THE INTERNATIONAL SOCIETY OF PRECISION AGRICULTURE PRESENTS THE
13th INTERNATIONAL CONFERENCE ON
PRECISION AGRICULTURE
July 31-August 4, 2016 • St. Louis, Missouri USA

DATA NORMALIZATION METHODS FOR DEFINITION OF MANAGEMENT ZONES

Kelyn Schenatto[1], Eduardo G. de Souza[2], Claudio L. Bazzi[3], Alan Gavioli[3], Nelson M. Betzek[3], Humberto M. Beneduzzi[4]

[1]Computer Science Department, Federal Technological University of Paraná (UTFPR), Rua Cerejeira s/n, 85892-000, Santa Helena, Paraná, Brasil;

[2]Technological and Exact Sciences Center, State University of West Paraná (UNIOESTE), Rua Universitária 2069, 85819-110, Cascavel, Paraná, Brasil;

[3]Computer Science Department, Federal Technological University of Paraná (UTFPR), Av. Brasil 4232, 85884-000, Medianeira, Paraná, Brasil;

[4]Analysis and systems development Department, Federal Institute of Paraná, Av. Araucária 780, 85860-000, Foz do Iguaçu, Paraná, Brasil.

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract. *The use of management zones is considered a viable economic alternative for the management of crops due to low cost of adoption as well as economic and environmental benefits. The decision whether or not to normalize the attributes before the grouping process (independent of use) is a problem of methodology, because the attributes have different metric size units, and may influence the result of the clustering process. Thus, the aim of this study was to use a Fuzzy C-Means algorithm to evaluate the performance of various data normalization techniques used in the data clustering process for generating MZs. The tests were conducted in three experimental areas. A Fuzzy C-Means clustering method using Euclidean distance was used to define MZs in sub-regions two, three, and four of each field, respectively. The conclusion was that the use of normalization techniques is essential for defining MZs with a Fuzzy C-Means algorithm using the Euclidean distance when there is a need to use more than one attribute with different metric units.*

Keywords. *precision agriculture, management units, Fuzzy C-Means.*

Introduction

A management zone (MZ) is defined as a subregion of a field that exhibits similar combinations of yield-limiting factors (Tagarakis et al. 2013). This facilitates the application of precision agriculture (PA) techniques by reducing the costs of its adoption and implementation. Clustering methods are appropriate for dividing data into groups with homogeneous characteristics. Of these, the Fuzzy C-Means is used most often (Li et al. 2007; Zhang et al. 2013). A methodological problem is the decision whether to normalize data before the clustering process because the attributes have different metric size units (meters, degrees, t ha^{-1}, for example), and may influence the result of the clustering process (regardless of use). The goal of this study was to evaluate the performance of different normalization techniques used in the data clustering process by the Fuzzy C-Means algorithm to generate MZs.

Material and Methods

This research was conducted in three commercial agricultural fields: Field A was 15 ha, located in Céu Azul (PR, Brazil), Field B was 9.9 ha, located in Serranópolis do Iguaçu/PR and Field C was 19.8 ha, located in Cascavel/PR. Selected for data considered to be stable, and the yield of soybeans and corn was analyzed using spatial correlation statistics. The selected attributes were interpolated using ordinary kriging to create a 5 x 5 m grid with greater detail than the layers of sampling grids that were normalized using standard score (Larscheid and Blackmore 1996), amplitude (Mielke and Berry 2007), and averaging (Swindel 1997) methods. A Fuzzy C-Means clustering method using Euclidean distance was used to define MZs in sub-regions two, three, and four of each field, respectively. Using the statistical methods of MPE (Modified Partition Entropy), FPI (Fuzziness Performance Index), VR (Variance Reduction), SI (Smooth Index), and Kappa, the normalization methods influenced the clustering process when using more than one layer.

Results and Discussion

MEP and FPI showed diverse results for each set of normalized data, indicating the normalization process had a significant influence on the grouping results. For all fields, regardless of non-normalization or the normalization method, the indexes showed that grouping into two MZs was best, except for Field A, where it was indicated that three MZs were best when non-normalized data or data normalized by average were used. The normalization method interfered with the analyzed indexes (Figure 2; VR, FPI, and MPE) when there was more than one variable. The behavior of the FPI and MPE were quite similar, providing the same interpretation (except for Field B with non-normalized data). Normalization by amplitude was the method (except for slight inconsistency in Field C) in which the choice of the best number of zones coincided in the three methods. The smoothness index (Figure 2) decreased as the number of MZs increased indicating the smoothness of the contour curves decreased, complicating the visual interpretation and site-specific management of agricultural inputs. Perfect agreement among normalization methods occurred for Field C (MZs defined using a single variable). Standard score and amplitude methods behaved similarly in all cases, with slight advantages for amplitude. Additionally, mean and non-normalized methods behaved more poorly than did the others.

