

RESEARCH NOTE

Classification of areas associated with soybean yield and agrometeorological variables through fuzzy clustering

Everton Coimbra de Araújo¹, Jerry A. Johann², Miguel A. Uribe-Opazo², and Eduardo C.G. Camargo³

¹Universidade Tecnológica Federal do Paraná/UTFPR, Campus Medianeira, PR, Brazil.

²Universidade Estadual do Oeste do Paraná/UNIOESTE, Campus Cascavel, Centro de Ciências Exatas e Tecnológicas/CCET, Programa de Pós-Graduação em Engenharia Agrícola/PGEAGRI, PR, Brazil.

³Instituto Nacional de Pesquisas Espaciais/INPE, São José dos Campos, SP, Brazil.

Abstract

E.C. Araújo, J.A. Johann, M.A. Uribe-Opazo, and E.C.G. Camargo. 2013. Classification of areas associated with soybean yield and agrometeorological variables through fuzzy clustering. Cien. Inv. Agr. 40(3): 617-627. This study aimed to apply an approach based on fuzzy clustering for the classification of areas associated with soybean yield combined with the following agrometeorological variables: rainfall, average air temperature and average global solar radiation. The study was conducted with 48 municipalities in the western region of Paraná State, Brazil, with data from the crop-year 2007/2008. Through the fuzzy c-means algorithm, it was possible to form groups of municipalities that were similar in soybean yield using the Method of Decision by the Higher Degree of Relevance (*MDMGP*) and Method of Decision by Threshold β (β *MDL*). Subsequently, the identification of the appropriate number of clusters was obtained using Modified Partition Entropy (*MPE*). To measure the degree of similarity for each cluster, the Cluster Similarity Index (*ISC_c*) was constructed and implemented. From the perspective of this study, the method used was adequate, allowing the identification of clusters of municipalities with degrees of similarities between 63 and 94%.

Key words: Agrometeorological variables, classification of areas, Fuzzy c-Means, methods of decision, similarity index, soybean yield.

Introduction

Soybeans have been the subjects of studies trying to understand the role that agrometeorological variables have in the formation of crop production (Dalacort *et al.*, 2006; Toledo *et al.*, 2010). In the western region of Paraná State, soybean produc-

tion has been responsible for contributing to the trade balance of Paraná State and consequently of Brazil. Soybean is one of the primary agricultural products of Brazil; in 2011 alone it generated a revenue of US\$ 24.9 billion (Conab, 2008).

The procedures for generating harvest predictions, as well as a knowledge of their spatial distribution, are important for the agricultural sector. Most of these procedures make use of meteorological

variables, considered as independent in statistical relation with the dependent variable, soybean yield. Some of these variables are rainfall, average temperature and global solar radiation (Berlato *et al.*, 1992; Dalacort *et al.*, 2006). Notably, some agrometeorological factors, such as drought, excessive rainfall, high temperatures and very low or low luminosity, can cause significant reductions in soybean yields, including restricting the areas where commercially important species can be grown (Farias, 2011).

For soybean development, water availability is of utmost importance, especially during two periods: germination-emergence and flowering-grain filling. During the former period, both an excess and lack of water are detrimental to obtaining good uniformity in plant population, with the water surplus's being more limiting than that of a deficit (Embrapa, 2008). Another agrometeorological factor affecting this process is temperature. The adaptation of soybean is best suited to regions where temperatures vary between 20 °C and 30 °C, and seeding should not to be performed when the soil temperature is below 20 °C, at which point germination and emergence of the plant is put at risk. Another important environmental component to the development of soybeans is solar radiation because in addition to providing light energy for photosynthesis, solar radiation provides environmental signals for a range of physiological processes for the crop. Thus, in addition to the radiation intensity, the duration and quality of the light spectrum are determinants of significant morphological and phenotypic responses, such as plant height, induction of flowering and ontogeny (Thomas, 1994).

To investigate the relationship of the soybean yield to meteorological variables, the fuzzy sets theory has been adopted. This approach is based on the characterization of classes that do not have rigid boundaries between them (Burrough and McDonnell, 1998). Fuzzy modeling is implemented when complex mathematical information is obtained in an environment that may be uncertain, vague,

ambiguous, abstract and ambivalent, as well as where diffuse boundaries are represented (Taylor *et al.*, 2007; Yan *et al.*, 2007).

Rodrigues Junior *et al.* (2011) made use of the Fuzzy c-Means to define management zones for coffee culture. The authors' research was based on determinations performed with chlorophyll sensing and leaf analysis. Odeh *et al.* (1992) identified classes of soil with samples obtained from two profiles, also utilizing the Fuzzy c-Means classifier.

In this context, the aim of this study was to classify areas associated with soybean yield ($t\ ha^{-1}$) in the western region of Paraná State, considering the following agrometeorological variables: rainfall (mm), average air temperature (°C) and average global solar radiation ($W\ m^{-2}$).

