SOIL ELECTRICAL CONDUCTIVITY MAP QUALITY: IMPACT OF INTERPOLATION SEARCH NEIGHBORHOOD PARAMETERS

B.G. Sears, T.G. Mueller, N. Pusuluri, and K. Isukapalli
University of Kentucky, Lexington, Kentucky

Abstract

Spatial predictions of soil electrical conductivity (EC) measurements may be improved by adjusting the search neighborhood criteria. The objective of this study was to investigate how varying search parameters impacted the quality of soil EC maps. The three fields chosen for this study were from the Inner and Outer Blue Grass physiographic regions of Kentucky. Soil EC was measured by direct contact at all locations. The prediction datasets included EC measurements along transects that were separated by 18 and 19 m. The transects of the validation datasets were obtained with 9 m between passes and were collected in directions that were perpendicular to the prediction dataset measurements. Inverse distance weighted (IDW) interpolations were conducted with varying search radii, minimum numbers of prediction points required for interpolation, maximum allowable coefficients of variation (CV) of search neighborhood data values, and distance exponent values. The interpolations were evaluated by examining prediction efficiency and the percentage of validation points for which predictions were made (i.e., those predicted values not eliminated because of too few data points or CV values exceeding the threshold). Prediction efficiency was only substantially and consistently improved by reducing the maximum allowed CV of search neighborhood data values. The percentages of points for which valid predictions were made varied substantially by search radius, minimum number of prediction points required for interpolation, and the CV threshold. By using a CV threshold of 11%, prediction efficiency was improved by 19% and the percentage of points retained was reduced by 80% across locations. For sensor measurements, an 80% loss of data may be acceptable if map quality is substantially improved. A CV threshold is not typically used as a search neighborhood criterion. However, in situations where data are abundant, restricting the CV of data values within a search neighborhood may be an effective technique for improving map quality.

Introduction

Soil EC data collected along transects are often interpolated for yield map analysis and interpretation (e.g., Triantafilis et al., 2001; Corwin et al., 2003) because yield and EC sensor data are not collected at identical locations. However, the use of these maps may lead to incorrect conclusions about factors impacting yield variability if the quality of interpolated values are poor (i.e., predicted and true EC values vary substantially). While many studies have examined prediction quality for points collected along regular grids, little research has been conducted for intensive measurements along transects (Mueller et al., 2004).

Guidelines for adjusting the radius and number of points that will be used for interpolation (e.g., Isaaks and Srivastava 1989; Goovarets, 1997) have not been rigorously tested. The accuracy of an interpolated value will likely be poor if the variation of the measurements within the search neighborhood is large. Removing these interpolated points may improve the quality of the data for analysis. However, this will only be practical if measurements are abundant (e.g., intensive EC measurements made for yield map interpretation). The leading commercial interpolation software programs do not allow the variance within a search neighborhood to be used as a threshold for excluding predicted values.

The objective of this study was to investigate the impact of three search neighborhood parameters (search radius, minimum numbers of prediction points required for interpolation, and maximum allowable CV of search neighborhood data values) for IDW interpolation on map quality.

Materials and Methods

The three fields chosen for this study were from the Inner [LeRoy (37° 58'6" N. 84°32'9" W.) and Spindletop (38°7'57" N. 84°30'6" W.) locations] and Outer [Ellis location (38°21'48" N. 85°11'48" W.)] Blue Grass physiographic regions of Kentucky. Within and across locations, soils varied by slope class and degree of historic erosion as described by (Mueller et al., 2004).

Shallow (0 to 30 cm) and deep (0 to 90 cm) bulk soil EC were measured with a Veris 3100 Mapping System (Veris Technologies, Salina, KS) at speeds of approximately 10.5 km hr^{-1}. Measurements were geo-referenced with an AgGPS 132 differential global positioning system receiver (Trimble, Sunnyvale, CA). Bulk soil EC was logged each second. This provided a sample spacing of approximately 2.9 m between measurements along transects. Prediction and validation EC datasets were collected using the Trimble AgGPS Parallel Swathing Option to ensure that distances between transects were uniform. There were 19 m between transects for the prediction datasets at the Ellis and LeRoy locations and 18 m between transects for the Spindletop location. The validation dataset measurements had 9-m transect spacings and were collected in directions that were perpendicular to the prediction dataset measurements. See Mueller et al. (2004) for a more thorough description of sampling.