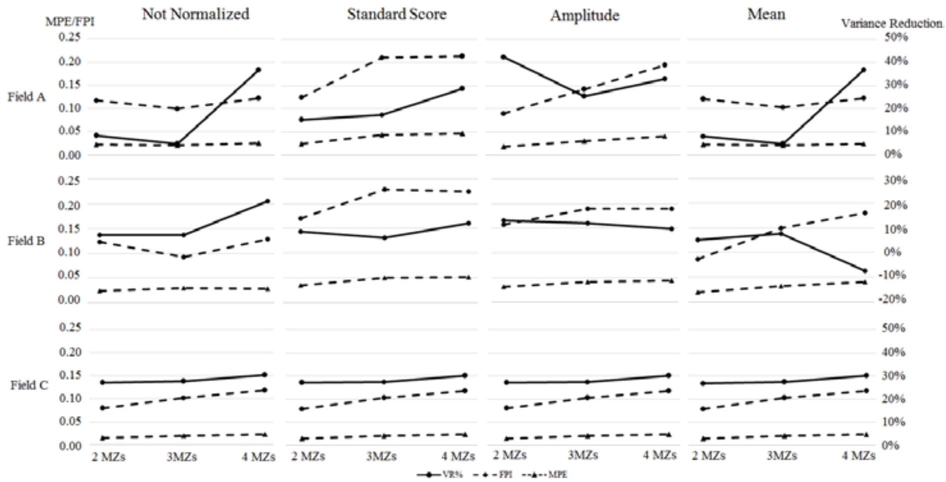

Fig. 1 Statistics of Variance Reduction (VR, %);Fuzziness Performance Index (FPI) and Modified Partition Entropy (MPE) obtained in the clustering of Fields A, B, and C.

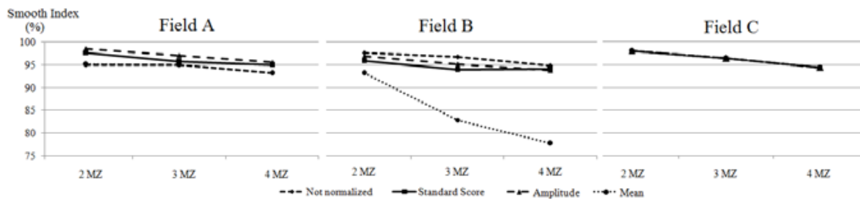

Fig. 2 Smooth Index calculated for Fields A, B, and C, as a function of normalization methods.

Conclusion

The use of normalization techniques is essential for defining MZs with a Fuzzy C-Means algorithm using the Euclidean distance when there is a need to use more than one attribute with different metric units. When using a single attribute, there is no need to use normalization methods against the input data for the algorithm. The amplitude method was considered the best normalization method.

References

Larscheid G., & Blackmore, B. S. (1996). Interactions between farm managers and information systems with respect to yield mapping. In 3rd International Conference on Precision Agriculture (pp.1153-1163). Minneapolis: Springer.

Li, Y., Shi, Z., Li, F.,& Li, H. (2007). Delineation of site specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture*, 56 (2),174-186.

Mielke, P. W., & Berry. K. J. (2007). Permutation methods: a distance function approach. New York: Springer.

Swindell, J. (1997). Mapping the spatial variability in the yield potential of arable land through GIS analysis of sequential yield maps. In 1st European Conference on Precision Agriculture (pp. 827-834). Warwick.

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., & Gemtos, T. A. (2013). Management zones delineation using fuzzy clustering techniques in grapevines. *Precision Agriculture*, 14 (1), 18-39.

Zhang, Z., Lü, X., Lv, N., Chen, J., Feng, B., Li, X. W., et al. (2013). Defining agricultural management zones using Gis techniques: Case study of Drip-irrigated cotton fields. *Information Technology Journal,* 12 (21), 6241-6246.