Term Ecological Research Site in Michigan. Total C was found to be significantly related to topography and 7-yr average standardized yield in all studied plots, with regression R^2 >0.5 in approximately half of the plots. Accounting for either topographical or yield information in regression kriging, however, produced only modest (<10%) improvement in mapping accuracy compared with ordinary kriging. Plots with promisingly strong relationships of total C with topography or yield were also found to be the ones where spatial distributions of total C were highly continuous, thus leading to no advantages in using regression kriging. The results indicated that under soil and topographical conditions similar to those of this study, dense topographical data or dense long-term yield data might not lead to substantial improvement in C mapping accuracy.

Obtaining efficient and accurate assessments of soil C has became increasingly important in recent years due to the need for better estimations of existing soil C pools, their vulnerability to change, and the impact of change on atmospheric CO_2. Soil C mapping at the field scale is essential for accurate evaluation of efforts to sequester soil C and for providing individual producers with accurate information on C sequestration potentials of their fields; however, intense and detailed soil sampling and soil C measurements are required to obtain accurate field maps of soil C. Even though such sampling efforts might be feasible on experimental research sites, sufficiently dense sampling at a typical agricultural field, farm, or county level is prohibitively expensive. Thus, easily and relatively inexpensively measured variables related to soil C are becoming more important as a means of mapping soil C with less intensive sampling at reduced expense.

If a densely sampled secondary variable(s) is correlated with a sparsely sampled primary variable, it can be used to improve the mapping accuracy of the primary variable via various geostatistical procedures, including cokring, external drift kriging, or regression kriging (Isaaks and Srivastava, 1989; Goovaerts, 1997; Webster and Oliver, 2001). Recent examples of geostatistical applications for predicting soil C and other soil properties using various sources of dense secondary information are given by Gessler et al. (2000), Mueller and Pierce (2003), Terra et al. (2004), and Simbahan et al. (2006); many of the earlier examples are referenced in the review by McBratney et al. (2003). Walter et al. (2003) used a combination of deterministic modeling with stochastic methodology in predicting soil C distribution across a terrain with diverse land uses. Topographical characteristics are among commonly used sources of additional information for improving the mapping accuracy of soil C in soils of the U.S. Midwest (Gessler et al., 2000; Mueller and Pierce, 2003; Terra et al., 2004).

Another source of dense spatial field information is crop yield data collected via combine monitors. The use of combine yield monitors that obtain georeferenced measurements of crop yield has grown tremendously in the last several years. Almost 20% of all corn (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.] yields in the USA currently are being collected via yield monitors (Economic Research Service, 2006). Yield maps obtained from such monitors for a number of years reflect areas within the field having different yield potentials due to variability of either topographical or soil characteristics. As such, they have been used extensively in precision agriculture research to characterize within-field soil variability and to derive management zones (Lark and Stafford, 1997; Lark and Stafford, 1998; Blackmore, 2000; Flowers et al., 2005). Webster and Oliver (2001) illustrated the application of yield-monitor data in improving soil P mapping. Bishop and McBratney (2001) have used yield-monitor data for predicting soil cation exchange capacity.

We hypothesized that both topographical information and the long-term yield monitor data could be of value in improving accuracy in soil C mapping. Given the large number of producers that collect yield monitor data, the long-term yield monitor data could be particularly valuable when other sources of soil and topographical information are unavailable. There is no information, however, on how useful the long-term yields could be in practice and what is the level of improvement in accuracy that could be reached by using yield as well as to-

A.N. Kravchenko, Dep. of Crop and Soil Sciences, Michigan State Univ., East Lansing, MI 48824-1325; and G.P. Robertson, W.K. Kellogg Biological Station and Dep. of Crop and Soil Sciences, Michigan State Univ., Hickory Corners, MI 49060-9516. Received 1 Sept. 2005. *Corresponding author (kravche1@msu.edu).