Materials and methods

The study area comprises 48 western municipalities of Paraná State, Brazil, located between longitudes W 52° 54' and W 54° 36' and latitudes S 23° 58' and S 25° 40', as shown in Figure 1. In this figure, the municipalities that have weather stations are identified (2, 8, 15, 16, 32, 36, 41 and 45).

The data organization in the field of Geographic Information Systems (GIS) and the considerations for Geographic Database (GDB) organization are applied according to the different types and the nature of the data that need to be gathered. Generally, and for the purposes of this study, the data for GDB are classified as spatial and non-spatial.

The spatial data are associated with geographical or spatial elements, such as the set of polygons that represent the map of municipalities covering the study area (see Figure 1). Currently, many of these maps are provided in digital formats. For this study, the map of municipalities in western Paraná was obtained from the IBGE (2012).

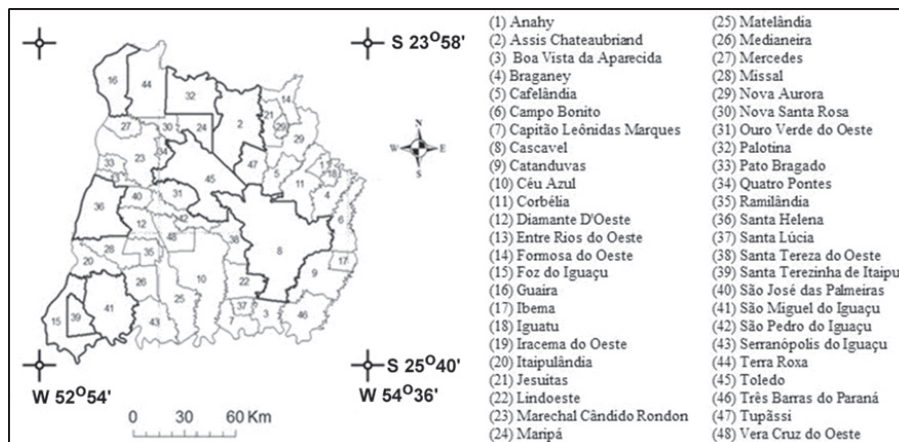


Figure 1. Western Paraná Region with highlighting of the municipalities with weather stations.

The non-spatial data refer to a set of attributes that complement the spatial data, describing what is associated with a point along a line or a polygon. In this work, the variables were organized into a table of 48 rows (identifications and variables) for 5 columns (number of attributes). Subsequently, this table was built according to the GDB and connected to the map of municipalities. The definition of attributes was based on the identification of the highest mean soybean yield of all the municipalities between crop-years of 2000/2001 and 2007/2008. The crop-year 2007/2008 was selected. Thus, for each of the 48 identified municipalities the following attributes were established:

- Polygon identifier (*PID*);
- Soybean yield (*PROD*), measured as (t ha⁻¹), provided by SEAB (2010);
- Three agrometeorological attributes arising from SIMEPAR (2010): rainfall (*PREC*), measured in mm, average air temperature (*TMED*) in °C and global solar radiation average (*RSGM*) in W m⁻². Initially, daily data for these variables were only available for eight municipalities in the study area with weather stations, as outlined in Figure 1. To measure the harvest period, these daily agrometeorological data were used between

1st October 2007 and 29 February 2008 for the eight municipalities. Rainfall (*PREC*) was calculated as the sum of daily data for the period, and the temperature (*TMED*) and global solar radiation averages (*RSGM*) were arithmetic means of the data period. For other counties without weather stations, estimates were obtained from agrometeorological data for the harvest period for the crop through the use of Thiessen Polygons (Andrade *et al.*, 2008) and Spatial Join (Jacox; Samet, 2007).

The method used in this study involves a 4-step procedure:

- Standardization of attributes (*PROD*, *PREC*, *TMED*, *RSGM*), as Equation (1):

$$Vatr_{pad} = \frac{(Vatr - Vatr_{min})}{(Vatr_{max} - Vatr_{min})} \quad (1)$$

where *Vatr_{pad}* is the standardized attribute, ranging from 0 to 1; *Vatr* is the attribute observed; *Vatr_{min}* and *Vatr_{max}* refer to the minimum and maximum values of the attribute observed, respectively. This form of standardization is required for the *FCM* use, ensuring that all attributes have the same order of magnitude, *i.e.*, ranging from 0 to 1 (Gomes *et al.*, 2011);

