POINT VS. AREA GRID SOIL SAMPLING IN THE GREAT PLAINS

T.L. Kastens, K.C. Dhuyvetter, and J.P. Schmidt
Kansas State University
tkastens@ksu.edu, (785) 532-5866

INTRODUCTION

Traditionally, soil sampling has relied on the approach of collecting multiple cores (or sub-samples) from the entire area of interest (e.g., a field), and then combining them into a single soil sample for analyses. Samples were collected from throughout the area with recognition that soil fertility varied spatially. Where greater spatial variability was anticipated, size of sampling area usually was reduced to provide a better estimate of soil fertility. Unfortunately, as Beckett and Webster (1971) pointed out, mean-estimation accuracy of localized soil properties may not dramatically improve as the sampling area size is reduced (small gains for large increases in cost). Coupled with the fact that, varying fertilizer rates within a field has been cumbersome (costly) due to farming patterns and equipment limitations, it should not be surprising that soil samples traditionally have been collected over large areas such as a whole field.

Eventually, the importance of statistically characterizing spatial relationships within a field became known. In particular, if the variogram (mathematically relates degree of correlation to distance of separation) were known, a geostatistical technique called kriging could be used to accurately estimate localized soil properties with fewer soil samples (large gains for small increases in cost) than by using classical statistics along with traditional sampling over sufficiently small sampling areas. Unfortunately, the variogram is never known with certainty; rather, it must be estimated. McBratney and Webster (1983) point out the obvious economic problem. That is, although the variogram can determine optimal sample spacing, the variogram itself often cannot be determined without intensive soil sampling. Moreover, a variogram for one field or soil property typically is not appropriate for another field or soil property.

Technological advances in fertilizer application equipment and in data handling and analysis have greatly increased the demand for site-specific soil sampling and fertilizer recommendations. Despite potential problems linking geostatistical techniques to practical applications, it seems that "point-based plus interpolation" grid soil sampling has emerged as a common practice in site-specific soil fertility management. Nonetheless, researchers continue to wonder about the most appropriate interpolation technique. More importantly, that issue can be even more disconcerting for practitioners, where profits can hang in the balance. Such decision-makers often resort to a combination of two approaches: 1) select the interpolation technique that provides the most visually appealing map by some subjective assessment, or 2) trust the interpolation software's default algorithm with its default parameters. Also, the point sampling/interpolation issue is clouded further with questions about the radius of soil core collection. Consequently, some site-specific practitioners are beginning to ask whether traditional area-based soil sampling procedures might have an inherent advantage to grid soil sampling. While this question is quite relevant to management decisions that impact the profitability of site-specific soil fertility management, empirical data for site-specific yield response to soil fertility variables (e.g., N and P) do not exist, so adequately answering this question relies on data simulations that accurately represent "real world" data. Our objective was to use simulated field data to evaluate expected returns to two methods of soil sampling: 1) collecting multiple sub-samples in a small radius near a cell centroid and interpolating between sampling locations, and 2) collecting multiple sub-samples from throughout the sampling area and using the soil test results from this sample to represent the entire sampling area (similar to whole-field sampling except on some smaller unit, e.g., 1-acre cells). This evaluation is completed in the context of a western Kansas non-irrigated crop producer managing N and P in a corn-fallow-wheat cropping sequence. As a side issue, we briefly examine the expected returns to site-specific over field-scale soil sampling, and whether soil-test proxy data (e.g., electrical conductivity, EC) might provide a feasible substitute for grid soil sampling.

METHODS

Yield models

A mathematical function characterizing expected yield response to fertilizer and fertility is crucial to any economic analysis of fertilizer decisions. We use Kansas State University (KSU) fertilizer recommendations (Leikam, et al., 2003) in conjunction with procedures outlined in Kastens et al., 2003a, 2003b) to generate wheat and corn yield response models from 10000 randomly simulated data points. KSU's N recommendations depend on yield goal, a 2-foot profile soil test N measure (*STN*), and percent soil organic matter (*OM*). KSU's sufficiency P recommendations depend on yield goal and a measure of soil test P (*STP*). Consequently, each simulated data point is comprised of a measure of crop yield, *OM*, *STN*, and *STP*. The asymptotic plateau yield model for each of corn and wheat is specified as

