BIOSYSTEMS ENGINEERING 143 (2016) 95–107

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Protocol for multivariate homogeneous zone delineation in precision agriculture

Mariano A. Córdoba [a], Cecilia I. Bruno [a], José L. Costa [b], Nahuel R. Peralta [b], Mónica G. Balzarini [a,*]

[a] Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ing Agr. Felix Aldo Marrone 746, Ciudad Universitaria, C.C. 509, 5000, Córdoba, Argentina

[b] Instituto Nacional de Tecnología Agropecuaria (INTA), Estación Experimental Balcarce and Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata, C.C.276, 7620, Balcarce, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 19 August 2015
Received in revised form
9 December 2015
Accepted 12 December 2015
Published online 6 February 2016

Keywords:
Spatial data
Spatial autocorrelation
Multivariate classification
Site-specific management
Software specifications

Uniform management of agricultural fields has been increasingly replaced with environmentally based management, which is benefited by the identification of homogeneous zones within crop fields. Such zones are based on the classification of field sites into groups of homogeneous features. Multiple causative agents of variability and the response of agricultural crops should be considered for zoning. Several correlated variables are usually measured and georeferenced for this purpose at multiple sites within the field. This paper presents an approach to promoting the integration of different statistical tools for identifying homogeneous zones based on site covariates. The methodological innovation of this work involves cleaning and re-scaling of spatial data, as well as multivariate and geostatistical analyses in a logical sequence (protocol). Statistical topics for further improvement and protocol applications are noted. The analytical process has been illustrated using a rain-fed wheat crop (60 ha) from the Argentine Pampas, with apparent electrical conductivity, elevation and soil depth as master variables for zoning, and yield, soil organic matter and clay to validate the created management zones; however, it may be applied to other production systems using georeferenced data. The R scripts and the sample file to run the proposed protocol are provided as electronic supplementary material.

1. Introduction

Farm machinery equipped with new technologies provides the opportunity for a more accurate measurement of spatial soil and terrain variability of crop fields. The recorded spatial data is used for multivariate classification of field sites into groups of homogeneous features. The classification is then used to delimit contiguous zones within the field aimed at site-specific management in precision agriculture (PA) (Yao et al., 2014). Management zones (MZs) are usually areas with similar characteristics, such as texture, topography, water status and soil nutrient levels (Moral, Terrón, & Marques Da

* Corresponding author. Tel.: +54 351 4334103x219; fax: +54 351 4334118x114.
E-mail address: mbalzari@agro.unc.edu.ar (M.G. Balzarini).

http://dx.doi.org/10.1016/j.biosystemseng.2015.12.008

Nomenclature

AIC	Akaike information criterion
ECa	Apparent electrical conductivity (mS m^{-1})
ECa30	Apparent electrical conductivity at 30 cm depth (mS m^{-1})
ECa90	Apparent electrical conductivity at 90 cm depth (mS m^{-1})
KM-sPC	Fuzzy k-means clustering from spatial PCA
Elev	Elevation (m)
Sd	Soil depth (m)
MZ	Management zone
MLM	Mixed Linear Model
I_i	Moran's local index
PCA	Principal component analysis
PA	Precision agriculture
sPCA	Spatial principal components analysis
SOM	Soil organic matter (%)

Silva, 2010). Physical and chemical soil properties are the most widely reported variables used for zoning, followed by landscape attributes (Khosla, Westfall, Reich, Mahal, & Gangloff, 2010). In environments in which water and nutrient availability are crop limiting factors, production largely depends on soil type, specifically, on its capacity to retain water and nutrients. In the last years, besides the introduction of yield monitors, there has been an increase in the use of proximal sensors that capture spatial data on apparent electrical conductivity (ECa) as a soil attribute. The use of soil ECa has gained attention as a good surrogate method for detecting spatial variation in soil chemical and physical properties (Arno, Martinez-Casasnovas, Ribes-Dasi, & Rosell, 2011; Corwin & Lesch, 2010; Moral et al., 2010; Peralta, Costa, Balzarini, & Angelini, 2013; Rodríguez-Pérez, Plant, Lambert, & Smart, 2011; Taylor, McBratney, & Whelan, 2007). Data captured by ECa sensors are georeferenced using high-precision GPS. Elevation is another topographic characteristic that influences water movement and soil development within the field, generating yield spatial variability. Effective soil depth is also a useful soil characteristic to delimitate MZs (Peralta, Costa, Balzarini, & Castro Franco, 2013) because this variable affects the water storage capacity and its spatial distribution, generating yield variability.