Interpolations were conducted with unpublished software (T.G. Mueller, personal communication). The Cartesian distance between points was calculated with the SearchbyDistance MapObjects LT ActiveX automation control object (ESRI, Redlands, CA). For each validation point, a search neighborhood was defined by the search radius. Only those interpolated values which had at least the minimum numbers of prediction points required for interpolation and less than the maximum allowable CV of search neighborhood data values. There was no restriction set on the maximum number of points allowed for interpolation. Interpolations for the validation points were computed by calculating the distance weighted averages (IDW approach) of the values within each search radius.

The MSE, which was defined as

$$\mathrm{MSE} = \frac{1}{\mathrm{n_v}} \sum_{i=1}^{n_v} v_i^2$$

where v_i was the difference between predicted and observed values at location s_i, i=1, ..., n_v, and n_v was the number of points in the validation data set. Prediction efficiency, a reduction-in-error index, was calculated as

$$\text{Prediction efficiency}\,(\%) = 100\left(\text{MSE}_{\text{field average}} - \text{MSE}\right)\left(\text{MSE}_{\text{field average}}\right)^{-1}$$

where $\text{MSE}_{\text{field average}}$ is the MSE obtained from using the field average as an estimate for all test data. Prediction efficiency, often referred to as goodness by Agterberg (1984), is a useful index for comparing prediction quality across variables and locations.

Prediction efficiency and percent of points retained were calculated for all combinations of the following sets of interpolation parameters: search radius values (from 2 to 20 m by 2 m), minimum numbers of prediction points required for interpolation (from 2 to 20 by 2), the maximum allowable CV of search neighborhood data values (from 5 to 50% by increments of 5%), and IDW distance exponent values (from 1 to 3 by 0.1). The averages of the parameters that resulted in maximum prediction efficiency while retaining 40, 50, and 60% of the points in the validation dataset were considered the *nominal parameter values*. Next we determined the prediction efficiency and percent point retention for each iteration as parameter values were varied while the remaining three parameters were held constant to the nominal parameter values).

Results and Discussion

The nominal parameter values were 17 m for the search radius, 8 for minimum numbers of prediction points required for interpolation, 19% for the maximum allowable CV of search neighborhood data values, and 1.6 for the IDW distance exponent. These nominal values were determined so that they could be used as constants for the examination of the individual influence of interpolation parameters on prediction efficiency and the retention of validation points.

The impacts of the search radius, minimum numbers of prediction points required for interpolation, maximum allowable CV of search neighborhood data values, and distance exponent values are shown in Fig. 1 and 2. Of the four parameters investigated, the maximum CV had the largest impact on prediction efficiency across locations. However, search radius, minimum numbers of prediction points required for interpolation, maximum allowable CV of search neighborhood data values each had a substantial impact on point retention (Fig. 1). The IDW distance exponent value did not impact point retention and, therefore, this relationship was not shown (Fig. 2).

Prediction quality was not stable and point retention was poor when search radius values were less than 10 (Fig. 1). At larger distances map quality generally improved with greater search radius values. Initially, the percentage of validation points retained was greater as search radius values increased. This occurred because an increasing number of validation points had search neighborhoods that contained at least the nominal number of points required for prediction (i.e., 8 points) as the search radius increased. Point retention diminished as search radius values exceeded 10 to 15 m because an increasing number of neighborhood CV values exceeded the nominal CV threshold (i.e., 19%).