$$\begin{aligned} Yield = {} & \mathrm{B0} * Z * \left(1 - \mathrm{B1} * \exp\left\{-\mathrm{B2} * fertP - \mathrm{B3} * STP\right\}\right) * \\ & \left(1 - \mathrm{B4} * \exp\left\{-\mathrm{B5} * fertN - \mathrm{B6} * STN - \mathrm{B7} * OM\right\}\right), \end{aligned} \quad [1]$$

where *Yield* is the expected crop yield in bu acre^{-1} conditional upon the causal factors specified on the right-hand side, and where the B expressions (parameters) denote numerical constants to be determined. B0 is the model's asymptotic yield plateau, which is the level of yield expected to be approached as all causal factors approach their yield-maximizing levels. The term exp denotes the exponential function. The terms *fertP* and *fertN* denote fertilizer P in lb P_2O_5 acre^{-1} and fertilizer N in lb N acre^{-1}, respectively; *STN* is soil test N in lb N acre^{-1}; *STP* is soil test P as ppm Bray P1; and *OM* is percent soil organic matter. The term *Z* represents unaccounted-for yield causal factors – a variable dynamically created during estimation that equals the simulated yield at each data point, divided by the maximum yield across all 10000 *Yield* values for a crop.

Simulated data used to estimate the parameters of the yield models are based on the following means and coefficients of variation (CV): corn yield (75, 25%), wheat yield (45, 25%), *STP* (16, 50%), *STN* (40, 50%), and *OM* (1.6, 25%). As described in Kastens et al., 2003a, certain economic assumptions also are needed by the procedure. We use a bank interest rate of 8%, an income tax rate of 40%, and prices as follows: wheat \$3.20 bu^{-1}, corn \$2.32 bu^{-1}, *fertP* \$0.24 lb^{-1}, and *fertN* \$0.21 lb^{-1}. The simulated data used to estimate the yield models were simply points that had no reference to space. But, to make meaningful inferences regarding soil sampling schemes, spatially-dependent soil test data are needed. Then, those soil test data, along with model-predicted yields and model-generated optimal fertilizer rates, can be used to characterize finitely-sized land areas over time to draw economic inferences.

Spatial structure

Though this research might apply to different scales, our basic spatial structure is a square 100-acre field managed by the operator as 1-acre units. That is, we assume that the operator can accurately measure crop yield and apply fertilizer at the 1-acre scale (1 rate for each acre for each crop). Though 1-acre crop management is assumed, finer- and coarser-scaled data are considered in making the 1-acre management decisions.

Each acre is considered overlain with an 8x8 grid, making 64 cells per acre. Thus, the 100-acre field, containing 6400 such cells, is considered overlain with an 80x80 grid. An 80x80 grid has a total of 81(81 nodes (points where grid lines intersect). Each of the 6561 nodes is a point in space from which distances to other points can be calculated, and to which simulated variable values are assigned, which means that a simulated spatially dependent variable will contain 6561 observations. Distance is considered measured in units equal to the length of one cell side (Dunits). Hence, since an acre contains 43560 ft^2, in this 1-acre framework, 1 Dunit is $43560^{0.5}/8 = 26.1$ ft. Finally, because points (nodes) are presumed measured but areas (1-acre cells) are presumed managed, means across nodal values depict area values of interest.

Each acre contains 49 interior nodes (excludes border nodes). The mean of simulated values at these 49 points, say for *STP*, is considered to be the acre's true *STP* mean (which is used in simulating the acre's true yield described later). The value at the acre's center node is the acre's centroid value. Starting at the node 1 Dunit inside the two borders of an acre's corner, and then proceeding to increments of 2 Dunits in the x and y direction, results in selecting 16 of the acre's 49 interior nodes. The mean of values at these 16 equally-spaced nodes is considered to be an estimate of the acre's true mean (e.g., representing 16 equally-spaced soil cores that are bulked to form a single soil sample). To examine coarser-scaled soil sampling, we also consider a 4-acre block, with a corresponding 4-acre centroid and a corresponding 16-point 4-acre composite.