- ii. Application of the Fuzzy c-Means (*FCM*) to obtain the similar clusters or groups within a data set. The *FCM* is an iterative algorithm, and new centers of clusters and degrees of relevance are calculated in each iteration, always seeking to minimize the Euclidean metrics from each data point and the center of the cluster. The responsibility to check this convergence is for the objective function. A more detailed *FCM* description can be seen in Dunn (1973), Bezdek (1981) and Bezdek and Pal (1992);
- iii. Data allocation in the clusters established by the *FCM* was performed using decision methods. For this study, two methods were used: Method of Decision by the Greater Degree of Relevance (*MDMGP*) and Method of Decision by Threshold β (β *MDL*) (Guiera *et al.*, 2005; Ng *et al.*, 2008; Wang and Fei, 2009). The *MDMGP* determines which group owns the data according to the highest degree of relevance. This ensures that all data are allocated. In the *MDL* β , determination is by means of a threshold β , *i.e.*, data belong to a cluster because its degree of relevance is greater or equal to β . Unlike the *MDMGP*, the *MDL* β can generate a data set not assigned to any cluster, precisely because they are below the established threshold β (Guiera *et al.*, 2005). From the data allocation into clusters, it becomes important to evaluate the quality of the results obtained. In this study, the Modified Partition Entropy (*MPE*) was used to indicate the best number of clusters (Budayn *et al.*, 2009; Sun *et al.*, 2012):

$$MPE = \frac{-\sum_{k=1}^n \sum_{l=1}^c u_{lk} \log(u_{lk}) / n}{\log(c)} \quad (2)$$

where c is the number of clusters; n is the number of polygons; u_{lk} corresponds to the degree of relevance of the polygon k of cluster l .

The *MPE* measures the degree of clutter between clusters of a dataset. Its value ranges from 0 to 1. Values near zero indicate distinct clusters, with a small degree of sharing among data. Values near one indicate there being no distinct clusters, with a high degree of sharing data (McBratney and Moore, 1985; Fridgen *et al.*, 2004). According to Boydell and McBratney (2002), the optimal number of clustering of a dataset is established based on the minimum *MPE* value;

- iv. Measuring the degree of similarity of each cluster. For such, the Index of Similarity of Clusters (*ISC*) was set up and used as Equation (3):

$$ISC_l = \left[\frac{\sum_{k=1}^{n_l} u_{lk}}{n_l} \right] 100 \quad (3)$$

where n_l is the number (n) of polygons in the clustering l and u_{lk} follows the same definition given in Equation (2).

Results of *FCM* were transported to GDB and connected to the map of polygons for further analysis. This work was performed using the following programs: ArcMap 9.3 (ESRI, Redlands/CA, USA) and Matlab R2010a (MathWorks, Natick/MA, USA).

Results and discussion

Initially an exploratory data analysis was performed. Table 1 summarizes the main descriptive statistics of observed variables, as well as their standard values.

From Table 1, it was found that the average productivity of soybeans obtained in crop year 2007/2008 was 3.27 t ha⁻¹, high in comparison with the national average yield of 2.82 t ha⁻¹ (Conab, 2008). The average productivity had a minimum yield 2.49 t ha⁻¹ and a maximum of

3.66 t ha⁻¹, demonstrating the production potential of the western region of Paraná State. The total rainfall for the study period ranged between 450.4 mm and 859.0 mm with an average of 752.2 mm and coefficient of variation of 11.94% (homogeneous data). This represented 1.5 mm of average daily water available. According to Embrapa Soja (2010), for soybean crops to obtain maximum productivity, the water availability for the entire crop cycle must be between 450 and 800 mm, depending on the length of the cycle and crop management. The average temperature of 24.33 °C is within the parameter which best fits soybeans (20 to 30 °C). The variables studied showed leftward asymmetrical distribution, except for the variable that showed *RSGM* right asymmetry. The variables *PROD* and *PREC* tend to a platykurtic distribution, and the remaining to a mesokurtic. A likely explanation for this pattern of non-normality of the data, especially for the meteorological variables, is due to estimated data for municipalities that did not have weather stations.

To investigate the clusters of municipalities with respect to soybean yield (*PROD*) and meteorological variables (*PREC*, *TMED*, *RSGM*), the study applied the *FCM* using the program Matlab R2010a (Matlab, 2011). Among the ten trials conducted for 4 clusters, the Fuzzy c-Means algorithm reached its best condition after stopping at 16 iterations. For

the first iteration, the objective function provided the value of 4.3246414 and for the last iteration the value of 1.644438. These values confirm the convergence of the data, minimizing the distance of any cluster data relative to its center (Chen *et al.*, 2009; Zhu *et al.*, 2012).

After running the *FCM*, the data allocation into the clusters was submitted to the *MPE* index to qualify the separation of clusters. After performing the *MPE* ten times, we identified four clusters. According to the characteristics of the *MPE*, it was aimed at validating the indication of an optimal number for the number of clusters. This was accomplished through the Method of Decision by Greater Degree of Relevance (*MDMGP*), which quantifies the degree of inclusion (overlap) between groups. The results are shown in Table 2, with $S(A, B)$ as the degree of overlap between the groups A and B to $A \neq B = 1, 2, 3, 4$.

Regarding the values of inclusion among groups (A, B), the lower the identified value ($S(A, B)$), the more defined the cluster is because the values do not match each other. In contrast, values over 0.5 identify cloudiness among clusters because they are closer. Thus, the quality of the four clusters, which is highlighted by the *MPE*, was considered satisfactory for this study because only one relevance was identified above 0.5, $S(1, 2) = 0.5252$ (i.e., groups 1 and 2).

