Spatial and Multivariate Analysis of Soybean Yield in the State of Paraná-Brazil

Denise Maria Grzegozewski

Federal University Souther Border

Rodovia BR 158 – Km 405, Laranjeiras do Sul, Paraná, 85301-970, Brazil

Tel: +55 (42) 3635-0000, E-mail: denisegrzegozewski@gmail.com

Elizabeth Giron Cima (Corresponding author)

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-3000 E-mail: egcima74@gmail.com

Miguel Angel Uribe-Opazo

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-3000 E-mail: miguel.opazo@unioeste.br

Luciana Pagliosa Carvalho Guedes

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-3000 E-mail:Luciana_pagliosa@hotmail.com

Jerry Adriani Johann

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná, 85819-170, Brazil

Tel: +55 (45) 3220-3000   E-mail: jerry.johann@hotmail.com

Abstract

In this work, the aim was to evaluate the existence of spatial association of the municipal average official soybean yield (t ha^{-1}) with agrometeorological data and vegetation indices. The information was observed by ten-day periods, in crop years 2010/2011, 2011/2012 and 2012/2013 in the State of Paraná. Local univariate spatial correlation (LISA index), as well as global bivariate correlation (L statistics) were calculated. With this study, we identified neighboring municipalities with high yield in the West as well as municipalities that are located

with low-low yield Northwestern, showing positive spatial autocorrelation ($(I_{MG} = 1)$, significative (p-value < 0.05). In addition, there were differences between seeding times in different regions, and climate irregularity during flowering periods and grain filling in crop year 2011/2012 throughout the state, which caused a large drop in production in all municipalities of the state of Paraná. The analysis of local spatial association showed that in the three crop years, the Northwest region presented a significant low yield potential of soybean (p-value < 0.05). In addition, it was observed that the period from the 3rd ten-day period of October to the 2nd ten-day period of January was essential for the soybean cycle in the different regions of the state, since this period encompasses the critical phases of crop. Differences were also observed between the crop years studied, regarding the agrometeorological variables, which affected soybean yield mainly in the Western region of Paraná – Brazil.

Keywords: climate irregularity, ten-day periods, correlation, clusters

1. Introduction

Brazil stands out in world agriculture, as the second largest soybean producer and the expansion of production of this oilseed in the country is due to the physical conditions of the soil, the favorable climate for its production, the advancing its cultivation in cerrado areas, and both the development and availability of scientific knowledge and technologies to the productive sector (Nunes, 2013).

Soybeans are used in agribusiness (vegetable oil production, animal feed), in the food industry (flour, extracts, beverages, sprouts), and in the chemical industry (adhesives, coatings, water emulsion paper for paints) (Bergamaschi, 1992; Berlato, 1999; Prudente et al., 2014).

Due to the complex interrelationships that soybean yield has with other factors of this agricultural activity, it is difficult to define standards for their respective production areas (Araújo et al., 2013). Studies in the literature demonstrate the importance of the investigation regarding the existence of a relationship or an effect of agrometeorological characteristics with

soybean yield, aiming to establish a better management of the production of this agricultural commodity (Araújo et al., 2013; Araújo et al., 2014; Bilbas et al., 2017).

One way to analyze the relationship between soybean yield and agrometeorological characteristics of the productive region is to consider the spatial variability of these variables and the spatial relationships between them. In this context, methods of spatial area statistics are used, which simultaneously consider information regarding the value and geographical location of variables.

Buss et al. (2019) verified the spatial variability of soybean productivity in relation to soil attributes using multivariate and geostatistical techniques. The result showed the efficiency of the statistical methods covered in the study.

The verification of a possible pattern of association between two georeferenced variables can be measured by measures of global bivariate autocorrelation, such as the bivariate Moran index and the L statistic (Lee, 2001; Lee, 2004). The purpose of these measurements is to show whether the values of one variable observed in one region reveal a relationship with the values of another variable observed in neighboring regions.

Studies using these indicators have shown for the state of Paraná: the soybean crop profile at different seeding dates (Dalposso et al., 2013, Cima et al., 2018); how municipalities relate spatially to soybean production; which are the main municipalities producing soy (Prudente et al., 2014); and the analysis of the spatial relationship of soybean yield with agrometeorological characteristics (Grzegozewski et al., 2017)

In this context, regarding the simultaneous analysis of a set of variables, the multivariate analysis techniques can be helpful in finding patterns generated by the set of variables. In the Western region of Paraná, Araújo et al. (2013) carried out a cluster analysis using information on the local Moran index (LISA) applied to agrometeorological and soybean yield data in crop year 2005/2006, and the formation of groups of similar municipalities was identified regarding their spatial distribution in relation to soybean yield and all agrometeorological elements analyzed.

Moreover in this focus of analysis, Vidigal et al. (2018), when analyzing the spatial arrangement of soybean productivity in the municipalities of Rio Grande do Sul in the period of 1990 to 2013 trough exploratory spatial analysis, concluded that soybean productivity was not homogeneous throughout the analysed period.

In addition to yield information, vegetation indices have been employed as indicators to discriminate areas with soybean cultivation (Grzegozewski et al., 2016), as well as areas with other types of land use (Araújo et al., 2014), and to analyze the conditions of soybean crop growth on a regional scale.

Thus, the objective of this work was to analyze the spatial variability of the official municipal average yield of soybean (t ha^{-1}), and its spatial relationship with agrometeorological data and vegetation indices, grouped by ten-day periods in the crop years 2010/2011, 2011/2012, and 2012/2013 in the State of Paraná, through an analysis of univariate spatial autocorrelation

(local) and bivariate spatial correlation (global), as well as using multivariate techniques.

2. Material and Methods

2.1 Determination of the Study Area and Data Acquisition

The study area comprises the 399 municipalities of the state of Paraná, Southern Brazil, limited by the coordinates 22º 29' S and 26º 43' S and 48º 2' W and 54º 38' W. This area was divided into 10 mesoregions of according to IBGE (Brazilian Institute of Geography and Statistics); these mesoregions are presented in Grzegozewski et al. (2017).