Simulating spatially-dependent data

In geostatistics, the spatial dependency of one point on another is considered related to distance and direction, and often to only distance (the isotropic case, and the only one considered here). Then, for a given distance (h) that two points (i and $i+h$) separate each other, one might plot the values associated with each such ith pair in x-y space. Then, the moment of inertia about the x = y line for the n points corresponding to that hth distance is calculated as

$$\gamma\left(h\right) = \frac{1}{2n} * \sum_{i=1}^{n}\left(x_i - y_i\right)^2 \quad [2]$$

Computing *(*(h) for successively greater h, followed by plotting *(*(h) against h, results in a graphical representation of spatial dependency referred to as a variogram. For kriging, graphical variogram data are generalized using one of several functional forms (only certain classes of functions allow the necessary mathematics). In variogram functions, again represented as *(*(h), the y-intercept is referred to as the nugget value. Strictly speaking, the calculated y-intercept is replaced with 0 for h exactly equal to 0; thus, *(*(0) = 0. This represents an expected discontinuity due to the small-scale variability expected in practical applications. An example of such a discontinuity is that, even ignoring laboratory error, two immediately adjacent soil samples likely will have significantly different test values. At some sufficiently high h (the range), variogram functions plateau (the sill) because spatial dependencies are expected to become 0. Assuming isotropy, the function's sill is expected to equate to the variance (Φ^2) of the data. Nugget information often is expressed in terms of the nugget-to-sill ratio (N/S), with lower N/S values associated with more spatially continuous (dependent) data. Finally, because of the sill = Φ^2 assumption, variogram functions readily can be recast as correlogram functions, $\Delta(h)$, which depict the expected correlation between data points based on separation distance. Beyond the range, point-to-point correlation is assumed to be 0.

To simulate data, we use a variant of the sequential multivariate Gaussian simulation exercise described by Deutsch and Journel (1998), where a variable value is drawn conditional upon all already-drawn values from other variables, so that the end result approximates the desired correlation matrix among the variables. In this spatial setting, "other variables" are actually other values of the same simulated variable, where the desired correlations depend on separation distance. That is, using a random sequence, each of the 6561 points in the field is visited and a value established based on already-simulated values, and so on. The expected correlations depend on a 6561x6561 distance matrix constructed so that the *ij*th matrix value is the distance between points *i* and *j*. The correlogram function converts the distance matrix to a correlation matrix.

In spatial interpolation, and in simulation, only "nearby" points are allowed to impact a point of interest. Thus, a search radius must be selected, which should be somewhat consistent with the range that was assumed in the underlying correlogram function. Thirty grid-sampled Corn Belt fields studied by Kravchenko and Bullock (1999) had an average sampling grid of 207 ft (about 1 acre). Also, they suggested that the traditional interpolation technique is IDW4 (inverse distance weighted to the fourth power) and a 12-point search. We too assume that practical sampling points in our field might be the 1-acre centroids. Thus, a 12-centroid search would mean a search radius greater than 16 Dunits. We arbitrarily used 18 Dunits (about 470 feet) as both our search radius and underlying simulation correlogram function range.

To ensure the believability of simulation results, it should be a goal of spatial data simulation to preserve certain features expected of actual data – had it been possible to empirically observe such data. One such feature is the N/S ratio. Kravchenko (2003) notes that the majority of empirical work in the literature reports variograms with N/S ratios ranging between 0.1 and 0.6. A second feature is typical accuracies reported when IDW methods are used to interpolate between practically scaled samples. A common measure of accuracy is G = (1&MSE/MSE_{avg}) (100, where MSE is the mean square error for some predictive method, and MSE_{avg} is the MSE of using the average value as a predictor everywhere (i.e., the population variance). G depicts the percent improvement in accuracy over simply knowing the average (negative values suggest the predictive method is worse than simply using the average). Using 12-centroid IDW4 predictions, Kravchenko and Bullock's (1999) 30 data sets resulted in an average G value of 21.2 for soil test P and 20.8 for soil test K. Although not directly comparable since they were derived from an optimal search radius rather than a 12-point one, the 15 Kansas data sets underlying work reported by Kastens and Staggenborg (2002) resulted in an average G value of 20.8 for soil test P using IDW4. Consequently, our simulations target similar predictive accuracies.