Table 1. Statistics of exploratory variables (soybean yield – *PROD*, rainfall – *PREC*, average temperature – *TMED* and average global solar radiation – *RSGM*) of 48 municipalities of western Paraná, from 1 October, 2007 to 28 February, 2008.

Attribute	Mean	SD	CV	Min.	Max.	Median	S	K
<i>PROD</i> (t ha ⁻¹)	3.27	0.22	6.82	2.49	3.66	3.26	-0.82	1.83
<i>PROD</i> std	0.67	0.19	28.57	0.00	1.00	0.66		
<i>PREC</i> (mm)	752.20	89.80	11.94	450.40	859.00	742.20	-1.41	3.73
<i>PREC</i> std	0.74	0.22	29.73	0.00	1.00	0.71		
<i>TMED</i> (°C)	24.33	0.88	3.63	22.90	25.50	24.60	-0.70	-1.00
<i>TMED</i> std	0.55	0.34	61.77	0.00	1.00	0.65		
<i>RSGM</i> (W m ⁻²)	439.91	23.49	5.34	411.20	515.20	441.20	0.80	0.94
<i>RSGM</i> std	0.28	0.23	81.82	0.00	1.00	0.29		

SD: Standard Deviation; CV: Coefficient of Variation; Min: lower value; Max: higher value; S: Skewness; K: Kurtosis; std: standardized.

Table 2. Overlap degrees among clusters (A,B) established by the MDMGP method.

(A,B)	S(A,B)	(A,B)	S(A,B)	(A,B)	S(A,B)	(A,B)	S(A,B)
1,2	0.5252 ¹	2.1	0.2982	3.1	0.1308	4.1	0.3756
1,3	0.1451	2.3	0.0885	3.2	0.1406	4.2	0.4186
1,4	0.3349	2.4	0.2119	3.4	0.1838	4.3	0.2287

¹clusters with a high degree of overlap (> 0.5). S(A, B) the degree of overlap between clusters A and B.

To form an established number of clusters and their respective degrees of inclusion, the municipalities were quantified according to their clusters. To this end, we employed the methods of decision *MDMGP* and *MDL* β . The results are shown in Table 3 wherein column (M) denotes the number of municipalities allocated to each cluster, and column (%) indicates the percentage of municipalities allocated from the total (48).

For the *MDMGP* method, 100% of municipalities were distributed into established clusters. This was due to the allocation of the data being determined by the value of the highest degree of membership found in each grouping. The distribution of municipalities in clusters by the method *MDL* β depends on the threshold value β . When $\beta = 0.8$ ($0.8 \leq \text{degree of relevance} \leq 1$), 50% of municipalities were allocated, and 50% were not allocated (24 municipalities). For $\beta = 0.65$ ($0.65 \leq \text{degree of relevance} \leq 1$) there were 75% of municipalities allocated, and 25% were unallocated (12 municipalities). For $\beta = 0.5$ ($0.5 \leq \text{degree of relevance} \leq 1$) approximately 90% of the municipalities were allocated, and only 10% (5 municipalities) were not allocated.

The distribution of municipalities imposed by the *MDMGP* and *MDL* β can be spatially displayed in the map, as illustrated in Figure 2 whose color tones from light to dark denote the level of relevance of the municipality for the cluster to which it is allocated. The shades were divided into 3 classes: $0 < \text{lighter} \leq 0.5$; $0.5 < \text{intermediate} \leq 0.75$; and $0.75 < \text{darker} \leq 1.0$. The values presented for each color represent the similarity levels obtained for each group, and the third most similar collation of the order is

95%, followed by groups 2, 4 and 1, respectively with similarities in the order of 77, 73, and 63%. These levels of similarity were obtained by the *ISC_i*. These levels of similarity can be justified because there is no weather station in the 10 municipalities of Group 3, there are only three cities with weather stations of the 19 municipalities in Group 2 (Foz do Iguaçu, Sao Miguel do Oeste and Santa Helena) and 11 municipalities in Group 1 (Assis Chateaubriand, Palotina and Guaira) and 2 counties (Toledo and Cascavel) of the 8 municipalities in Group 4.

Among the four maps shown in Figure 2, we chose the map in Figure 2 (a), which represents the classification generated by the method *MDMGP*, allowing the classification of all municipalities in the study area. The other maps deriving from the *MDL* β method allowed the verification of the levels of belonging of municipalities to the clusters identified. This situation allows for the following assumption: for example, when the threshold is 0.8, the grouped municipalities are more similar or less fuzzy. However, in this case, 50% of the municipalities failed to be allocated to any grouping.

By Table 4, which summarizes the main descriptive statistics within the clusters found by the method *MDMGP*, we sought to analyze the soybean yield and the specified agrometeorological variables.