Yield data of the average municipal soybean [Yield] (t ha^{-1}) were used for crop years 2010/2011 (362 municipalities), 2011/2012 (361) and 2012/2013 (368), obtained from Paraná's Department of Agriculture and Supply (SEAB, 2015). The national average soybean yield was obtained from the Brazilian Institute of Geography and Statistics (IBGE, 2015).

The images used to obtain the vegetation index (EVI) were obtained from the MODIS product database (*Moderate Resolution Imaging Spectroradiometer*) of the Brazilian Agricultural Research Corporation (EMBRAPA, 2010) which is related to the MOD13Q1.5 product coupled to the Terra and Aqua satellite, "*Tile*" h13v11 from EVI (*Enhanced Vegetation Index*), which has a resolution of 250 m and a temporal resolution of 16 days.

The mapping of soybean cultivated areas for crop years 2010/2011 and 2011/2012 was provided by Souza et al. (2015), and for crop year 2012/2013 was provided by Grzegozewski et al. (2016). From these mappings, the pixel values of the average EVI vegetation index of the crop years studied were extracted for each municipality, and the details of the procedure adopted to calculate this average EVI for each municipality are described in Grzegozewski et al. (2017).

For each crop year under study, the agrometeorological variables obtained on the ten-day period scale were: climatological water balance [Cw] (mm), global radiation [Gr]

($\mathrm{MJ\ m^{-2} dia^{-1}}$), mean air temperature [Te] (°C), in the period from September to March. These

are at a spatial resolution of longitude and latitude grade, have been pre-processed and made into 0.25-degree ($\pm$ 25 x 25 km) grids and made available for free at *Joint Research Center* (JRC) *(ECMWF, 2012) website* in Shape format (Figure 1A).

It was necessary to adjust the agrometeorological data of 303 virtual stations for a spatial scale of 399 municipalities. This procedure was adopted to provide information on agrometeorological characteristics in the same municipalities for which soybean average yield data were collected. Thus, for each ten-day period and for each municipality under study, an average value was calculated, using the values of the virtual stations of each agrometeorological variable, contained in the center, near or around the polygon, as exemplified in Figure 1B.
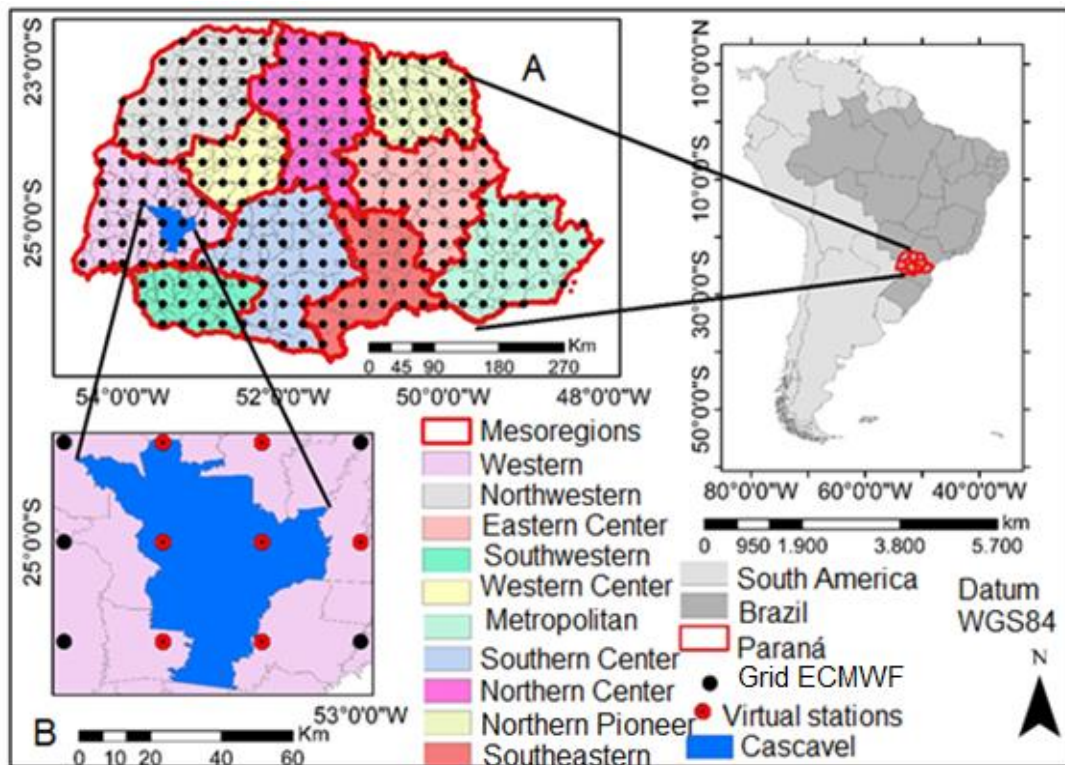
Figure 1. Regular grid representation of the 303 virtual stations of the *European Center for Medium-Range Weather Forecasts* (ECMWF) (A) and virtual stations used to obtain the municipal average of each agrometeorological variable (B)

2.2 Analysis of Spatial and Multivariate Statistics

In order to verify the existence of spatial dependence in each study variable, the Moran Global Index (I_{MG}- Equation 1) (Moran, 1950) was calculated and, based on the obtained value, a global statistical test was performed, in which the null hypothesis is the existence of random distribution of the variable under study, and the alternative hypothesis defines the existence of the association of similar or different values to a defined significance level (Anselin, 1995; Bailley, 1995; Nunes, 2013). Moran index is a value in the range [-1, 1], in which: $I_{MG} = -1$ it indicates a perfect negative autocorrelation, $I_{MG} = 0$ indicates the absence of autocorrelation, and $I_{MG} = 1$ indicates a perfect positive autocorrelation.

$$I_{MG} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{1}$$

wherein n is the sample size, x_i and x_j are the attribute values observed respectively in areas

A_i and A_j $(i, j = 1, \ldots, n)$; $\bar{x}$ is the average value of the attribute in the region under study; and

w_{ij} is a measure of spatial proximity between populations (municipalities) A_i and A_j

$(i, j = 1, \ldots, n)$ (Moran, 1950; Anselin et al., 2004). Given that in this work we adopted the

Queen type contiguity, so that $w_{ij} = 1$ A_i a common boundary is shared with A_j, by

analyzing this through the nodes (vertices) (Figure 1B). Otherwise, $w_{ij} = 0$; to

$i \neq j = 1, \cdots, n$.