In simulating the data, the typically-used exponential and spherical variograms did not adequately meet the desired accuracy criteria. In particular, for the resultant IDW4 predictions of 1-acre centroids, to achieve a G value around 20 would have required a N/S ratio that was less than 0.1, implying an extremely high degreee of spatial continuity at close distances. On the other hand, a simple linear variogram easily met the desired criteria. For distances (*h*) less than the range, the associated linear correlogram is

$$\rho(h) = \mathrm{C1} * \left(1 - \frac{h}{\text{range}}\right), \qquad [3]$$

where, given a range (we use 18 Dunits throughout), the only parameter requiring modification to depict different degrees of spatial continuity is C1. In Eq. [3], it is easy to see that the y-intercept is C1, implying that, as h approaches 0, the correlation among data pairs approaches C1. The N/S ratio is 1&C1. Though not shown, that $\Delta(h) = 0$ when h > range is implied.

To sufficiently cover the degree of spatial continuity expected of actual data, our simulations consider C1 values of 0.9, 0.8, 0.7, 0.6, and 0.5, corresponding to N/S ratios ranging from 0.1 to 0.5. To ensure adequate statistical confidence in our randomization-based simulations, 100 standard normal vectors (6561 observations each) are simulated for each C1, representing 100 alternative data characterizations (or maps) for each degree of spatial continuity. The standard normal vectors are transformed to variables of interest by using the means and CV's discussed in the Yield models section. In particular, for a given N/S and a given variable, say *STP*, a randomly-selected 20 of the 100 possible vectors are drawn (selecting 20 was a result of trading off computer time against confidence in results). Then, 20 of the 100 vectors are drawn for *STN*, followed by *OM*. The end result is 20 alternative 3-variable data sets to be used in the economic analysis of soil sampling schemes – for each N/S ratio considered. In this setting, the 3 variables are considered independent of each other, but contain the same degree of spatial continuity.

Simulating a crop production framework

In this research, because the managed unit is 1 acre, inferences (predictions) need to be made about average soil test values for each acre. How such information is collected (e.g., through soil sampling) and used (e.g., through spatial interpolation) is referred to as a soil information system, or SIS. An SIS leads to a unique N and P management strategy which results in a unique set of fertilizer rates and yields, and hence changes in *STP* and *STN* over time (*OM* is assumed constant over time). To capture a sufficient part of this dynamic system, we consider 10 corn-fallow-wheat crop sequences (20 crops, 30 years) for each SIS examined.

Given an N/S ratio, each of the 20 soil-test data sets provides what we consider to be true soil tests (*STP*, *STN*, and *OM*) in year 1 for each of the 6561 points in the field – hence, also the 100 true 1-acre soil tests (mean of an acre's 49 interior points), and so on. However, each SIS will result in a different evolution of true *STP* and *STN* in subsequent years. Regardless of the presumed soil sampling density, a particular SIS results in a *predicted* soil test level for each acre. Setting the partial derivative of yield with respect to fertilizer rate in the yield model (Eq. [1]) equal to the fertilizer/crop price ratio results in a posited optimal (hence, the presumed actual) fertilizer N and P rate for each acre. It should be noted that, though the fertilizer rate calculation depends on a predicted (not necessarily the true) soil test for each acre, such predictions could be the same everywhere e.g., when only a single field soil sample is used to generate a uniform fertilizer rate for the whole field.

Given the true soil test values for an acre (49-point means) and fertilizer rates, the yield model determines the *expected* 1-acre yield in a given scenario. However, to establish what we consider to be the true yield, a random error is added to model predictions. The error is structured so that only around 25% of the 1-acre yield variation (across both time and space) is explained by fertilizer and soil test. Also, to depict random annual weather events, a random annual yield shifter (a year "dummy") is included (not shown in Eq. [1]) that proportionately adjusts each acre's yield across all SIS scenarios for a given year. The year "dummy"is structured so that its information, together with that of fertilizer and soil test, will explain about 67% of the yield variation. Of course, fertilizer rates are determined using only expected or average year-dummy information, since it would not be known until after the year is over.