The hypothesis that the productivity gap of crops is influenced by the interaction of soil and landform characteristics has been present in several works. Pennock (2003) and Kravchenko, Robertson, Thelen, and Harwood (2005) related topography to productivity. Siqueira, Marques, and Pereira (2010) and Sanchez, Marques Jr., Siqueira, Camargo, and Pereira (2013) used multivariate analysis and geostatistics to analyse the potential use of site covariates to predict the variability of important agronomic traits. The combined application of multivariate and geostatistical analysis together has already been proposed, but using different programs and non-spatially restricted multivariate methods. In this work we propose a logical sequence of statistical analyses that can be implemented using a protocol. The protocol application is illustrated using data on ECa, elevation and effective soil depth of a 60-ha wheat field cultivated under PA in the southern Argentine Pampas as master variables for zoning. Agricultural fields in the south-eastern Pampas frequently have multiple soil map units within them, despite their sometimes relatively small size, and wide range of soil textures and properties, causing high soil spatial variability (Peralta, Costa, Balzarini, & Angelini, 2013). The adoption of ECa sensors in Argentina is increasing because the ECa signal provides an integration of several effects at a site. In the southern Argentine Pampas ECa measurements successfully delimited homogeneous soil zones associated with spatial distribution of clay, soil moisture, CEC, SOM content and pH (Peralta, Costa, Balzarini, & Angelini, 2013).

From the analytical point of view, Taylor et al. (2007) set the basis for a protocol to delineate management zones. However, as the computing capability increases, new statistics can be used to handle multivariate spatial data. New statistical techniques are available for handling spatial data at the clean-up step, in the multivariate classification, and in predictive and validation steps.

A first step in any quantitative protocol is the removal of artefacts within data before analysis. Local spatial autocorrelation indices which are particularly useful for data debugging, such as Moran local index (Anselin, 1995) and Getis-Ord index (Getis & Ord, 1992), can integrate analytic protocols given their current availability in free software. On the other hand, the availability of several types of variables requires combining data from different sources of information and usually of different spatial resolutions. Several spatial interpolation techniques (Oliver & Webster, 2014) can be used to perform a re-scaling of original measurements to associate data from the different variables from each site onto a common grid. The issue of different scales (grid sizes) for different variables should be taken into account to proceed with multivariate classification. The spatial correlation structure for each variable is a first guide to follow, but the spatial covariance of variables is also meaningful. Other practical aspects as the minimum area needed for crop management are usual constraints. In choosing the grid size, there is a trade-off between maintaining spatial precision by selecting a fine grid and reducing noise and making the data more manageable by selecting a coarser grid (Long, 1998). Since variability may be studied at any spatial scale, the choice of grid size depends on the aims of the investigation. In making this choice, the investigator is aided by knowledge of how much variability is lost for each variable in moving from one scale to a larger one (Roel & Plant, 2004). Grid sizes of 10- to 50-m range are common for mapping in site-specific agriculture (Peralta et al., 2015; Ping & Dobermann, 2003; Roel & Plant, 2004), and this is because 108 variation in soil properties appears to occur at a much near scale than the 1 ha strata (McBratney & Pringle, 1999). Choosing a coarse resolution (>16 m) for spatial interpolation may result in biased aggregated data sets (Ping & Dobermann, 2003). In practice, the selection of an appropriate sampling cell size requires understanding of the relationships between grid size, yield variability accounted for, and the resulted spatial map fragmentation.

After re-gridding measurements for all variables the next step, in a protocol of analysis, consists of multivariate site classification. Several clustering algorithms (Anderberg, 1973),