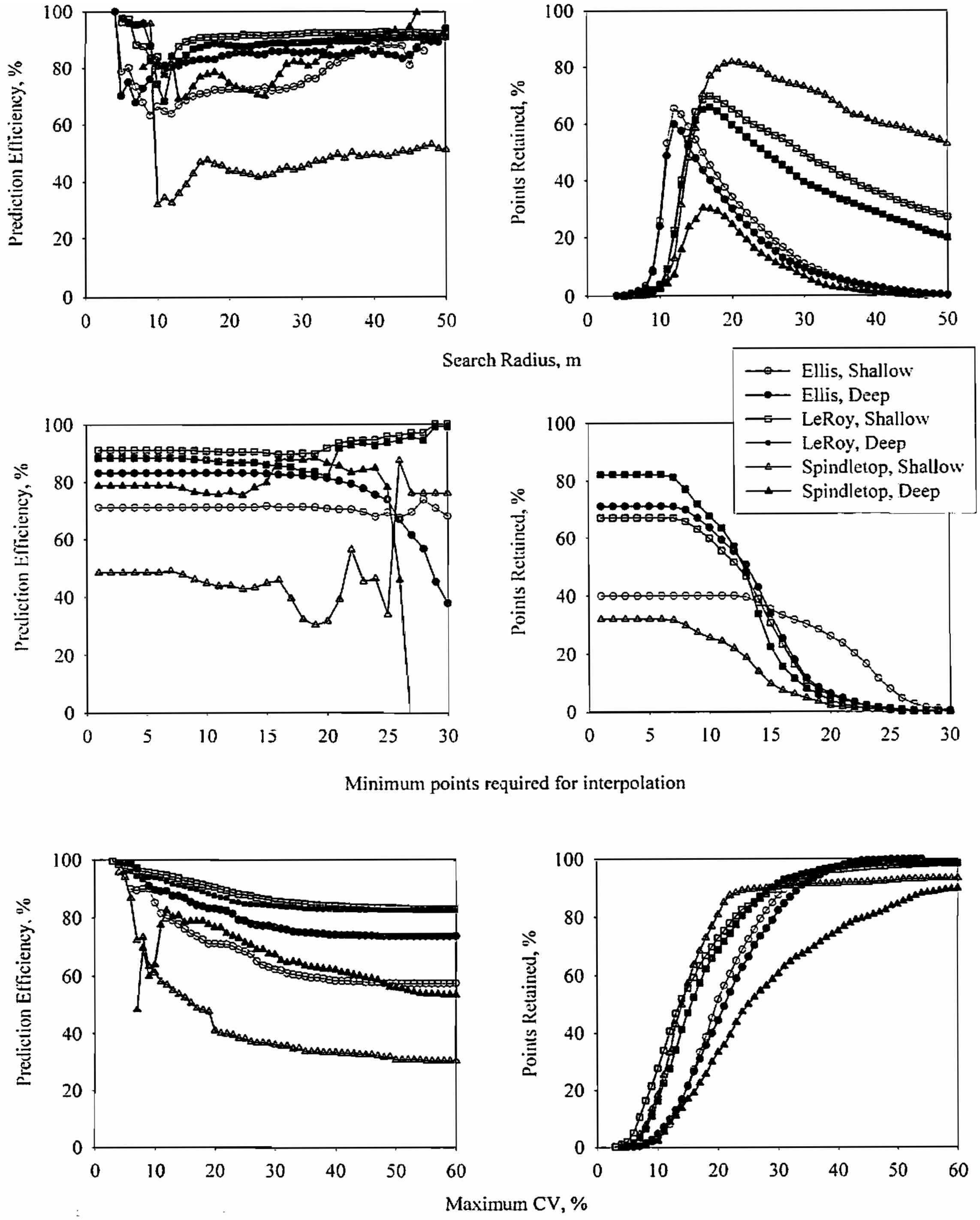

Fig. 1. The impact of varying the search radius, minimum number of prediction points required for interpolation, maximum allowable CV of search neighborhood data values on prediction efficiency and point retention. As each parameter was varied, the other three parameters were set to the nominal values (i.e., 17 m for the search radius, 8 for the minimum number of points, 19% for the maximum CV value, and 1.6 for the IDW distance exponent).

