

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v39n6p707-714/2019>

## IDENTIFICATION OF HOMOGENEOUS RAINFALL ZONES DURING GRAIN CROPS IN PARANÁ, BRAZIL<sup>1</sup>

Allan R. Lopes<sup>2\*</sup>, Jonatas Marcolin<sup>3</sup>, Jerry A. Johann<sup>4</sup>, Márcio A. Vilas Boas<sup>4</sup>,  
Adilson R. Schuelter<sup>5</sup>

<sup>2\*</sup>Corresponding Author. Western Paraná State University (UNIOESTE)/ Cascavel - PR, Brazil.  
E-mail: [allanremorlopes@gmail.com](mailto:allanremorlopes@gmail.com) | ORCID ID: <https://orcid.org/0000-0002-3151-654X>

### KEYWORDS

data mining, clusters, expectation-maximization, weka, soybean, maize, wheat.

### ABSTRACT

The aim of this study is to identify homogeneous rainfall zones in the winter and summer 1<sup>st</sup> and 2<sup>nd</sup> crops, in the state of Paraná, Brazil. The zones were defined by clustering using the expectation-maximization (EM) algorithm to transform seasonal rainfall series. Monthly average rainfall data collected from 157 weather stations for 20 years (1996 to 2015) were employed. The results show that the number of homogeneous zones varied among growing seasons. The summer crop presented two clusters, with rainfall averages of 1489 and 1925 mm; the second crop presented four clusters, with averages of 1849, 1004, 1454, and 1182 mm; and the winter crop had three clusters, with averages of 969, 1498, and 1171 mm. Clustering was a useful instrument to identify geographical regions with similar rainfall regimes during different growing seasons in the state of Paraná. Rainfall distribution was more homogeneous in the summer crop. In all crops analyzed, the clusters with the lowest rainfall rate were present in the northwestern, northern center, and northern pioneer of the state of Paraná, whereas the clusters with the highest rainfall rate were found in the coastal regions.

### INTRODUCTION

The state of Paraná has the highest agricultural production in Brazil. In 2016, the agricultural gross domestic product was R\$ 88.83 billion. Agricultural production accounted for 50% of this amount, and grain exceeded 38 million tons (Seab, 2015).

The climatic conditions of Paraná allow three distinct growing seasons: summer crop, from September to December; second crop, from January to March; and winter crop, from March to July (Iapar, 2017a). The three main grain crops grown in Paraná are soybean, maize and wheat. Maize is sown from January to April and harvested from May to September; soybean is sown from October to December and harvested from January to April; and wheat is sown from April to July and harvested from August to December (Conab, 2017).

Climate has a significant impact on agricultural production, especially rainfall (Pereira et al., 2014), which can produce varying results, from high yields to partial or total crop losses. Therefore, the knowledge of rainfall

dynamics is crucial in decision-making to mitigate the damaging effects of droughts and floods (Romani et al., 2010).

Rainfall in Paraná is higher than that in the rest of Brazil and is variable throughout the state (Ely & Dubreuil, 2017). The average annual rainfall varies between 1200 and 3500 mm, and most of the state presents averages between 1400 and 2000 mm (Iapar, 2017b). However, the high annual rainfall, which also exhibits a seasonality that is unsuitable for crop development, limits crop yield and causes agricultural losses during heavy rainfall.

Considering the different growing seasons and high variability in rainfall in Paraná, systems that accurately analyze climate information are essential. Fayyad et al. (1996) proposed the application of data mining. This technique is the primary step in the Knowledge Discovery in Databases (KDD) process.

Among data mining techniques, clustering uses rainfall time series to define homogeneous rainfall zones, in which each rainfall station corresponds to a data time series, and a homogeneous zone is formed by grouping similar time

<sup>1</sup> Research developed in the discipline of "Data Mining and Knowledge Discovery", in 2016, in Graduate Program in Agricultural Engineering - PGEAGRI, from Western Paraná State University (UNIOESTE) - Cascavel - PR, Brazil.

<sup>3</sup> Western Paraná State University (UNIOESTE)/ Marechal Cândido Rondon - PR, Brazil.

<sup>4</sup> Western Paraná State University (UNIOESTE)/ Cascavel - PR, Brazil.

<sup>5</sup> Dynamics University Center of Falls (UDC)/ Foz do Iguaçu - PR, Brazil.

Received in: 5-8-2018

Accepted in: 9-26-2019

series (Dourado et al., 2013). Oliveira-Júnior et al. (2017) reported that the identification of homogeneous zones was fundamental for agricultural planning. Marzban and Sandgathe (2006) showed that clustering is an effective algorithm when using observed meteorological data.

Clustering has been used to define similar climatic zones in different regions of Brazil (André et al., 2008; Comunello et al., 2013; Moreira et al., 2016), including the state of Paraná (Fritzens et al., 2011; Pansera et al., 2015). However, no studies have defined homogeneous rainfall zones in winter and summer 1<sup>st</sup> and 2<sup>nd</sup> crops in Paraná. Owing to the high production of soybean, maize, and wheat in Paraná, Sampaio et al. (2006) emphasized that a rainfall study based on the crops cultivated in each region should be carried out. Overall, the objective of this study is to identify homogeneous rainfall zones in the in the winter and summer 1<sup>st</sup> and 2<sup>nd</sup> crops in state of Paraná, Brazil.

