

Soybean Productivity and Agrometeorological Variables Assessed from the perspective of Spatial and Circular Statistics

Edilza Martins da Silva (Corresponding author)

Western Paraná State University – UNIOESTE

1619 Universitária St., Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: edilza_martins@hotmail.com

Priscila Pigatto Gasparin

Federal University of Technology – Paraná (UTFPR)

4232 Brasil Avenue, Medianeira, Paraná, 85884-000, Brazil

Tel: +55 (45) 3240-8101 E-mail: priscilap@utfpr.edu.br

Alex Paludo

Western Paraná State University – UNIOESTE

1619 Universitária St., Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: paludo.alex@hotmail.com

Willyan Ronaldo Becker

Western Paraná State University – UNIOESTE

1619 Universitária St., Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: willyan.becker@unioeste.br

Luciana Pagliosa Carvalho Guedes

Western Paraná State University – UNIOESTE

1619 Universitária St., Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: luciana.guedes@unioeste.br

Jerry Adriani Johann

Western Paraná State University – UNIOESTE

1619 Universitária St., Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-7320 E-mail: jerry.johann@unioeste.br

Received: June 1, 2021 Accepted: August 10, 2021 Published: August 12, 2021

doi:10.5296/jas.v9i3.18697

URL: <https://doi.org/10.5296/jas.v9i3.18697>

Abstract

Climate change can affect the development of soybean cultivation, impacting your productivity. Thus, agrometeorological information is essential in order to improve productivity strategies. The objective of the paper was to analyze the influence and occurrence of seasonality of the following agrometeorological variables on soybean productivity: mean air temperature [TMean] (°C), accumulated rainfall value [Rain] (mm), global solar radiation [Sr] (MJ m⁻² day⁻¹), and potential evapotranspiration [ETp] (mm), in ten-day variations of the the maximum vegetative development date (MVDD), in the 2011/2012 and 2013/2014 harvest years in the state of Paraná. The study was based on spatial distribution of variables, using univariate and bivariate Global Moran's Indexes, and multivariate clustering analysis. To verify seasonality in the time distribution of the agrometeorological variables in the ten-day variations close to soybean MVDD, we used the circular statistics, through the mean vector length (R). Result it was identified regions of the state that have higher and lower rainfall and seasonality, also have higher and lowest productivity, respectively. That the variation in soybean productivity between harvest years was correlated with the agrometeorological variables, and rainfall volume is an important factor in productivity. The other agrometeorological variables occurred uniformly, especially in 2011/2012 harvest year, in the Northwest, Central-northern and West mesoregions. Furthermore, there was clustering of regions with similar spatial distribution of the evapotranspiration and rainfall variables in 2aDMDV2d in the 2011/2012 and 2013/2014 harvest year, showed the same spatial distribution of the agrometeorological variables and the productivity variable.

Keywords: phenological stages, Paraná, Geoprocessing, Moran's Index, seasonality

1. Introduction

Soybean has been the protagonist in the increase of grain production in Brazil in recent years. Paraná is the second state with the highest soybean production, with a mean production of 15 million tons in the last 10 years and with a standard deviation of 3 million tons, that is, the mean productivity across the harvest years showed moderate dispersion; this was confirmed by the value of 20% of the variation coefficient (VC). If Paraná were a country, it would be

the fourth largest soybean producer, with 16 million tons in the 2018/2019 harvest year, producing more than China, which totaled 15.7 million tons (USDA, 2021).

Official data showed great variability in interannual productivity in the state, for example, with a production of 9 million tons (2009/2010) to 19 million tons (2017/2018) in Paraná according to the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística*, IBGE, 2018). The main aspects that affect soybean productivity are agrometeorological factors such as water availability, temperature, photoperiod and evapotranspiration (Klosowski, 1997), as they influence the development of plants in all their phenological phases.

However, water availability is one of the most critical factors for soybean development, especially during germination-emergence and flowering-filling periods. In addition to the need for an adequate water volume in these periods, it must also occur uniformly throughout the crop cycle so as to reach maximum productivity, according to its genetic potential (Farias et al., 2007).

Several studies have evaluated the influence of agrometeorological variables on soybean productivity (Araújo et al., 2014; Radin et al., 2017; Grzegozewski et al., 2017). Araújo et al. (2014) identified a spatial association between soybean productivity and agrometeorological variables such as rainfall, temperature and global solar radiation.

Radin et al. (2017) verified the correlation between soybean productivity and rainfall accumulation during the crop's vegetative cycle in municipalities of Rio Grande do Sul, from 2005 to 2012. The results showed that the variation in productivity could be explained by rainfall, where the occurrence of consecutive days without rain during the crop cycle exerted a negative impact on soybean productivity.

In addition, one study showed that temporal changes (seasonality or uniformity) of the agrometeorological variables significantly affect soybean productivity (Moura et al., 2018). In this context, circular statistics techniques make it possible to verify seasonality of the agrometeorological variables analyzed in a certain period of time (Silva, 2018). Studies that analyzed the relationship between seasonality of the agrometeorological variables and soybean productivity are restricted.

The objective of the paper was to analyze the influence of the following agrometeorological variables on soybean productivity: mean air temperature [TMean] (°C), accumulated rainfall value [Rain] (mm), global solar radiation [Sr] ($\text{MJ m}^{-2} \text{day}^{-1}$), and potential evapotranspiration [ETp] (mm), in ten-day variations of the maximum vegetative development date (MVDD), in a pixel scale, in the 2011/2012 and 2013/2014 harvest years in the state of Paraná. The objective was also to evaluate the occurrence of seasonality of these agrometeorological variables, and their influence on soybean productivity.

2. Materials and Methods

2.1 Determination of the Study Area and Data Acquisition

The study area comprises the state of Paraná, Brazil, limited between parallels 22° 29'S and 26° 43'S and meridians 48° 2'W and 54° 38'W, with a territorial area of 199,298,982 km², larger than Uruguay (176,215 km²), and with a Gross Domestic Product (GDP) of US\$ 89 billion, higher than countries such as Guyana and Montenegro (IBGE 2018; IPARDES, 2020) (Figure 1).

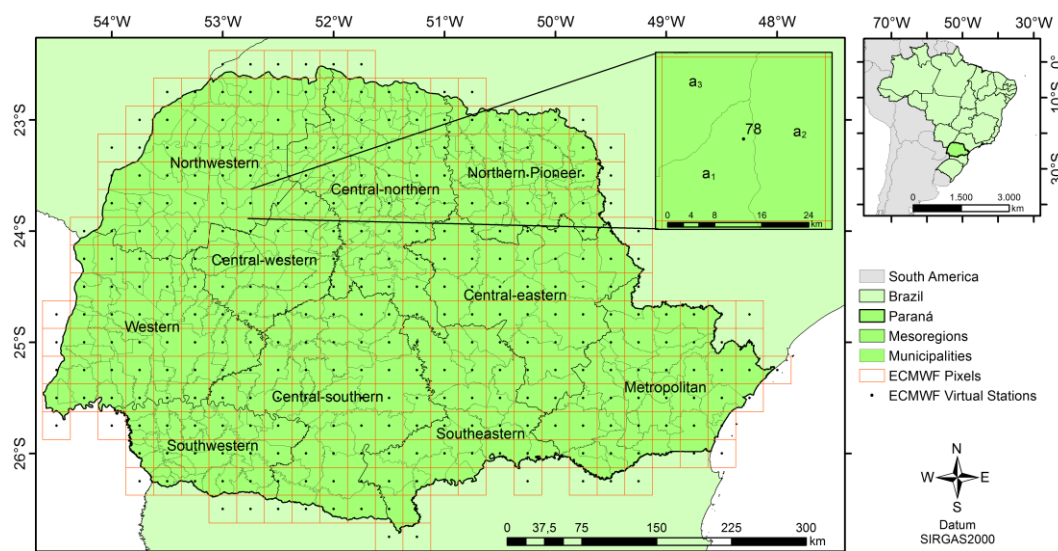


Figure 1. Study area with the 10 mesoregions of the state of Paraná and details of a virtual station (VS) of the European Centre for Medium-Range Weather Forecasts (ECMWF), with the percentage of the area from the 3 municipalities that comprise it (a₁, a₂, a₃)

Data from two harvest years (2011/2012 and 2013/2014), which represent different agrometeorological scenarios from a historical series between 2000/2001 and 2015/2016, were used in this study. The 2011/2012 harvest year was dry, with the lowest mean values of rainfall and productivity in this historical series. The 2013/2014 harvest year presented mean rainfall and productivity, close to the mean of the period analyzed. In addition, it is important to emphasize that the soybean harvest year under analysis was comprised between the 3rd ten-day interval of August (sowing) and the 1st ten-day interval of April of the following (harvest) year. This period was chosen because we want to verify the influence of temperature on a few ten-day periods before planting, since this is determined in the state of Paraná through ordinance No. 202, in the second half of each year (ADAPAR, 2017).

