Computers and Electronics in Agriculture 136 (2017) 202–209

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Interpolation selection index for delineation of thematic maps

Vanderlei Artur Bier, Eduardo Godoy de Souza *

PGEAGRI, Technological and Exact Sciences Center, Western Paraná State University (UNIOESTE), Cascavel, Paraná, Brazil

A R T I C L E   I N F O

Article history:
Received 14 June 2016
Received in revised form 7 March 2017
Accepted 10 March 2017
Available online 23 March 2017

Keywords:
Cokriging
Geostatistics
Inverse distance weighting
Ordinary kriging
Precision agriculture

A B S T R A C T

Precision agriculture aims at the precise management of agricultural inputs toward increasing profits, decreasing losses, and preserving the environment. Thus, the use of thematic maps to understand the behavior of involved attributes is important, and the construction of these maps usually involves some type of interpolation. Choosing the best interpolation method can be difficult, and the use of cross-validation is not trivial. Therefore, the purpose of this study is to propose a selection index of interpolation methods that will help in choosing the best deterministic and stochastic model among those evaluated. The study was conducted within a 15.5-ha area in Southern Brazil, and an interpolation selection index (ISI) was applied to data on clay, copper, and manganese content, and apparent electrical conductivity of the soil using four interpolation methods: inverse distance, inverse distance squared, ordinary kriging, and cokriging. Using the ISI, choosing between deterministic and stochastic interpolation methods is simplified and less subjective. In cases where the deterministic interpolator (inverse of distance squared) was chosen, the spatial dependency was moderated. Note that the proposed statistic (ISI) does not quantify the difference between the analyzed methods.

1. Introduction

In addition to representing the terrain, thematic maps are used to illustrate themes. Generally, thematic maps are used to identify different cartographic representations, and they represent not only the land but also associated characteristics. The development of thematic maps is linked to data collection, analysis, interpretation, and representation of the information on a map. This facilitates the identification of similarities and enables the visualization of spatial correlations. One specific case of thematic maps is contour maps, which are built by connecting points of the same value, and are applicable to geographical phenomena that show continuity in the geographic space. Contour maps can be constructed from absolute data (elevation, temperature, precipitation, humidity, and atmospheric pressure) or relative data (density, percentages, and indices). Based on samples collected during a harvest, thematic maps are generated in order to identify the variability of properties of the soil, plants, and yield. However, first it is necessary to interpolate the data into a dense and regular grid to provide values for locations that were not sampled. This task is performed with the aid of interpolation methods.

Geostatistic analysis is the most used method of interpolation. The best geostatistical model for a series of georeferenced data is selected by comparing theoretical values with those obtained from sampling, and then analyzing the estimation errors and choosing the best model (Arlot and Celisse, 2010; Kohavi, 1995). This technique, called cross-validation, was chosen by Faraco et al. (2008) as the best way to evaluate the adjustment of spatial theoretical models, and was deemed better than Akaike's and Filiben's information criteria and the maximum logarithm value of the likelihood function.

Cross-validation allows for the evaluation of estimation errors by comparing predicted values with values determined through samples. The average error (AE) is calculated as the arithmetic average difference between the original values and those simulated by the interpolation, temporarily discarding any sample taken from the same location where the estimate is made by the interpolator (Isaaks and Srivastava, 1989). Other measures indicating the accuracy of the estimation are then calculated using the reduced average error ($\overline{RE}$), standard deviation of the average error (S_{AE}), and standard deviation of the reduced error (S_{RE}) (Cressie, 1993; McBratney and Webster, 1986).

According to non-tendentiousness criteria, to choose the best-adjusted model, the values for AE and $\overline{RE}$ should be as close to zero as possible, the value of S_{AE} should be as small as possible, and the value of S_{RE} should be close to 1 (Cressie, 1993; McBratney and Webster, 1986). Because cross-validation makes it possible for

---

* Corresponding author at: Rua Universitária, 1619, Jd. Universitário, Cascavel, Paraná CEP: 85819-110, Brazil.

E-mail address: eduardo.souza@unioeste.br (E.G. de Souza).

http://dx.doi.org/10.1016/j.compag.2017.03.008

ambiguous situations to occur, Souza et al. (2016) proposed an estimation called the error comparison index (ECI), which uses $\overline{\mathrm{RE}}$ and S_{RE}. The best semivariogram model is the one with the lowest ECI.

When the spatial dependence is moderate (Cambardella et al., 1994) or worse, the use of a deterministic interpolator can be more appropriate. The ECI methodology does not include methods that involve deterministic interpolators for inverse distance weighting, in which it is impossible to determine the standard deviation to calculate $\overline{\mathrm{RE}}$ and S_{RE}.

In this context, the aim of the present study is to propose an interpolation selection index (ISI) that advances the research on tools that can assess the actions of deterministic and stochastic interpolators. Using case studies, this study selects the best of four interpolation methods: inverse distance, inverse distance squared, ordinary kriging, and cokriging.