**MATERIAL AND METHODS**

The study area comprised the state of Paraná (Figure 1). According to Thornthwaite’s classification, the climate of Paraná is divided into 12 classes. Class C1 (C<sub>1d</sub>A’a’ and C<sub>1w</sub>A’a’) is more common in higher latitudes (north, northwestern, and northern pioneer) whereas classes B2 and B3 (B<sub>2r</sub>B’<sub>4a</sub>’ and B<sub>3r</sub>B’<sub>3a</sub>’) are predominant in lower latitudes and in the east coast (south-central, southwest, and southeast). The most prevalent class in Paraná is C<sub>1r</sub>A’a’, corresponding to 30.74% of the state. Class B<sub>1r</sub>A’a’ is the second most represented class in the state (21.28%) and is predominant in the west and central-west regions (Aparecido et al., 2016).

The study followed the KDD process described by Fayyad et al. (1996), with five steps: data selection, preprocessing, formatting/transformation, data mining, and interpretation.

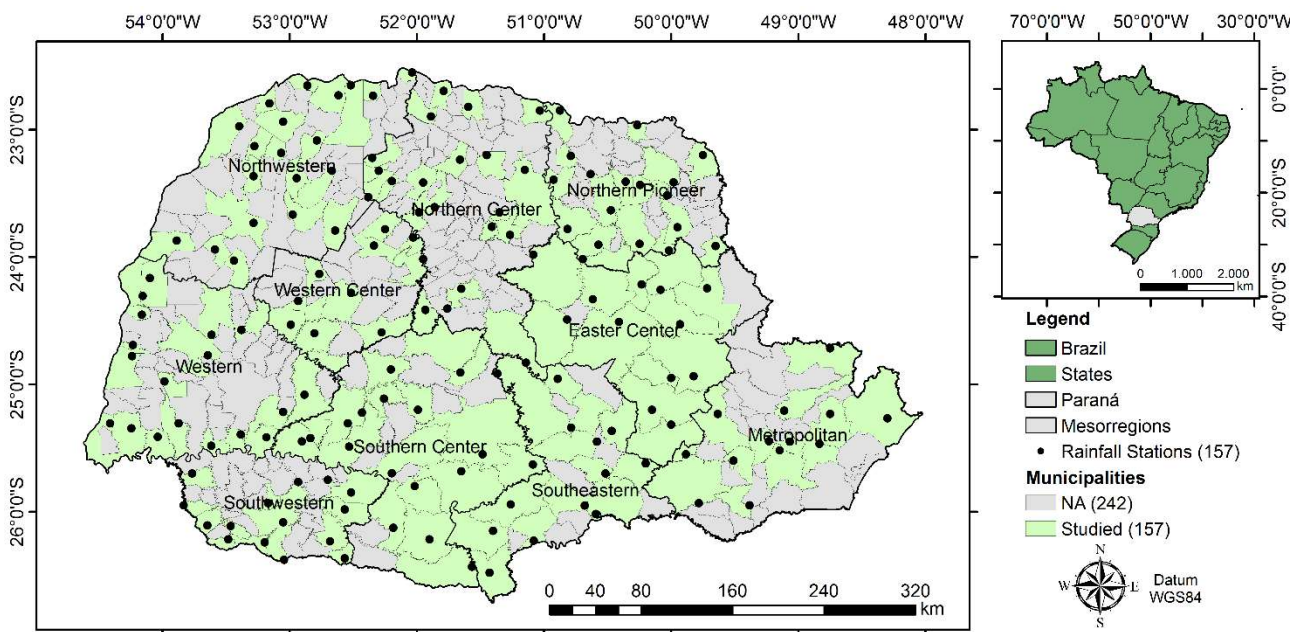


FIGURE 1. Location of the state of Paraná, studied municipalities and the analyzed rainfall stations.

Rainfall data were obtained from the Paraná Water Institute database maintained by the Hydrological Information System. We choose homogeneous and continuous time series and selected 157 weather stations that collected daily rainfall data from January 1996 to June 2016 covering most of the state of Paraná, corresponding to 20 crops (Figure 1). The 157 stations represent different municipalities, and the obtained maps were expressed at municipality level for better visualization and analysis of the results (Figure 1).

A monthly rainfall database was constructed for each weather stations using the obtained data. Rainfall rates for the months of each growing season and year were summed, forming a series of accumulated rainfall data for each season (summer crop, second crop and winter crop).

Various techniques were applied to identify outliers (clustering). However, the outliers were not excluded because they were observed values. There were neither flaws nor inconsistent values in the database.

Data normalization was applied to the entire database: the highest value of each attribute was assigned a value of 1, whereas the lowest value of each attribute was

assigned a value of 0. The other values were proportionally allocated within this interval. This technique was applied using Weka software version 3.8.0. This software contains visualization tools and algorithms for data analysis and predictive modeling, as well as graphical interfaces to facilitate access to these features (Arora & Suman, 2012).

Clustering was used for data mining in Weka software version 3.8.0. Cluster analysis involves the organization of a set of patterns commonly represented as vectors of attributes or points in a multidimensional space in groups according to a measure of similarity (Pontes Júnior et al., 2018). The similarity between attributes was measured by the Euclidean distance. For cluster analysis, a partitioning method was adopted using two algorithms: K-means and expectation-maximization (EM).

K-means is a very popular method (Dhanachandra et al., 2015) in which data are grouped according to their proximity to each other as a function of the Euclidean distance. The goal of this method is to divide a set of N data vectors into K non-overlapping clusters ( $K \leq N$ ) using K centroids adequately positioned in the data space. Then, each data vector is associated with a centroid by similarity.