For each harvest year, the agrometeorological data were obtained from the virtual stations (VSs) of the European Centre for Medium-Range Weather Forecasts (ECMWF) model, which is available free of charge on the JRC website (ECMWF, 2018). These data are in shapefile format, with an approximate spatial pixel resolution of 25 x 25 km (Figure 1) and a ten-day time resolution.

Agrometeorological variables, in the most critical phenological phases of the soybean cycle, were analyzed, that is, in the intervals close to the maximum vegetative development date

(MVDD), defined by Becker et al. (2017) as the crop’s reproductive stages (flowering, pod formation and filling). For the two harvest years, the MVDD was obtained from Becker et al. (2020). Thus, two time intervals were defined, in which the first corresponds to a one ten-day variation around the MVDD (1aMVDD1d), totaling three ten-days in this interval. While the second time interval corresponds to a two ten-day variation around the MVDD (2aMVDD2d), totaling five ten-days in the intervals (Figure 2).

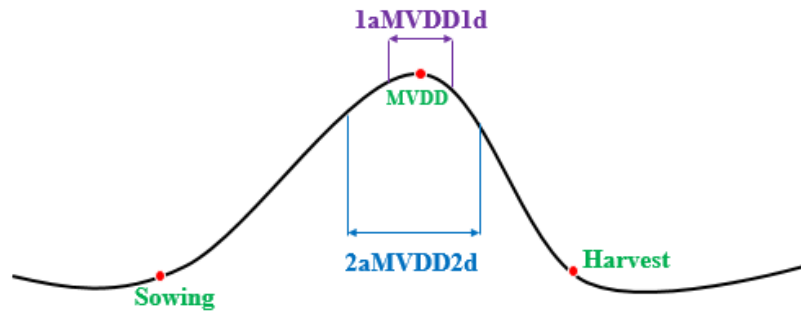


Figure 2. Representation of the first interval around the Maximum Vegetative Development Date (1aMVDD1d) and second interval around the MVDD (2aMVDD2d)

For each interval, measures associated with agrometeorological variables were calculated, namely: mean air temperature [TMean] (°C), accumulated rainfall value [Rain] (mm), global solar radiation [Sr] (MJ m⁻² day⁻¹), and potential evapotranspiration [ETp] (mm).

The soybean productivity data (t ha⁻¹) for the 399 municipalities in the state of Paraná were obtained from the Brazilian Institute of Geography and Statistics (IBGE, 2018). There were no productivity data in 38 and 29 municipalities for the 2011/2012 and 2013/2014 harvest years, respectively, since they did not produce grains.

As there is no spatial coincidence between the perimeters of each municipality and the data pixel grid of the ECMWF (Figure 1), the ECMWF grid was considered as the spatial reference. Hence, it was necessary to determine the soybean weighted mean productivity (WMP) in each VS (Equation 1):

$$WMP = \frac{\sum_{j=1}^n a_j p_j}{\sum_{j=1}^n a_j}, \quad (1)$$

where a_j and p_j are, respectively, the area and productivity of the j -th municipalities within the VS, with $(j = 1, \dots, n)$ where n is the number of municipalities that compose the VS area. In addition, the municipalities that compose the VS and that did not show productivity ($p_j = 0$) were disregarded in the calculation of WMP. Figure 1 represents a VS, as an example, which contains an area percentage of three municipalities.

2.2 Descriptive Statistical Analysis

Initially, a descriptive analysis was performed for each variable and compared between the harvest years; when on average, the agrometeorological and productivity variables were statistically significant at 5% probability by the Z test.

2.3 Spatial Statistics

The spatial distribution of the productivity and agrometeorological variables in the state of Paraná was analyzed by comparing the maps of the two harvest years, using similarity measures of Global Accuracy (GA), Kappa (K) and Tau (T) (Duarte & Silva, 2019).

In addition, the spatial autocorrelation analysis of soybean productivity and agrometeorological variables was performed with Moran's Global Univariate Index (Equation 2) (Câmara et al., 2004). The spatial correlation between soybean productivity and the agrometeorological variables was determined by Moran's Global Bivariate Index (Equation 3) (Araújo et al., 2014). For each variable, the Jones test (Jones, 1969) was applied, with a 5% significance level, to evaluate normal distribution of the probabilities. For those that did not present this pattern, Johnson's transformation was performed (Yeo & Johnson, 2000) or Box-Cox (Box & Cox, 1964).

$$I = \frac{v \sum_{i=1}^v \sum_{j=1}^v z_i z_j w_{ij}}{S_0 \sum_{i=1}^v z_i^2}, \quad (2)$$

$$I_{xy} = \frac{\sum_{i=1}^v \sum_{j=1}^v u_i z_j w_{ij}}{S_0 \sqrt{S_u^2} \sqrt{S_z^2}}, \quad (3)$$

where v is the number of VSs considered in this study, in the state of Paraná, $z_i = (x_i - \bar{x})$ and $z_j = (x_j - \bar{x})$, where $i \neq j$ and $i, j = 1, \dots, v$, are the observed values of the same variable (X) in the i -th and j -th VSs, respectively, centered on the mean (\bar{x}) of this variable under study; $u_i = (y_i - \bar{y})$ are the values of a variable Y observed in the i -th and j -th VSs, respectively, centered on their respective mean (\bar{y}); S_u^2, S_z^2 are the variances of X and Y, respectively; w_{ij} represents the element located in the i -th row and j -th column ($i, j = 1, \dots, n$), the symmetric spatial weight matrix $\mathbf{W}_{n \times n}$, in which its elements represent the proximity between the i -th and the j -th VSs. If the VSs have a common border, then $w_{ij} = 1$; and, when they have no common border, then $w_{ij} = 0$. In addition to that, S_0 is the sum of elements w_{ij} ($i, j = 1, \dots, n$). For this study, the symmetric matrix of spatial weight (Torre) was

considered (Câmara et al., 2004).

Moran's univariate index varies from -1 to 1, with its positive values (from 0 to 1) indicating a direct spatial autocorrelation of a variable, that is, regions with high values have neighbors with high values as well. The same occurs with the low values of this variable (Câmara et al., 2004); whereas the negative values of this index indicate an inverse spatial autocorrelation of a variable, that is, regions with high values are surrounded by regions with low values, or vice versa (Anselin et al., 2005).

Moran's bivariate index aims at verifying whether the values of one variable observed in one region are related to those of another variable observed in a neighboring region, with its values also ranging from -1 to 1 (Câmara et al., 2004).

After calculating Moran's bivariate and univariate indexes, the pseudo significance Moran's index test was performed. In this case, different permutations of the observed values associated with the pixels were generated and the statistics were recalculated; this process was repeated, thus generating a reference distribution. In this way, if the Moran's index value calculated initially corresponds to an extreme of the generated distribution, then the index value has statistical significance (Câmara et al., 2004).

2.4 Circular Statistics

Due to the nature of the agrometeorological variables, which were observed in a certain period of time, these can be called circular data and can be analyzed using circular statistical techniques (Pewsey et al., 2013).

For this, a descriptive measure was used to verify the existence of uniformity or seasonality of an event in a given period of time, called mean vector length (R) (Beskow et al., 2014).

Thus, for each harvest year and each agrometeorological variable, the R value was determined, considering the two ten-day intervals (1aMVDD1d and 2aMVDD2d) as period of time.

To use the R value, some procedures must be performed first, such as organizing the data (Zar, 2010). Each period analyzed corresponded to the angular interval of a circle $[0^\circ, 360^\circ]$. In this way, the entire ten-day period was transformed into an angle. Thus, the 3 ten-day periods of the 1aMVDD1d interval turned into angles 0° , 120° , and 240° and the 5 ten-day periods of the 2aMVDD2d interval were converted into angles 0° , 72° , 144° , 216° , and 288° . Figure 3 shows a fictitious example of a circular data set, used only to illustrate the procedure for calculating the mean vector length.

