Computers and Electronics in Agriculture 146 (2018) 1–11

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Rectification methods for optimization of management zones

Nelson Miguel Betzek[a,b,*], Eduardo Godoy de Souza[b], Claudio Leones Bazzi[a], Kelyn Schenatto[c], Alan Gavioli[a]

[a] Department of Computer Science, Federal University of Technology of Paraná (UTFPR), Medianeira, Paraná, Brazil
[b] PGEAGRI, Technological and Exact Sciences Center, Western Paraná State University (UNIOESTE), Cascavel, Paraná, Brazil
[c] Department of Computer Science, Federal University of Technology of Paraná (UTFPR), Santa Helena, Paraná, Brazil

ARTICLE INFO

Keywords:
Interpolation
Precision agriculture
Smoothing
Thematic maps

ABSTRACT

The use of management zones (MZs) is an approach to precision agriculture that considers spatially contiguous subregions of the field, within which effects on the crop due to differences in soil, topography, and other abiotic factors are expected to be nearly uniform. Delimiting regions within the fields with similar yield potential and yield-limiting factors can lead for field management optimization. Regardless of the method used to delimit these zones, patches or isolated pixels generally appear. To smooth the MZs and improve their contiguity, a computational rectification function was implemented, allowing the analysis of 8 (3 × 3 mask) or 24 (5 × 5 mask) neighboring pixels using the statistical median and mode, to evaluate whether each pixel in the map should be reassigned to a different MZ. After being interpolated and normalized, sample data from three experimental fields were used to create clusters through fuzzy c-means algorithm, generating maps with two, three, four, and five classes. Then, the rectification function was applied five times on each map, which eliminated isolated pixels and virtually all patches, smoothing the boundaries between classes. The smoothness index showed higher variation in the first rectification as well as with an increase in the number of classes. The best performance was obtained with the 5 × 5 mask regardless of the statistical method used (median or mode). Our results show that these techniques are an effective way to increase the contiguity and smoothness of MZs, thereby improving their effectiveness, and are suitable for application in precision agriculture.

1. Introduction

The variability of nutrients in the soil directly affects crop yield; it may be related to factors such as climate, topography, organic matter, vegetation, geological processes, and soil management practices (Mallarino and Wittry, 2004). This influence occurs on different scales, complicating soil management and reducing the effectiveness of fertilizers if applied on a uniform scale (Mohammadi, 2002). As a result, different fertilization requirements may be needed for the same area. Since crop yield is influenced by soil characteristics, the study and understanding of the sources of variation is paramount to identify appropriate site-specific management (Rodrigues et al., 2016; Mzuku et al., 2005).

Spatial variability in crops is determined through monitoring and measurements, making it possible to create a plan for the correction of any deficiencies, particularly when site-specific management is intended, in order to improve soil quality, and, consequently, increase production (Davidson, 2014; Mzuku et al., 2005). Timlin et al. (1998) showed that topography and related factors such as soil depth and organic matter have a large effect on the variation of crop yields and can be used to identify management zones.

Normally, soil samples are analyzed to determine the levels of nutrients in the soil. The sampling should be dense enough to allow the determination of the variability of nutrients in the soil so that fertilizers can be used profitably and in an environmentally sustainable manner (Ferguson and Hergert, 2009; Franzen et al., 2002).

A management zone (MZ) is a subregion of a field that expresses a functionally homogeneous combination of yield-limiting factors for which a uniform rate of a specific crop input is appropriate (Bobryk et al., 2016; Doerge, 2000; Moral et al., 2010; Moshia et al., 2014). After delineation of the MZs, the number of samples needed to delineate the field soil variability can be reduced to one composite sample per zone, a technique suggested by Wollenhaupt et al. (1994) whereby subsamples are collected around georeferenced points, ensuring

Abbreviations: CV, coefficient of variation; FPI, fuzziness performance index; IDW, inverse distance weighting; KRI, kriging; MPE, modified partition entropy index; MZ, management zone; PA, precision agriculture; SDUM, software for definition of management zones; SI, smoothness index; SPR, soil penetration resistance; VR, variance reduction index