EM involves expectation (E) and maximization (M). M uses the probabilities employed in phase E to refine the model parameters. E and M constitute an iterative process in which the new probabilities calculated in phase M are used to perform inferences in phase E. As a result, EM allows learning the parameters that govern the distribution of data in which some characteristics are missing. The algorithm is based on statistical calculations that estimate maximum-likelihood parameters in cases in which equations cannot be directly solved because the model depends on unobserved latent variables (Torres et al., 2017).

K values (number of groups) from 2 to 10 were tested to obtain a number of clusters most consistent with the actual distribution of data. The classifiers J48, Naive Bayes, and sequential minimal optimization (SMO) were used to select the optimal number of clusters and the algorithm that analyzed the data more accurately.

The J48 algorithm is a modified version of the c4.5 and ID3 algorithms, which are used for constructing classification trees (Quinlan, 1996). Classification trees are generated from a training dataset, and at each node, the algorithm chooses an attribute that most efficiently subdivides the sample set into homogeneous subsets according to their class (Giasson et al., 2013).

Naive Bayes is a simple but highly practical classifier with a wide range of applications. It generates a group of probabilities that are estimated by calculating the frequency of the characteristic value of each training data instance. Given a new instance, this classifier estimates the

probability that the instance belongs to a specific class (Gao et al., 2018).

The SMO is a variant of the SMV algorithm (Zhang et al., 2018) and was created to solve the quadratic programming problem presented by the SMV algorithm. SMO analyzes sparse datasets as well as binary or non-binary input data. SMO also analyzes missing data via global substitution and nominal attributes (Witten & Frank, 2005).

The percentage of accuracy of each clustering algorithm according to the classifiers, the Kappa index, and the domain expert was calculated to determine the ideal number of clusters during each growing season.

Descriptive statistics were applied using the ideal number of clusters to better analyze the data. The clusters were spatialized in ArcGIS environment version 10.3 for better visualization of the results.

**RESULTS AND DISCUSSION**

The EM clustering algorithm was the most adequate for analyzing rainfall time series, with higher success rates and Kappa indices (Table 1) than the k-means algorithm (Table 2). EM was previously used for analyzing climatic variables (Kalaiselvi & Geetha, 2016; Vrac & Yiou, 2010), and the K-means algorithm was efficient in analyzing groups of spatiotemporal rainfall characteristics in the state of Rio Grande do Sul (Boschi et al., 2011). Richetti et al. (2018) evaluated the regionalization of the state of Paraná from 1989 to 2013 and concluded that K-means obtained the best results and fewer errors when compared with the EM algorithm.

TABLE 1. Efficiency of the EM algorithm for grouping three growing seasons according to success rate (%) and Kappa index using the J48, SMO, and Naive Bayes classifiers.

Crops	Number of clusters	EM					
		J48		SMO		Naive Bayes	
		%	Kappa	%	Kappa	%	Kappa
Summer 1 <sup>st</sup>	2	96.17	0.92	96.18	0.92	100	1.00
Summer 2 <sup>nd</sup>	4	94.27	0.91	94.27	0.91	96.81	0.95
Winter	3	88.54	0.83	96.82	0.95	99.36	0.99

TABLE 2. Efficiency of the K-means algorithm for grouping three growing seasons according to the success rate (%) and the Kappa index using the J48, SMO, and Naive Bayes classifiers.

Crops	Number of clusters	K-means					
		J48		SMO		Naive Bayes	
		%	Kappa	%	Kappa	%	Kappa
Summer 1 <sup>st</sup>	2	96.17	0.92	96.18	0.92	100	1.00
Summer 2 <sup>nd</sup>	4	94.27	0.76	93.63	0.91	94.91	0.93
Winter	3	86.62	0.80	96.18	0.94	96.82	0.95

The boxplot of the rainfall clusters in each growing season is shown in Figure 2. There were three outliers in cluster 1 during the summer 1<sup>st</sup> crop and 1 outlier in cluster 2 during the summer 2<sup>nd</sup> crop. However, during the winter crop, data had a lower dispersion than in other seasons and

no outliers at all. Average rainfall data in regions with discrepant values should be carefully used for irrigated agriculture, hydraulic works, and climatic zoning because the high variability in rainfall may cause severe damage by the lack or excess of water (Lima et al., 2008).