In addition, for each variable, Moran scattering maps were constructed to identify points with a positive or non-positive spatial association (Anselin, 1995). This map was constructed based on the Moran scattering diagram, which is divided according to four quadrants: High-High [Q (+/+)], Low-Low [Q (-/-)], Low-High [Q (-/+)] and High-Low [Q (+/-)].

Although able to point to the general tendency of data clustering, Moran's *I* is a global measure and therefore does not reveal local patterns of spatial association; that is, they are generally hidden by global autocorrelation statistics.

Thus, in order to capture local association patterns (spatial *clusters* or *outliers*), the value of the indicator in the subarea composed of neighbors was also calculated for each study variable, using the function *Local Indicator of Spatial Association* (LISA) (Anselin, 1995) (Equation 2).

$$I_{Local} = \frac{x_i - \bar{x}}{S_i^2} \sum_{j=1, j \neq i}^{n} W_{ij}\left(x_j - \bar{x}\right);$$

$$(2)$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^{n} x_j}{n-1} - \bar{x}^2;$$

wherein n is the sample size, x_i is the value of the attribute observed in area A_i $(i = 1, \ldots, n)$;

$\bar{x}$ is the average value of the attribute in the region under study; and w_{ij} is a measure of spatial

proximity between populations (municipalities) A_i and A_j ($i, j = 1, \dots, n$), calculated by the

Queen contiguity criterion type, described in Equation 1.

Using the local Moran index values, a map named *LISA MAP* (Nunes, 2013) was prepared for each variable, in which it can be considered that there is no spatial autocorrelation if the *p-value* is greater than 0.05. Otherwise, autocorrelation is significant. Then, the areas were classified considering the significance levels equal to 0.05, 0.1, and 0.01.

We also calculated a bivariate spatial association measure called L statistics (Equation 3) (Lee, 2001). This measurement integrates Pearson's linear correlation coefficient with the univariate global spatial association measure of the Moran Index (Equation 1) and aims to observe whether there is spatial correlation between two variables, that is, whether the values of a variable in a given region are associated with the values of another variable observed in neighboring regions (Lee, 2001; Anselin et al., 2004).

$$L(x,y) = \frac{n}{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij}\right)^2} \frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})\right)\left(\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})\right)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{j=1}^{n}(y_j - \bar{y})^2}}, \qquad (3)$$

wherein n is the sample size, x_i and y_i are the values respectively of the attributes X and Y observed in area A_i ($i = 1, \dots, n$); $\bar{x}$ and $\bar{y}$ are the average values of attributes X and Y, respectively, in the region under study; and w_{ij} is a measure of spatial proximity between populations (municipalities) A_i and A_j ($i, j = 1, \dots, n$), calculated by the *Queen* contiguity criterion type.

In addition, for each crop year under study and considering yield, agrometeorological variables and vegetation indices in each ten-day period, a principal component analysis was performed. This analysis was performed in order to understand the relationship between these variables, as well as reducing the dimensionality of the number of variables. The selection criterion for choosing the main number of components was to obtain a minimum percentage of 70% of the total variability of the variables (Furtado, 1996).

For the selected main components, their respective scores were calculated in each municipality and with these values the analysis of clustering of the municipalities was performed. The *k-means* cluster method was used, and the number of formed groups was defined by the following measures: C and Dunn indices, and Silhouette coefficient (Dunn, 1973; Rousseuw and Silhouette 1986).

All statistical analyses were performed using R Studio version 3.3.5 software (R Development

Core Team, 2018). The representation on maps was performed in the software *ArcGIS* version 10.3 (ESRI, 2018).

3. Results and Discussions

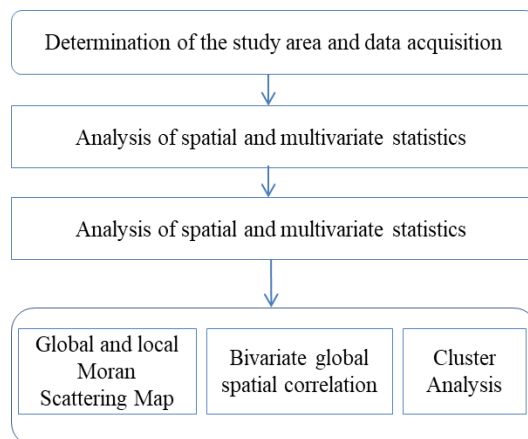The data referring a flowchart of the work, showing the process (Figure 2)

Figure 2. Flowchart of the work

3.1 Global and Local Moran Scattering Map

Analysis of global spatial autocorrelation by the Moran scattering map (Figure 3) for the three crop years showed similarity in the spatial distribution of yield between the 2010/2011 and 2012/2013 crop years (Figure 3). These two crop years had the highest yields concentrated in part of the Southwest, West, Midwest and East, and North Central and Pioneer regions. Of these regions, only a few municipalities had a significant Local Moran index (Figure 3), indicating that these municipalities form small *clusters* with high yield.

In addition, in these two crop years, the global Moran scattering map indicates that there is a low-low distribution of yield in the Center-South and Northwest of the state. This result indicates that the lowest yield values were concentrated in these regions of the state.

Moreover, for the three crop years, the Northwest region had *clusters* of municipalities with a significant similarity in yield (p-value $<$ 0.05, Figure 3). Concomitant with Moran scattering map, this result indicates that this region is formed by municipalities with low productive potential, as well as a region susceptible to summer, which harms soybean yield, corroborating Cima et al. (2018).

