FROM PLOT TO POINT: UNCOVERING PHOSPHORUS SPATIAL VARIABILITY

S. Cominelli, A. Arbelaez, T. Hefley, and D. Ruiz Diaz
Kansas State University, Manhattan, KS
scominelli@ksu.edu (785)317-7541

ABSTRACT

Soil test phosphorus (P) in field trials is usually reported as one value per plot: a composite of cores mixed together, with core locations rarely recorded. This average hides the variation inside the plot. Our objective was to quantify, using simulated data, how much of that variation can be recovered from plot averages, and whether the 95% prediction range can be trusted. We hypothesized that both depend on the distance over which P stays similar (similarity distance) relative to plot width. We simulated P on the layout of a field trial with 72 plots of 10 × 40 ft. In each plot, 10 random cores with unknown locations were averaged into one composite, and random noise was added. For similarity distances of 5, 20, and 80 ft (500 fields each), the plot average, centroid kriging, and area-to-point (ATP) kriging, which treats each value as the average of the whole plot, rebuilt the field from the plot values. Both kriging methods estimated the spatial parameters by REML, and predictions were compared with the simulated truth. With random noise of 0.1, ATP kriging reduced prediction error by 2–20% versus the plot average, but recovered at most 19% of the variation inside plots, and none at 5 ft. Its 95% range held the truth in 75% of cells at 5 ft and 93–94% at 20 and 80 ft. In these simulations, plot averages retained little of the pattern inside plots; maps from composite samples should be reported with their uncertainty.

INTRODUCTION

In field research trials, composite soil samples are commonly used to characterize soil fertility at the field, block, or plot level. Although practical, a composite is an average of the sampled area and cannot represent the variability within it. This matters because ignoring variability can inflate experimental error, confound treatments with pre-existing soil patterns, and make blocks less effective at grouping similar soil.

To address this limitation without additional sampling, statistical methods can estimate values at points from averages over an area (Gotway and Young, 2002; Wikle and Berliner, 2005). Kriging is a spatial interpolation method that estimates values at unsampled locations from nearby samples, giving more weight to closer ones. Area-to-point (ATP) kriging is a version that accounts for the fact that each composite is the average of an entire area (Kyriakidis, 2004). It has been applied to soil P composites when the location of the cores was recorded, in zones of 0.6 to 10 acres, and gave smaller prediction errors than treating each composite as a single point at the center of its zone (Schirrmann et al., 2012). Composite kriging was later extended to samples with different numbers and locations of cores, with the caution that composites carry little information to separate short-distance variation from measurement error (Orton et al., 2015).

These applications used sampling units the size of fields and farms. Experimental plots are much smaller (for example, 10 × 40 ft in this study), and the

distance over which P stays similar can be comparable to, or shorter than, the plot itself. Whether variation inside plots can be recovered from plot composites at this scale, and whether the uncertainty reported with the predictions can be trusted, has not, to our knowledge, been quantified. Our objective was to quantify how much of the variation inside plots can be recovered from plot averages using kriging methods, and whether the 95% prediction range can be trusted. We hypothesized that both depend on the distance over which P stays similar (similarity distance) relative to plot width.

MATERIALS AND METHODS

Simulated fields

To evaluate the methods, we simulated P over a field trial with 72 plots of 10 × 40 ft, forming a 180 × 160 ft field divided into 2.5-ft cells (4,608 cells; 64 per plot) (Figure 1 A). True soil test P in each cell was generated as a random field with a mean of 12 mg P/kg, a variance of 1 ppm ^{2}, and a correlation between cells that decreases with their distance d as $\exp(-d/\varphi)$, where φ is the similarity distance, the distance over which P stays similar. We used φ = 5, 20, and 80 ft: shorter than the plot width (10 ft), intermediate, and longer than the plot length (40 ft).

Sampling

Within each simulated plot, 10 cells were selected at random to represent soil cores. The composite was the mean of the true P in these cells plus random noise with variance 0.1 or 0.5 ppm^2, representing laboratory error. Core locations were treated as unknown. We simulated 500 fields for each of the six combinations of φ and noise level.