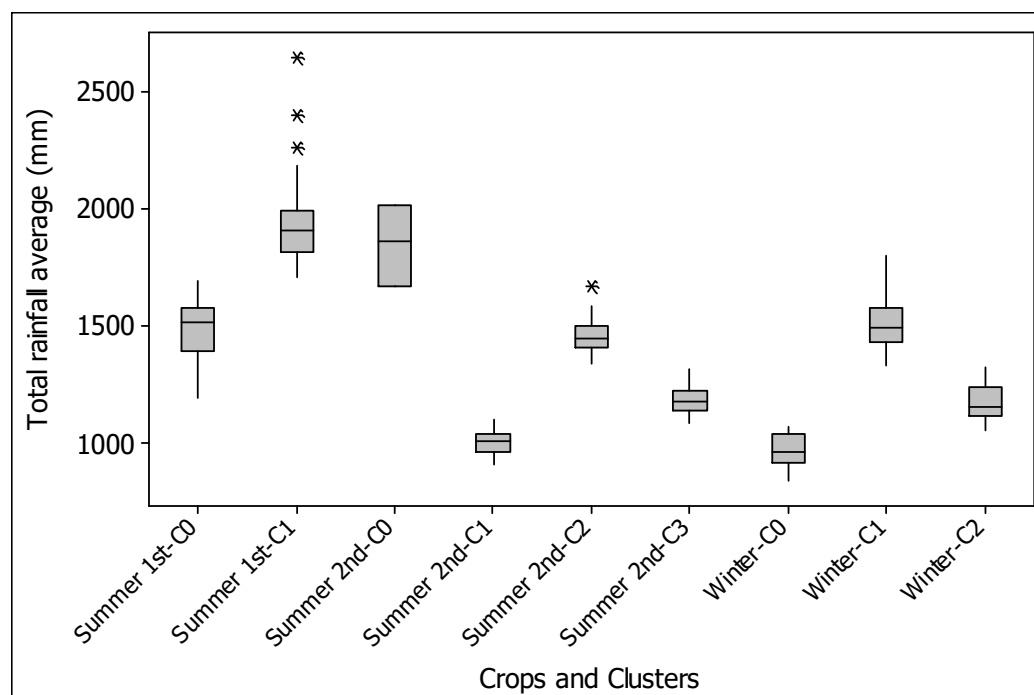


FIGURE 2. Boxplot of total rainfall average (mm) per cluster (C0: cluster 0; C1: cluster 1; C2: cluster 2; C3: cluster 3) during different growing seasons (winter, summer 1<sup>st</sup> and 2<sup>nd</sup> crops) in the state of Paraná, Brazil, from 1996 to 2016.

For summer 1<sup>st</sup> crops, clustering generated two homogeneous rainfall zones, designated as clusters 0 and 1. A total of 96 stations were grouped in cluster 0, with an average rainfall of 1489 mm, a maximum of 2695 mm, a minimum of 634 mm, and the lowest rainfall (Table 3). Cluster 0 was predominantly found in the northwestern,

northern, eastern center regions, and in the metropolitan region of Curitiba (Figure 3). These regions presented similar rainfall regimes and were strongly affected by continentality. Farias et al. (2001) found that the risk of droughts in the northwestern of Paraná was high during the most critical phases of soybean cultivation.

TABLE 3. Clustering and descriptive statistics of rainfall during summer 1<sup>st</sup> crops in the state of Paraná, Brazil, between 1996 and 2015.

Cluster	Number of stations	Average	Standard deviation	Coefficient of variation (%)	Median	Maximum	Minimum
0	96	1489	306	20,5	1458	2695	634
1	61	1925	388	21,2	1867	3269	1077

Sixty-one weather stations were grouped in cluster 1, with an average rainfall in summer 1<sup>st</sup> crops of 1925 mm, a maximum of 3269 mm, a minimum of 1077 mm (Table 3), and the highest rainfall rate for all crops. Both clusters presented rainfall rates (450–850 mm) above the requirement for soybean crops (Carvalho et al., 2013).

The rainfall rate was lowest in the summer 2<sup>nd</sup> (Sampaio et al., 2006), and four clusters were formed during this crop (Figure 4).

Thirty-seven weather stations were grouped in cluster 0, which had the lowest average rainfall (1004 mm), lowest minimum rainfall (510 mm), and lowest maximum rainfall (2006 mm). Cluster 0 predominantly included the north side of the state of Paraná (Figure 4), where water deficit occurs (Fritzens et al., 2011). Gonçalves et al. (2002) studied sowing dates for maize in Paraná in summer 2<sup>nd</sup> as well as its climatic risks and concluded that the risk of drought in the region of cluster 0 exceeded 50%.

Seventy weather stations were grouped in cluster 1, which was the most representative in the state, with an

average rainfall of 1182 mm, a maximum of 2108 mm, and a minimum of 610 mm. This cluster included the western center, eastern center, western, and the metropolitan region of Curitiba (Figure 4). Variability was higher in the south, which can be attributed to the interaction between the relief and the constant development of frontal systems associated with subtropical jet streams. Shiroga & Gerage (2010) found that there was water deficit in maize crops in the region covered by cluster 1.

Cluster 2 comprised 47 stations, with an average rainfall of 1454 mm. This cluster had the highest maximum rainfall (2612 mm) and primarily comprised the southwestern and eastern center regions (Figure 4). Cluster 3 presented the highest rainfall rates (Table 4) but contained only three weather stations, indicating that this rainfall regime was restricted to a small portion of the state. The high rainfall rate is due to the influence of the orographic factor (Silva et al., 2012).