Aiming to analyze the clusters identified by the *FCM* (Figure 2a) and the statistics shown in Table 4, the map in Figure 3 shows the soybean yield (t ha⁻¹) highlighted for each municipality. Figure 3 also shows the clusters obtained by the standardized values from Equation (1).

Table 3. Distribution of municipalities in the clusters according to the MDMGP and MDLβ.

Methods	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Not allocated	
	(M)	(%)	(M)	(%)	(M)	(%)	(M)	(%)	(M)	(%)
MDMGP	11	22.92	19	39.59	10	20.83	8	16.66	0	0.00
$MDL\beta - \beta=0.80$	2	4.16	10	20.83	9	18.76	3	6.25	24	50.00
$MDL\beta - \beta=0.65$	6	12.50	15	31.26	10	20.83	5	10.41	12	25.00
$MDL\beta - \beta=0.50$	8	16.67	18	37.50	10	20.83	7	14.58	5	10.42

M: municipalities.

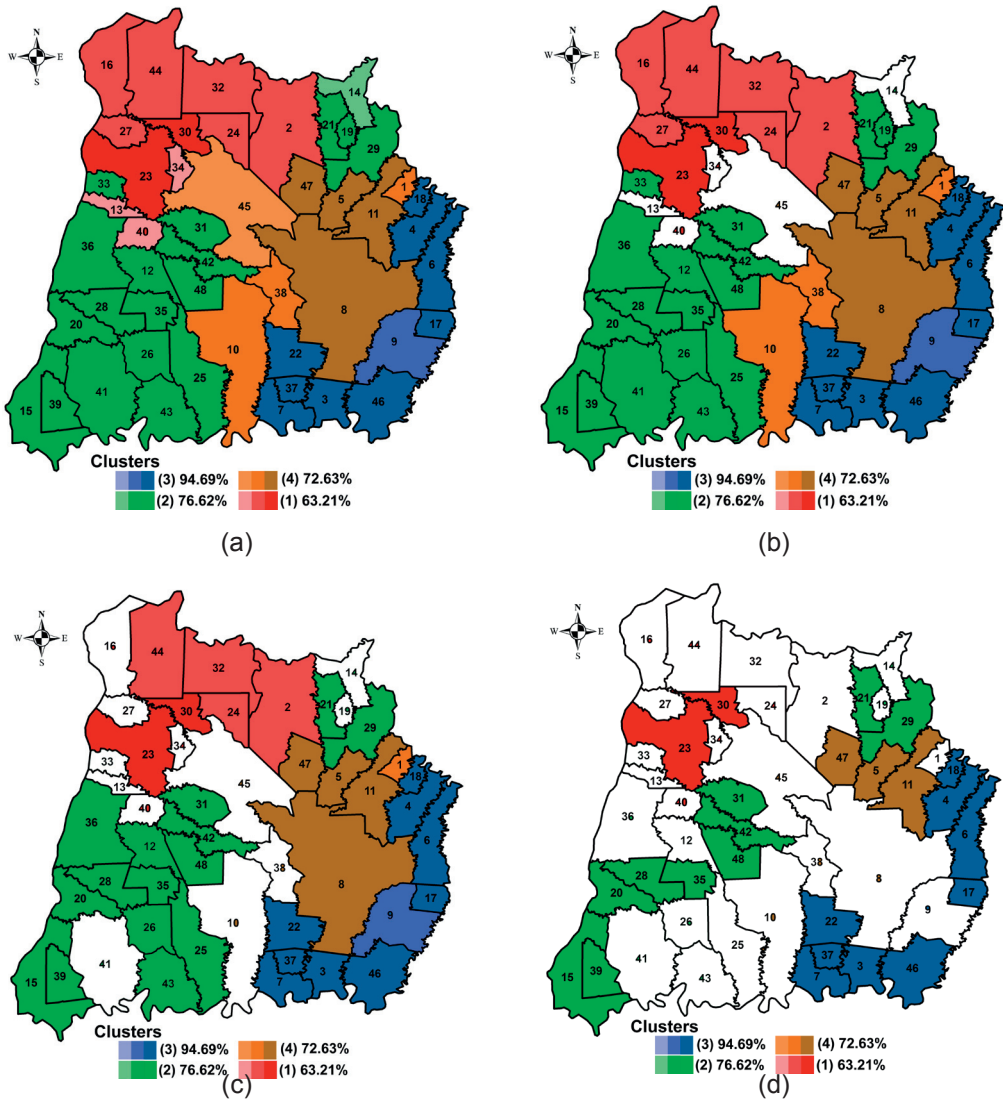


Figure 2. Distribution of clusters imposed by FCM resulting from the methods: MDMGP (a) MDLβ=0.5 (b), MDLβ=0.65 (c) and MDLβ=0.8 (d).

It was found that the highest average productivity (3.38 t ha^{-1}) and the second lower amplitude (0.51 t ha^{-1}) and coefficient of variation (5.80%) in productivity were achieved for Cluster 4. It also had the second highest average total rainfall (762.92 mm) and the third highest average mean temperature ($23.94 \text{ }^\circ\text{C}$) and global solar radiation (425.97 W m^{-2}). This group was formed by the municipalities located in the central region of the study area (Figure 3).