2. Materials and methods

Independent of the interpolation method to be used, the number of samples in the data collection must be chosen carefully. Many studies related to sampling density have been performed (Journel and Huijbregts, 1978; Wollenhaupt and Wolkowski, 1994; Doerge, 2000; Franzen et al., 2002; Ferguson and Hergert, 2009; Demattê et al., 2014). All of these studies suggest a minimum density from 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 sample ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000), but with at least eight samples at each sampling point. Studies such as Kerry et al. (2010) show that a sampling density of 2.5 sample ha^{-1} is widely used in many countries and is considered viable for the farmer. Forty points were selected to satisfy the recommendation of 2.5 sample ha^{-1} (Journel and Huijbregts, 1978) and the restriction of 30 pairs of data in the semivariogram definition (Isaaks and Srivastava, 1989). The sampling points were located along an imaginary line among intermediate contour lines with distances of 65 and 130 m (Fig. 1). These alternated distances provided a better fit at the smallest lag distances, which are important in kriging.

Data on clay, copper, and manganese content were used, along with the apparent electrical conductivity of the soil (AECS). The AECS measurements were performed one every two seconds, as they move through the area over the points of collection of other attributes. The study was conducted within a 15.5 ha area in the municipality of Céu Azul, Paraná, Southern Brazil, with the geometric center at coordinates 25°06′33″S; 53°49′55″W, and an average altitude of 660 m, using no-till. At this particular site, a succession of different crops was cultivated for more than a 10-year period. Soybeans and corn were grown during the summer harvest or off-season period, and wheat or oats were used as a cover crop during the winter. The soil of the area was classified as typical distroferric Red Latosol, according to Embrapa (2006), with a clay content ranging between 600 and 740 g kg^{-1} of soil.

Soil samples (from 0 to 0.20 m deep) were taken during the month of October in 2011, 2012, and 2013. The soil water content was approximately 25%, calculated according to Embrapa's methodology (Embrapa, 1997). Around each sampling point (located using a GPS Trimble GeoExplorer XT-2005) and using a 3-m radius, eight subsamples were randomly collected, two per quadrant, within a symmetrical circle divided into four quadrants (adapted from Wollenhaupt et al., 1994).

In October of 2011, the measurements of the AECS were obtained using a Geonics EM38-MK2 conductivimeter (Geonics Ltd., 2009) close to the ground, with measurements taken under

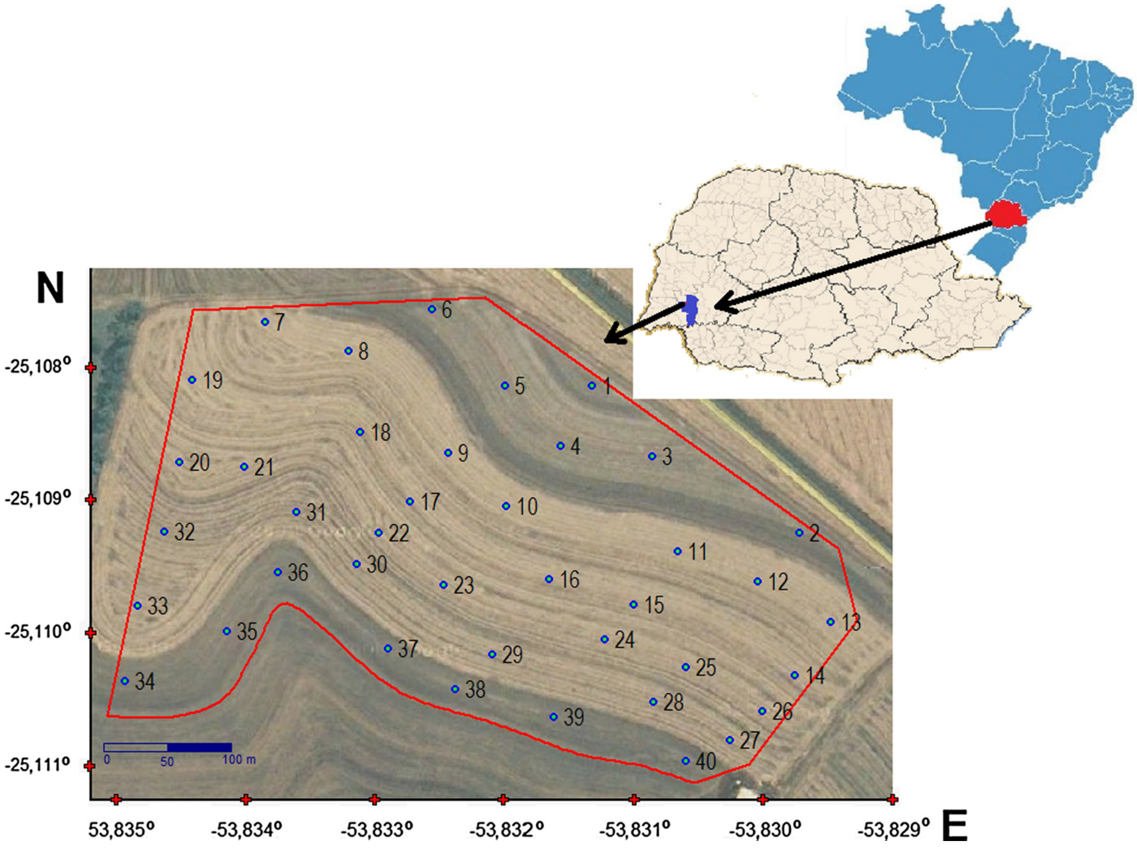

Fig. 1. Location of experiment and sampling points in experimental field in the municipality of Céu Azul, Southern Brazil. Red contour delineates the area used. Coordinates are in degrees (WGS 1984). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)