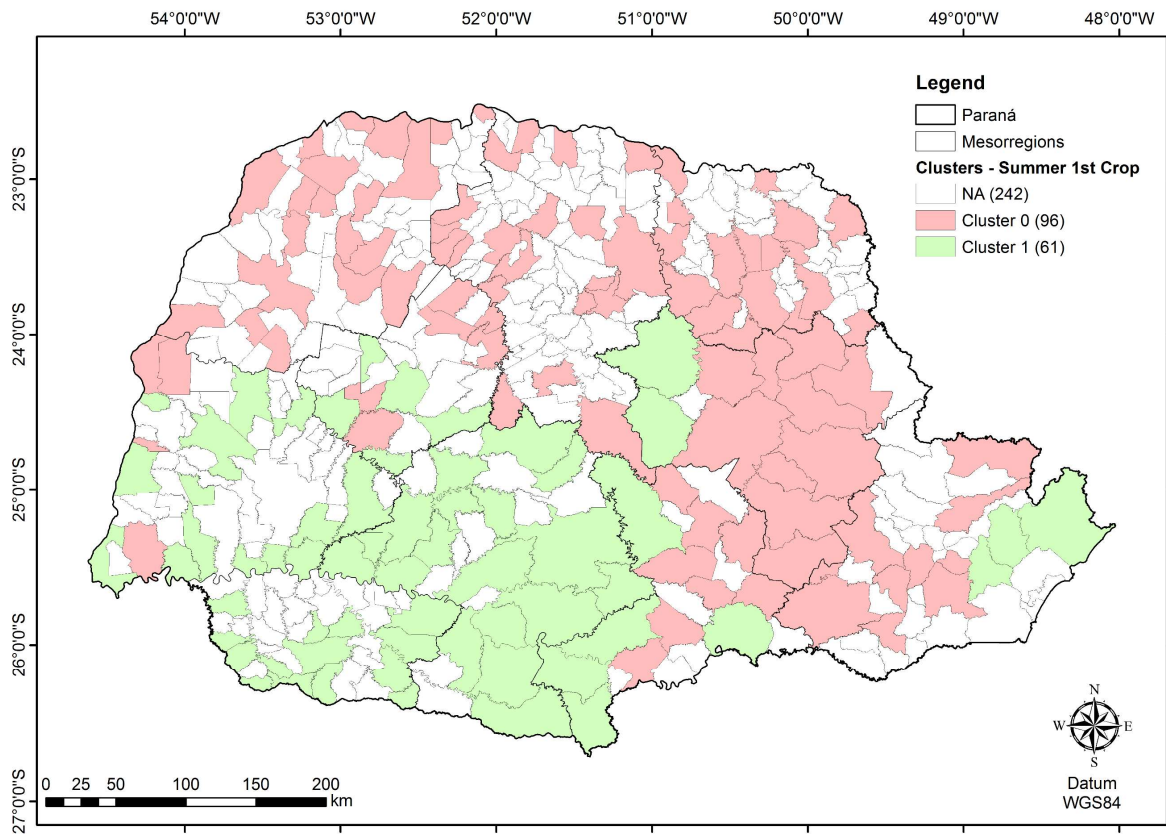


FIGURE 3. Clustering of homogeneous rainfall zones during summer 1<sup>st</sup> crops in the state of Paraná, Brazil.

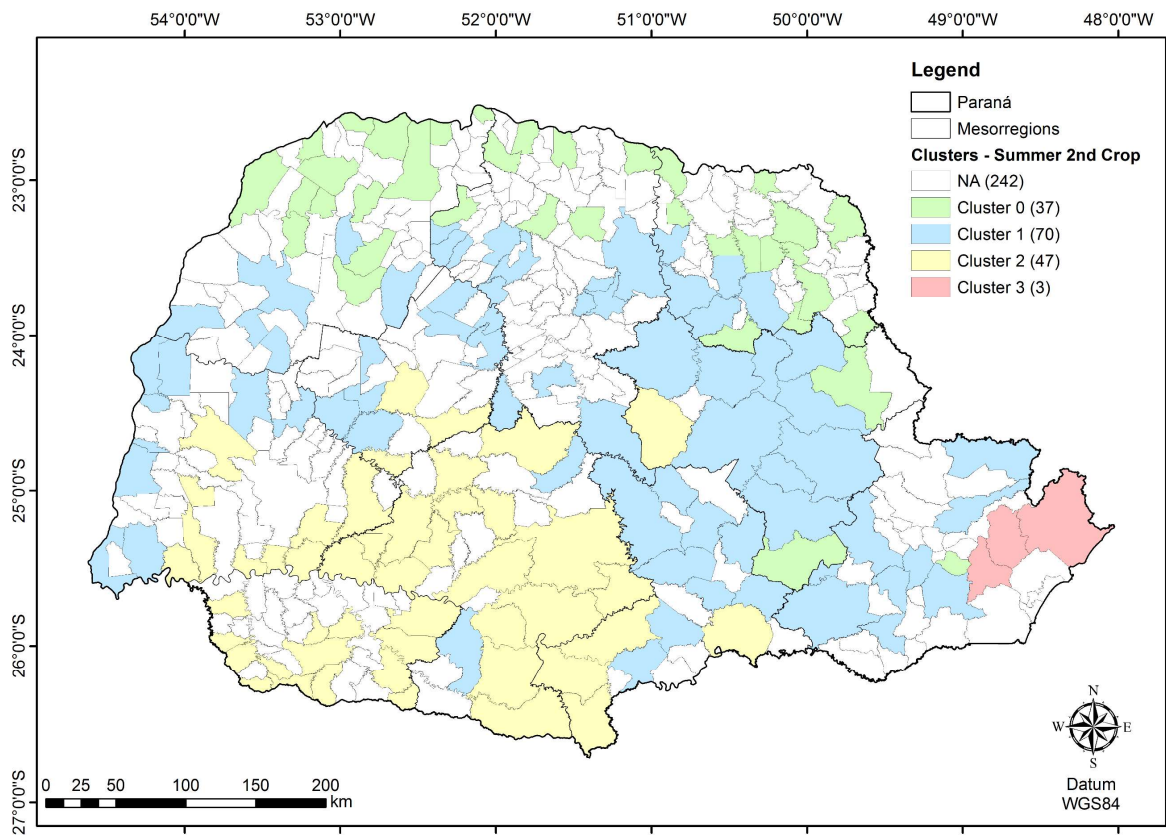


FIGURE 4. Clustering of homogeneous rainfall zones during summer 2<sup>nd</sup> crops in the state of Paraná, Brazil.