Published in Agron. J. 99:12–17 (2007).
Spatial Variability
doi:10.2134/agronj2005.0251

Abbreviations: CT, chisel-plowed management with conventional chemical inputs; CTcover, chisel-plowed management with a winter leguminous cover crop and no chemical inputs; NT, no-till management with conventional chemical inputs; N/S, nugget to sill ratio.

pographical information. The objective of this study was to assess the potential for dense topographical data and crop yield monitor data to improve mapping of total soil C. To achieve this objective, we first evaluated the strength of the relationships of soil C with auxiliary data, that is, topographical information, such as relative elevation, terrain slope, terrain curvature, and aspect, and crop grain yield information collected at multiple sites with different management practices. Second, we quantified the improvement achieved in soil C mapping when the auxiliary information was used, and third, examined factors that potentially affect the usefulness of auxiliary information in C mapping. The studied factors included the strength of the relationships between the primary and secondary variables and the strength of the spatial correlation of the primary variable. We used data from 12 60- by 60-m field sites, which allowed assessment of the variability in strength of the relationship between C and yield across a diverse landscape. Because of time- and labor-intensive efforts needed for intensive soil sampling and C measurements, in this preliminary study we only obtained total soil C data at the 0- to 5-cm depth.

MATERIALS AND METHODS

Data Collection

Data for this study were collected at the Long-Term Ecological Research (LTER) site, Kellogg Biological Station, in southwest Michigan (85°24′W, 42 °24′N). Soils are well-drained Typic Hapludalfs of the Kalamazoo (fine-loamy, mixed, mesic) and Oshtemo (coarse-loamy, mixed, mesic) series, developed on glacial outwash (Mokma and Doolittle, 1993). The complete site description, experimental design, and management protocols are available at http://lter.kbs.msu.edu (verified 21 Sept. 2006).

The 12 plots selected for this study were four replications of three LTER treatments: chisel-plowed management with conventional chemical inputs (CT), no-till with conventional chemical inputs (NT), and chisel-plowed management with a winter leguminous cover crop and no chemical inputs (CTcover). Soil sampling was conducted in June of 2003. At each plot, approximately 100 georeferenced soil samples were collected from the 0- to 5-cm depth. The sampled areas of each plot were 60 by 60 m. Twenty of the samples were collected on a regular triangular grid with distance between grid points in the east–west direction of 13 m and in the north–south direction of 15 m. The remaining samples were taken at varying distances from the regular grid sampling points. At each sampling location, the sample was taken from between the plant rows and was composited from five 2.5-cm-diameter cores collected within a 0.2-m radius. Total C was measured using a Carlo-Erba (Milan, Italy) CN analyzer. As determined by triplicate analysis runs on approximately one-fifth of the samples, the measurement error varied around 3 to 5%. A detailed description of the sampling scheme, soil analyses, and elevation data and terrain attribute processing was presented by Kravchenko et al. (2006).

The elevation measurements were recorded every 2 m using a 12-channel Leica SR530 real-time kinematic DGPS receiver (Leica Geosystems, St. Gallen, Switzerland). The measurements were converted into a cell-based terrain map on a 4- by 4-m grid using ArcGIS 9.0 Spatial Analyst (Environmental Systems Research Institute, 2004). The grid size was selected so that almost all grids included at least one of the elevation measurement points, thus not affecting derivation of other terrain attributes. Then terrain slope, aspect, and curvature were derived from the elevation data using the surface hydrologic analysis functions of ArcGIS 9.0 Spatial Analyst. Aspect was reclassified into two categories: southern orientation and all other orientations.

Crop yield data included corn, soybean, and wheat (*Triticum aestivum* L.) grain yields collected via yield monitors from the studied plots during 1996 through 2002. For a detailed description of the collection, processing, and cleaning of these yield data, see Kravchenko et al. (2005). The number of yield data points that remained within the studied area of each plot after data processing ranged from at least 500 for wheat to as many as 1600 for corn and soybean. Point yield data were converted into cell-based yield maps on a 4- by 4-m grid (ArcGIS 9.0 Spatial Analyst) using inverse distance weighting with power of two and 15 nearest neighbors. Then, yield data from each year were standardized in each plot as $(Z_i - Z_m)/s$, where Z_i is the yield at location i, and Z_m and s are the plot mean yield and the standard deviation, respectively.

Data Analysis

The relationship of total C with 7-yr average standardized yield was studied using simple linear regression. The relationship of total C with topography was studied using multiple regression that included the linear effects of relative elevation, terrain slope, curvature, and reclassified aspect. The need for quadratic terms for elevation, terrain slope, and curvature was investigated separately in each plot and the quadratic terms were added to the regression equation whenever significant at the 0.05 level of significance. To assess the contribution of the 7-yr average crop yield to total C prediction along with topography, the standardized average yield was added to the previously selected topographical regression model of each studied plot. The regression analyses were performed separately in each studied plot as well as in the data set of all 12 plots combined and were conducted using PROC REG (SAS Institute, 2001).

