Computers and Electronics in Agriculture 121 (2016) 74–83

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

An improved method of delineating rectangular management zones using a semivariogram-based technique

Zhang Xiaohu [a], Jiang Li [a], Qiu Xiaolei [a], Qiu Jianxiu [b], Wang Juan [c], Zhu Yan [a,*]

[a] National Engineering and Technology Center for Information Agriculture, Jiangsu Key Laboratory for Information Agriculture, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, 1 Weigang Road, Nanjing, Jiangsu 210095, PR China

[b] Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, PR China

[c] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form 11 November 2015
Accepted 15 November 2015
Available online 24 December 2015

Keywords:
Management zones
Semivariogram
Kriging interpolation
Binary integer linear programming

ABSTRACT

Management zone delineation is of great importance for precision agriculture applications. Compared with oval management zones, rectangular management zones are more practical for variable rate technology and fertilization machinery, and they are also easy to use for farmers in developing regions. This research proposes an improved method for delineating rectangular management zones, which applies a semivariogram analysis to interpolating grid data with an optimal grid size. By using a well-designed grid size, optimal grids are considered when generating instances and solving binary integer linear programming (BILP) problems. These improvements greatly reduce the computational time as well as the total variance of the delineated management zones. The experimental results indicate that the proposed method provides a practical method of applying rectangular management zone delineation that performs better and is more efficient compared with conventional algorithms.

1. Introduction

Management zone analyses (MZAs) provide approaches for segmenting a field into subfield regions with respect to the within-field variability (Doerge, 1999a). These subfield regions (management zones) typically represent areas of the field that are similar in terms of environmental parameters, and they denote the basic datasets for achieving precision agriculture, i.e., site-specific management. MZA is a spatial data processing technique that relies on spatially referenced data and spatial analysis tools. Based on the study of Fridgen et al. (2004), successful implementation of an MZA should address the following three crucial questions: (1) What parameters should be considered during the implementation of the MZA? (2) What is the optimal technique for segmenting a field into harmonious management zones? (3) How many sub-regions are suitable for a specific field?

The majority of previous investigations approached these three questions through theoretical analyses and experimental verification. With respect to the first question, two main answers arise. As mentioned by Doerge (1999b), the most useful and meaningful parameters are factors that have a direct effect on crop yield, such as soil pH, soil moisture, and soil nutrient levels. In addition, researchers have documented that crop yield patterns are also effective for determining field segmentation (Diker et al., 2004; Pedroso et al., 2010). However, researchers and end users prefer to integrate these two parameter groups for more accurate management zone delineation. Because the main concept underlying an MZA is to categorize a field into several groups, various clustering algorithms and approaches are applied in MZAs. Most MZAs are performed by an unsupervised cluster procedure, such as the C-means or Fuzzy C-means (FCM) methods (Ortega et al., 2002; Franzen et al., 2003), and the number of sub-regions is determined by an index describing the performance of the cluster algorithms. Because of the characteristics of unsupervised algorithms, the management zones produced by this method are always fragmental with oval shapes (Frogbrook and Oliver, 2007). In practice, however, the traverse pattern of the application equipment in the field should be considered when delineating the management zones (Kvien and Pocknee, 2000). To address this issue, Cid-Garcia and his colleagues developed a new framework for MZAs that included a rectangular-shaped management zone delineation using binary integer linear programming (BILPMZ) (Cid-Garcia et al., 2013). With BILPMZ, the original field is made rectangular and then segmented into zones with maximum homogeneity. The BIMPMZ consists of two critical stages: generating instance and solving BILP (Fig. 1).

* Corresponding author. Tel.: +86 25 84396598.
E-mail address: yanzhu@njau.edu.cn (Y. Zhu).

http://dx.doi.org/10.1016/j.compag.2015.11.016
0168-1699/

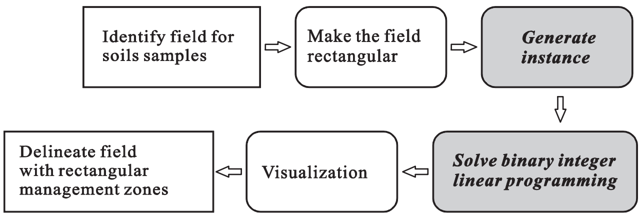

Fig. 1. Diagram of the BILPMZ method. BILPMZ consists of two critical stages: generating instance and solving binary integer linear programming.

The rectangular shape assists farmers by reducing the difficulties of adopting variable technologies as well as by facilitating the use of agriculture machinery. Moreover, management zones with rectangular shapes are more applicable for farming in underdeveloped areas because farmers can easily apply these management zones to reduce fertilizer input, labor costs, and environment waste without using advanced agriculture machinery. However, during the process of our experimental application of the BILPMZ method, two problems arose. First, in most situations, sufficient sample sites are not available for instance generation. Second, different instances with different grid sizes lead to different management zones; therefore, the optimal grid size is unknown.