For each i -th angle, there is an amount of the analyzed variable that is represented by the frequency (f_i) (Figure 3), where $i = 1, \dots, k$, k being the number of angles of the ten-day interval. In this way, after organizing the data, it is possible to determine the value of the

mean vector length (Equation 4) (Beskow et al., 2014).

$$R = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (4)$$

where (\bar{x}, \bar{y}) are cartesian coordinates defined as $\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i \cos \theta_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^k f_i \sin \theta_i$,

with f_i being the angular frequency observed at the i -th angle θ_i , with $i = 1, \dots, k$, such that

$$n = \sum_{i=1}^k f_i.$$

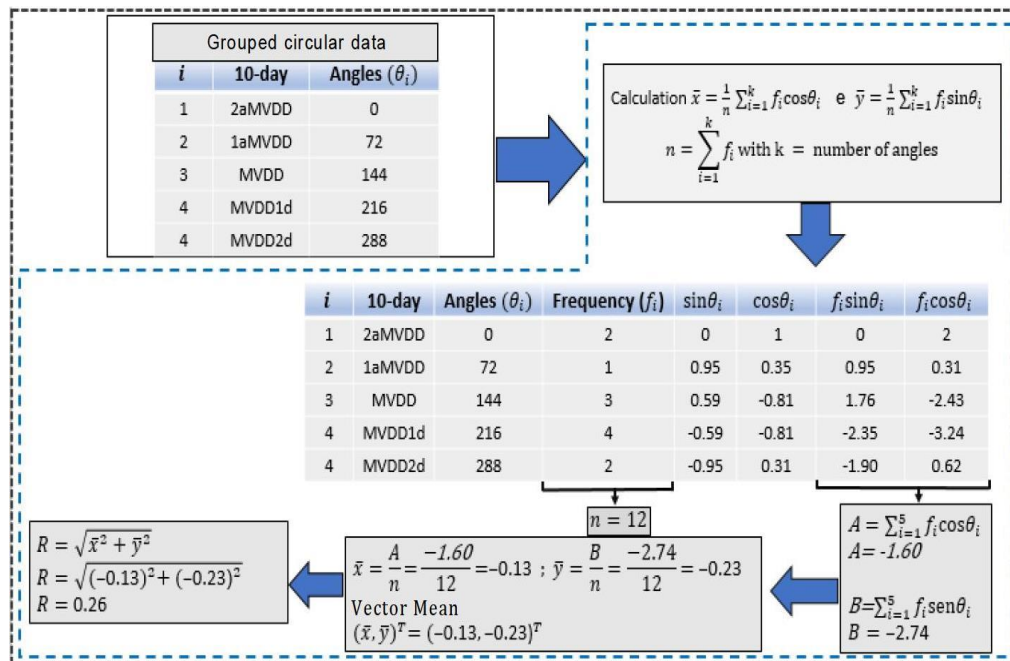


Figure 3. Example of grouped circular data and procedures for calculating mean vector length R , considering the 2aMVDD2d 10-day interval

This circular measure varies from 0 (it indicates uniform distribution of the agrometeorological variable, in the ten-day variation) to 1 (it indicates occurrence of seasonality of the agrometeorological variable in the ten-day variation) (Beskow et al., 2014).

In the sequence, for each ten-day interval and each agrometeorological variable that presented seasonality, the circle graph that describes the ten-day variation and the mean vector was created. Figure 4-A shows the representation of the mean vector (in orange) of a simulated circular data set. The length of this vector indicates data uniformity, and its direction indicates that, on average, the data occur close to a ten-day period before the MVDD (1aMVDD). In Figure 4-B, the length of the mean vector (in blue) indicates data seasonality and its direction indicates that, on average, data concentration is close to two ten-day periods before the MVDD (2aMVDD).

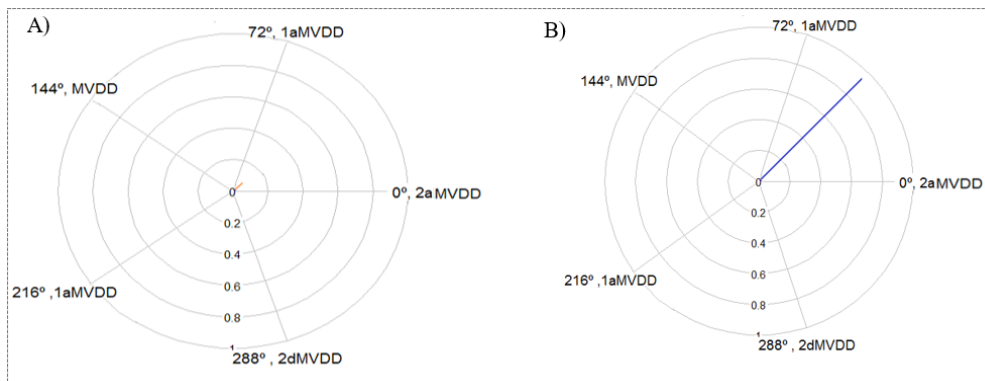


Figure 4. Graphical representation of the mean vector for two fictitious examples that present: A) uniformity in data distribution (vector length in orange); B) seasonality in data distribution (vector length in blue)

For each agrometeorological variable and considering the *R* value, the Chi-square test was applied at a 5% significance level to confirm the existence of uniformity, for grouped circular data (Pewsey et al., 2013).

2.5 Multivariate Analysis

A cluster analysis was also carried out, for the VSs of the ECMWF model, considering productivity and the measures associated with the agrometeorological variables in each harvest year and the length of the *R* vector of the rainfall variable. For grouping, the K-means method was used, which is one of the non-hierarchical grouping methods and consists of allocating each sample element in the cluster, whose centroid (sample mean vector) is the closest to the observed value of the vector of the respective element (Mingoti, 2013). In addition, the number of groups was defined by the Calínski criterion (Calínski & Harabasz, 1974).

In Figure 5 we summarize the steps performed in the research.

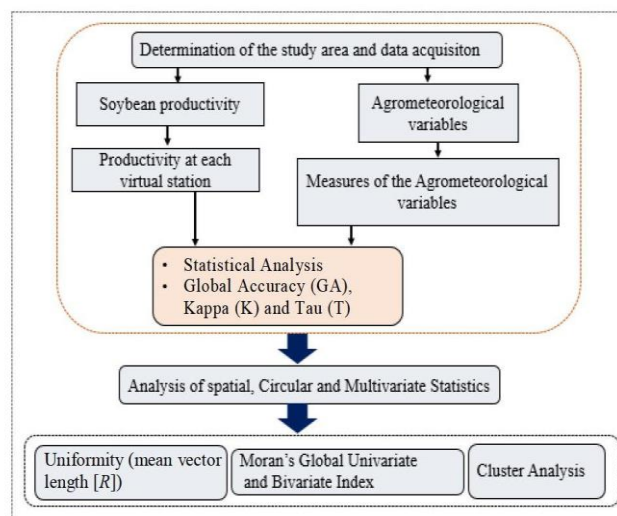


Figure 5. Flowchart of the paper

For the spatial analysis of the area and preparation of maps, the Geoda 1.12 (Anselin et al., 2005) and ArcGIS 10.0 (ESRI, 2015) software programs were used. The R software (R DEVELOPMENT CORE TEAM, 2018) was used for the following statistics: descriptive analyses, calculation of the mean vector length, performance of the Chi-square test, elaboration of the graph that describes the mean vector, and cluster analysis.

3. Results and Discussion

3.1 Descriptive and Spatial Analysis of Soybean Productivity and Agrometeorological Variables

The weighted mean of soybean productivity in Paraná in the 2011/2012 harvest year was 2.6 t h^{-1} , similar to a Brazilian national productivity study which was (IBGE, 2018). In the 2013/2014 harvest year, weighted mean productivity was 2.9 t h^{-1} , higher than the national mean productivity, which was 2.8 t h^{-1} (IBGE, 2018).

Thus, the 2013/2014 harvest year showed, on average, higher soybean productivity in relation to the 2011/2012 harvest year. For this confirmation, the Z test was used, which showed that the mean productivity between the mean harvests is statistically different at 5% significance ($p\text{-value} < 0.05$). Thus, the smaller dispersion (VC) of the agrometeorological variables in relation to the 2011/2012 harvest year and the amount in the ten-day intervals of the agrometeorological variables (Table 1) influenced this difference in productivity.

