Analysis of Spatial Factors Influencing Crop Yield

K. A. Sudduth, United States Department of Agriculture
S. T. Drummond, University of Missouri
Stuart J. Birrell, University of Missouri
N. R. Kitchen, University of Missouri
Nutrient Mapping Implications of Short-Range Variability

S.J. Birrell
Dept. of Agricultural and Biological Engineering
University of Missouri
Columbia, Missouri

K.A. Sudduth
USDA-Agricultural Research Service
Cropping Systems and Water Quality Research Unit
Columbia, Missouri

N.R. Kitchen
Dept. of Soil and Atmospheric Sciences
University of Missouri
Columbia, Missouri

ABSTRACT

Successful site specific nutrient application depends on accurate soil nutrient maps, which are generally developed from grid samples. The implication of short range variability for soil nutrient mapping is investigated. Interpretation of soil nutrient maps must consider the level of confidence associated with estimated values.

INTRODUCTION

The economic and environmental benefits of site specific nutrient application will only be realized if spatial variability across the field is accurately determined. Generally the spatial variability of soil nutrients is determined by grid sampling. While increasing the sampling intensity is the best way to improve the accuracy of soil nutrient maps, the maximum sampling intensity is limited by cost. Studies have addressed the effect of different grid sampling sizes, determining the degree of misapplication by comparing maps developed from coarser grids containing a subset of the original sampling grid to the reference map developed using all the samples in the original grid (Motz and Searcy, 1993). However, this method assumes that the recorded values for the finest grid are an accurate representation of the actual value at each grid intersection, and does not account for micro- and meso-variability of soil nutrient levels.

Although the interpolation method (Kriging, inverse distance, etc.) can have a significant effect on the maps developed, varying interpolation parameters such as number of neighbors and search radius can have a greater effect than using different interpolation procedures. Nutrient recommendation decisions based on soil nutrient maps should not only consider the magnitude of the value for each cell, but also the reliability of the estimate, and whether any differences are significant or merely a result of uncertainty in the estimates.

Published 1996
METHODS

A 28 ha field was sampled on a 25 m grid, with additional samples taken at the intersection of every fourth row and column (100 m grid) of the original sample grid (Figure 1). Two additional samples were taken co-incident with the original 100 m samples (for a total of 3 samples). Three other samples were obtained 2 m, 5 m and 10 m from the grid point in a straight line on a random bearing (Figure 1). Each sample consisted of 8 cores to a depth of 6-7 inches obtained within a radius of 50 cm from the sample point. Twenty randomly selected samples were split into sub-samples (duplicates) before submission to the University of Missouri Soil Testing Laboratory for analysis. Although the samples were collected in a sequential manner, the sample identification numbers were randomized to remove the possibility of systematic bias in laboratory analysis. Unfortunately, results from two additional fields sampled in a similar manner had to be eliminated because statistical analysis proved that the small scale samples (additional co-incident samples, 2 m, 5 m and 10 m samples) were biased. This was traced to systematic error in the laboratory analysis on the particular day the small scale samples were tested.

The data was used to create 6 different data sets for analysis. The first data set consisted of all the data (25 m grid), including all small scale samples. A second set consisted of the 100 m grid samples with all small scale samples. The third data set consisted of all 3 co-incident samples of the 100 m grid. The final three data sets were obtained by assigning one of the 3 co-incident samples on the 100 m grid to each set.

STATISTICAL ANALYSIS

Classical Statistics

General ANOVA methods are based on the assumption that the data samples are independent, but this assumption is violated when there is a spatial
relationship between samples. However, the small scale samples were taken on a 100 m grid which is almost at the limit of spatial dependence. Therefore, classical statistical methods were used during preliminary data analysis on the different data sets. During classical analysis, all small scale samples taken at a particular grid point were considered to have the same “location”. The 2, 5, and 10 m points were considered to have different “distances”, the co-incident samples were different “samples”, and the duplicate sub-samples were termed “sub-samples”. The different data sets were analyzed using unbalanced ANOVA models (location, distance, sample, sub-sample) to investigate the relative contribution of the different sources of variation.

**Geostatistical Analysis**

The different data sets were analyzed using geostatistical methods. The semi-variograms of the different data sets displayed no drift provided the semi-variogram was restricted to 2/3 of the maximum lag to prevent edge effects. Therefore, the data was assumed to satisfy the intrinsic hypothesis. The distribution of the complete data set was slightly skewed.

The classical semi-variogram (Matheron, 1963) is affected by skewed data.

\[
\gamma(h) = \frac{1}{2N(h)} \sum (Z(s_i) - Z(s_j))^2
\]

Therefore, the robust semi-variogram estimator developed by Cressie and Hawkins (1980) was used.

\[
\gamma(h) = \frac{\left( \frac{1}{2N(h)} \sum \left| (Z(s_i) - Z(s_j)) \right|^{1/2} \right)^4}{0.457 + \frac{0.0484}{N(h)}}
\]

The semi-variograms were developed using various lag interval widths to compare the effect on semi-variogram estimation. Theoretical semi-variogram models were fitted using linear (linear model) and non-linear (spherical and exponential models) least squares estimation procedures in SAS. The model parameters were estimated using both a non-weighted and a weighted procedure where weighting was based upon the number of paired observations at each lag distance. The model with the smallest error sum of squares was selected as the best fit model.