Regression kriging (Odeh et al., 1995; Hengl et al., 2004) was used to evaluate potential improvement in C mapping when using auxiliary information compared with mapping based on available C measurements via only ordinary kriging. For that, in each plot data set we selected 20 C samples located on a regular triangular grid to be a model (or training) data set. The choice of the model data was based on the following considerations: First, sampling on a regular or semiregular grid is among the most common practices in soil sampling and a triangular grid is known to be somewhat superior to a rectangular grid (Webster and Oliver, 2001). Second, the distances between the grid samples within the plots used in this experiment are obviously much smaller than those of sampling on a field or a farm scale, which is often done on the basis of one sample per hectare or one sample per acre; however, the total number of samples (20) is comparable with the number of samples typically obtained per field and used in mapping in field- or farm-scale sampling situations.

The model data set for the combined 12-plot data was created by using two samples from each plot, one from the southwest corner of the plot and one from the northeast corner of the plot. This produced a relatively regularly sampled model data set with 24 points. The sampling density of this model data set was more similar to that of typical field- or farm-scale sampling than the model data sets from the individual plots. In plot data sets and in the combined data set, all remaining samples were used for independent testing. For individual plots, the independent test data sets consisted of approximately 70 to 80 observations. For the combined data, the independent test data set consisted of >1100 observations.

We considered regression kriging with multiple regression equations with topographical variables to assess mapping improvement due to topographical information, and the regression kriging with simple linear regressions with the 7-yr average standardized yield to assess mapping improvement due to crop yield information. Both regression kriging and ordinary kriging were performed using the model data sets and then were used to obtain predictions of the independent test data values.

Sample variograms for the original model data sets for ordinary kriging and for the residuals for regression kriging were fitted with variogram models using the weighted least squares approach, with weights proportional to the number of data pairs in the sample variogram value and inversely proportional to the fitted variogram value (Cressie, 1985). The three most common variogram model types, i.e., spherical, exponential, and Gaussian, were used. A preliminary choice of the model type was based on the visual examination of the sample variogram, then the MSE values of the models that appeared suitable were compared and the one that produced the lowest MSE was selected for kriging. We recognized that the small number of model data points in this study substantially limits the reliability of sample variograms and their fitting.

Sample variogram calculations and fitting were also performed for the complete data sets of both individual plots and the combined data set. A ratio between the variogram model nugget and the sill (sum of nugget and partial sill) was used as a characteristic of the spatial correlation strength in total C distribution across studied landscapes. Expressed as a percentage, the nugget to sill ratio (N/S) reflects the contribution to the overall variability of random variation occurring at distances shorter than those of the minimal sampling distance. Sample variogram calculation, variogram model fitting, and kriging were performed using PROC VARIOGRAM, PROC NLIN, and PROC KRIGE2D, respectively (SAS Institute, 2001).

For each test data set, RMSEs were calculated based on predicted and observed test data values. Relative improvement over ordinary kriging due to using regression kriging was assessed using the respective RMSE:

$$\text{Relative improvement} = \frac{\text{RMSE}_{\text{OK}} - \text{RMSE}_{\text{RegK}}}{\text{RMSE}_{\text{OK}}} 100\%$$

where RMSE_{OK} is the RMSE obtained based on ordinary kriging predictions and $\text{RMSE}_{\text{RegK}}$ is the RMSE obtained based on regression kriging predictions.

A relatively small number of C data sets (only 12) used in this study were found to be insufficient for an in-depth assessment of possible sources of influence on relative improvement due to regression kriging over ordinary kriging. Thus, we used multiple simulated data sets. A data set from one of the experimental plots was used as a conditioning data set. Gaussian simulations were performed to create multiple sets of 2500 simulated observations of primary and secondary variables (GSLIB, Deutsch and Journel, 1998). The simulated data sets varied in the strength of the relationship between the primary and secondary variables, with R^2 values between primary and secondary variables in the simulated data sets ranging from 0.10 to 0.95 in approximately 0.10 increments. The simulated data sets also varied in the strength of spatial correlation in the primary variable with the N/S ratio ranging from 1% (primary variable with strong spatial correlation) to 70% (primary variable with weak spatial correlation) in approximately 10% increments. From each complete set of 2500 simulated observations, we randomly selected 400 data points that were used as a model data set and 100 data points to be used as a test data set. Regression kriging and ordinary kriging predictions for test data were obtained using model data sets and RMSE and relative improvements of regression kriging over ordinary kriging were calculated as described above. Then, based on the results from all simulated data sets, we assessed the effect of the strength of the relationship between the primary and secondary variables, i.e., R^2, and the effect of the strength of the spatial correlation of the primary variable, i.e., N/S ratio, on the relative improvement in mapping accuracy due to regression kriging over ordinary kriging. The relationship between the obtained relative improvement, R^2, and N/S data were fitted with a polynomial regression equation. A hierarchical approach was used in polynomial regression fitting with higher order and interaction terms being added sequentially to the initial first-order equation until further higher order additions were not significant at the 0.05 level.