When comparing the means of each agrometeorological variable, between the harvest years and through the Z test, only rainfall in the 1aMVDD1d ten-day interval was statistically equal in the two harvest years (5% significance) (Table 1). This difference between the harvest years also occurred when the maps generated for each variable under study (Figures 6 and 7); and by similarity measures of Global Accuracy, Kappa and Tau concordance indexes (Table 2), only rainfall in the two ten-day intervals showed similarities between the harvest year maps (Duarte & Silva, 2019).

Soybean productivity in the state of Paraná in the 2011/2012 harvest year (Figure 6-A) presented a value close to the state's mean in 27% of the state's pixels. The highest yields were between 2.76 t ha^{-1} and 3.86 t ha^{-1} , which accounted for 35.74% of the state (mainly in the Central-eastern and Southeastern mesoregions). On the other hand, the lowest yields (from 0.87 t ha^{-1} to 2.17 t ha^{-1}) occurred in 28% of the state's pixels, mainly in the Northwestern, Southwestern, and most of the Western mesoregions.

The 2013/2014 harvest year presented values close to or above the state's mean in 54% of the state's pixels, with predominance of the Western, Southeastern, Central-western and Central-southern regions, as well as part of the Central-eastern, Central-northern and Metropolitan Region. On the other hand, the lowest productivity in the state (18%) was concentrated in the northern part of the state (Northwestern and Northern Pioneer mesoregions and part of the Central-northern region). When compared to 2011/2012, the 2013/2014 harvest year showed a higher mean soybean yield, greater amplitude, and less data dispersion (VC) (Table 1, Figure 6-B). This trend in the spatial distribution of soybean productivity for the 2013/2014 harvest year was similar to that found by Franchini et al.

(2016) when considering the spatial and time distribution of soybean productivity in this state, between the 1999/2000 and 2012/2013 harvest years.

Table 1. Descriptive statistics of the weighted mean productivity (WMP t ha⁻¹) of soybean and agrometeorological variables in the intervals around the ten-day intervals around the maximum vegetative development date (MVDD)

Statistics	WMP t ha ⁻¹	TMean °C	1aMVDD1d		2aMVDD2d		TMean °C	Sr MJ m ⁻² day ⁻¹	ETp mm	Rain mm
			Sr MJ m ⁻² day ⁻¹	ETp mm	Rain mm	Sr MJ m ⁻² day ⁻¹				
2011/2012	Minimum	1.1	19	0.5	56	49	27	0.9	165	124
	Mean	2.6	23	0.7	221	153a	23	1.2	271	246
	VC (%)	19	9	5	13	35	8	4	18	20
2013/2014	Minimum	0.1	16	0.5	56	41	17	1.3	136	93
	Mean	2.9	21	0.7	129	152a	21	1.1	222	237
	VC (%)	26	11	9	24	46	10	6	21	29
	Maximum	3.6	25	0.8	196	308	25	1	303	388

Note. TMean: Mean temperature; Sr: Accumulated solar radiation; ETp: Accumulated potential evapotranspiration; Rain: Accumulated rainfall; 1aMVDD1d: first ten-day interval around the MVDD; 2aMVDD2d: second ten-day interval around the MVDD; VC: Variation Coefficient; the letter “a” on the table values represents that there is no statistically difference between the harvest years; in relation to the mean for each variable ($p \geq 0.05$), the Z test was used.

Table 2. Agreement indexes for the 2011/2012 and 2013/2014 harvest year maps of the agrometeorological variables

Indexes	TMe °C	Sr MJ m ⁻² day ⁻¹	ETp mm	Rain mm	TMean °C	Sr MJ m ⁻² day ⁻¹	ETp mm	Rain mm
Global Accuracy	0.16	0.69	0.42	0.90	0.29	0.61	0.49	0.83
Kappa	-0.04	0.43	0.13	0.83	0.08	0.29	0.04	0.58
Tau	0.16	0.69	0.42	0.90	0.29	0.61	0.49	0.74

Note. TMean: Mean temperature; Sr: Accumulated solar radiation; ETp: Accumulated potential evapotranspiration; Rain: Accumulated rainfall; 1aMVDD1d: first ten-day interval around the MVDD; 2aMVDD2d: second ten-day interval around the MVDD.

This difference in productivity between the harvest years (Figures 6-A and 6-B) was possibly due to the lack of rainfall throughout the crop cycle in the 2011/2012 harvest year, which presented 40% less rainfall when compared to 2013/2014. In addition, lack of rainfall in the phenological stages that are located in the 2aMVDD2d interval were compromised, as this interval was statistically different on average (Table 1). This result is explained by the influence that rainfall exerts on soybean productivity, especially in the grain filling phase (Radin et al., 2017).

In addition, the mean rainfall values in the state were 827 mm (2011/2012 harvest year) and 1,340 mm (2013/2014 harvest year). This is one of the main causes of productivity variability

from one year to the other (Ferreira et al., 2007).

The agrometeorological variables: mean temperature, solar radiation and evapotranspiration, showed higher values in the 2011/2012 harvest year when compared to 2013/2014, which also justifies the difference in productivity between the two harvests. Studies conducted by Grzegozewski et al. (2017) in the state of Paraná in the 2010/2011, 2011/2012 and 2012/2013 harvest years, show that different soybean yields between harvest years are influenced by agrometeorological variables such as lack of rainfall, high temperature values, global radiation, and evapotranspiration during the crop cycle.

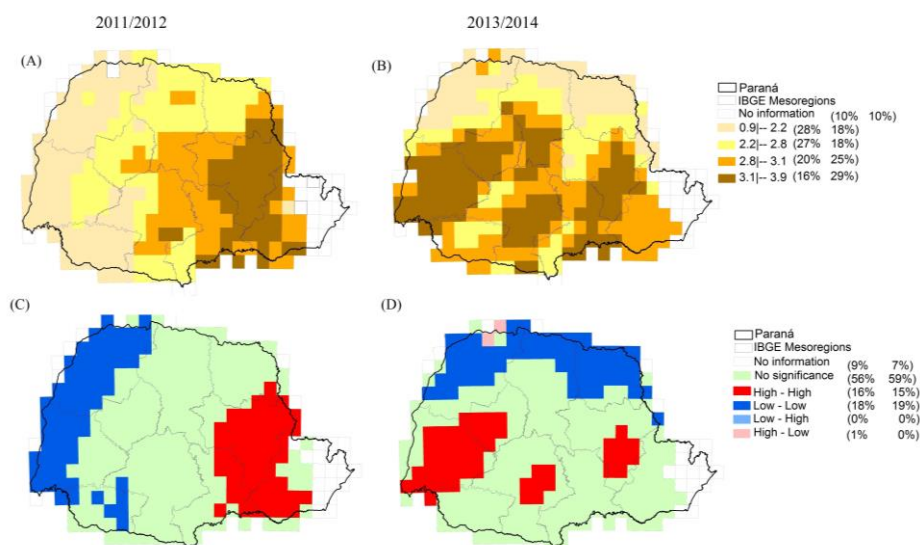


Figure 6. Maps of soybean productivity (t ha⁻¹) (A) and (B); and Local Indicator of Spatial Association (LISA) grouping maps (C) and (D) in the state of Paraná for each harvest year

Note. *I* = Moran's Global Univariate Index; pseudo significance test at 5% probability was used; (b) = Transformed data to show normal distribution behavior.

A positive and significant spatial autocorrelation was observed ($I = 0.81$) in productivity in the two harvest years. That is, pixels with high (or low) productivity have neighboring pixels that follow the same pattern of high (or low) productivity, respectively. This is evidenced by the formation of two clusters (Figures 6-C and 6-D) of High-High pixels (in approximately 15% of the area) and Low-Low pixels (in approximately 18% of the area) in the two harvest years, which follow the same pattern of low and high soybean yields (Figures 6-A and 6-B), although spatially significant. These results corroborate those found by Felema et al. (2016) who, when analyzing soybean productivity in the state between 2000 and 2010, identified univariate spatial autocorrelation of productivity with regions of high or low productivity being neighbors of regions with similar characteristics.

In the two harvest years and in all the ten-day intervals, there was a positive and significant univariate spatial autocorrelation for all agrometeorological variables (Figure 7). For all agrometeorological variables, this indicates the existence of regions of high and low values,

surrounded by neighboring regions with the same characteristics.