In relation to crop year 2011/2012, Moran scattering map identified larger regions of municipalities with high-high distribution and low-low distribution of yield near municipalities of equal scale (Figure 3). Given that this year it was identified that, for yield, low-low distribution in the Moran scattering map and significant local Moran index (p-value $<$ 0.05, Figure 3) have also occurred in part of the West and Southwest regions. These results indicated that there was low yield in these regions. In addition, there was a drop in yield in 2011/2012 in

the major producing regions (Southwest, West, and Midwest), when compared with the others in this study.

In contrast, in crop year 2011/2012, the Central-East, Southeast, as well as parts of North-Central and Center-South regions had municipalities with high-high distribution and significant local Moran index values (p-value $<$ 0.05, Figure 3), representing *clusters* of a municipality with higher yield values de according to Zeffrin et al. (2017).

In relation to crop year 2012/2013, we identify municipalities in the West and Central-East regions that presented in the Moran scattering map high-high distribution with significant local Moran index (p-value $<$ 0.05, Figure 3), indicating that the Western and central Eastern regions presented the formation of clusters of municipalities with high yield values.

Global Moran Scattering

2010/2011   **2011/2012**   **2012/2013**
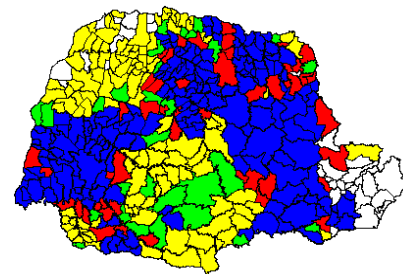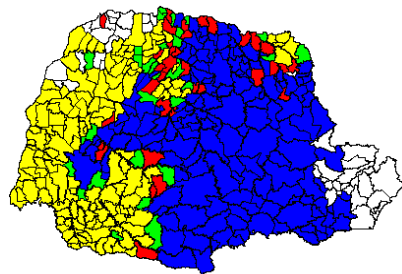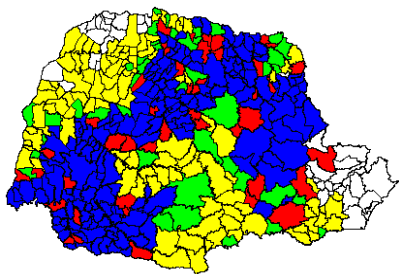
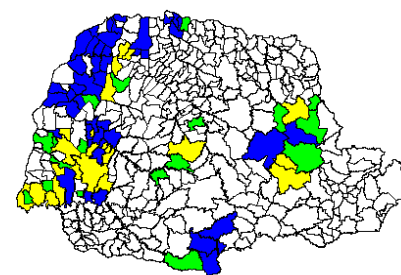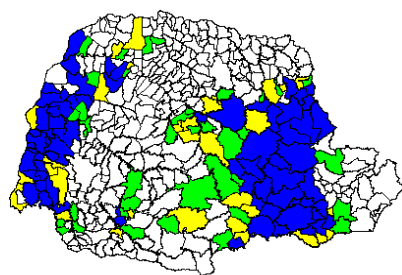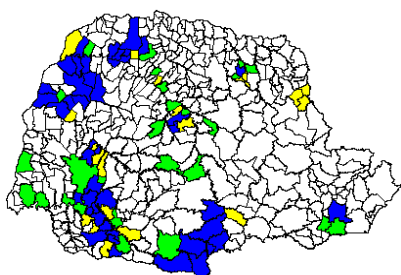☐ No Information   ■ Q(+/+)   ☐ Q(-/-)   ■ Q(+/-)

■ Q(-/+)

Local Moran Index Significance Level

■ 0.01%   ☐ 1%   ■ 5%   ☐ ns

Figure 3. Global Moran spatial scattering map and significance level of local Moran index of soybean yield in Paraná State for crop years 2010/2011, 2011/2012, and 2012/2013. Autocorrelation: Q (+/+) high/high, Q (+/-) high/low, Q (-/+) low/high and Q (-/-) low/low. ns = not significant

The global Moran scattering maps and the calculated significance level for the local Moran index, which were prepared for the agrometeorological variables and EVI vegetation index (Figures 4 to 6), indicated that almost the entire state of Paraná had a positive spatial association (low-low and high-high distribution), as well as large regions of Paraná with significant local association (p-value < 0.05). In other words, large areas in the state were formed to represent *clusters* of municipalities with similar values to each other and close to each other, in relation to the EVI vegetation index and agrometeorological variables.

Felema et al. (2016), when analyzing the spatial behavior of productivities in the crops of beans, corn and soybeans in the years between 2000 and 2010, through spatial exploratory analysis,

found that there is a positive and significant spatial interdependence in the regions of the State, and that the local standards are different.

The spatial variations found for the EVI variable in the crop years (Figure 4) can be explained by the differences between seeding dates and vegetation cover during the period under analysis. This variability is evident in the State, which has almost half of its municipalities with positive spatial autocorrelation containing crops in these periods, and the other half of the municipalities with negative spatial autocorrelation in the municipalities with no high vegetation index. In all crop years, the positive spatial autocorrelation of the EVI in the municipalities of the Western region occurred from the second decade of November, with the formation of *clusters* with significant local spatial association (Figures 4 and 5). The municipalities belonging to the West region presented similar characteristics for the EVI (Figure 4).

This result is associated with the early start of seeding, due to the increased levels of vegetation indices in these municipalities, demonstrating the difference between this and other regions of the state.

However, the variation of vegetation indices in the state is associated with water balance, rainfall, light intensity, and temperature. This can be verified by spatial autocorrelation for such variables during the period studied (Figures 4 to 6). It was also verified that soybean seeding occurred in municipalities with Cw, Te, and Gr had positive spatial autocorrelation [Q (+/+)] from 2nd ten-day period of October (Figure 4 to 6).