A significant advantage of Kriging is that the estimation variance is provided as well as the estimated value. Assuming a normal population distribution, the confidence interval (95%) around the estimate can be determined using

\[
(\bar{Z} + (1.96)\sigma_o/\sqrt{n}, \bar{Z} - (1.96)\sigma_o/\sqrt{n})
\]

Simple Kriging was used to develop maps of the Kriged estimate, upper confidence interval and lower confidence interval on a 10 m grid, for each of the different data sets. The correlation between the different maps was determined on a cell by cell basis.
RESULTS

All the pH values for co-incident soil samples were within 10% of their mean value at each location. Approximately 80 percent of the potassium samples were within 10% of their mean value and the rest were within 25% of the mean. However, for phosphorus only 50% of the samples were within 10% of the mean value, 42% of the samples were within 20% and the rest were within 30% of their respective means. Therefore, short range variability can significantly affect individual cell estimates and the maps developed.

The frequency distributions of the different nutrients for the complete data set were slightly skewed. However, the distribution curves for the reduced data sets varied from slightly skewed distributions to highly irregular distribution patterns, although they were samples from the same population. Therefore, even the frequency distributions were affected by the number of samples and short range variability.

Classical Statistics

The assumptions upon which classical ANOVA methods are based were violated. However, the results provided an indication of the relative contribution of the different sources of variation. As expected location within the field was highly significant (<1% level). Distance was a significant parameter for potassium at the 5% level, and not significant for most other soil measurements (pH, neutralizable acidity, calcium, magnesium, potassium). The sample was significant for phosphorus at the 1% level and distance was only significant at the 5% level. Therefore, for phosphorus the location of the individual coring positions for each sample (meso-variability) was more significant than the relative position (macro-variability) of the small scale sampling site.

Geostatistical Analysis

The semi-variograms of pH, P and K for the full data set (Figure 2), were calculated using the classical and robust semi-variogram estimation equations. The range of spatial influence was approximately 100 m for this data set. However, the range of the spatial structure was not maintained when the reduced data sets were used. Many of the reduced data sets exhibited pure nugget variance, and semi-variograms produced from the co-incident data sets did in some cases display significantly different characteristics. Therefore, the co-incident data sets did not have enough data points for the reliable characterization of semi-variograms.

The classical semi-variogram consistently exhibited a higher sill value than the robust semi-variograms. However, the latter tended to have a greater spatial range. The width of the lag interval did not have a significant effect on the determination of the semi-variogram parameters when the weighted least squares procedure was used. However, lag interval width did have a significant effect on the error sum of squares and correlation coefficient between the experimental and fitted semi-variogram model. The correlation between the experimental and theoretical semi-variogram increased as the lag interval width increased. The lag
Figure 2 Estimation of semi-variograms using classical and robust procedures (complete data set).

interval did affect the estimated semi-variogram parameters with the non-weighted least squares procedure, particularly with short lag widths, when the number of paired observations could be very low. The relative effect of lag width on the correlation coefficient was more pronounced for the non-weighted least squares procedure than the weighted methods, due to the increased amount of averaging at each lag interval which occurred as the interval width increased.
Figure 3  Kriged estimates (top) and 95% confidence intervals (middle and bottom) for soil potassium, developed from the complete data set (left) and the three co-incident 100 m grid sample sets (right).
Figure 4  Kriged estimates (top) and 95% confidence intervals (middle and bottom) for soil phosphorous, developed from the complete data set (left) and the three co-incident 100 m grid sample sets (right).
Figure 5 Kriged estimate (top) and estimation standard deviation (bottom) for soil pH, developed from the complete data set (left), 100 m grid samples including all small scale samples (2nd left) and the three co-incident 100 m grid sample sets (right).
### Table I  Correlations between Kriged maps developed from the 6 data sets.