RESULTS AND DISCUSSION

The studied plots had similar general topography that could be characterized as moderately to very flat. The elevation ranges in the studied plots were <2.0 m and the average terrain slopes were around 1 to 2°.

Total C was significantly ($P < 0.05$) positively correlated with yield in all but one studied plot (CT-r4; Table 1). The percentage of total C variability explained by the 7-yr average yield ranged from 7% (CTcover-r3) to 74% (CT-r3; Table 1). In only two of the studied plots did the yield explain >50% of the total C variability.

The percentage of total C variability explained by topography ranged from 11% (CT-r4) to 81% (CTcover-r1), with only six of the studied plots where topography

Table 1. Results of simple linear regression for total C at the 0- to 5-cm depth and 7-yr average standardized crop grain yield, and results of multiple regressions for total C and topographical variables alone or along with 7-yr average standardized yield. Unless stated otherwise, the regressions were statistically significant at $P < 0.01$.

Plot	Soil samples	R^2 for yield	R^2 for topography†	R^2 for (topography + yield)/slope‡	Nugget/ sill ratio§
	no.				%
CT-r1	99	0.29	0.54	0.54	22
CT-r2	94	0.35	0.32	0.45/0.10	19
CT-r3	97	0.74	0.80	0.83/0.17	2
CT-r4	97	NS¶	0.11*	0.12*	100
NT-r1	96	0.12	0.27	0.28	52
NT-r2	60	0.15	0.26	0.28	71
NT-r3	100	0.34	0.54	0.56/0.09	11
NT-r4	101	0.42	0.74	0.80/0.24	1
CTcover-r1	98	0.56	0.81	0.81	0
CTcover-r2	94	0.48	0.20	0.59/0.19	17
CTcover-r3	97	0.07*	0.23	0.25	40
CTcover-r4	98	0.34	0.67	0.67	5
All plots	1128	0.15	0.26	0.35/0.17	8

* Significant at $P < 0.05$.
† Topographical regression models of all plots included linear effects of relative elevation, terrain slope, curvature, and aspect. Quadratic terms were added to the models when found to be significant at the 0.05 level.
‡ Yield was added to the previously selected topographical regression model. Regression slopes for yield in the topography + yield model are shown only for the data sets where contribution of yield to the topographical regression model was statistically significant at the 0.05 level.
§ Reported nugget/sill ratios were obtained based on the variogram fitting for complete data sets.
¶ Not significant at the 0.05 level.

explained >50% of the total C variability (Table 1). The topographical regression model was significant with $P < 0.01$ in all but one (CT-r4) plot, where it was significant with $P < 0.05$. These results are consistent with other studies (Moore et al., 1993; Gessler et al., 2000; Mueller and Pierce, 2003; Terra et al., 2004).

When 7-yr average yield data were included in the topographical regression models, the yield contribution was statistically significant in only five of the studied plots as well as in the combined data set (Table 1). The regression slopes for yield in these data sets were greater than zero, indicating that even when topographical characteristics were controlled for, the areas within the fields with above-average long-term standardized yield were also the areas with higher total C.

By adding standardized average yield to topographical regression models, we assessed the hypothesis that areas with consistently high or low yields contribute different biomass inputs to the topsoil and that these differences will explain a substantially large additional amount of total C variability besides that already accounted for by topographical characteristics. This hypothesis, however, was not well supported by the data. One reason is that the dense topographical data of this study were more accurate than the yield-monitor data, which despite cleaning and checking could still contain a significant amount of spatial error (Drummond and Sudduth, 2003). The other reason for better performance of topographical data in predicting total C is that topography is a factor that substantially influences yields itself as well as total C. Indeed, in this study the R^2 values for regression between 7-yr average yield and topographical features ranged from 0.22 to 0.80 (data not shown). These results are consistent with information on the importance of topography in crop yield variability reported by Kravchenko and Bullock (2000), Cox et al. (2003), Jiang and Thelen (2004), Kaspar et al. (2004), and others. Also, it is possible that corn, soybean, or wheat grain yield is not a good indicator of the actual amounts of biomass inputs to the soil.