The spatial autocorrelation of the EVI in 2011/2012 was similar to the previous year (Figures 3 and 4). In addition, it is observed that the Western region began soybean cultivation before the other regions, due to the occurrence of appropriate water balance for seeding. However, the positive spatial autocorrelation distribution of Cw, conducive to crop development, was between the 2nd and 3rd ten-day period of November, initial phase and implementation of seeding in the rest of the state. In December, a negative and significant spatial autocorrelation of Cw prevails (Figures 4 and 5) in regions where the crop is more advanced in this period, therefore being caused by water deficit.
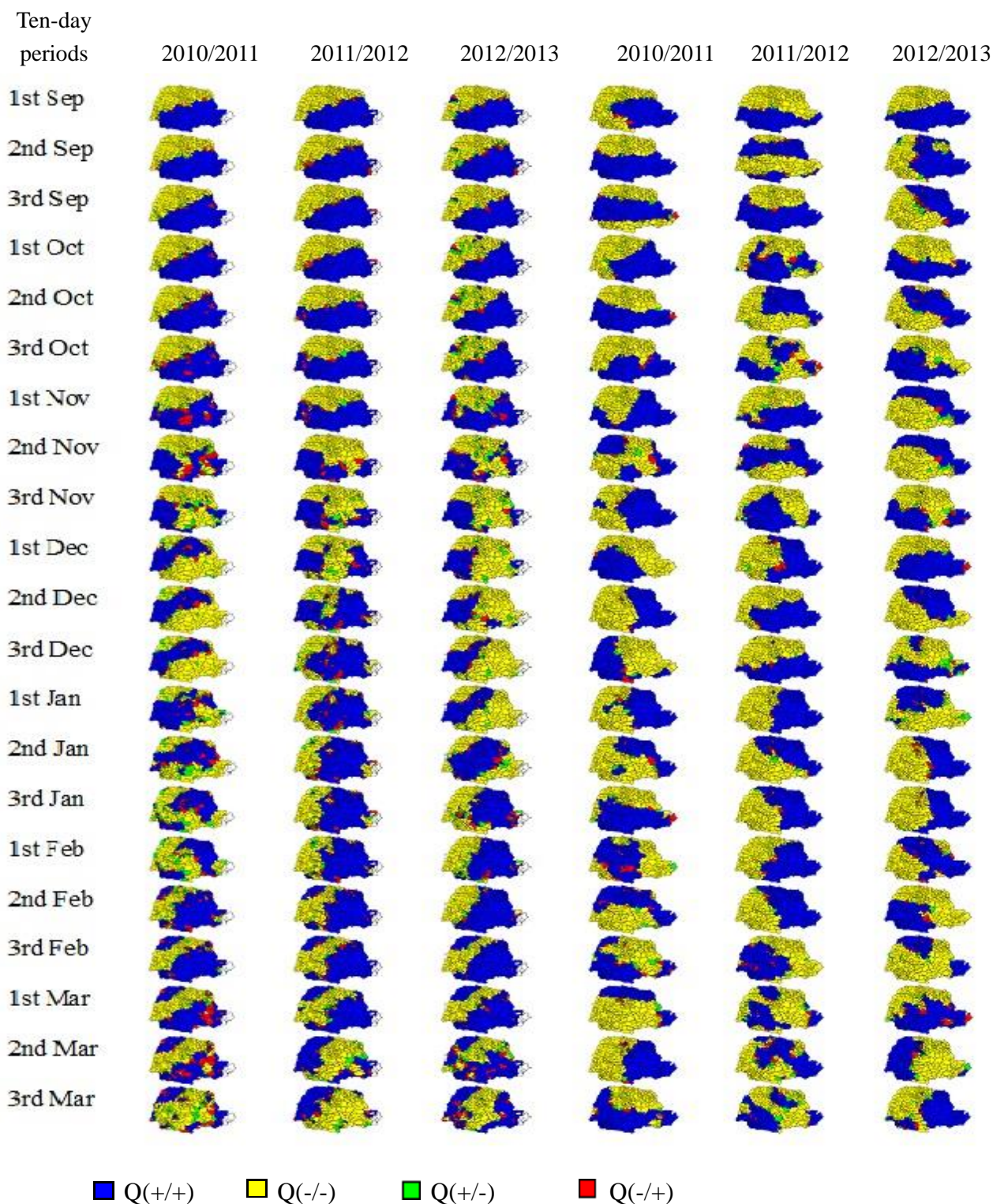
EVI   Cw

Figure 4. Spatial global Moran scattering map for EVI and Cw for each ten-day period and crop year. Autocorrelation: Q (+/+) high/high, Q (+/-) high/low, Q (-/+) low/high and Q (-/-) low/low

Regarding the EVI in 2012/2013 (Figures 3 and 4), there was similarity in its spatial

distribution with the 2010/2011 crop year, as well as in the identification of early seeding in the Western and part of the Southwest regions. It is also noted that the Cw (Figure 5) was significant the entire vegetative cycle of culture for the West region, providing optimal plant development conditions and consequently not affecting the final productivity.