* Corresponding author at: Av. Soledade, 1090, Medianeira, Paraná CEP: 85884-000, Brazil.
E-mail addresses: nmbetzek@utfpr.edu.br (N.M. Betzek), eduardo.souza@unioeste.br (E.G.d. Souza), bazzi@utfpr.edu.br (C.L. Bazzi), kschenatto@utfpr.edu.br (K. Schenatto), alan@utfpr.edu.br (A. Gavioli).

https://doi.org/10.1016/j.compag.2018.01.014
Received 25 January 2017; Received in revised form 12 January 2018; Accepted 15 January 2018

superior evaluation of nutrients in the area. This approach to sampling is likely to reduce laboratory costs while maintaining the level of reliability (Ferguson and Hergert, 2009; Mallarino and Wittry, 2004), and it has been shown to improve the efficiency of nutrient use, maintaining or increasing the yield and potentially reducing the overloading of nutrients into the environment (Khosla et al., 2002; Moshia et al., 2014). Many studies related to sampling density have been performed (Journel and Huijbregts, 1978; Demattê et al., 2014; Wollenhaupt et al., 1994; Franzen et al., 2002; Ferguson and Hergert, 2009; Doerge, 2000), resulting in a suggested minimum density of 1 sample ha^{-1} (Ferguson and Hergert, 2009) to 2.5 sample ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000), which should be composed of at least eight individual samples (Wollenhaupt et al. (1994)). It is reported in the literature that the use of MZs is commercially viable, but there is a need for further research to improve delineation techniques (Nawar et al., 2017).

Several kinds of data can be used to delineate MZs; however, it is advantageous to use a set of multivariate data attributes that do not vary significantly over time (topography, electrical conductivity, physical properties of the soil) and that are correlated with crop yield, thus producing more stable MZs (Buttafuoco et al., 2010; Doerge, 2000). Some researchers, such as Taylor et al. (2007), Guastaferro et al. (2010), and Nawar et al. (2017), have also included crop yield as a variable in the delineation of MZs using clustering methods. According to Kitchen et al. (2005), the identification of areas of similar productivity potential called "productivity zones" may be useful when some key management decisions depend on reliable estimates of expected yield, such as the application of N fertilizer rate or seedling rate. In precision agriculture (PA), the terms "management zone" and "management class" are often used as synonyms. However, a management class is the area to which a particular treatment may be applied, whereas a management zone is a spatially contiguous area to which a particular treatment may be applied. Thus, a management class may consist of numerous zones, whereas a management zone can correspond to only one management class (Taylor et al., 2007).

Several approaches have been developed to delineate MZs; these are often classified as empirical or clustering according to the technique used (Córdoba et al., 2016; Guastaferro et al., 2010). The clustering approach is the most typical, and it has been used to obtain satisfactory results by many researchers (Arno et al., 2011; Moral et al., 2010; Saleh and Belal, 2014; Tagarakis et al., 2013). However, regardless of the procedure used to generate the MZs, small patches generally arise within a class. In order to resolve this issue, Lowrance (2014) created the EZZone software package, which smooths zones by merging small polygons having an area smaller than a certain threshold with larger polygons of different zones. Other researchers, such as Pramanik et al. (2013), have proposed that fields be merged by assigning weights to pixels, that is, by analyzing the asymmetry of neighboring pixels. To eliminate isolated cells or patches, Xiang et al. (2007) used post-classification majority filtering (mode). According to Lark (1998), Ping and Dobermann (2003), and Córdoba et al. (2016), the use of spatial filters applied to the results of such classification is also recommended to improve zone contiguity. Nonetheless, none of these studies compared the mode and median statistics or used the smoothness index or Cohen's kappa coefficient to evaluate the quality of the smoothing process.

The objective of the present study was to develop and apply computational techniques based on mode and median statistics to rectify and thus smooth MZs. The expectation was that this would eliminate the patches, making the zones more contiguous and thus more viable operationally.