<table>
<thead>
<tr>
<th></th>
<th>25m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</th>
<th>100m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</th>
<th>100m Grid Samples&lt;sup&gt;2/&lt;/sup&gt;</th>
<th>100m Grid Set 1</th>
<th>100m Grid Set 2</th>
<th>100m Grid Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Soil pH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>1.00</td>
<td>0.56</td>
<td>0.70</td>
<td>0.28</td>
<td>0.27</td>
<td>0.33</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>0.56</td>
<td>1.00</td>
<td>0.63</td>
<td>0.05</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;2/&lt;/sup&gt;</td>
<td>0.70</td>
<td>0.63</td>
<td>1.00</td>
<td>0.30</td>
<td>0.28</td>
<td>0.37</td>
</tr>
<tr>
<td>100m Grid Sample Set 1</td>
<td>0.28</td>
<td>0.05</td>
<td>0.30</td>
<td>1.00</td>
<td>0.98</td>
<td>0.96</td>
</tr>
<tr>
<td>100m Grid Sample Set 2</td>
<td>0.27</td>
<td>0.05</td>
<td>0.28</td>
<td>0.98</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>100m Grid Sample Set 3</td>
<td>0.33</td>
<td>0.11</td>
<td>0.37</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Phosphorous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>1.00</td>
<td>0.67</td>
<td>0.48</td>
<td>0.42</td>
<td>0.31</td>
<td>0.35</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>0.67</td>
<td>1.00</td>
<td>0.83</td>
<td>0.75</td>
<td>0.56</td>
<td>0.63</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;2/&lt;/sup&gt;</td>
<td>0.48</td>
<td>0.83</td>
<td>1.00</td>
<td>0.89</td>
<td>0.68</td>
<td>0.73</td>
</tr>
<tr>
<td>100m Grid Sample Set 1</td>
<td>0.42</td>
<td>0.75</td>
<td>0.89</td>
<td>1.00</td>
<td>0.72</td>
<td>0.76</td>
</tr>
<tr>
<td>100m Grid Sample Set 2</td>
<td>0.31</td>
<td>0.56</td>
<td>0.68</td>
<td>0.72</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>100m Grid Sample Set 3</td>
<td>0.35</td>
<td>0.63</td>
<td>0.73</td>
<td>0.76</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Potassium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>1.00</td>
<td>0.63</td>
<td>0.61</td>
<td>0.60</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;1/&lt;/sup&gt;</td>
<td>0.63</td>
<td>1.00</td>
<td>0.90</td>
<td>0.81</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>100m Grid Samples&lt;sup&gt;2/&lt;/sup&gt;</td>
<td>0.61</td>
<td>0.90</td>
<td>1.00</td>
<td>0.92</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>100m Grid Sample Set 1</td>
<td>0.60</td>
<td>0.81</td>
<td>0.92</td>
<td>1.00</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>100m Grid Sample Set 2</td>
<td>0.54</td>
<td>0.84</td>
<td>0.93</td>
<td>0.75</td>
<td>1.00</td>
<td>0.89</td>
</tr>
<tr>
<td>100m Grid Sample Set 3</td>
<td>0.58</td>
<td>0.84</td>
<td>0.92</td>
<td>0.75</td>
<td>0.89</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<sup>1/</sup> Includes all small scale samples (3 samples at 100m grid and 2, 5, 10m from original point)

<sup>2/</sup> Includes small scale samples at grid point (3 samples at 100m grid)
The potassium Kriged maps using the different data sets showed the greatest similarity. The maps from the co-incident 100 m data sets displayed similar trends (Figure 3), and were consistent with the patterns obtained using the complete data set. The range between the upper and lower confidence limits were also similar using the different data sets.

The phosphorous maps (Figure 4) were the least similar. The Kriged maps from the co-incident data sets were significantly different from the map developed using the complete data set and from each other. Similarly, the confidence limits were not consistent among the maps. The lower confidence limit map shows that a significant portion of the field would require a high fertilizer application rate, while the upper map shows that a relatively small portion of the field would respond to fertilizer. Therefore, there is a high probability that fertilizer application based on the Kriged estimates would result in misapplication in some areas.

Soil pH was relatively less variable than phosphorous. The maps based on 100 m grids were similar to the original and to each other (Figure 5). The 100 m grid map which included the small scale samples identified most of the trends seen in the original grid. The maps based on the co-incident samples showed the same basic soil pH levels as the original, but considerable local variability was missed. Although the estimation maps were similar, the different sampling patterns resulted in very different estimation variances. The estimation standard deviation for the complete data set generally was less than 0.14, which is of the same approximate magnitude as the standard deviation of the laboratory analysis. For the 100 m data, the estimation standard deviation depended on the data set used, which means that the semi-variogram developed was not adequately modeled from this data.

The correlation between the different maps was determined on a cell by cell basis (Table 1). The correlation between the original map and those from 100 m grid maps was highest for potassium, whereas the correlations for pH and phosphorous were more variable and generally lower than those for potassium.

**CONCLUSIONS**

The interpretation of soil nutrient maps must consider the level of confidence associated with estimated values, and the maps must not be automatically accepted as a true representation of the actual conditions in the field. The acceptance of interpolated maps without questioning their quality could compromise the nutrient recommendation procedure. Also, it will not be possible to accurately evaluate the spatial relationship of soil nutrients and yields if the nutrient patterns are not accurately determined.

**REFERENCES**


Motz, D.S., and S.W. Searcy. 1993. Interpolation methods for spatially variable data. ASAE paper 93-3561, ASAE, St Joseph, MI