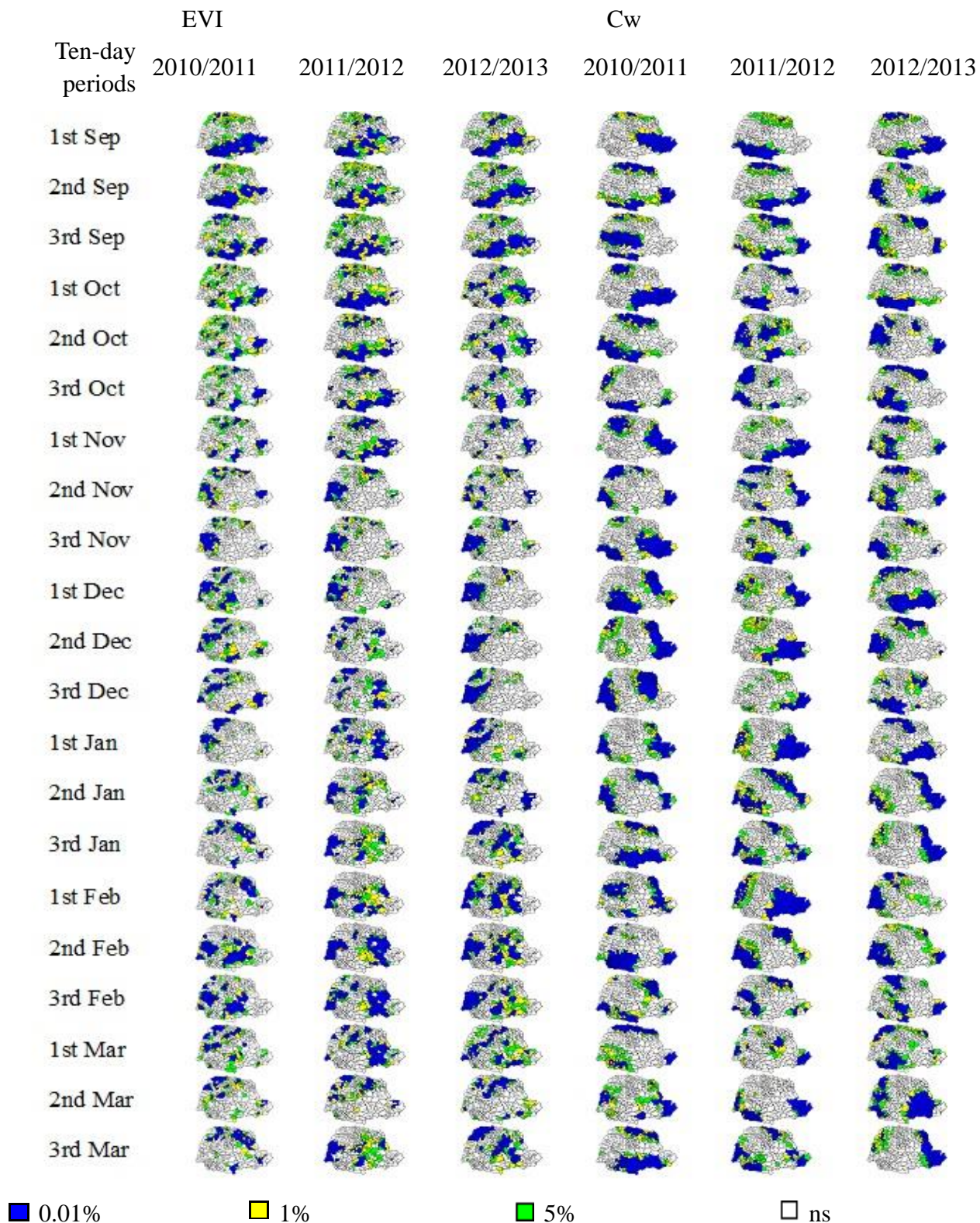| Ten-day periods	EVI			Cw		
	2010/2011	2011/2012	2012/2013	2010/2011	2011/2012	2012/2013
1st Sep						
2nd Sep						
3rd Sep						
1st Oct						
2nd Oct						
3rd Oct						
1st Nov						
2nd Nov						
3rd Nov						
1st Dec						
2nd Dec						
3rd Dec						
1st Jan						
2nd Jan						
3rd Jan						
1st Feb						
2nd Feb						
3rd Feb						
1st Mar						
2nd Mar						
3rd Mar						

■ 0.01%　　　□ 1%　　　■ 5%　　　□ ns

Figure 5. LISA map with significance level of the local Moran index of variables EVI and Cw, by ten-day period and crop year. Autocorrelation: Q (+/+) high/high, Q (+/-) high/low, Q (-/+) low/high and Q (-/-) low/low

Due to the low rainfall during the crop cycle, in the first two ten-day periods of November, there was negative Cw and spatial autocorrelation with distribution (Figure 5), especially in regions where soy was already in a vegetative stage. However, the irregularities that occurred did not harm the crop in its critical moments (emergence-germination and flowering-fruiting). Therefore, it was not considered detrimental to the final yield in almost the entire state, except for the Northwest region, where significant low-low spatial autocorrelation was found in several ten-day period of the three crop years, combined with the low productive potential of the region, which directly affected yield.

Regarding solar radiation (Gr) in 2010/2011 and 2012/2013 (Figure 6), there were significant low-low (Q (-/-)) spatial autocorrelation values in part of the State, i.e., due to the positive soil water balance, evapotranspiration was lower with low light. In general, Gr presented high-high spatial autocorrelation (Q (+/+)) in the Northwest, West, Midwest, and North-Central regions in almost all three-day periods, which corroborates SEAB (2015).

In crop year 2011/2012, the autocorrelation values of Gr had adequate levels for seeding from the 3rd ten- days period of September (Figure 5). However, due to the Cw [Q (-/-)] that occurred between the 2nd and 3rd three-day periods of November homogenized the seeding in the State regions, besides the high light intensity causing high evapotranspiration, which affected the photosynthetic activity and, consequently, the development of the crop.

In each region of the state, temperatures for the three crop years studied were similar between the ten-day periods (Figure 6). The Northwest and Western parts throughout the period showed high/high spatial autocorrelation [Q (+/+)], with several regions forming significant clusters (Figure 7). These results indicate that these regions present municipalities with high temperature values surrounded by municipalities with the same characteristic.

In the rest of the state, the temperature presented low/low spatial autocorrelation [Q (-/-)], with several regions forming significant clusters.   That is, they form a *cluster* of municipalities with low temperature values. This finding corroborates temperature data obtained by IAPAR (2011) for Paraná.