To resolve these two problems, we modified the BILPMZ method using optimal instances generation based on a semivariogram analysis. The remainder of this paper is organized as follows. In Section 2, we introduce our approaches to improving the BILPMZ method and resolving the problems mentioned above, and in Section 3, we present the evaluation results of the proposed method. Finally, in Section 4, we present the conclusions.

2. Methodology

BILPMZ has two main steps (Cid-Garcia et al., 2013): the first is to generate the rectangular instances, which are the potential zones for management zones; and the second is to identify a feasible solution to the BILP problem. The proposed method in this work improves upon the conventional method because we apply a semivariogram analysis to determine the optimal grid size during instance generation.

2.1. Semivariogram-based grid size selection

A semivariogram, $\gamma(\vec{h})$ is a function describing the degree of spatial dependence of a spatial field $Z(\vec{x})$ in spatial statistics (Cressie, 1993). According to the spatial stationary hypothesis, the semivariogram only depends on the separation vector, lag $\vec{h}$, not on the location, $\vec{x}$. When referring to an isotropic semivariogram, $\gamma(h)$ denotes the semivariogram at a given distance $h = \|\vec{h}\|$ and is defined as follows:

$$\gamma(h) = \frac{1}{2} Var(Z(x) - Z(x+h)) \tag{1}$$

where $N(h)$ is the total number of pairs of the field sample value $Z(x)$ and $Z(x+h)$. A well-regularized semivariogram has 3 parameters (Fig. 2). In theory, the nugget refers to the variability of $Z(x)$ that cannot be explained by the distance between the samples. The nugget is attributed to measurement errors or spatial sources of variation at distances smaller than the sampling interval or both. The sill refers to the maximum observed variability of $Z(x)$. The sill corresponds to the variance of the data as normally estimated in statistics. The difference between the sill and the nugget represents the amount of observed variation that can be explained by the distance between observations. The range is the point at which the semivariogram stops increasing. If h is within the range, then spatial autocorrelation occurs among the samples; otherwise, the spatial autocorrelation is not obvious. Hence, the range of the semivariogram is the maximum distance of two sample locations that are spatially autocorrelated and indicates the maximum reasonable grid size. This strategy has been applied in many studies with similar objectives (Hengl, 2006; Ming et al., 2012; Mohammadi, 2013; Bottega et al., 2014). Moreover, selecting larger grid sizes reduces the computational time required for BILP. Hence, to generate grid data with homogeneous soil properties and reduce the computational cost of solving BILP, we should select the range of semivariograms for soil properties as the grid size. However, most of the study field is not dividable by the range 'a'; therefore, the extent of the study field should be resized. The new extent is the parameter of interpolation, which generates the initialized grid data of the study field. The new extent can be determined by the following formulae:

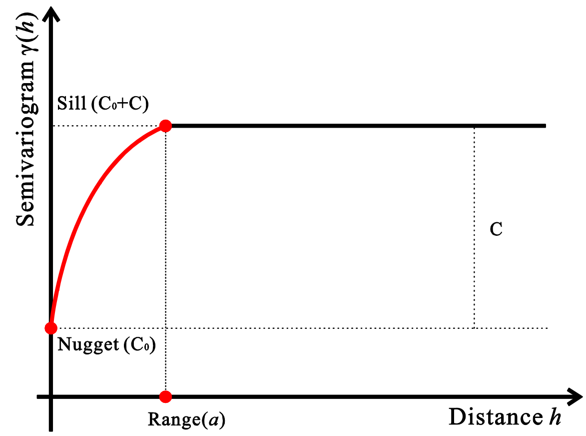

Fig. 2. Regularized semivariogram model.

$$WidthC_{new} = a\left(Fix\left(\frac{WidthS}{a}\right) + 1\right) \tag{2}$$

$$LengthC_{new} = a\left(Fix\left(\frac{LengthS}{a}\right) + 1\right) \tag{3}$$

where $WidthC_{new}$ and $LengthC_{new}$ are the width and length of the new extent, $WidthS$ and $LengthS$ are the width and length of the study field, $Fix(\cdot)$ is a function that returns the integer portion of a number, and a is the range of the semivariogram.

2.2. Optimal instance generation

Instances are potential zones that are rectangular and homogeneous with respect to the properties of the soil. The objective of instance generation is to calculate all of the potential zones and their variances for the soil properties. The process includes two main stages.

(a) *Interpolation.* In this stage, we propose using the Kriging method to calculate the properties of all grids. With the hypothesis of spatial stationarity, the Kriging method gives the best linear unbiased prediction of the intermediate values. The value $Z(x_0)$ at location x_0 can be predicted as follows:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{4}$$