TABLE 4. Clustering and descriptive statistics of rainfall during summer 2<sup>nd</sup> crops in the state of Paraná, Brazil, between 1996 and 2015.

Cluster	Number of stations	Average	Standard deviation	Coefficient of variation (%)	Median	Maximum	Minimum
0	37	1004	190	18,9	980,5	2006	510
1	70	1182	251	21,2	1144	2108	610
2	47	1454	337	23,2	1412	2612	645
3	3	1849	351	18,9	1800	2586	1108

Bergamaschi et al. (2001) analyzed soils of Rio Grande do Sul and found that early-cycle maize requires an average rainfall of 650 mm throughout the growth cycle to express its maximum potential. None of the analyzed clusters presented values lower than the averages, indicating that, when considering the averages during the summer 2<sup>nd</sup> crops, the lack of rainfall was not a limiting factor for maize production in Paraná. However, compared with the minimum rainfall values, the summer 2<sup>nd</sup> crops grouped in clusters 0, 1, and 2 presented rainfall levels

below the averages, and rainfall might limit productivity in these areas.

Three homogeneous clusters were generated during winter crop (Figure 5). Cluster 0, containing 46 municipalities, presented the lowest average rainfall (969 mm), the lowest maximum rainfall in one crop (1965 mm), and the lowest minimum rainfall (476 mm) (Table 5). This Cluster comprised the northern pioneer and center, northwestern (Figure 5) and was characterized by the lack of rainfall (Salton et al., 2016). Fritzon et al. (2011) showed that droughts occur in this region during winter seasons.

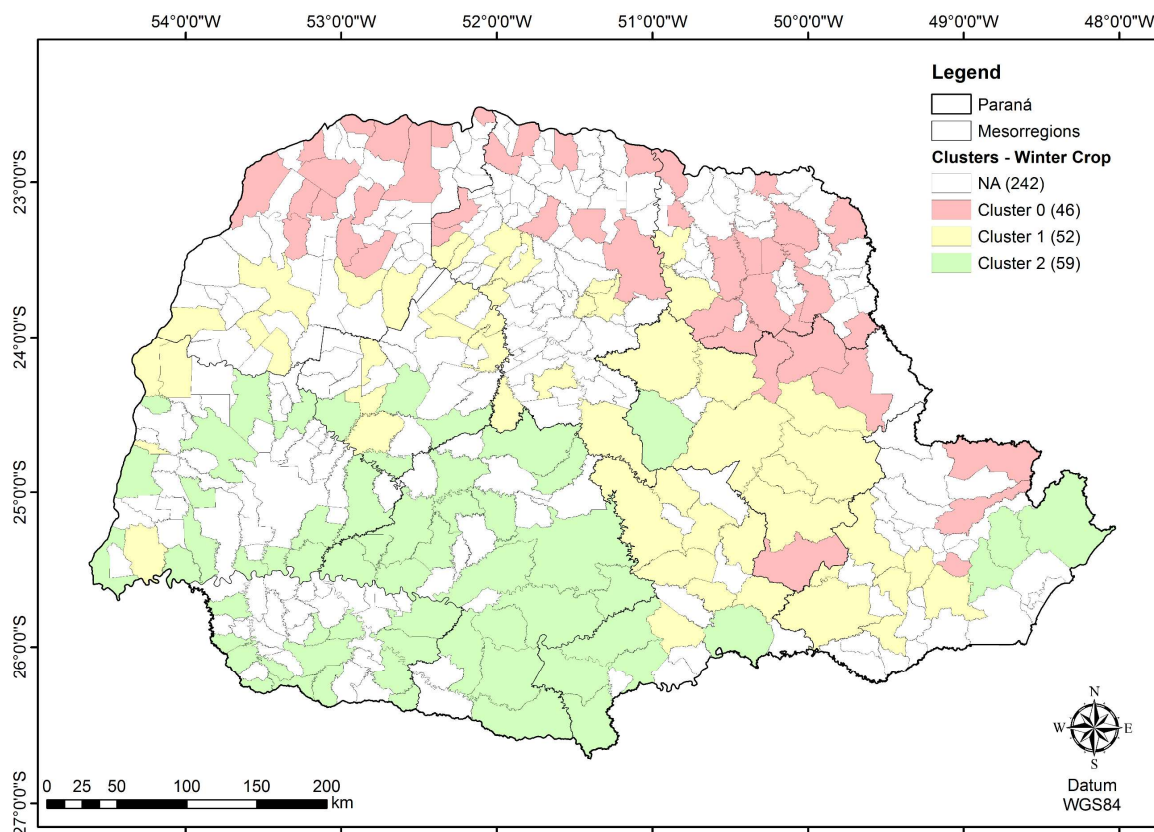


FIGURE 5. Clustering of homogeneous rainfall zones during winter crops in the state of Paraná, Brazil.

TABLE 5. Clustering and descriptive statistics of rainfall during winter crops in the state of Paraná, Brazil, between 1996 and 2015.