The problems that occurred with the climate in the 2011/2012 crop year caused yield losses in almost the entire state, except for the Central-Eastern region (cities of Ponta Grossa, Castro, and Tibagi) whose production was higher than the national average, with the following results: 3.35, 3.55, and 3.30 ha^{-1}. In these municipalities, none of the agrometeorological variables varied during the cycle (Figures 3 and 5), so soybean yield was not compromised as in other regions of Paraná.
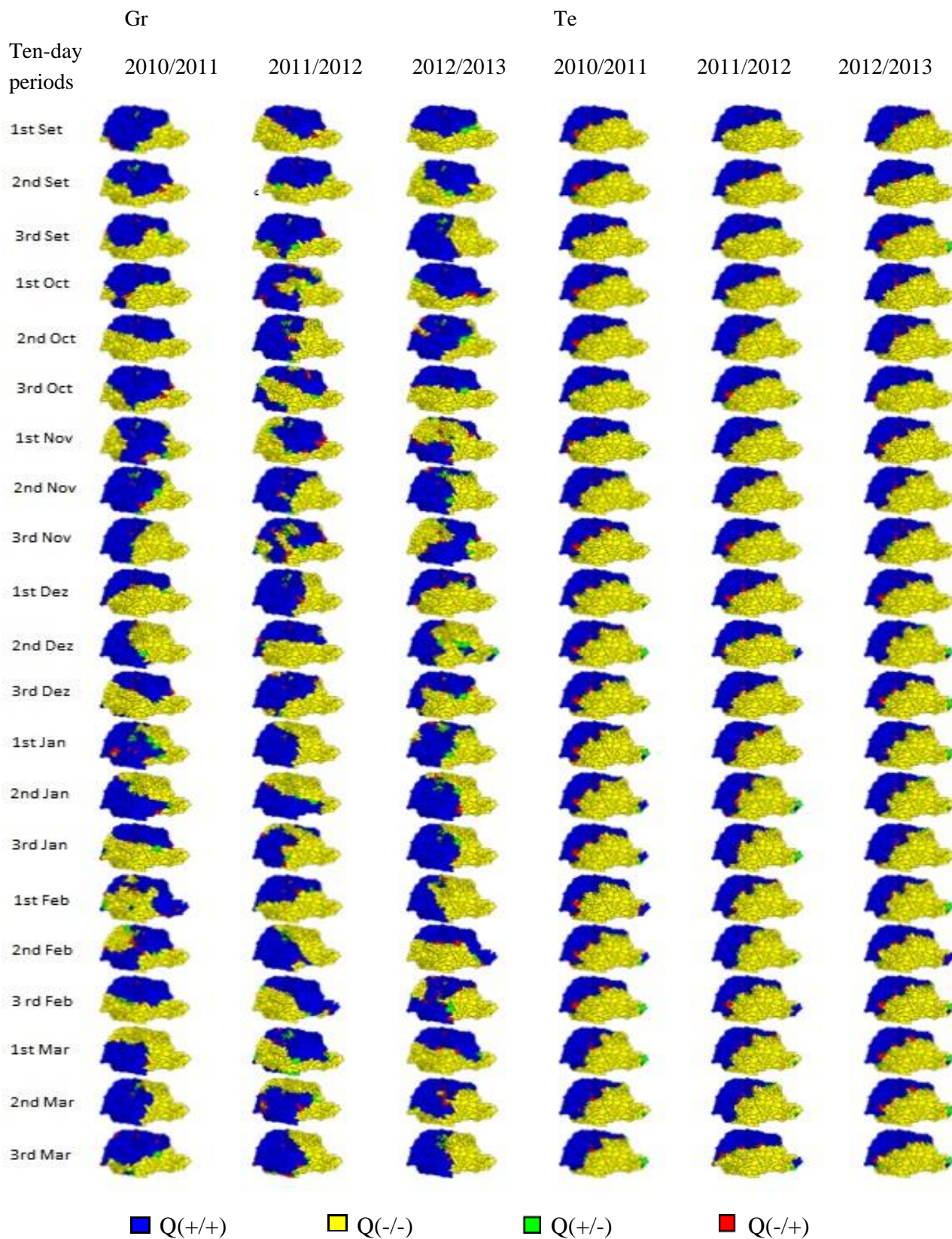
Figure 6. Spatial global Moran map for Gr and Te for each ten-day period of crop years 2010/2011, 2011/2012 and 2012/2013. Autocorrelation: Q (+/+) high/high, Q (+/-) high/low, Q (-/+) low/high and Q (-/-) low/low

The results showed that the climatic problems that occurred in the 2011/2012 crop years compromised soybean productivity in this period, which can influence the profitability of that culture (Figures 6 and 7).
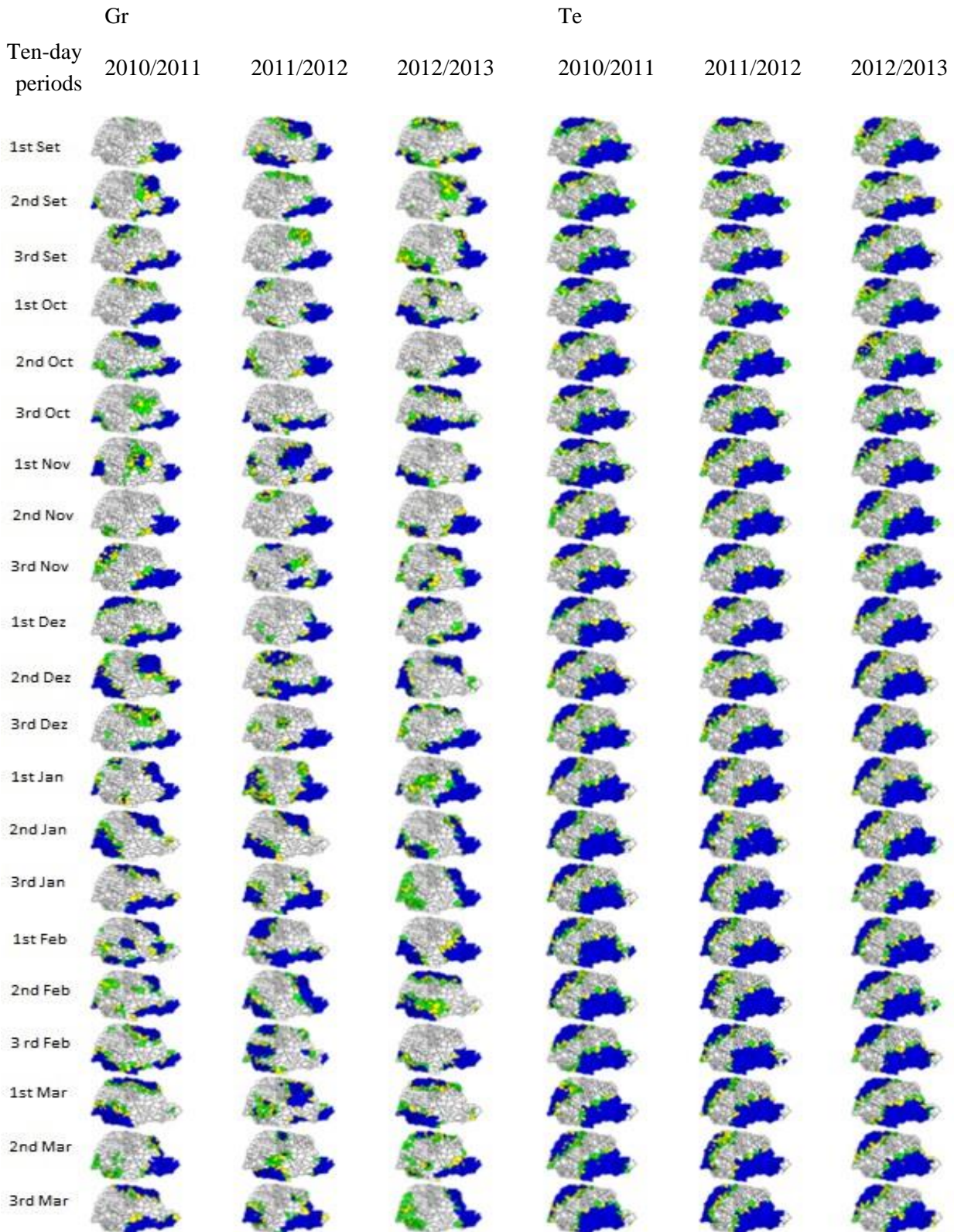
■ 0.01% □ 1% ■ 5% □ ns

