Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers and Electronics in Agriculture 50 (2006) 97–108
Computers and electronics in agriculture
www.elsevier.com/locate/compag

Testing the performance of spatial interpolation techniques for mapping soil properties

T.P. Robinson *, G. Metternicht

Department of Spatial Sciences, Curtin University of Technology, GPO Box U 1987, Perth WA 6845, Australia

Received 9 August 2004; received in revised form 30 June 2005; accepted 27 July 2005

Abstract

In this paper, we implement and compare the accuracy of ordinary kriging, lognormal ordinary kriging, inverse distance weighting (IDW) and splines for interpolating seasonally stable soil properties (pH, electric conductivity and organic matter) that have been demonstrated to affect yield production.

The choice of the exponent value for IDW and splines as well as the number of the closest neighbours to include was decided from the root mean squared error (RMSE) statistic, obtained from a cross-validation procedure. Experimental variograms were fitted with the exponential, spherical, Gaussian and linear models using weighted least squares. The model with the smallest residual sum of squares (RSS) was further interrogated to find the number of neighbours that returned the best cross-validation result.

Overall, all of the methods gave similar RMSE values. On this experimental field, ordinary kriging performed best for pH in the topsoil and lognormal ordinary kriging gave the best results when applied to electrical conductivity in the topsoil. IDW interpolated subsoil pH with the greatest accuracy and splines surpassed kriging and IDW for interpolating organic matter.

In all uses of IDW, the power of one was the best choice, which may due to the low skewness of the soil properties interpolated. In all cases, a value of three was found to be the best power for splines. Lognormal kriging performed well when the dataset had a coefficient of skewness larger than one. No other summary statistics offered insight into the choice of the interpolation procedure or its parameters. We conclude that many parameters would be better identified from the RMSE statistic obtained from cross-validation after an exhaustive testing.

Keywords: Interpolation; Spatial prediction; Geostatistics; Soil properties; Precision agriculture; Cross-validation

1. Introduction

Implementation of variable-rate technology can provide considerable financial gains to the farming industry. However, its effectiveness relies on the accuracy of the spatial interpolation used to define the spatial variability of soil properties.

The accuracy of interpolation methods for spatially predicting soil properties has been analysed in several studies. Kravchenko and Bullock (1999) compared inverse distance weighting (IDW), ordinary kriging and lognormal ordinary kriging for soil properties (phosphorous (P) and potassium (K)) from 30 experimental fields. They found that if the underlying dataset is lognormally distributed and contains less than 200 points, lognormal ordinary kriging generally

* Corresponding author. Tel.: +61 8 9266 3935; fax: +61 8 9266 2707.
E-mail addresses: T.Robinson@curtin.edu.au, pfrass@bigpond.net.au (T.P. Robinson), G.Metternicht@curtin.edu.au (G. Metternicht).

0168-1699/$ – see front matter
doi:10.1016/j.compag.2005.07.003

outperforms both ordinary kriging and IDW; otherwise, ordinary kriging is more successful. Further, Laslett et al. (1987) also found ordinary (isotropic) kriging to be a better method than IDW for interpolating pH. In fact, Laslett et al. (1987) judged splines to be a better than both IDW and kriging. In contrast, Gotway et al. (1996) observed better results than kriging for soil organic matter and nitrogen when using IDW. Weber and Englund (1992) also found that IDW produced better results than kriging (with lognormal kriging outperforming ordinary kriging).

There have been many conflicting reports concerning the use of basic statistics to predetermine both interpolation methods and their parameters. For example, Kravchenko and Bullock (1999) report a significant improvement in accuracy of soil properties interpolated using IDW by manipulating the exponent value. They found that data with high skewness (>2.5) were often best estimated with a power of four (five out of eight datasets) and for most of the soil properties with low skewness (<1), a power of one yielded the most accurate estimates (9 out of 15 datasets). Alternatively, Weber and Englund (1994) report that IDW with a power of one resulted in a better estimation for data with skewness coefficients in the range of four to six when interpolating blocks of contaminant waste sites. Likewise, a larger exponent produced better estimations when the data had low skewness.

For organic matter, in particular, Gotway et al. (1996) found that the accuracy of the inverse distance method increased as the exponent value increased. Their findings show that properties with a low coefficient of variation (<25%) were better explained by a higher power, in most cases a power of four. In addition, datasets with a high coefficient of variation gave best results when a power of one was used. On the contrary, Kravchenko and Bullock (1999) found no significant correlation between the exponent value used for IDW and the coefficient of variation.

Given the variability of results obtained by these previous studies the research reported hereafter aims to:

- Assess the accuracy of various well-known interpolation techniques for mapping soil pH, electrical conductivity and organic matter through manipulation of the various parameters attributable to each technique;
- Determine if non-spatial statistics could assist in determining the best interpolation method to implement without using exhaustive test parameters;
- Identify the spatial prediction method that best illustrates the spatial variability of the soil properties studied. This would enable the identification of areas where remediation is required to improve crop growth.

2. Spatial prediction methods

2.1. Kriging

The presence of a spatial structure where observations close to each other are more alike than those that are far apart (spatial autocorrelation) is a prerequisite to the application of geostatistics (Goovaerts, 1999). The experimental variogram measures the average degree of dissimilarity between unsampled values and a nearby data value (Deutsch and Journel, 1998), and thus can depict autocorrelation at various distances. The value of the experimental variogram for a separation distance of h (referred to as the lag) is half the average squared difference between the value at $z(x_i)$ and the value at $z(x_i + h)$ (Lark, 2000b):

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{i=1}^{N(h)}[z(x_i) - z(x_i + h)]^2 \quad (1)$$

where $N(h)$ is the number of data pairs within a given class of distance and direction. If the values at $z(x_i)$ and $z(x_i + h)$ are autocorrelated the result of Eq. (1) will be small, relative to an uncorrelated pair of points. From analysis of the experimental variogram, a suitable model (e.g. spherical, exponential) is then fitted, usually by weighted least squares, and the parameters (e.g. range, nugget and sill) are then used in the kriging procedure.

2.2. Inverse distance weighting

Similar to kriging, inverse distance weighting directly implements the assumption that a value of an attribute at an unsampled location is a weighted average of known data points within a local neighborhood surrounding the unsampled