Prediction methods

Three methods rebuilt the soil test P in all 4,608 cells from the 72 composites. (i) Plot average: every cell received the composite value of its plot, i.e., no spatial prediction. (ii) Centroid kriging: each composite was treated as a measurement at the plot center and interpolated to all cells. (iii) ATP kriging: each composite was treated as the average of the P values over its whole plot, plus an error; this accounts for the fact that the composite represents an area and not a point. Because the average is taken over the whole plot and the core locations are unknown, the variability due to which cores were selected is absorbed in the error term. For both kriging methods, the spatial parameters (the overall variance, the similarity distance, and the error variance) were estimated from the 72 composites by restricted maximum likelihood (REML; Patterson and Thompson, 1971), and a constant mean was assumed. As a benchmark, ATP kriging was also run with the true parameters.

Evaluation

Predictions were compared with the true P in all cells using (i) the average prediction error, as the root mean squared error (RMSE); (ii) within-plot skill, the share of the variation inside plots that was recovered (0 = no better than giving every cell the plot average; 1 = perfect); and (iii) coverage of the 95% prediction range, the proportion of cells whose true value fell inside the range. Results are means over the simulated fields, with ± 2 Monte Carlo standard errors.

RESULTS AND DISCUSSION

What a plot average keeps

We first compare the true P of one simulated field with what each method reconstructs (Figure 1). The true P surface (Figure 1A) varies at a scale finer than the plots, but the plot averages (Figure 1B) show only blocks of constant color: the broad gradient survives and the detail inside each plot is gone. Kriging the 72 plot values back to the 4,608 cells produced smooth surfaces (Figures 1C and 1D) that follow the plot averages and interpolate between plots, and they are visually close to each other. The question is how much of the true detail these maps actually recover.

More accurate maps, little recovery inside plots

The average prediction error was lower with kriging than with the plot average, and the difference increased with the similarity distance (Figure 2A). With random noise of 0.1, the average prediction error (RMSE) of ATP kriging was 2.5% lower than that of the plot average at φ = 5 ft (0.97 vs. 1.00 ppm), 8.7% lower at 20 ft (0.71 vs. 0.77 ppm), and 19.8% lower at 80 ft (0.42 vs. 0.52 ppm). Centroid kriging gave reductions of 3.3, 7.4, and 18.8%, and the two kriging methods differed by less than 0.02 ppm. Most of the gain therefore comes from borrowing information among neighboring plots, not from the way the plot area is modeled.

A lower prediction error does not mean that the pattern inside plots was recovered. Within-plot skill, which looks only at the deviations from each plot average, was 0.00 for ATP kriging at 5 ft, 0.09 at 20 ft, and 0.19 at 80 ft (Figure 2B); centroid kriging reached 0.00, 0.08, and 0.18. Even with the true parameters, skill was only 0.01, 0.10, and 0.20, so the ceiling is set by the information in 72 plot averages and not by the estimation of the spatial parameters. When P changes within a plot (5 ft), practically nothing of the variation inside plots could be recovered, and the improvement in prediction error came entirely from the differences among plots.

Can the 95% prediction range be trusted?

The answer depends on the same distance. Because the plot average has no prediction range, coverage could be evaluated only for the kriging methods (Figure 2C). At 20 and 80 ft, the range of ATP kriging contained the true value in 93 and 94% of the cells, close to the 95%. At 5 ft, it contained it in only 75%, and centroid kriging performed similarly (72%) and also fell short at 20 ft (86%). With the true parameters, coverage was 95% in every scenario, which points to the estimation of the spatial parameters as the cause. When P changes within a plot, plot averages are almost independent of each other and carry little information on how P varies with distance, and the prediction range does not account for the uncertainty of the estimated parameters (Zimmerman and Cressie, 1992). Maps derived from composites are thus most overconfident exactly where the plot hides the most.

