COMPARISON OF MANAGEMENT ZONES GENERATED BY THE K-MEANS AND FUZZY C-MEANS METHODS

C. L. Bazzi

Department of Computer Science
Federal Technological University of Paraná
Medianeira, Paraná, Brazil

E. G. Souza, K. Schenatto, F. Rodrigues, D. M. Rocha

Technological and Exact Sciences Center
Western Paraná State University
Cascavel, Paraná, Brazil

ABSTRACT

The clustering methods are highly suggested for the definition of MZ (TAYLOR et al., 2007; YAN et al., 2007). The most used clustering methods for defining management units correspond to the K-Means (FRIDGEN, et al., 2004; RIBEIRO, et al., 2011) and the Fuzzy C-Means algorithm (STAFFORD et al., 1998; JAYMES et al., 2003; PING & DOBERMANN, 2003; YAN et al., 2007). This study aimed to evaluate if there is an output difference between the clustering methods K-Means and Fuzzy C-Means. It has been done the selection of layers according Bazzi et al. (2013) and the evaluation was performed with the MZs mean comparison tests (ANOVA) and variance reduction. To evaluate the difference between MZs, Kappa and Tau maps indices comparison were used. The layer altitude was selected as best option, and 2, 3, 4 and 5 MZs were generated with both clustering methods. Comparing the thematic maps generated that represent the MZs, it was found that for two MZs, the divisions may be classified as excellent agreement (LANDIS & KOCH (1977, p.165) as well as the division into 3 classes. For the division into 4 classes, the agreement degree can be considered as substantial and moderate for the division into 5 classes. It could be concluded that although there are differences between the methods used to generate MZs, the results were the same for both methods. It was found that the division is valid for both as to set different levels of yield in field and as to perform division of the field to use as a source of recommendation and analysis.

Keywords: Precision agriculture, cross-correlation, Kappa index, Tau index.

INTRODUCTION

A research to define management zones (MZ) aims at dividing producing areas in smaller MZ and should be treated differently, serving as a source of recommendation and analysis. The most commonly used clustering methods to define MZ correspond to the K-means algorithm (RIBEIRO et al., 2011; ORTEGA & SANTIBÁÑEZ, 2007) and Fuzzy C-Means (YAN et al., 2007).

The Fuzzy C-Means method is more difficult that K-Means and spend more time to execution. The aim this work was evaluated the differences about these two methods of clustering. It were used the kappa and Tau indexes to evaluated the differences between maps and Anova and variance reduction to evaluated if the divisions showed different potential of production.

MATERIAL AND METHODS

The sample data were collected in 2010 on an area of 20 ha, in Cascavel/PR - Brazil, with geographic coordinates 24º 57' 08" S; 53º 33' 59" O. data of soybean Yield, chemical and physical attributes in 87 points were sampled. To select the attributes to generated the MZ the technique of Bazzi et al (2013) was used, and the software SDMZ (Software of Definition to Management Zone) was used to interpolate the samples and generate and evaluated the MZ. Two techniques (ANOVA and reduction of variance) were used to evaluate the MZ. The Kappa and Tau Index were used to compare the MZ generated with K-Means and Fuzzy C-Means methods. .

RESULTS AND DISCUSSION

The layer elevation was selected as the best to generate the MZ. Four maps (with 2, 3, 4 and 5 regions) were created to represented de MZ to K-Means method and other four maps to Fuzzy C-Means method (Figure 1).

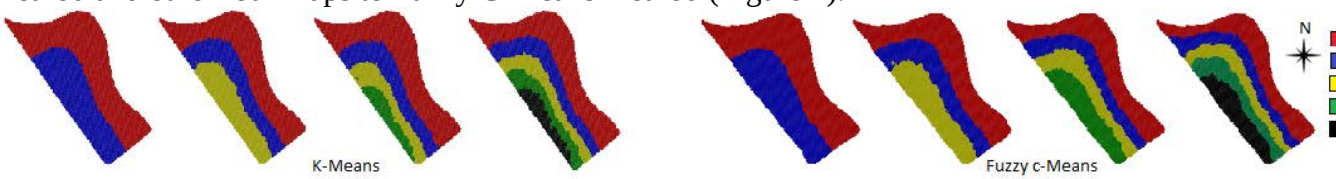

Figure 1. Management zones maps of K-Mens and Fuzzy C-Means Algorithm.

The Kappa and TAU index showed that the concordance between maps were lower when was increased the number of MZ. To division in two zones, the division was virtually the same (Table 1). The Anova and reduction of variance showed that the result of evaluated about Yield potential between zones was valid to division in two zones, independent of algorithm of clustering used. The same conclusions in relation that Yield potential were taken when the number of division increased, indicating that both algorithms showed the same results.

Table 1. Kappa and Tau index of comparation between maps

KAPPA				TAU			
2 zones	3 zones	4 zones	5 zones	2 zones	3 zones	4 zones	5 zones
1	0.97	0.77	0.43	1	0.97	0.78	0.45

CONCLUSION

Both algorithms showed similar results to evaluated zones with Anova and variance reduction, but the difference between maps were biggest when increased of the number of zones.

ACKNOWLEDGEMENTS

Araucária Foundation (Fundação Araucária), Technological Foundation Park from Itaipu and the National Council for Scientific and Technological Development (CNPq), Federal Technological University of Paraná, CAPES Foundation – Ministry of Education of Brazil.