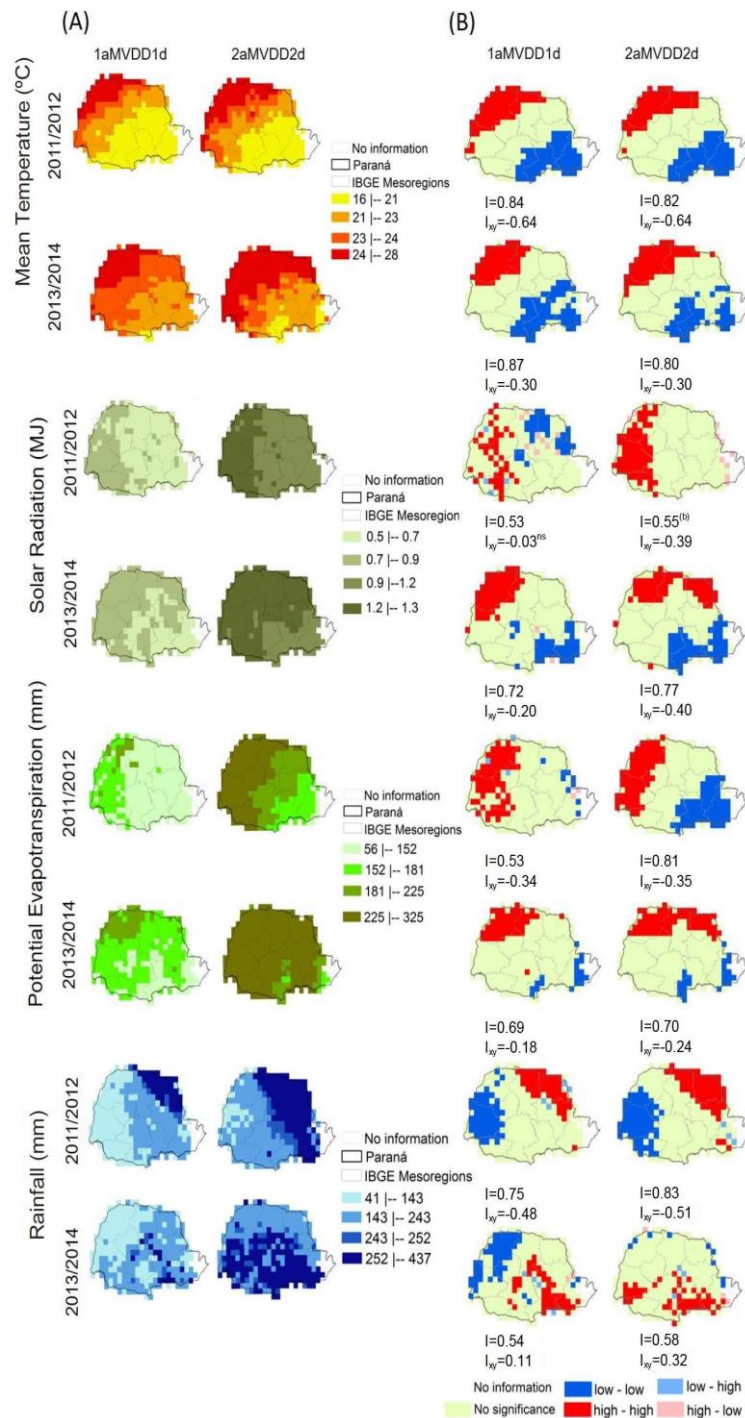


Figure 7. Spatial distribution of the agrometeorological variables (A) and Local Indicator of Spatial Association (LISA) cluster map (B), in the state of Paraná, in the 2011/2012 and 2013/2014 harvest years

Note. I = Moran's Global Univariate Index; I_{xy} = Moran's Global Bivariate Index between the productivity variable and the other variables, in which for both indexes, the pseudo significance test at 5% probability was used, (ns) = Not significant at 5% significance, (b) = Data that was not necessary to transform to show normal distribution behavior.

In the 2011/2012 harvest year and in the two ten-day intervals, mean temperature, accumulated solar radiation, and accumulated potential evapotranspiration presented higher values and formation of the High-High cluster (Figure 7) mainly in the following mesoregions: Northwestern, Western and Southwestern, with these regions showing lower productivity. In addition, the temperature variable obtained the highest spatial autocorrelation values, indicating similarity of temperatures across the municipalities, that is, low spatial variability (Grzegozewski et al., 2017). This is due to the fact that the state has a mean annual temperature between 15°C and 24°C (Aparecido et al., 2016).

In 2011/2012, there was formation of low-low clusters for the temperature variable in the two ten-day intervals and evapotranspiration in the 2aMVDD2d ten-day interval, in most of the following mesoregions: Southeastern, Central-southern, Curitiba Metropolitan and Central-eastern.

In addition, in the 2011/2012 harvest year, in the two ten-day variations, for the rainfall variable, there was a spatial grouping of the highest rainfall values (high-high) in parts of the Central-western, Southeastern, and Western mesoregions (Figure 7-B). In the 2013/2014 harvest year, in the 1aMVDD1d interval, the highest rainfall values (high-high) occurred in parts of the following mesoregions: Northwestern, Central-eastern, and Western. In both ten-day variations for the 2013/2014 harvest year, rainfall showed a direct bivariate spatial correlation with productivity (Figure 7-B).

The spatial distribution maps and the LISA map of productivity and rainfall (Figures 6 and 7) evidenced these results, which indicate regions of Paraná that simultaneously have high rainfall and productivity, as well as regions that simultaneously have low rainfall and productivity. This result is explained by the influence that rainfall exerts on soybean productivity, especially in the grain filling phase (Silva et al., 2014).

The Bivariate Moran's Index (Figure 7-B) shows the occurrence of an inverse and significant spatial correlation ($-0.40 \leq I_{xy} \leq -0.20$ and $p\text{-value} < 0.05$) between productivity and agrometeorological variables (temperature, potential evapotranspiration, and solar radiation), for the two harvest years in all the ten-day intervals considered, since the only value that was not significant of the bivariate Moran's index occurred for solar radiation, in the first ten-day variation (1aMVDD1d) in 2013/2014 harvest year. These results, as well as the analysis of the spatial distribution maps and LISA Map (Figures 6 and 7), show that regions which presented lower productivity had higher temperatures, solar radiation and evapotranspiration, and vice versa. In this context, the Northwestern and Western mesoregions, with high-high type clusters for temperature, solar radiation and evapotranspiration, and the lowest productivity values, concentrating in the December intervals of the 2011/2012 harvest year, focusing clusters of the low-low type for productivity.

In contrast, the Southeastern and Metropolitan regions, in 2011/2012, presented the highest productivity, with high-high clusters for productivity, as well as low-low clusters for temperature (in the two ten-day intervals) and evapotranspiration (in 2aMVDD2d).

In general, in regions with lower productivity there were higher temperatures (from 23°C to

28°C) and, consequently, higher radiation and evapotranspiration, thus reaching their productive potential (Alberto et al., 2006; Grzegozewski et al., 2017). However, the rainfall amount in these regions was not enough for better grain development according to Farias et al. (2007).

In addition, the type of soil also influenced soybean productivity, as in regions with lower yields the soil has low clay content and, consequently, low water retention in the soil, requiring more rainfall. On the other hand, in regions with higher productivity, with higher clay content, there is greater retention of water in the soil, requiring less rainfall (Farias et al., 2005). The spatial distribution of productivity is similar to the spatial distribution of favorable and unfavorable regions for soybean cultivation (Farias et al., 2007).

3.2 Uniformity of Agrometeorological Variables and Their Spatial Association with Productivity

The spatial distribution of the mean vector length (R) to mean temperature, solar radiation, and potential evapotranspiration presented, throughout the state, values close to zero (Figure 8), with mean temperature and solar radiation showing R values that were not significant at 5% probability (Chi-square test) in the distribution of their values over the ten-day variations considered (Figure 9), indicating uniformity.