These results support our hypothesis. Both the recovery of variation inside plots and the reliability of the 95% range improved as the similarity distance grew relative to the plot width, and when P changed over distances shorter than the plot, neither method could return what the composite had averaged away.

Effect of larger random noise

A larger random noise (0.5) made the plot average less accurate, so kriging reduced the prediction error by 11, 19, and 36% relative to it. However, kriging recovered less of the variation inside plots (at most 11%) and its 95% range was less reliable (coverage between 69 and 93%). Centroid kriging had slightly lower RMSE than ATP kriging (0.99 vs. 1.05, 0.80 vs. 0.82, and 0.51 vs. 0.52 ppm at 5, 20, and 80 ft), so the choice between the two kriging methods was secondary to the main finding.

Implications for field research and next steps

Taking more cores per plot represents the plot better, but it costs field and laboratory work. Our simulations show that kriging of plot averages can recover part of the variation inside plots (up to 19%) at no extra sampling cost, but only when P stays similar over distances longer than the plot width (20 to 80 ft here). When P changes within a plot (5 ft), little was recovered and the prediction range was too narrow, so maps built from plot averages should not be interpreted at a scale finer than the plot and should be reported with their uncertainty. These results come from simulated fields on one trial layout, with 10 cores per plot and no treatment or block effects, and real soil P is often skewed. We did not vary the number of cores per plot. The next step is to apply the method to the P trial at Scandia, Kansas (72 plots), where the true point-scale P is unknown.

CONCLUSIONS

Plot averages retain little of the pattern inside plots. ATP kriging recovered at most 19% of the variation inside plots, and almost none when P changed within a plot; in that case, its 95% range was also too narrow. Maps derived from composite samples should therefore be reported with their uncertainty.

REFERENCES

Cahn, M.D., J.W. Hummel, and B.H. Brouer. 1994. Spatial analysis of soil fertility for site-specific crop management. Soil Sci. Soc. Am. J. 58:1240-1248.

Gotway, C.A., and L.J. Young. 2002. Combining incompatible spatial data. J. Am. Stat. Assoc. 97:632-648.

Kyriakidis, P.C. 2004. A geostatistical framework for area-to-point spatial interpolation. Geogr. Anal. 36:259-289.

Mallarino, A.P. 1996. Spatial variability patterns of phosphorus and potassium in no-tilled soils for two sampling scales. Soil Sci. Soc. Am. J. 60:1473-1481.

Mallarino, A.P., and D.J. Wittry. 2004. Efficacy of grid and zone soil sampling approaches for site-specific assessment of phosphorus, potassium, pH, and organic matter. Precis. Agric. 5:131-144.

Orton, T.G., M.J. Pringle, D.E. Allen, R.C. Dalal, and T.F.A. Bishop. 2015. A geostatistical method to account for the number of aliquots in composite samples for normal and lognormal random variables. Eur. J. Soil Sci. 66:1023-1032.

Patterson, H.D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545-554.

Schirrmann, M., R. Herbst, P. Wagner, and R. Gebbers. 2012. Area-to-point kriging of soil phosphorus composite samples. Commun. Soil Sci. Plant Anal. 43:1024-1041.

Wikle, C.K., and L.M. Berliner. 2005. Combining information across spatial scales. Technometrics 47:80-91.

Zimmerman, D.L., and N. Cressie. 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. Ann. Inst. Stat. Math. 44:27-43.