The first cluster, which also had the lowest similarity (63.21%) among the groups (Figure 2a), was also the cluster that presented with the lowest average yield (3.05 t ha^{-1}), the lowest average total rainfall (663.15 mm), and the largest amplitude total rainfall (326.60 m) for the study period. These results indicate that the amount of water in the growing cycle contributes to lower yields. Moreover, in this region (Figure 3), higher values of global solar radiation (469.90 W m^{-2}) were observed; in addition, the largest range of solar radiation between municipalities within the same cluster was found. The more intense solar radiation increases the evaporation of the cultures that are associated with lower amounts of precipitation between the clusters, explaining the lower productivity encountered.

Cluster 2 (19 municipalities) had the majority of municipalities in the southeast of the study area (Figure 3). For these municipalities, as shown in Table 4, we obtained the highest mean temperature ($25.09 \text{ }^\circ\text{C}$) among groups, which is justified by the municipalities' being located near the Itaipu Lake area, which is known to be a warmer region (between $30 \text{ }^\circ\text{C}$ and $40 \text{ }^\circ\text{C}$). Moreover, the region had the second highest average yield (3.34 t ha^{-1}), the third highest total rainfall averaged over the crop cycle (731.14 mm) and the second largest amplitude (40.50 W m^{-2}) and average global solar radiation (444.02 W m^{-2}) between clusters.

The third cluster was formed by 10 municipalities without weather stations to the east of the study area (Figure 3). Although they presented the third highest average yield (3.30 t ha^{-1}), the total rainfall of 859 mm cycle attributed to these municipalities was actually the precipitation of the weather station from Cascavel, due to the methodology used to estimate data by the Thiessen polygon. This methodology justifies equal minimum and maximum values for all meteorological variables presented in Table 4.

In general, coherence was observed in the formation of the clusters. The great advantage of evaluating clusters instead of each municipality separately is that a better understanding of regional similarities can be obtained.

The classification of municipalities with degrees of similarities in the order of 63 to 94% was obtained through the application of Fuzzy c-Means algorithm. Of the two methods used for classifying municipalities, *i.e.*, the Decision Method for Higher Degree of Relevance (*MDMGP*), showed better results because it ensured all data belonged to one group.

With the classifications obtained, it was possible to identify different similarities, in both municipalities comprising each cluster, and between clusters obtained. Both the Index of Similarity Clustering

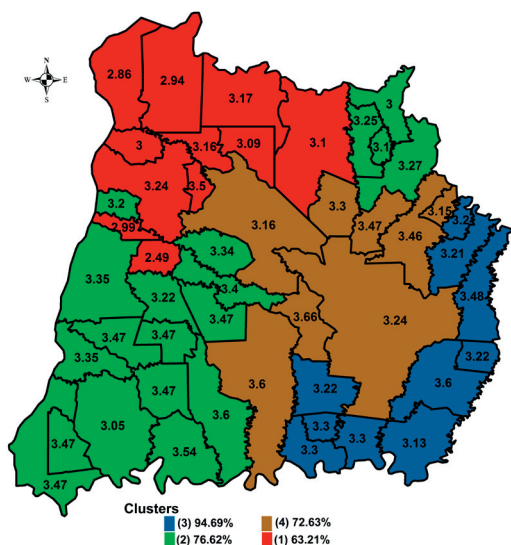


Figure 3. Thematic map of the soybean yield (t ha^{-1}) according to the performed groups.

Table 4. Exploratory statistical variables (soybean productivity – PROD, rainfall – PREC, average temperature – TMED and average global solar radiation – RSGM) for each of the four clusters generated by the classification method MDMGP.

Cl.	Var.	Min.	Max.	R.	Mean	Med.	SD	CV	S	K
1 (11)	PROD	2.49	3.50	1.01	3.05	3.09	0.25	8.28	-0.62	2.26
	PREC	450.40	777.00	326.60	663.15	700.00	115.97	17.49	-1.22	0.37
	TMED	24.30	25.20	0.90	24.60	24.45	0.29	1.17	1.10	0.31
	RSGM	441.20	515.20	74.00	469.05	466.90	20.63	4.40	0.98	1.67
2 (19)	PROD	3.00	3.60	0.60	3.34	3.35	0.17	5.07	-0.59	-0.51
	PREC	700.00	817.80	117.80	731.14	729.00	27.62	3.78	1.85	4.80
	TMED	24.90	25.50	0.60	25.09	25.10	0.17	0.69	0.47	-0.11
	RSGM	427.40	467.90	40.50	444.02	441.30	10.22	2.30	0.24	0.71
3 (10)	PROD	3.13	3.60	0.47	3.30	3.26	0.14	4.31	1.32	1.34
	PREC	859.00	859.00	0.00	859.00	859.00	0.00	0.00	*	*
	TMED	22.90	23.90	0.00	22.90	22.90	0.00	0.00	-1.19	-2.57
	RSGM	411.20	411.20	0	411.20	411.20	0.00	0.00	1.19	-2.57
4 (8)	PROD	3.15	3.66	0.51	3.38	3.38	0.20	5.80	0.20	-1.60
	PREC	746.60	859.00	112.40	792.92	798.40	33.83	4.27	0.73	1.66
	TMED	22.90	24.60	1.70	23.94	23.90	0.51	2.15	-1.00	2.10
	RSGM	411.20	434.30	23.10	425.97	426.20	7.00	1.64	-1.35	2.79