Total C predictions of the independent test data sets using the data from 20 grid samples based on the regression kriging with 7-yr average standardized yield resulted in R^2 for the regression between predicted and observed values to exceed 0.50 in five of the studied data sets (Table 2). In regression kriging with topography, R^2 exceeded 0.50 also in five of the 12 studied data sets. For the combined data set, the R^2 values between observed and predicted test data were equal to 0.40 and 0.31 for regression kriging with yield and topography, respectively. Overall, prediction performances of topography- and yield-based regression kriging were relatively similar. Using yield produced lower RMSE and higher R^2 values than those of topography in five plot data sets and in the analysis of the combined data set, while topography performed somewhat better in the remaining seven plot data sets.

Despite relatively high R^2 values in regressions of total C with topographical variables or with crop yield observed in a number of plots (Table 1), however, only a modest improvement in mapping accuracy was observed when the regression kriging results were compared with those of the ordinary kriging. Relative improvement in the RMSE of yield-based regression kriging over the RMSE of ordinary kriging exceeded 10% in only two of the plot data sets and in the combined data set. In four of the data sets, the relative improvement was negative, indicating somewhat better performance of ordinary kriging. The topography-based regression kriging produced a 14% improvement in RMSE in one of the studied plots and was negative also in one of the studied plots. In the remaining plot data sets and in the combined data set, the relative improvement did not exceed 10%. These results are comparable with modest improvements in mapping accuracy of soil properties based on either topography or yield in a number of reported studies (Baxter and Oliver, 2005; Webster and Oliver, 2001; Hengl et al., 2004).

We hypothesize that smoothly varying terrain and, related to it, smoothly varying spatial distributions of crop yields and soil C were a partial reason for the lack of apparent advantage in using topography or yield for C mapping in this study, even in those plots where the relationships between the auxiliary variables and C were relatively strong. When strength of spatial correlation of the primary variable is relatively high (low N/S ratios), the improvement in mapping accuracy due to using secondary variables is known to be relatively small when accounting for secondary information via cokriging procedures (Goovaerts, 1997). Our assessment of regression kriging performance in the simulated data sets at varying R^2 values and N/S ratios demonstrated that it is also true in accounting for secondary information in regression kriging procedures. The polynomial regression equation relating relative improvement with R^2 and primary variable's N/S ratios indicates that even though

Table 2. Prediction results for independent test data sets using regression kriging based on multiple regression with topography and simple regression with 7-yr average standardized yield.

	Regression kriging with 7-yr avg. standardized yield			Regression kriging with topography‡		
Plot	RMSE†	R^2	Relative improvement over ordinary kriging	RMSE	R^2	Relative improvement over ordinary kriging
			%			%
CT-r1	0.099	0.27	−10	0.083	0.48	7
CT-r2	0.100	0.45	12	0.112	0.30	1
CT-r3	0.112	0.82	6	0.128	0.76	−8
CT-r4	0.091	NS§	1	0.088	0.10	5
NT-r1	0.151	0.13	5	0.136	0.30	14
NT-r2	0.203	0.11	3	0.195	0.18	7
NT-r3	0.169	0.56	8	0.175	0.52	5
NT-r4	0.154	0.83	10	0.162	0.81	6
CTcover-r1	0.180	0.69	−13	0.150	0.78	5
CTcover-r2	0.128	0.52	9	0.140	0.42	0
CTcover-r3	0.159	0.29	−2	0.151	0.36	3
CTcover-r4	0.107	0.43	−14	0.091	0.58	3
All	0.233	0.40	14	0.251	0.31	7

† Root mean square error between observed and predicted values of the test data sets.

‡ Topographical regression models of all plots included linear effects of relative elevation, terrain slope, curvature, and aspect. Quadratic terms were added to the models when found to be significant at the 0.05 level.