Cluster	Number of stations	Average	Standard deviation	Coefficient of variation (%)	Median	Maximum	Minimum
0	46	969	239	24,66	933,5	1965	476
1	52	1171	260	22,17	1144	2159	545
2	59	1498	298	19,93	1467	2517	662

Cluster 1, with 52 municipalities, presented an average rainfall of 1171 mm, a maximum of 2159, and a minimum of 545 mm. This cluster was considered a transition zone between clusters 0 and 2. Fritzens et al. (2011) emphasized that a transition zone is located in the central region of the state. Fifty-nine municipalities were grouped in cluster 2 and were primarily located in the south western, western, and coastal region; the latter had the highest rainfall in the state of Paraná (Figure 5) (Fritzens et al., 2011), good rainfall distribution throughout the year, and no dry season. However, rainfall was lowest in winter crop (Vanhoni & Mendonça, 2008). The southwest is characterized by heavy rainfall (Machado et al., 2013), which is explained by cold fronts and a decrease in temperature, increase in atmospheric pressure, and presence of southerly winds (Waltrick et al., 2015). This cluster had the highest average rainfall (1498 mm), highest maximum rainfall (2517 mm), and highest minimum rainfall (662 mm), as well as the lowest coefficient of variation (19.93%), constituting the region with the highest rainfall homogeneity during winter season. Aparecido et al. (2016) reported that only the southeast and southwest regions had annual rainfall higher than 1800 mm.

## CONCLUSIONS

The EM algorithm was slightly more effective than the k-means algorithm for clustering homogeneous rainfall zones during winter, summer 1<sup>st</sup> and 2<sup>nd</sup> crops in the state of Paraná, Brazil.

Cluster analysis showed that each growing season had a different number of homogeneous zones. Rainfall distribution in the state of Paraná was more homogeneous during summer 1<sup>st</sup> crop, with only two clusters, and was more variable in summer 2<sup>nd</sup> crop, with four clusters.

The clusters with the lowest rainfall rates were the northwestern, northern pioneer, and north, and the coastal region was represented in all clusters with the highest rainfall. Clustering was an appropriate instrument to identify geographical regions with similar rainfall regimes during different growing seasons in the state of Paraná.

These results can be used to plan growing seasons and manage water resources in the state of Paraná and can serve as support for future research.

## REFERENCES

- André RGB, Marques VS, Pinheiro FMA, Ferraudo AS (2008) Identificação de regiões pluviometricamente homogêneas no estado do Rio de Janeiro, utilizando-se valores mensais. *Revista Brasileira de Meteorologia* 23(4): 501-509.
- Aparecido LEO, Rolim GS, Richetti J, Souza PS, Johann JA (2016) Köppen, Thornwaite and Camargo climate classifications for climatic zoning in the state of Paraná, Brazil. *Ciência e Agrotecnologia* 40(4):405-417.
- Arora R, Suman S (2012) Comparative analysis of classification algorithms on different datasets using WEKA. *International Journal of Computers Applications* 54(13):21-25.
- Bergamaschi H, Radin B, Rosa LMG, Bergonci JI, Aragonés RS, Santos AO, França S, Langensiepen M (2001) Estimating maize water requirements using agrometeorological data. *Revista Argentina de Agrometeorologia* 1(1):23-27.
- Boschi RS, Oliveira SEM, Assad ED (2011) Técnicas de mineração de dados para análise da precipitação pluvial decenal no Rio Grande do Sul. *Engenharia Agrícola* 31(6):1189-1201.
- Carvalho IR, Korcelski C, Pelissari G, Hanus AD, Rosa GM (2013) Demanda hídrica das culturas de interesse agrônomo. *Enciclopédia Biosfera* 9(17):969.
- Comunello E, Araújo LB, Sentelhas PC, Araújo MFC, Dias CTS, Fietz CR (2013) O uso da análise de cluster no estudo de características pluviométricas. *Sigmae* 2(3):29-37.
- CONAB – Companhia Nacional de Abastecimento (2017) Calendário de plantio e colheita de grãos no Brasil. Available: file:///C:/Users/User/Downloads/CalendarioZAgricola.pdf. Accessed Aug 22, 2019.
- Dhanachandra N, Mangle K, Chanu YJ (2015) Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science* 54:764-771.
- Dourado CS, Oliveira SRM, Avila AMH (2013) Análises de zonas homogêneas em séries temporais de precipitação no Estado da Bahia. *Bragantia* 72(2):192-198.
- Ely DF, Dubreuil V (2017) Análise das tendências espaço-temporais das precipitações anuais para o estado do Paraná – Brasil. *Revista Brasileira de Climatologia* 21:553-569.
- Farias JRB, Assad ED, Almeida IR, Evangelista BA, Lazzarotto C, Neumaier N, Nepomuceno AL (2001) Caracterização de risco de déficit hídrico nas regiões produtoras de soja no Brasil. *Revista Brasileira de Agrometeorologia* 9(3):415-421.
- Fayyad U, Piastetsky-shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI Magazine* 17(3):37-54.
- Fritzens E, Mantovani LE, Wrege MS, Chaves Neto A (2011) Análise da pluviometria para a definição de zonas homogêneas no estado do Paraná. *RA'E GA* 23(1):555-572.
- Gao C, Cheng Q, He P, Willy S, Li J (2018) Privacy-preserving naive bayes classifiers secure against the substitution-then-comparison attack. *Information Sciences* 444:72-88.
- Giasson, E, Hartemink, AE, Tornquist CG, Teske R, Bagatini T (2013) Avaliação de cinco algoritmos de árvores de decisão e três tipos de modelos digitais de elevação para mapeamento digital de solos a nível semidetalhado na Bacia do Lageado Grande, RS, Brasil. *Ciência Rural* 43(11):1967-1973.
- Gonçalves SL, Caramori PH, Wrege MS, Shioiga P, Gerage AC (2002) Épocas de semeadura do milho “safrinha”, no estado do Paraná, com menores riscos climáticos. *Acta Scientiarum Agronomy* 24(5):1287-1290.