For a given acre, the true *STP* and *STN* values are assumed to evolve in a nutrient budgeting framework, where fertilizer above or below crop-removal levels changes soil tests through a transformation rate. For P, we use the transformation function set forth in Kastens, et al., 2003, where corn and wheat crops remove 0.33 and 0.50 lb P_2O_5 bu^{-1}, respectively (Leikam, et al., 2003). To obtain true 1-acre *STP* values for the next year, an error was added to 1-acre crop-removal based *STP* predictions, reflective of expected crop-removal-based *STP* prediction accuracy. In particular, the error was selected such that predictions would have an R^2 of around 0.50. Though some calculations depend on true *STP* at every data point, true fertilizer rates and true crop yields are simulated for only 1-acre areas. Consequently, true 1-acre *STP* was converted to true all-points *STP* by multiplying last year's point values times the ratio of the new 1-acre measure to last year's 1-acre measure (a simple proportional proration). KSU N recommendations at 1.6% *OM*, and yield goals 10% above the 75 and 45 bu $acre^{-1}$ expected corn and wheat yields, imply crop removal rates of 1.17 and 2.04 lb N bu $acre^{-1}$ for corn and wheat, respectively. The errors added to 1-acre crop-removal-based *STN* predictions to determine true next-year *STN* are chosen to result in an R^2 of 0.20 (assumes some, albeit poor, prediction accuracy for *STN*).

Table 1. Description of soil information systems (SIS).

SIS	Method of assigning soil test results to 1-acre cells
1Aarea	16-point 1-acre composite
4Aarea	16-point 4-acre composite
1Apoint	1-acre centroid
4Apoint	4-acre centroid
1AID4	1-acre centroid with IDW4 interpolation
4AID4	4-acre centroid with IDW4 interpolation
Farea	16-point 100-acre composite
EC	49-point 1-acre composite based on EC proxy

Eight SIS's were considered in this research and are summarized in Table 1. *1Aarea* (1-acre area) assumes each acre's soil tests are predicted via a single 16-point composite soil sample taken from each acre (each sampling point always assumed to be multi-core). *4Aarea* assumes a 4-acre block is sampled with a 16-point composite soil sample from the area; each acre in the block is assigned the same soil test value. *1Apoint* assumes each acre's soil tests are predicted by assigning a single-point soil sample (the centroid) to the whole acre. *4Apoint* assumes each 4-acre block's soil test is predicted using only the 4-acre centroid soil sample. *1AID4* assumes IDW4 is used to interpolate between soil samples taken from 1-acre centroids in the search area. The 49 interior points of each acre that are predicted with IDW4 are averaged to provide a 1-acre prediction. *4AID4* assumes similar calculations, except that only the 4-acre centroids in the search area are used in the IDW4 interpolations. To allow comparisons with traditional field-scale sampling, we included *Farea* (field area), where each acre is assigned the 16-point field-composite soil sample value. Finally, we consider a proxy for soil-test data by simulating predictive variables for each crop with R^2 values of 0.20, 0.0625, and 0.20 for *STP*, *STN*, and

OM, respectively (all 6561 points are predicted; an acre measure is the 49-interior-point mean). The correlations assumed for the proxy variables are taken from EC and soil test data from a study farm in northwest Kansas; hence, this SIS is called *EC*. An annual *Farea* soil sample is assumed to accompany the EC data.

RESULTS

To demonstrate sufficient coverage of the expected degree of spatial continuity arising from previous research, row 13 of Table 2 reports the 20-data-set average G statistic for *STP* associated with *IDW4* interpolation – for each N/S value considered (i.e., predicting each centroid value using interpolation among other centroids in the range, vs. using the 100-centroid average). G statistics for *STN* and *OM* were comparable. Clearly, given that the 18-Dunit range assumption is appropriate, these G values and N/S ratios should provide sufficient coverage of the span of spatial continuity possibilities believed possible for western Kansas soil tests. The N/S = 0.3 or N/S = 0.4 simulations likely are the closest to what might be expected in real data.
Economic results of this research are presented in rows 1-12 of Table 2 as mean *differences* between competing SIS's, where the means are across the 20 alternative runs for each N/S ratio. Differences in profit were first computed as annually amortized (across 30 years) \$ acre^{-1}, and then multiplied by 1.5 for reporting as \$ acre^{-1} crop^{-1} (since the 30 year span contains 20 crops). Positive table values indicate that the method named first in that row was more profitable. For example, for N/S = 0.1, *1Aarea* is \$0.58 acre^{-1} *more* profitable than *1AID4* (row 1), which is \$0.27 acre^{-1} *less* profitable than *1Apoint* (row 2).