FIGURES

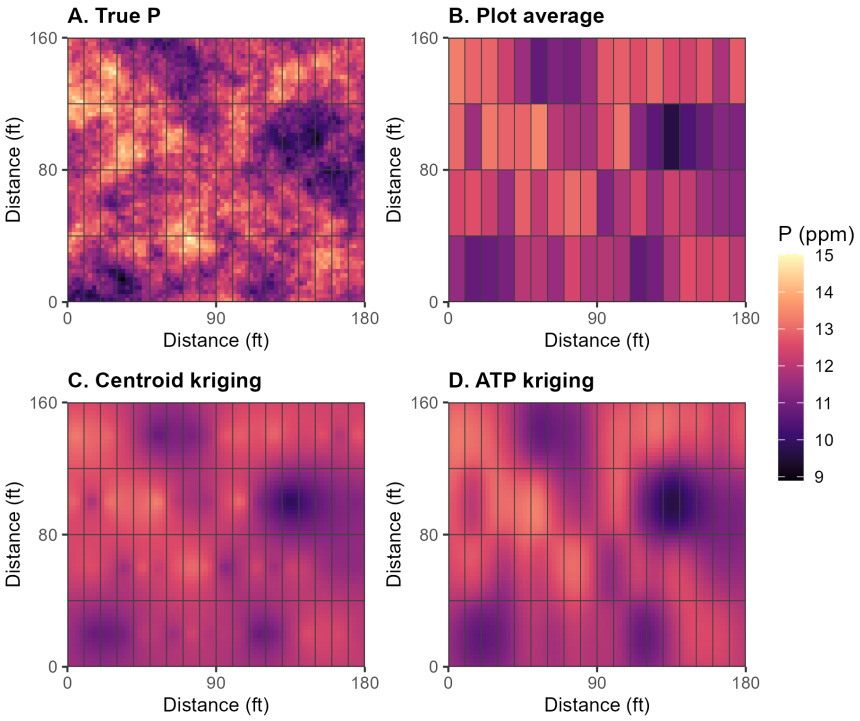


Figure 1. One simulated field (similarity distance 20 ft, random noise 0.1). A, true P; B, plot average (mean of 10 random cores per plot); C, centroid kriging; D, ATP kriging. Same color scale in all maps; lines show plot boundaries (10 × 40 ft).

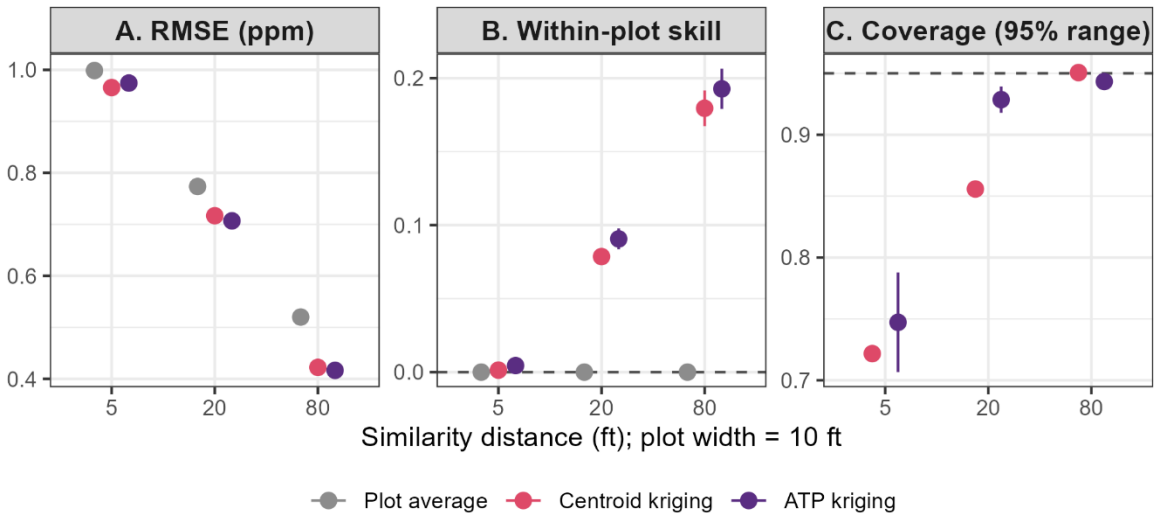


Figure 2. Performance of the three methods over 500 simulated fields per similarity distance (random noise 0.1; bars, ± 2 Monte Carlo SE). A, average prediction error (RMSE); B, within-plot skill; C, coverage of the 95% prediction range. Dashed lines mark 0 (no better than the plot average) and 0.95 (target coverage). The plot average has no prediction range, so it has no coverage. Plot width is 10 ft.