- IAPAR – Instituto Agronômico do Paraná (2017a) Zoneamento agrícola. IAPAR. Available: <http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=1044>. Accessed Nov 10, 2017.
- IAPAR – Instituto Agronômico do Paraná (2017b) Cartas climáticas do Paraná. IAPAR. Available: <http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=595>. Accessed Nov 10, 2017.
- Kalaiselvi P, Geetha D (2016) Weather prediction using J48, EM and K-Means clustering algorithms. *International Journal of Innovative Research in Computer and Communication Engineering* 4(12):20889-20895.
- Lima JSS, Silva SA, Oliveira RB, Cecílio RA, Xavier AC (2008) Variabilidade temporal da precipitação mensal em Alegre-ES. *Revista Ciência Agronômica* 39(2):327-332.
- Machado CB, Brand VS, Capucim MN, Martins LD, Martins JÁ (2013) Eventos extremos de precipitação no estado do Paraná. *Ciência e Natura* 35(2):81-83.
- Marzban C, Sandgathe S (2006) Cluster analysis for verification of precipitation fields. *Weather and Forecasting* 21(5):824-838.
- Moreira PSP, Galvanin EAS, Dallacort R, Neves RJ (2016) Análise de agrupamento aplicado ao ciclo diário das variáveis meteorológicas nos biomas do estado de Mato Grosso. *Acta Iguazu* 5(1):80-94.
- Oliveira-Júnior JF, Xavier FMG, Teodoro PE, Gois G, Delgado RC (2017) Cluster analysis identified homogeneous regions in Tocantins State, Brazil. *Bioscience Journal* 33(2):333-340.
- Pansera WA, Gomes BM, Vilas Boas MA, Queiroz MMF, Mello EL, Sampaio SC (2015) Regionalization of monthly precipitation values in the state of Paraná (Brazil) by using multivariate clustering algorithms. *Irriga* 20(3):473-489.
- Pereira VGC, Gris DJ, Marangoni T, Frigo JP, Azevedo KD, Grzesiuck AE (2014) Exigências climáticas para a cultura do feijão (*Phaseolus vulgaris* L.). *Revista Brasileira de Energias Renováveis* 3:32-42.
- Pontes Júnior, AP, Santana Júnior, CJS, Bastos-Filho, CJA (2018) Uso de técnicas de clusterização em uma base de dados financeira. *Revista de Engenharia e Pesquisa Aplicada* 3(3):87-95.
- Quinlan JR (1996) Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research* 4:77-90.
- Richetti J, Cima EG, Johann JA, Uribe-Opazo MA (2018) Data mining techniques for rainfall regionalization in Parana state. *Acta Iguazu* 7(1):1-8.
- Romani LAS, Ávila AMH, Zullo Júnior J, Traina Júnior C, Traina AJM (2010) Mining relevant and extreme patterns on climate time series with CLIPSminer. *Journal of Information and Data Management* 1(2):245-260.
- Salton FG, Morais H, Caramori PH, Borrozzino E (2016) Climatologia dos episódios de precipitação em três localidades no estado do Paraná. *Revista Brasileira de Meteorologia* 31(4):626-638.
- Sampaio SC, Longo AJ, Queiroz MMF, Gomes BM, Vilas Boas MA, Suszek M (2006) Estimativa e distribuição da precipitação mensal provável no Estado do Paraná. *Acta Scientiarum Human and Social Sciences* 28(2):267-272.
- SEAB - Secretaria De Estado Da Agricultura e do Abastecimento (2015) Valor bruto da produção agrícola paranaense em 2015. Available: [http://www.agricultura.pr.gov.br/arquivos/File/deral/VBP\\_2015\\_AnaliseCompletaVD.pdf](http://www.agricultura.pr.gov.br/arquivos/File/deral/VBP_2015_AnaliseCompletaVD.pdf). Accessed Jun 12, 2017.
- Shiroga PS, Gerage AC (2010) Influência da época de plantio no desempenho do milho safrinha no estado do Paraná, Brasil. *Revista Brasileira de Milho e Sorgo* 9(3):236-253.
- Silva AMA (2012) Avaliação do comportamento da precipitação entre o primeiro planalto paranaense e o litoral do Paraná no ano hidrológico 2010/2011. *Revista Geonorte* 2(5):967-974.
- Torres E, Macedo H, Chella M, Matos L (2017) Otimização do algoritmo expectation maximization para o modelo de misturas gaussianas usando cuda. *Revista de Sistemas e Computação* 7(1): 15-20.
- Vanhoni F, Mendonça F (2008) O clima do litoral do estado do Paraná. *Revista Brasileira de Climatologia* 3:49-63.
- Vrac M, Yiou P (2010) Weather regimes designed for local precipitation modeling: Application to the Mediterranean basin. *Journal of Geophysical Research* 115(12):1-23.
- Waltrick PC, Machado MAM, Dieckow J, Oliveira D (2015) Estimativa da erosividade de chuvas no estado do Paraná pelo método da pluviometria: Atualização com dados de 1986 a 2008. *Revista Brasileira de Ciência do Solo* 39(1):256-267.
- Witten IH, Frank E (2005) *Data mining – Practical machine learning tools and techniques*. San Francisco, Elsevier. 506p.
- Zhang Q, Wang J, Lu A, Shouyang W, Ma J (2018) An improved SMO algorithm for financial credit risk assessment –evidence from China’s banking. *Neurocomputing* 272:314-325.