§ Not significant at the 0.05 level.

the R^2 positively affects the relative improvement, it also significantly interacts with N/S ratio ($\alpha = 0.05$):

$$\begin{aligned}\text{Relative improvement} = & -10.3 - 78.1(R^2) + 1.03(N/S) \\ & + 147.1(R^2)^2 - 0.0186(N/S)^2 \\ & + 0.0000905(N/S)^3 \\ & + 1.23(R^2)(N/S) \\ & - 1.15(R^2)(N/S)^2\end{aligned}$$

Figure 1 shows a plot of the relationship between relative improvement and R^2 and N/S ratios as expressed by the above equation. A combination of a primary variable having strong spatial correlation (N/S ratios <30%) and primary and secondary variables being relatively weakly related ($R^2 < 0.40$) appeared to be particularly unfavorable to obtaining any improvement due to using regression kriging (left lower corner of Fig. 1). When the R^2 was <0.30, the maximum improvement due to regression kriging did not exceed 10% across all ranges of N/S values. When the N/S ratio of the primary variable was very low (<10%), however, no improvement or very small improvement (<10%) could be expected even at relatively high R^2 values of 0.6 to 0.7. Indeed, in cases like that, an excellent performance in predicting the primary variable can be expected from the primary variable data itself with little to no improvement possible due to secondary information. This explains the poor performance of regression kriging compared with ordinary kriging observed in the plots that had relatively strong relationships with either topography or yield but also very low N/S ratios, e.g., Plots CT-r3, NT-r4, CTcover-r1, and CTcover-r4.

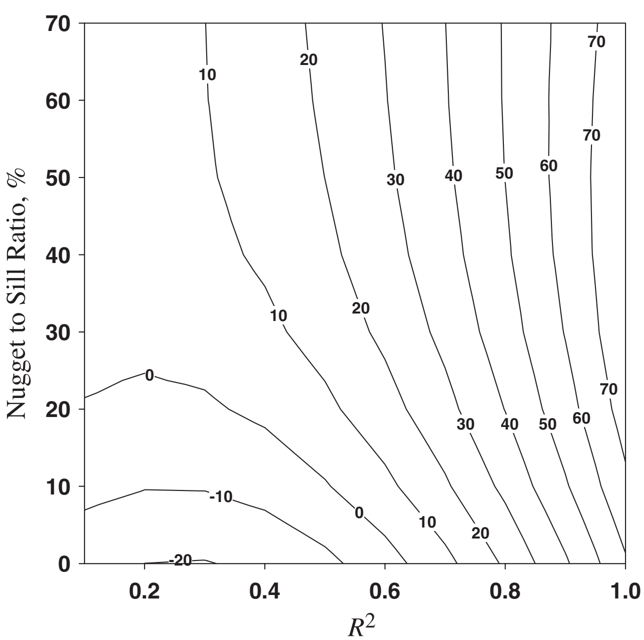

Fig. 1. Percentage of relative improvement in RMSE from using regression kriging over ordinary kriging as a function of the strength of correlation between primary and secondary variables expressed as R^2, and the strength of the spatial correlation of the primary variable expressed as the nugget/sill ratio. The results were obtained based on the simulated data sets.

A more detailed look at the relationship between regression with topography or yield and N/S values of total C data revealed that the strength of the spatial correlation in total C distribution, i.e., N/S ratio, was inversely related to the strength of the relationship between total C and topography (Fig. 2a) as well as to that between total C and long-term average yield (Fig. 2b). It appeared that plots where total C had the strongest relationship with either topography or yield were also the plots where total C data had the most continuous spatial distribution. In plots with less continuous spatial distribution of total C (N/S >40%) performance of regressions with either topography or yield were found to be too poor to be useful for predictive purposes. The observed relationship reflects the fact that, in the studied soils, topography is among the main driving forces of C distribution across the landscape (as well as, discussed above, the spatial distribution of crop yields). The plots where topography and C were strongly related also had a smooth continuous C distribution, probably in response to smoothly varying topographical features of