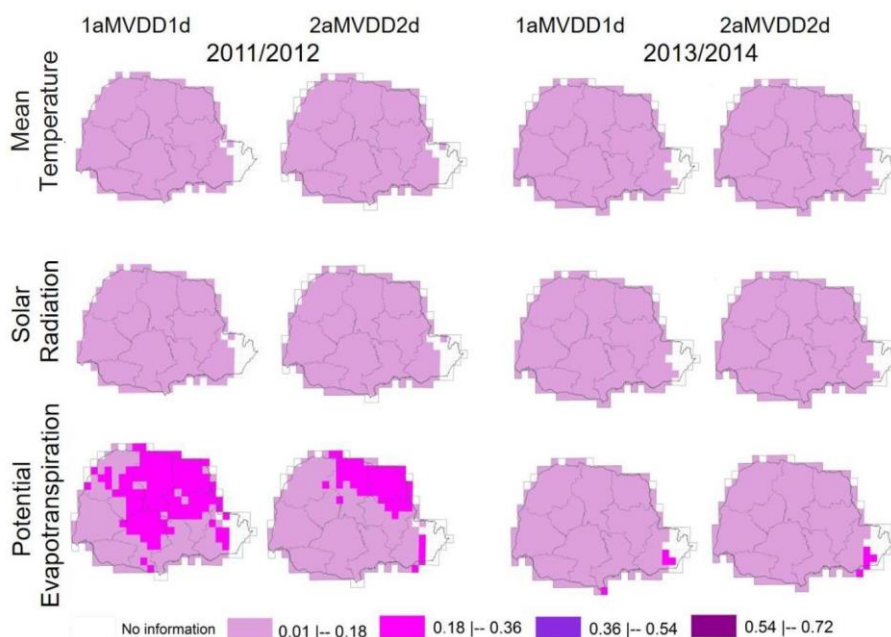


Figure 8. Spatial distribution of the mean vector length (R) of the mean temperature, solar radiation, and potential evapotranspiration variables of the ten-day variations in the harvest years under study

Although the evapotranspiration variable had significant R values by the Chi-square test, in

some pixels (Figure 9), this variable was still considered as a uniform distribution of its values in the ten-day variations. Hence, the values of the mean vector length varied from 0.18 to 0.36 (Figure 8); these values are considered low (on a scale from 0 to 1).

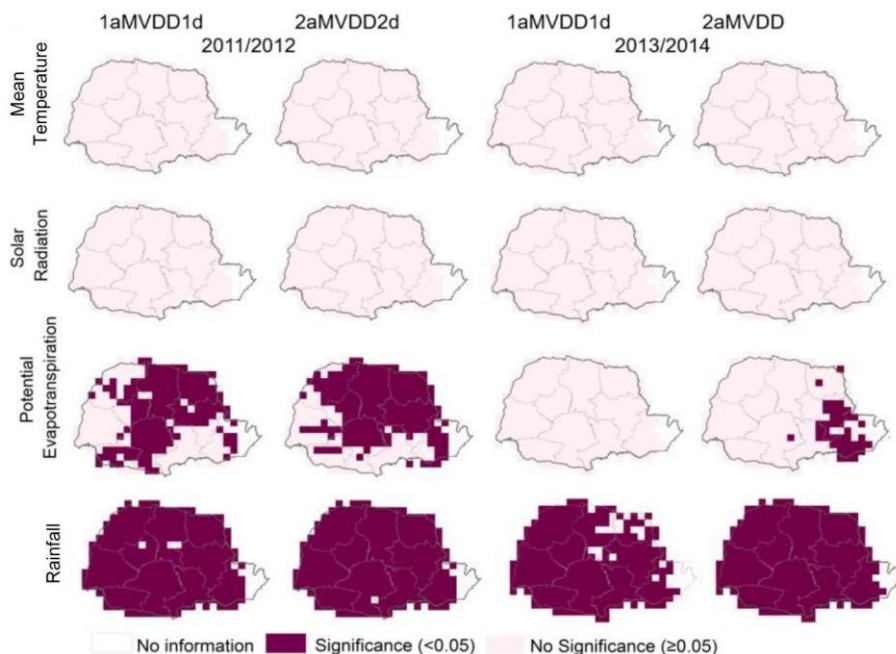


Figure 9. Spatial distribution of the Chi-square test used to evaluate significance of uniformity of the values of agrometeorological variables over the ten-day intervals

Therefore, the only variable that did not show uniformity, that is, which had seasonality, was rainfall (Figures 8 and 9), in almost all regions of Paraná. This is because, in the 1aMVDD1d ten-day variation of the 2011/2012 harvest year (Figure 10), the R values are between 0.36 and 0.54 in 60.36% of the state, with greater concentration of the rainfall peak close to 1aMVDD.

In addition, when considering the 2aMVDD2d ten-day variation, in that same harvest year (Figure 10), the R value was between 0.18 and 0.36 in 53% of the state, with the rainfall peak close to the MVDD until 2aMVDD. The highest R values occurred in 31% of the state, concentrated in some mesoregions with higher productivity (Figure 6-A), such as Central-eastern. It is noteworthy that there was no pixel or VS with the lowest group (Figure 10).

In the 2013/2014 harvest year, the 1aMVDD1d ten-day interval showed R values between 0.36 and 0.72 in 48% of the state, with the rainfall peak occurring between the MVDD and immediately after 1aMVDD (Figure 10). This was the case in the mesoregions with the highest productivity and rainfall volume (Figures 6-B and 7-A), such as the Western, Southeastern, and Central-southern regions. When considering the 2aMVDD2d interval, it shows that 53% of the state had R values between 0.36 and 0.72, also in mesoregions with higher productivity (Figure 6-B), such as Central-eastern, Central-southern, and Northern Pioneer.

Thus, in the 2aMVDD2d interval of the two harvest years, it was observed that, in the regions that presented seasonality and more rainfall, higher productivity also occurred. This is possibly due to the need for water during the crop's flowering and grain filling periods (Farias et al., 2007).

The mean moment of the greatest rainfall volume (Figure 10) occurred close to 1aMVDD, except in the Western and Curitiba Metropolitan regions, for the 2011/2012 harvest year and in the 1aMVDD1d ten-day interval. In the 2aMVDD2d variation of that same year, the mean moment of the greatest rainfall volume was in practically the entire ten-day variation.

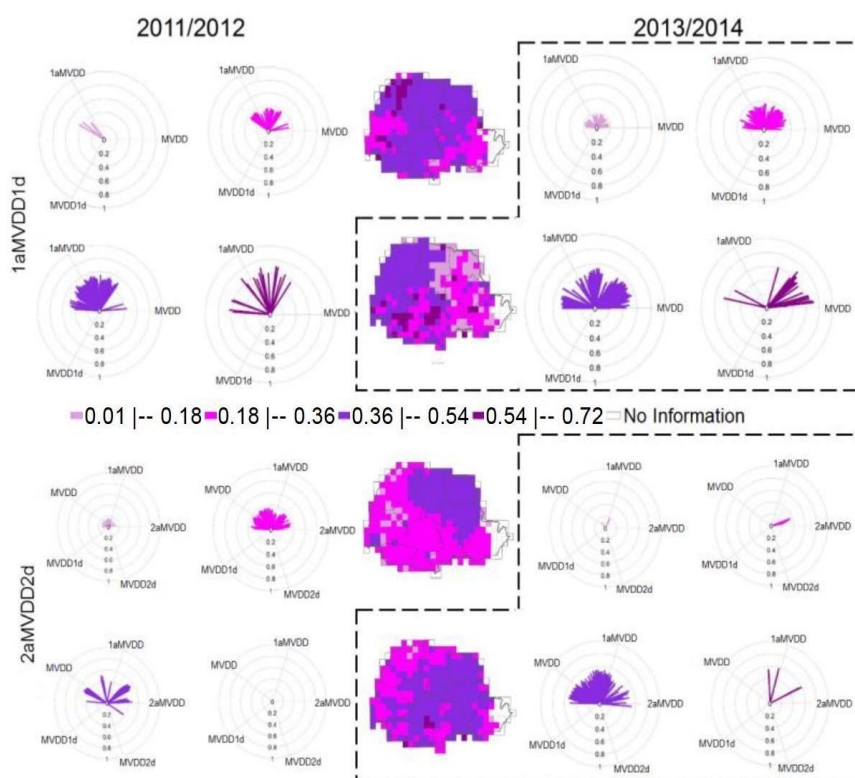


Figure 10. Spatial distribution of mean vector length R and graphical representation for each class of the mean vector length in the circle that corresponds to each ten-day interval, in the 2011/2012 and 2013/2014 harvest years

In the two harvest years and in the two ten-day intervals, the analysis of the global spatial autocorrelation of the mean rainfall vector length was positive and significant, with the formation of two clusters, one of the high-high type, which in this context refers to the presence of seasonality; and the other of the low-low type, which similarly refers to the presence of uniformity (Figure 11).