Cl.: Cluster with number of municipalities; Var.: Variable; Min.: Minimum; Max.: Maximum; R.: Range; Med.: Median; SD: Standard Deviation; CV: Coefficient of variation (%); S: Skewness; K: Kurtosis; *: No data.

(ISC_i), which measures the similarity between the municipalities of each cluster, as the indicator ($S(A, B)$) that measures the degree of similarity between clusters, were shown to be useful tools in the grouping analysis.

These results facilitate future studies with methodologies that can, for example, consider the spatial correlation between unit areas (municipalities).

Acknowledgements

To Universidade Tecnológica Federal do Paraná (UTFPR), Universidade do Oeste do Paraná (UNIOESTE), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação Araucária, for granting scholarships and financial support.

Resumen

E.C. Araújo, J.A. Johann, M.A. Uribe-Opazo e E.C.G. Camargo. 2012. Clasificación de las áreas asociadas con la producción de la soya y de las variables agrometeorológicas, usando el agrupamiento difuso. Cien. Inv. Agr. 40(3): 617-627. Este trabajo tuvo como objetivo aplicar un enfoque basado en el análisis de agrupamiento *fuzzy* para la clasificación de áreas asociadas con la productividad de la soya, juntamente con las variables meteorológicas: nivel de precipitaciones, temperatura media del aire y la media de la radiación solar. El estudio se llevó a cabo con la participación de 48 municipios de la región oeste del Estado de Paraná, Brasil, con los datos de la temporada de cultivo del año 2007/2008. Mediante el algoritmo *Fuzzy C-Means*, fue posible formar grupos de municipios similares al rendimiento de la soya, utilizando el método de decisión de mayor grado de relevancia (*MDMGP*) y el método de decisión por *Threshold β (MDL β)*. Seguidamente, se obtuvo la identificación del número apropiado de conglomerados utilizando la entropía de particiones modificada. Para medir el grado de similitud de cada grupo, se definió el Índice de Similitud de Agrupamiento (ISC_i). Dentro de la perspectiva de este estudio, el método utilizado se presentó adecuado, lo que permitió identificar grupos de municipios con grados de similitudes en el orden de 63 a 94%.

Palabras clave: Clasificación de áreas, *Fuzzy C-Means*, índice de similitud, métodos de decisiones, rendimiento de soya, variables meteorológicas,.