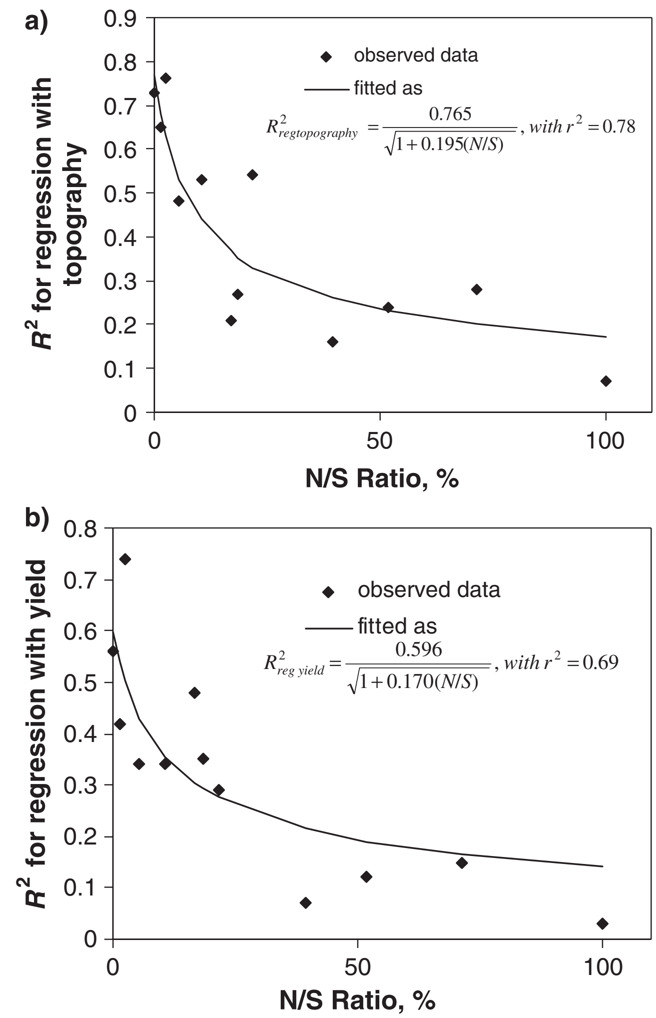

Fig. 2. Values of R^2 for (a) multiple regression of total C with topography and (b) for simple regression of total C with 7-yr average standardized yield plotted vs. the nugget to sill (N/S) ratio of the total C variograms in the 12 studied plots.

the studied plots. Possibly, in the plots where C and topography were not noticeably related, the spatial distribution of C was also affected by varying patterns of glacial outwash parent material or patterns of longer lived perennial plants of preagricultural vegetation that could be present in the studied area (Robertson et al., 1997). Both of these possible sources of variation in total C spatial distribution might be responsible for distributions of total C across landscape being less continuous and thus difficult to predict.

CONCLUSIONS

Total C at the studied depth of 0 to 5 cm was found to be related to topography and the 7-yr average standardized yield in more than half of the studied plots. Accounting for either topographical or yield information in regression kriging, however, produced only a modest (<10%) improvement in mapping accuracy compared with ordinary kriging. Regression kriging was demonstrated to perform poorly in data sets with strong spatial correlation in the primary variable, i.e., total C, even when the regression with a secondary variable was relatively strong. It was observed that, in the studied soil and topographical conditions, the strength of the spatial correlation in total C and the strength of regression of total C with either topography or yield were inversely related. That is, the data sets with promisingly strong regressions with topography or yields were also the ones where strong spatial correlation of the total C itself made topography or yield only barely useful as sources of secondary information. The results indicate that in soil and topographical conditions similar to those of this study, even dense topographical data or dense long-term yield data might be of little value for purposes of improving accuracy in C mapping.

ACKNOWLEDGMENTS

This research was sponsored by a USDA Sustainable Agriculture grant and MSU Intramural Research Grant; it was also supported by the NSF Long-Term Ecological Research, Project no. DEB98-10220, at the Kellogg Biological Station and by the Michigan Agricultural Experiment Station. We thank Matias Ruffo and Ignacio Colonna (Department of Crop Sciences, Univ. of Illinois at Urbana-Champaign) for conducting elevation measurements; Andrew Corbin and the LTER team for help with collecting soil samples; and Suzanne Sippel (Kellogg Biological Station, Michigan State Univ.) for georeferencing soil sample locations.

REFERENCES

Baxter, S.J., and M.A. Oliver. 2005. The spatial prediction of soil mineral N and potentially available N using elevation. Geoderma 128: 325–339.

Bishop, T.F.A., and A.B. McBratney. 2001. A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma 103:149–160.

Blackmore, S. 2000. The interpretation of trends from multiple yield maps. Comput. Electron. Agric. 26:37–51.

Cox, M.S., P.D. Gerard, M.C. Wardlaw, and M.J. Abshire. 2003. Variability of selected soil properties and their relationships with soybean yield. Soil Sci. Soc. Am. J. 67:1296–1302.

Cressie, N. 1985. Fitting variogram models by weighted least squares. Math. Geol. 17:563–586.