The additional benefit to using a 1-acre area soil sample over a "1-acre point sample with IDW4 interpolation" is positive for all spatial dependencies considered (row 1). This indicates that, ignoring differences in soil sampling and analysis costs, area sampling is more profitable than point-with-interpolation sampling. On the other hand, row 2 of the table shows that, using 1-acre centroids, the manager probably would be better off *assigning* the centroid value to the acre rather than using IDW4 interpolation. This was unexpected. Apparently, given the simulations, more weight yet should be given to nearby points – suggesting either a smaller search area or a higher-powered inverse distance interpolation method. Consequently, row 3 shows that the additional gains to using a 1-acre area sample over merely using the centroid point directly are modest. On the other hand, row 6 shows substantial gains to area sampling if the sample is for 4 acres rather than for 1 acre. Now, interpolation does result in small gains over merely using the centroids directly (row 5). Regarding the apparent trends across columns in the table, only those in rows 1 through 3 are statistically different from 0; they support the theoretical expectation that less spatial continuity should lead to increased profits for area- over point-based soil sampling.

For convenience, rows 7 to 9 report the corresponding averages of 1-acre and 4-acre sampling differences – perhaps as an indicator of what might be expected of 2.5-acre sampling, the most typical resolution used by practitioners. Row 8 shows virtually no gains to using interpolation when sampling by centroid, especially if the additional software and computer-time costs associated with interpolation would be assessed. Assuming that the average of the N/S = 0.3 and N/S = 0.4 columns provides the best guess of real-world data, row 9 suggests a benefit to area- over point-sampling of about \$2.18 acre^{-1}. The *additional* costs for acquiring such a gain would be associated with the additional labor required of walking or driving to the multiple spots in the sample area rather than simply pulling cores around the centroid. Based on conversations with a Kansas provider of grid soil sampling, our best guess of the additional time required of 2.5-acre area- over point-sampling, assuming area cores would not need to be geo-referenced, is 11 minutes per 2.5-acre cell, or 4.4 minutes acre^{-1}. Using an arbitrary labor charge of \$15 hour^{-1}, this comes to \$1.10 acre^{-1}, which leaves a net gain of \$1.08 acre^{-1} for area- over point-sampling.

Table 2. Twenty-run average difference in profit (yield revenue less fertilizer cost) in \$ acre^{-1} crop^{-1} for competing soil information systems.

		N/S Ratio (higher N/S ratios imply less spatial dependency)				
Row#	SIS	0.1	0.2	0.3[a]	0.4[a]	0.5
	spatial comparisons for 1-acre and 4-acre sampling		\$ acre^{-1}			
1[b]	*1Aarea - 1AID4*	\$0.58[c]	\$0.71[c]	\$0.83[c]	\$0.92[c]	\$1.05[c]
2[b]	*1AID4 - 1Apoint*	-\$0.27[c]	-\$0.28[c]	-\$0.25[c]	-\$0.25[c]	-\$0.23[c]
3[b]	*1Aarea - 1Apoint*	\$0.31[c]	\$0.43[c]	\$0.57[c]	\$0.67[c]	\$0.81[c]
4	*4Aarea - 4AID4*	\$3.17[c]	\$2.50[c]	\$3.27[c]	\$3.52[c]	\$3.15[c]
5	*4AID4 - 4Apoint*	\$0.41[c]	\$0.44[c]	\$0.31[c]	\$0.38[c]	\$0.58[c]
6	*4Aarea - 4Apoint*	\$3.58[c]	\$2.95[c]	\$3.58[c]	\$3.90[c]	\$3.73[c]
	average of *1A* and *4A* above (perhaps indicative of 2.5 acre sampling)[d]		\$ acre^{-1}			
7	*area - ID4*	\$1.88	\$1.60	\$2.05	\$2.22	\$2.10
8	*ID4 - point*	\$0.07	\$0.08	\$0.03	\$0.07	\$0.17
9	*area - point*	\$1.95	\$1.69	\$2.07	\$2.29	\$2.27
	"might site-specific fertilizer management pay?" comparisons		\$ acre^{-1}			
10	*1Aarea - Farea*	\$12.29[c]	\$11.34[c]	\$10.88[c]	\$12.18[c]	\$10.64[c]
11	*4Aarea - Farea*	\$2.84[c]	\$1.74	\$1.26	\$2.46[c]	\$1.18
12	*EC - Farea*	\$9.31[c]	\$8.36[c]	\$7.93[c]	\$9.26[c]	\$7.74[c]
	accuracy improvement for *STP* (compare with Kravchenko and Bullock, 1999; and Kravchenko, 2003)		%			
13[b]	mean *STP* G-statistic	44.8	37.7	26.7	19.6	4.16