2. Materials and methods

2.1. Datasets

Data collection was based on an irregular sampling grid of three experimental fields located in rural areas of the state of Paraná, Brazil (Fig. 1) (Field A: 25°24′28″S, 54°00′17″ W; Field B: 25°26′49″S, 54°04′59″W; Field C: 25°06′32″S, 53°49′55″W). These fields have been cultivated under a no-tillage system for over 10 years, rotating soybean (*Glycine* max L.) with corn (*Zea mays* L.). Fields A, B, and C measured 9.9 ha, 19.8 ha, and 15.5 ha, respectively. They included 42 (4.2 sample ha^{-1}), 58 (2.9 sample ha^{-1}), and 40 (2.6 sample ha^{-1}) sampling points, respectively, set by means of irregular grids. This density of sampling points is sufficient to identify the variability of the attributes since it is greater than 2.5 sample ha^{-1} (Journel and Huijbregts, 1978; Doerge, 2000). Only stable attributes, that is, those recommended for studying the delineation of MZs (Doerge, 2000), were collected and analyzed (Table 1).

Locations of the sampling points were obtained by Global Navigation Satellite System (GNSS) receiver (Juno SB, Trimble Navigation Limited, Sunnyvale, CA, USA), and the elevations were obtained using a total station (GPT-7505, Topcon Corporation, Tokyo, Japan). Soil penetration resistance (SPR) measurements were taken around each point delineated on the sampling grid, using an electronic penetrometer (PenetroLOG, Falker, Porto Alegre, Brazil). The means of the measurements were subsequently calculated to represent the sampling value average at depths of 0–0.1, 0.1–0.2, and 0.2–0.3 m. At the same locations, eight subsamples of soil were collected at a depth of 0–0.2 m within a radius of 3 m from the point determined on the grid (adapted from Wollenhaupt et al., 1994). Subsequently, the samples were forwarded for laboratory analysis and to obtain data on soil texture (clay, silt, and sand).

Soybean and corn yield were determined at the same points at which the soil samples were taken, the harvest and threshing of which occurred manually in an area of 0.9 m^2. Subsequently, the yield values were calculated and were converted to a 13% moisture content. The yield data were normalized by amplitude range (Eq. (1)) with the objective of removing seasonal and crop variability, transforming the value obtained at each sampling point in each of the five crop cycles (P_{ij}) used in this study into a single normalized value ($P_{ij\_\mathrm{Amplitude}}$).

$$P_{ij\,\mathrm{Amplitude}} = \frac{P_{ij}-\widetilde{x}_j}{A_j} \tag{1}$$

where $P_{ij\_\mathrm{Amplitude}}$ is the yield normalized by the amplitude at point *i* in year *j*, P_{ij} is the yield at point *i* in year *j*, $\widetilde{x}_j$ is the median yield in year *j*, and A_j is the amplitude range of the sample value in year *j*.

2.2. Variable selection

To assess the spatial correlations among the analyzed attributes, Moran's bivariate spatial autocorrelation statistic (Czaplewski and Reich, 1993) was used, creating the spatial correlation matrix, which allows the determination of those attributes that influence yield positively or negatively and whether each variable is correlated spatially (spatial autocorrelation). Then, the variables to be used in the clustering process were chosen by using the approach proposed by Bazzi et al. (2013): (1) elimination of variables with non-significant spatial dependence at a 5% probability level, (2) elimination of variables having no correlation with the dependent variable, and (3) elimination of redundant variables (those that are correlated with each other), giving preference to the maintenance of variables having a higher correlation with the dependent variable.

2.3. Interpolation of the selected variables

Inverse distance weighting (IDW) and kriging (KRI) are the interpolation methods most commonly used in PA. They are differentiated by the way in which the weights are attributed to the sample values, which may influence the estimated values (Reza et al., 2010; Souza et al., 2016) as well as the smoothness of the delineated MZs. However, regardless of the interpolation method, there could be many isolated