References

- Andrade, N.L.R. de, F.V. Xavier, E.C.R. de F. Alves, A. Silveira, and C.U.R. Oliveira. 2008. Caracterização morfométrica e pluviométrica da bacia do Rio Manso – MT. *Revista Brasileira de Geociências* 27:237-248.
- Berlato, M.A., D.C. Fontana, and H.M. Gonçalves. 1992. Relação entre rendimento de grãos de soja e variáveis meteorológicas. *Pesquisa Agropecuária Brasileira* 27:695-702.
- Bezdek, J.C. and S.K. Pal. 1992. *Fuzzy Models for Pattern Recognition*, IEEE Press, New York. 544 pp.
- Bezdek, J.C. 1981. *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers Norwell, MA, USA. 256 pp.
- Boydell, B., and A.B. McBratney, 2002. Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture* 3(2):9-23.
- Budayan, C., I. Dikmen, and M.T. Birgonul. 2009. Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Systems with Applications* 36:11772–11781.
- Burrough, P.A., and R.A. McDonnell. 1998. *Principles of GIS*. Oxford University Press, Oxford, UK. 333 pp.
- Chen, W., and M. Wang. 2009. A fuzzy c-means clustering-based fragile watermarking scheme for image authentication. *Expert Systems with Applications* 36:1300–1307.
- CONAB-Acompanhamento da Safra Brasileira. 2008. Available online at: http://www.conab.gov.br/conabweb/download/safra/12_levantamento_set2008.pdf (Website accessed November 14, 2012).
- Dallacort, R., P.S.L. Freitas, R.T. de F. Faria, A.C.A. Gonçalves, R. Rezende, and A. Bertonha. 2006. Utilização do modelo Cropgro-soybean na determinação de melhores épocas de semeadura da cultura da soja, na região de Palotina, Estado do Paraná. *Acta Scientiarum. Agronomy. Maringá* 28:583-589.
- Dunn, J.C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J. Cybernetics. Journal of Cybernetics* 3(3):32-57.
- EMBRAPA-Empresa Brasileira de Pesquisa Agropecuária. 2010. *Tecnologias de produção de soja Região Central do Brasil 2011*. Londrina: Embrapa Soja, Out. 2010, 255 pp. Available online at: http://www.cnpso.embrapa.br/download/Sistema_Producao14_VE.pdf (Website accessed June 29, 2013).
- EMBRAPA-Centro Nacional de Pesquisa de Soja. 2008. *Tecnologias de produção de soja – região central do Brasil*. Londrina: EMBRAPA-CNPSO: EMBRAPA-CPAC: EMBRAPA-CPAO. Available online at: <http://www.embrapa.br/> (Website accessed June 24, 2012).
- Farias, J.R.B. 2011. Limitações climáticas à obtenção de rendimentos máximos de soja. *Mercosoja* 2011. Quinto Congreso de la Soja en el Mercosur. Rosario, Argentina, September, 14, 2011.
- Fridgen, J.J., N.R. Kitchen, K.A. Sudduth, S.T. Drummond, W.J. Wiebold, and C.W. Fraisse. 2004. Management Zone Analyst (MZA): software for sub-field management zone delineation. *Agronomy Journal* 96:100–108.
- Gomes, A. da S., M. de M. Pires, V.M. de Almeida, P.L. Rosado, P.R.P Santos, and A.R. São José. 2011. Análise dos territórios da região sudoeste da Bahia na perspectiva do desenvolvimento rural. *Revista Desenharia* 1:59-82.
- Guiera, A.J.A., T.M. Centeno, M.R. Delgado, and M. Muller. 2005. Segmentação por Agrupamentos Fuzzy C-means em Imagens LiDAR Aplicados na Identificação de Linhas de Transmissão de Energia Elétrica. *Espaço Energia* 3:24-31.
- IBGE - Instituto Brasileiro de Geografia e Estatística. 2012. *Mapas Digitais*. Available online at: <http://www.ibge.gov.br/home/download/geociencias.shtm> (Website accessed September 20, 2012).
- Jacox, E.H., and H. Samet. 2007. *ACM Transactions on Database Systems (TODS)*. Spatial join techniques, [s.l.] 32(1):7-es.
- Mcbratney, A.B. and A.W. Moore, 1985. Application of fuzzy sets to climatic classification. *Agr. Forest Meteorol.* 35:165–185.
- Ng, H.P., S.H. Ong, K. Weng, C. Foong, P.S. Goh, N.L. Wieslaw. 2008. Fuzzy c-means algorithm with local thresholding for gray-scale images.

- International Journal on Artificial Intelligence Tools 17:765–775.
- Odeh, I.O., A.B. Mcbratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-mean: application to classification and soil-landform interrelationships. *Soil Science Society American Journal*, 56:505-516.
- Rodrigues Junior, F.A., D.M. de Queiroz, and N.T. Santos, 2011. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. *Revista Brasileira de Engenharia Agrícola e Ambiental* 15:778-787.
- SEAB-Secretaria da Agricultura e Abastecimento. 2010. Agropecuária - Estatísticas, Produção agropecuária, Produção Agrícola Paranaense por Município - últimas 5 safras. Available online at: <http://www.agricultura.pr.gov.br/> (Website accessed December 28, 2010).
- SIMEPAR-Sistema Meteorológico do Paraná. 2010. Condições do Tempo. Available online at: <http://www.simepar.br> (Website accessed July 10, 2013).
- Sun, X., Y. Zhao, H. Wang, L. Yang, C. Qin, A. Zhu, G. Zhang, T. Peid, and B. Li. 2012. Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters. *Geoderma* 171-172:24-34.
- Taylor, J.A., A.B. Mcbratney, and B.M. Whelan. 2007. Establishing management classes for broadacre agricultural production. *Agronomy Journal* 99:1366-1376.
- Thomas, J.F. 1994. Ontogenetic and morphological plasticity in crop plants. *In*: Boote, K.J., Sinclair, T. (Comp.). *Physiology and determinations of crop yield*. Madison/WI, USA: ASA/CSSA/SSSA. Cap. 7B. p. 181-185.
- Toledo, N.T., A.G. Muller, J.L. Berto, and C.E.S. Mallmann. 2010. Ajuste do modelo fototérmico de estimativa do desenvolvimento e do índice de área foliar de soja. *Revista Brasileira de Engenharia Agrícola e Ambiental* 14(3):288-295.
- Wang, H., and Fei, B. 2009. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. *Medical Image Analysis* 13:193-202.
- Yan, L., S. Zhou, L. Feng, and L. Hong-Yi. 2007. Delineation of sitespecific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture* 56:174-186.
- Zhu, W., J. Jiang, C. Song, and L. Bao. 2012. Clustering Algorithm Based on Fuzzy C-means and Artificial Fish Swarm. *Procedia Engineering* 29:3307–3311.