[a] real data are expected to fall between N/S = 0.3 and N/S = 0.4 because a G statistic between 20 and 22 is expected
[b] trend across successive columns is statistically different from 0 at 0.95 confidence
[c] statistically different from 0 with a paired t-test (n = 20) at 0.95 confidence level
[d] statistical significance not calculated for this section

Rows 10 and 11 show the benefits of grid over whole-field soil sampling. Taking the average of rows 10 and 11 for N/S = 0.3 and N/S = 0.4 columns as a reasonable indicator of gains of 2.5-acre- over whole-field sampling, results in a benefit of \$6.70 acre^{-1}. KSU's soil test laboratory currently (2003) charges \$10.50 to test N, P, and OM, which is \$4.20 acre^{-1} for a 2.5 acre sample. For typical 2.5-acre grid sampling, Whipker and Akridge (2003) report a charge of \$6.19 acre^{-1}. These 2 costs add to \$10.39, making it clear that 2.5-acre sampling ahead of each crop would not be profitable. On the other hand, soil samples might be used for more than 1 crop, which surely is reasonable for at least *STP* and *OM*; then, grid soil sampling would be profitable.

Though not shown, using inverse distance squared interpolation (rather than 4th power) would increase the expected benefits to area- over point-based sampling (all row-7 values in

Table 2 would be greater). Finally, perhaps the most surprising result uncovered in this research is the one associated with the simulated soil-test proxy variables reported in row 12. Substantial profits would be expected from such information after accounting for costs. For example, EC information can be collected for around \$5.00 acre^{-1} and likely would be used for many years.

SUMMARY

From a simulation of spatially-dependent data representative of non-irrigated corn-fallow-wheat cropping practices in western Kansas we conclude the following. In grid soil sampling, modest net gains, in the order of \$1.08 acre^{-1}, are expected for area- over point-sampling. If only cell centroids are used in grid sampling, the benefits for using a spatial interpolation technique such as inverse distance are negative to small, suggesting that a manager might just as well assign the centroid value to the whole cell. Grid soil sampling is expected to be profitable as long as soil samples can be used for two or more crops. Using soil-test proxy information such as electrical conductivity, along with a field-scale soil sample for each crop, is expected to result in substantial profits that likely are greater than grid soil sampling profits.

REFERENCES

Beckett, P.H.T., and R. Webster. 1971. Soil variability: a review. Soils and Fertilizers. 34:1-13.

Deutsch, C.V., and A.G. Journel. 1998. GSLIB geostatistical software library and user's guide. 2nd ed. Oxford University Press. New York.

Kastens, T.L., and S. Staggenborg. 2002. Spatial Interpolation Accuracy. [online] www.agmanager.info (verified 3 Jan. 2004).

Kastens, T.L., J.P. Schmidt, and K.C. Dhuyvetter. 2003a. Yield models implied by traditional fertilizer recommendations and a framework for including non-traditional information. Soil Sci. Soc. Am. J. 67,1:351-64.

Kastens, T.L., J.P. Schmidt, and K.C. Dhuyvetter. 2003b. Managing fertilizer programs on leased land in the Eastern Corn Belt. [online] http://www.agmanager.info (verified 3 Jan. 2004)

Kravchenko, A.N. 2003. Influence of Spatial Structure on Accuracy of Interpolation Methods. Soil Sci. Soc. Am. J. 67:1564-1571.

Kravchenko, A., and D.G. Bullock. 1999. A comparative study of interpolation methods for mapping soil properties. Agron. J. 91:393-400.

Leikam, D.F., R.E. Lamond, and D.B. Mengel. 2003. Soil Test Interpretations and Fertilizer Recommendations. Kansas State University MF-2586 [online]. http://www.oznet.ksu.edu (verified 3 Jan. 2004).

McBratney, A.B., and R. Webster. 1983. How many observations are needed for regional estimation of soil properties? Soil Sci. 135,3:177-183.

Whipker, L.D. and J.T. Akridge. 2003. 2003 Precision Agricultural Services Dealership Survey Results. Staff Paper No. 3-10 [online]. http://www2.agriculture.purdue.edu/ssmc/ (verified 3 Jan. 2004).