Deutsch, C.V., and A.G. Journel. 1998. Geostatistical software library and user's guide. Oxford Univ. Press, New York.

Drummond, S., and K. Sudduth. 2003. From sensors to information: Cleaning yield monitor data. *In* Proc. InfoAg 2003 Conf., Indianapolis, IN. 29 July–1 Aug. Potash and Phosphate Inst., Norcross, GA.

Economic Research Service. 2006. Agricultural resources and environmental indicators. Ch. 4.7: Information technology management. Available at www.ers.usda.gov/publications/arei/eib16/Chapter4/4.7/ (verified 28 Sept. 2006). USDA-ERS, Washington, DC.

Environmental Systems Research Institute. 2004. ArcGIS Spatial Analyst. ESRI, Redlands, CA.

Flowers, M., R. Weisz, and J.G. White. 2005. Yield-based management zones and grid sampling strategies: Describing soil test and nutrient variability. Agron. J. 97:968–982.

Gessler, P.E., O.A. Chadwick, F. Chamran, L. Althouse, and K. Holmes. 2000. Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Sci. Soc. Am. J. 64:2046–2056.

Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.

Hengl, T., G.B.M. Heuvelink, and A. Stein. 2004. A generic framework for spatial prediction of soil variables based on regression kriging. Geoderma 120:75–93.

Isaaks, E.H., and R.M. Srivastava. 1989. Applied geostatistics. Oxford Univ. Press, New York.

Jiang, P., and K.D. Thelen. 2004. Effect of soil and topographic properties on crop yield in a north-central corn–soybean cropping system. Agron. J. 96:252–258.

Kaspar, T.C., D.J. Pulido, T.E. Fenton, T.S. Colvin, D.L. Karlen, D.B. Jaynes, and D.W. Meek. 2004. Relationship of corn and soybean yield to soil and terrain properties. Agron. J. 96:700–709.

Kravchenko, A.N., and D.G. Bullock. 2000. Correlation of corn and soybean grain yield with topography and soil properties. Agron. J. 92:75–83.

Kravchenko, A.N., G.P. Robertson, X. Hao, and D.G. Bullock. 2006. Management practice effects on surface total carbon: Differences in spatial variability patterns. Agron. J. 98:1559–1568.

Kravchenko, A.N., G.P. Robertson, K.D. Thelen, and R.R. Harwood. 2005. Management, topographical, and weather effects on spatial variability of crop grain yields. Agron. J. 97:514–523.

Lark, R.M., and J.V. Stafford. 1997. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. Ann. Appl. Biol. 130:111–121.

Lark, R.M., and J.V. Stafford. 1998. Information on within-field variability from sequences of yield maps: Multivariate classification as a first step of interpretation. Nutr. Cycling Agroecosyst. 50:277–281.

McBratney, A.B., M.L. Mendonca Santos, and B. Minasny. 2003. On digital soil mapping. Geoderma 117:3–52.

Mokma, D.L., and J.A. Doolittle. 1993. Mapping some loamy Alfisols in southwestern Michigan using ground-penetrating radar. Soil Surv. Horiz. 34:71–77.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57: 443–452.

Mueller, T.G., and F.J. Pierce. 2003. Enhancing spatial estimates of soil carbon with terrain attributes. Soil Sci. Soc. Am. J. 67:258–267.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: Heterotropic cokriging and regression kriging. Geoderma 67:215–226.

Robertson, G.P., K.M. Klingensmith, M.J. Klug, E.A. Paul, J.C. Crum, and B.G. Ellis. 1997. Soil resources, microbial activity, and primary production across an agricultural ecosystem. Ecol. Appl. 7:158–170.

SAS Institute. 2001. SAS user's guide. Version 8.2. SAS Inst., Cary, NC.

Simbahan, G.C., A. Dobermann, P. Goovaerts, J. Ping, and M.L. Haddix. 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132:417–489.

Terra, J.A., J.N. Shaw, D.W. Reeves, R.L. Raper, E. van Santen, and P.L. Mask. 2004. Soil carbon relationships with terrain attributes, electrical conductivity, and a soil survey in a Coastal Plain landscape. Soil Sci. 169:819–831.

Walter, C., R.A. Viscarra Rossel, and A.B. McBratney. 2003. Spatio-temporal simulation of the field-scale evolution of organic carbon over the landscape. Soil Sci. Soc. Am. J. 67:1477–1486.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Chichester, UK.