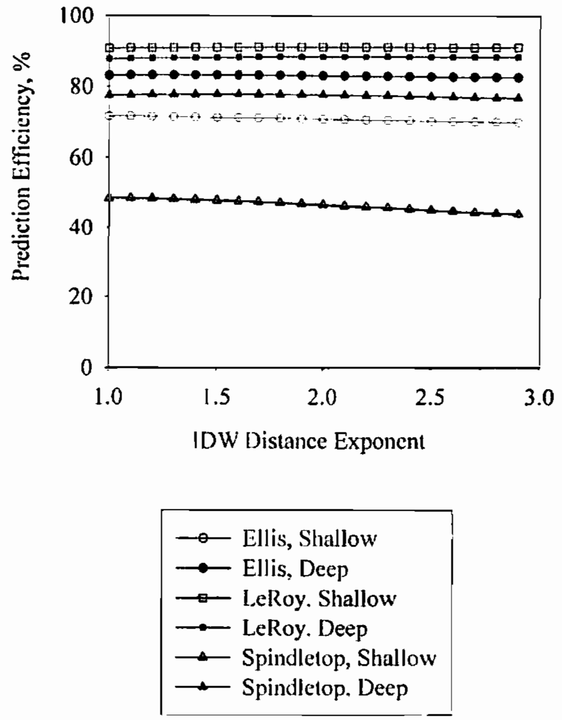

Fig. 2. The impact of IDW distance exponent values on prediction efficiency. As the distance exponent was varied, the remaining three parameters were set to the nominal values (i.e., 17 m for the search radius, 8 for the minimum number of points, 10% for the maximum CV value).

As the minimum number of points required for interpolation increased, prediction efficiency values generally decreased slightly and became increasingly unstable particularly at high values. The minimum number of points had little impact on retention for values less than 5 because there were at least 5 points from the prediction datasets within the search neighborhoods. For large values, few of these neighborhoods contained sufficient points for retention. Isaaks and Srivastava recommended that search ellipses be manipulated so that they would contained at least 12 points. Goovaerts (1997) recommendation that a minimum of 10 points be used for interpolation would not have worked well with the 17-m nominal radius used in this study. However, it may have been adequate with larger radius values.

Prediction efficiency generally increased as the maximum CV threshold was reduced. Unfortunately, point retention diminished rapidly as CV threshold values dropped below 20 to 30%. However, because sensor data involves the collection of large datasets, an adequate number of points may remain for analysis and interpretation despite a substantial reduction in interpolated values. For example, consider a CV threshold value of 11%, a radius of 17-m, a minimum of 8 points to be used for interpolation, and an IDW distance exponent of 1.6. For these parameters, average prediction efficiency would have been 83% for all three locations and across EC depths. For the same parameter combination without a CV threshold, the average prediction efficiency would have only been 64%. On average, imposing a CV threshold resulted in a 19% improvement in prediction quality with an 80% reduction in the number of interpolated points.

Conclusions

A CV threshold can be used to substantially improve soil EC interpolation quality. Unfortunately, the increase in quality can result in the removal of many predicted values. However, given the large number of points collected with sensors, the improvement in quality may justify the elimination of interpolated values.

Although, the search radius and minimum points required for interpolation parameters had only a slight impact on prediction efficiency, they can substantially impact the number of points retained for analysis. Therefore, the minimum number of points with a search neighborhood should be set to a low value. The search radius value chosen should be large if no CV threshold is used. If used, however, the search radius value that results in the largest number of valid predictions should be used.

References

Arterberg, F.P. 1984. Trend surface analysis. P. 174-171. In G.L. Gaile and C.J. Willmott (ed.) Spatial Statistics and Models. Reidel, Dordrecht, the Netherlands.

Corwin, D.L., S.M. Lesch, P.J. Shouse, R. Soppe, and J.E. Ayars. 2003. Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity. Agron. J. 95:352-364.

Goovaerts, P. 1997. Geostatistics for Natural Resource Evaluation. Oxford University Press. New York, NY.

Isaaks, E.H. and R.M. Srivastiva. 1989. An Introduction to Applied Geostatistics. Oxford University Press. New York, NY.

Mueller, T.G., B. Mijatovic, B.G. Sears, N. Pusuluri, and T.S. Stombaugh. 2004. Soil electrical conductivity map quality. Soil Sci. (In Press).

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated corn. Soil Sci. Soc. Am. J. 65:869-878.

PROCEEDINGS OF THE

THIRTY-FOURTH NORTH CENTRAL EXTENSION-INDUSTRY SOIL FERTILITY CONFERENCE

Volume 20

November 17-18, 2004
Holiday Inn Airport
Des Moines, IA

Program Chair:
Bob Hoeft
University of Illinois
Urbana, IL 61801
(217) 333-9480

Published by:
Potash & Phosphate Institute
772 – 22nd Avenue South
Brookings, SD 57006
(605) 692-6280
Web page: www.ppi-ppic.org