In the 2011/2012 harvest year and in the 1aMVDD1d interval, the first high-high type occurred in parts of the Northwestern, Central-northern and Northern Pioneer mesoregions, that is, higher R values (seasonality) occurred with lower productivity.

The low-low type (uniformity) occurred in part of the Central-southern and Curitiba Metropolitan mesoregions. For the 2aMVDD2d variation, the high-high type occurred in the Northern Pioneer, Central-northern, and Central-eastern mesoregions (Figure 11).

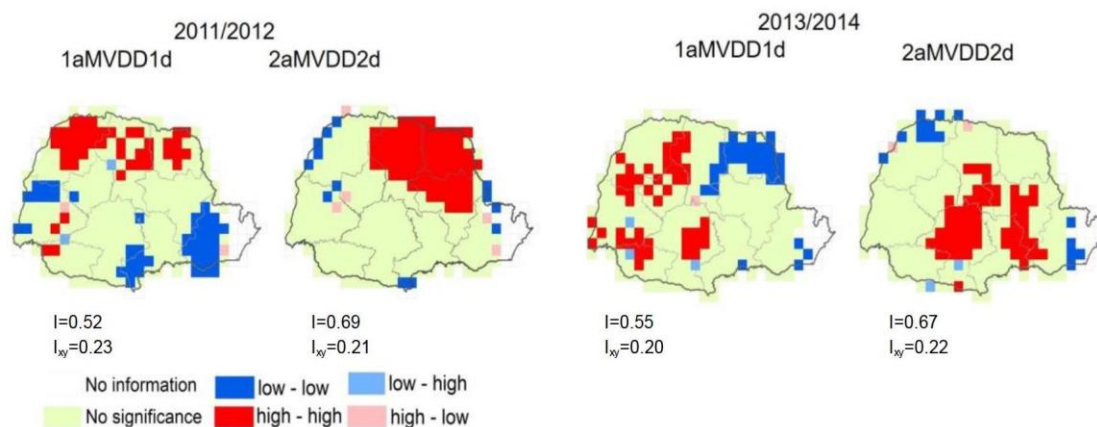


Figure 11. Local Indicator of Spatial Association (LISA) cluster map of the mean vector length (R) of the rainfall agrometeorological variable

Note: I = Moran's Global Univariate Index; I_{xy} = Moran's Global Bivariate Index between the productivity variable and the rainfall variable, both indexes used the significant test at 5% significance.

In the 2013/2014 harvest year and in the 1aMVDD1d variation, there was formation of the high-high type in parts of the Central-southern, Central-western, Northwestern, Southwestern and Western mesoregions, and the low-low type occurred in a larger number of pixels or VSs, mainly in Northern Pioneer. For the 2aMVDD2d variation, the high-high type occurred in the mesoregions of the center of the state, and the low-low type only in some VSs in the Northwestern and Curitiba Metropolitan area (Figure 11).

Analyzing the bivariate spatial correlation between of soybean yield and mean vector length for the two harvest years, we observed that there was a positive ($0.20 \leq I_{xy} \leq 0.23$) and significant at 5% probability.

In other words, in the regions with the highest productivity, there was also seasonality in rainfall distribution over the ten-day variations analyzed. On the other hand, regions with lower productivity showed uniform rainfall distribution over the ten-day intervals, probably due to the fact that the rainfall amount was not sufficient in the phenological phases analyzed, and due to other factors, that also influenced the type of soil (Farias et al., 2005; Farias, et al., 2007).

3.3 Grouping of Agrometeorological Variables, Productivity and R Values

Simultaneously considering the two ten-day variations, each harvest year, VS grouping was performed using the agrometeorological variables: productivity and rainfall mean vector length (R), as these variables presented high and significant R values, indicating seasonality practically in the entire state. Thus, groups that represent regions of similarity of the variables

considered were obtained (Figure 12).

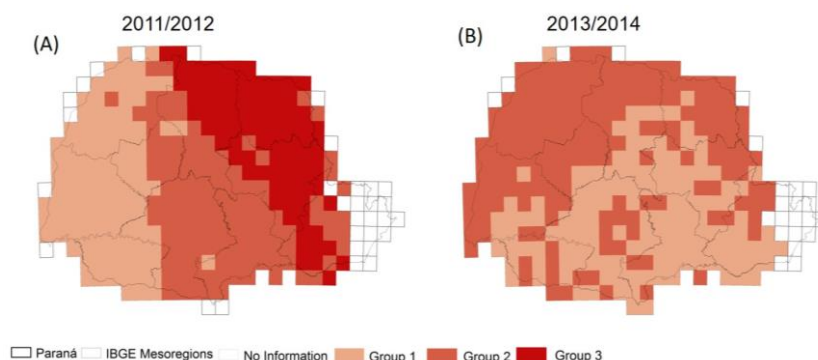


Figure 12. Spatial distribution of the groupings obtained for the 2011/2012 (A) and 2013/2014 (B) harvest years

Thus, for 2011/2012 (Figure 12-A), three groups were formed, the first being in the Western, Southwestern, and Northwestern regions. On the other hand, the second group is located in the Central-southern, Southeastern and part of the Western Center and Central-northern mesoregions, and the third group comprises the Northern Pioneer and part of the Central-northern and Central-eastern mesoregions. That is, in each group, the variables considered occur similarly in these mesoregions. In addition, these groups are distributed practically in the same regions with similar evapotranspiration and rainfall values (Figure 7-A).

On the other hand, in the 2013/2014 harvest year (Figure 12-B), two groups were formed, one concentrated in the northern part of the state and the other in the southern part of the state. Similarity was also found of the same regions with the spatial distribution of agrometeorological and productivity variables in this harvest year (Figures 6-B and 7-A).

In studies carried out on soybean yield and agrometeorological variables in the state of Paraná, similar results were observed, as reported by Grzegozewski et al. (2020) in the 2010/2011, 2011/2012 and 2012/2013 harvest years, concluding that, due to climatic variations, the 2011/2012 harvest year had better soybean yield, and some municipalities in the Northwest mesoregion were identified with low yields, due to the low potential of the region.

4. Conclusions

The results showed that the variation in soybean productivity between harvest years was correlated with the agrometeorological variables. It was identified that, in the regions of the state that present greater and lower rainfall volumes, the regions with the highest and lowest productivity were found, respectively.

In addition, rainfall was the only variable that showed seasonality along the ten-day intervals close to the MVDD practically in the entire state, verifying that rainfall volume in the ten-day intervals analyzed was an important factor for greater productivity.

On the other hand, the mean temperature, solar radiation and evapotranspiration variables

showed an inverse spatial relationship with productivity, that is, their greater intensity caused lower yields in the state, especially in 2011/2012. In addition, for these variables, there was uniformity in the ten-day intervals close to the MVDD, causing lower productivity in the Northwest, Central-northern and West mesoregions, especially in the 2011/2012 harvest year.

In addition, in the 2011/2012 harvest year, there was formation of a cluster of regions in the state of Paraná with the same spatial trend of the evapotranspiration and rainfall variables in 2aDMDV2d and, in the 2013/2014 harvest year, the clusters of the regions of the state of Paraná were similar in the spatial distribution of the agrometeorological and the productivity variables.

Acknowledgements

The authors would like to acknowledge the Coordination of Superior Level Staff Improvement (CAPES) for its financial support, Financing Code 001, and National Council for Scientific and Technological Development (CNPq), and the Applied Statistics Laboratory (LEA) members of the Western Paraná State University (UNIOESTE) – Cascavel – Paraná - Brazil.

References

- Adapar, Agência de Defesa Agropecuária do Paraná (2017). Estabelece o período do vazio sanitário, as datas limites para a semeadura e colheita e outras medidas de controle da ferrugem asiática (*Phokopsora pachyrhizi*) no estado do Paraná. n°202, de 19 de julho de 2017. [Online] Available: http://www.aenda.org.br/wp-content/uploads/2020/04/cir1208-anexo-parana_port-202-2017.pdf (January 20, 2020).
- Alberto, C. M., Streck, N. A., Heldwein, A. B., Buriol, G. A., & Medeiros, S. L. P. (2006). Água no solo e rendimento do trigo, soja e milho associados ao El Niño Oscilação Sul. *Pesquisa agropecuária brasileira*, 41(7), 1067-1075. <https://doi.org/10.1590/S0100-204X2006000700001>
- Anselin, L., Syabri, I., & Kho, Y. (2005). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38(1), 5-22. <https://doi.org/10.1111/j.0016-7363.2005.00671.x>
- Aparecido, L. E. O., Rolim, G. S., Richetti, J., Souza, P. S., & Johann, J. A. (2016). Koppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. *Ciência e Agrotecnologia*, 40(4) 405-417. <https://doi.org/10.1590/1413-70542016404003916>
- Araújo, E. C., Uribe-Opazo, M. A., & Johann, J. A. (2014). Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorológicas na região do oeste do Estado do Paraná. *Revista Engenharia Agrícola*, 34(2), 286-299. <https://doi.org/10.1590/S0100-69162014000200010>
- Becker, W. R., Johann, J. A., Richetti, J., & Silva, L. C. A. (2017). Data mining techniques for separation of summer crop based on satellite images. *Engenharia Agrícola*, 37(4), 750-759.

<https://doi.org/10.1590/1809-4430-eng.agric.v37n4p750-759/2017>

Becker, W. R., Johann, J. A., Richetti, J., Mercante, E., Esquerdo, J. C. D. M., Da Silva Junior, C. A., & Paludo, A. (2020). Agricultural Soybean and Corn Calendar Based on Moderate Resolution Satellite Images for Southern Brazil. *Semina: Ciências Agrárias*, 41(2), 2419-2428. <https://doi.org/10.5433/1679-0359.2020v41n5supl1p2419>

Beskow, S., Mello, C. R., Faria, A. C., Simoes, M. C., Caldeira, T. L., & Nunes, G. S. (2014). Índices de sazonalidade para regionalização hidrológica de vazões de estiagem no Rio Grande do Sul. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 18(7), 748-754. <https://doi.org/10.1590/S1415-43662014000700012>

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, 26(2), 211- 252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>

Calínski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1-27. [Online] Available:
<https://www.tandfonline.com/doi/abs/10.1080/03610927408827101> (April 22, 2019).

Câmara, G., Carvalho, M. S., Cruz, O. G., Correa, V., Análise Espacial de Área. In: Fucks, S. D., Carvalho, M. S., Câmara, G., & Monteiro, A. M. V. (2004). Análise Espacial de Dados Geográficos. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - Divisão de Processamento de Imagens. [Online] Available:
<http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf> (January 30, 2021).

Duarte, M. L., Silva, T. A. (2019). Avaliação do desempenho de três algoritmos na classificação de uso do solo a partir de geotecnologias gratuitas. *Revista de estudos Ambientais*. 21(1), 6-16. <https://doi.org/10.7867/1983-1501.2019v21n1p6-16>

ECMWF – European Centre for Medium-Range Weather Forecasts. Webpage da European Centre for Medium-Range Weather Forecasts (2018). Available:
<http://spirits.jrc.ec.europa.eu/download/downloaddata/downloadmeteodata/> (January 13, 2018).

ESRI. (2015). ArcGIS for Windows Version 10.3. License type ArcInfo. ESRI - Environmental Systems Research Institute.

Farias, J. R. B., Nepomuceno, A. L., & Neumaier, N. (2005). XXVII Reunião de Pesquisa de Soja da Região, Central do Brasil. Embrapa Soja. Documentos, 265. [Online] Available:
<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/468751> (June 24, 2021).

Farias, J. R. B., Nepomuceno, A. L., & Neumaier, N. (2007). Ecofisiologia da Soja. Circular Técnica, 48. Londrina, PR: Embrapa Soja. [Online] Available:
<http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/470308> (April 14, 2021).

Felema, J., Medeiros, E. R., Ferreira, C. R., Camara, M. R. G., & Nascimento, S. P. (2016). Um estudo da produtividade de feijão, do milho e da soja na agricultura paranaense, nos anos de 2000 e 2010: uma análise espacial. *Ensaios FEE*, 36(4), 817-842. [Online] Available:
<https://patriciasiqueira.github.io/arquivos/3-felema-2016-feijao-milho-soja-PR.pdf>

(December 18, 2019).

Ferreira, W. P. M., Costa, L. C., & Souza, C. F. (2007). Teste de um modelo agrometeorológico para estudo da influência da variabilidade climática na cultura da soja. *Revista Ceres*, 54(312) 207-214. [Online] Available: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/50741/1/Teste-modelo.pdf> (April 10, 2019).

Franchini, J. C., Balbinot Junior, A. A., Nitsche, P. R., Debiasi, H., & Lopes, I. O. N. (2016). Variabilidade espacial e temporal da produção de soja no Paraná e definição de ambientes de produção, n. 374, Londrina, PR: Embrapa Soja. [Online] Available: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/147327/1/Doc-374.pdf> (May 26, 2019).

Grzegozewski, D. M., Cima, E. G., Uribe-Opazo, M. A., Guedes, L. P. C., & Johann, A. J. (2020). Spatial and Multivariate Analysis of Soybean Yield in the State of Paraná-Brazil. *Journal of Agricultural Studies*, 8(1), 387-412. <https://doi.org/10.5296/jas.v8i1.16303>

Grzegozewski, D. M., Uribe-Opazo, M. A., Johann, A. J., & Guedes, L. P. C. (2017). Spatial correlation of soybean productivity, enhanced v vegetation index (EVI) and agrometeorological variables. *Engenharia Agrícola*, 37(3), 541-555. <https://doi.org/10.1590/1809-4430-eng.agric.v37n3p541-555/2017>

IBGE – Instituto Brasileiro de Geografica e Estatística (2018). Relatórios metodológicos - Pesquisas Agropecuárias - Departamento de Agropecuária. [Online] Available: <https://sidra.ibge.gov.br/> (September 06, 2018).

IPARDES – Instituto Paranaense de Desenvolvimento Econômico e Social (2020). [Online] Available:http://www.ipardes.pr.gov.br/sites/ipardes/arquivos_restritos/files/documento/2019-12/pib_trim_nota_de_divulgacao_20.12.19.pdf (July 2020).

Jones, T. A. (1969). Skewness and kurtosis as criteria of normality in observed frequency distributions. *Journal of Sedimentary Research*, 39(4), 1622-1627. [Online] Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.907.790&rep=rep1&type=pdf> (September 19, 2018).

Klosowski, E. S. (1997). Estimativa da Produtividade de Cultivares de Soja por meio do Modelo Soygro para Londrina, Estado do Paraná. *Revista UNIMAR*, 19(3) 751-765. [Online] Available <http://periodicos.uem.br/ojs/index.php/RevUNIMAR/article/view/4557> (May 27, 2019).

Mingoti. S. A. (2013). *Análise de dados através de métodos de estatística multivariada uma abordagem aplicada*. Belo Horizonte: UFMG.

Moura, M. N., Vitorino, M. I., & Adami, M. (2018). Análise de Componentes Principais da Precipitação Pluvial associada à produtividade da soja na Amazônia Legal. *Revista Brasileira de Climatologia*, 22(1), [Online] Available: <https://revistas.ufpr.br/revistaabclima/article/view/55109/35656> (May 20, 2020).

Pewsey, A, Neuhäuser M., & Ruxton, G. D. (2013). *Circular Statistics in R*. Oxford

University Press, p.11-56.

R Development Core Team. (2018). R: a Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Radin, B., Schönhofen, A., & Tazzo, I. F. (2017). Impacto da quantidade e frequência de chuva no rendimento da soja. *Revista Agrometeoros*, 25(1), 19-26. <https://doi.org/10.31062/agrom.v25i1.26263>

Silva, P. O. (2018). Comportamento fenológico de *Emmotum nitens* (Benth.) Miers (Metteniusaceae) em fragmento de cerradão. *Revista Nativa*, 6(1), 33-38. <https://doi.org/10.31413/nativa.v6i1.4667>

Silva, R. R., Santori, M. G. B., & Wollmann, C. A. (2014). Relação entre precipitação pluviométrica e produtividade da cultura de soja, no município de Ibirubá- Rs. *Revista do Departamento de Geografica*, 27(1), 296-314. <https://doi.org/10.11606/rdg.v27i0.510>

USDA, U. S. D. OF. A. World Agricultural Production, (2021). Online] Available: <http://apps.fas.usda.gov/psdonline/circulars/production.pdf%5Cnhttp://www.fao.org/faostat/en/#data> (June 22, 2021).

Yeo, K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Journal Biometrika, Oxford*, 87(4) 954-959. <https://doi.org/10.1093/biomet/87.4.954>

Zar, J. H. (2010). *Bioestatistical Analysis*, (5rd ed.). New Jersey: Prentice Hall. p. 609.

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).