Figure 7. LISA map with significance level of the local Moran index of variables Gr and Te, by ten-day periods and crop year. ns: not significant.

3.2 Bivariate Global Spatial Correlation

Analyzing the bivariate spatial correlation of yield with agrometeorological variables and vegetation index EVI (Figure 8) for the three crop years, we observed that there was a negative spatial correlation between yield and EVI of the first ten-day period of August until November 1st, which means there were municipalities with low and/or high productivity in relation to the EVI surrounded by municipalities with similar characteristics. This occurs because vegetation index levels are too low or absent. From the 2nd ten-day period of November, the spatial correlation between these variables is positive. However, for 2011/2012, for almost every ten-day period, the spatial correlation index between yield and EVI was lower, compared to other crop years, indicating a problem in crop development.
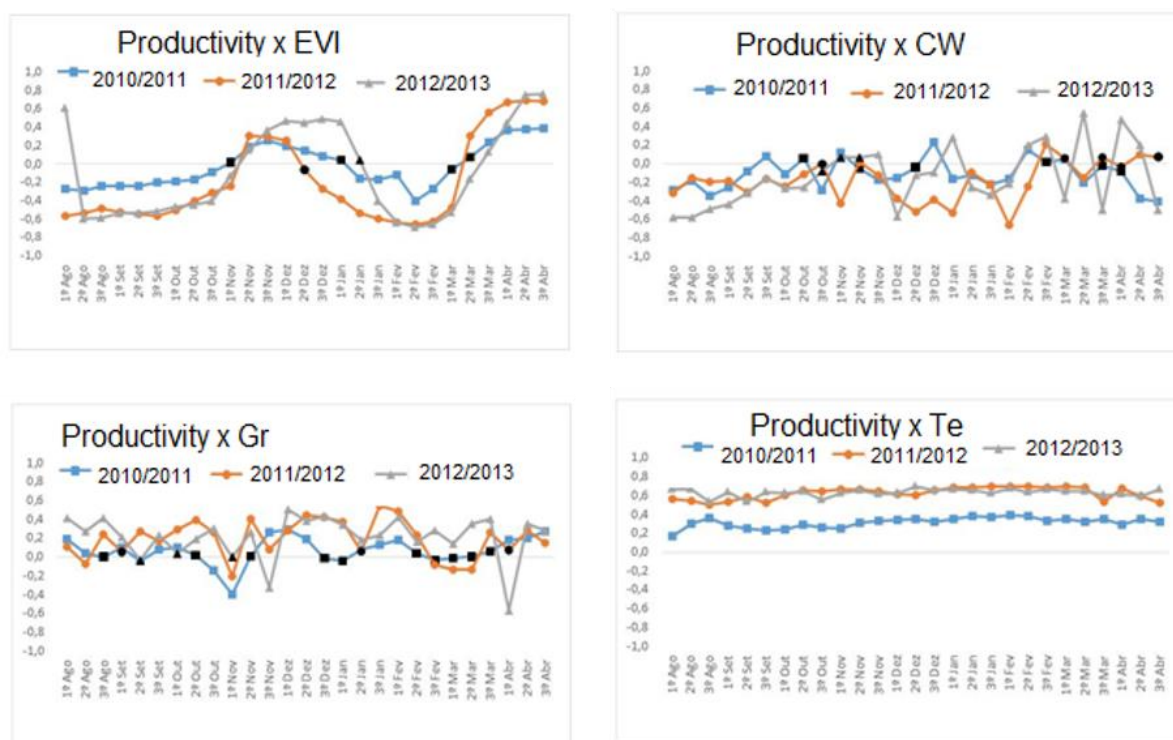

Figure 8. Bivariate spatial correlation between soybean yield and vegetation index EVI, Water balance (Cw), Global radiation (Gr) and Average temperature (Te) for each ten- day period and each crop year. Black dots indicate non-significant autocorrelation at 5% level

Similarities were also observed between 2010/2011 and 2012/2013 in relation to the values of vegetation index with yield in the period from the 2nd ten-day period of November until the 1st of January. However, when the 2010/2011 indices were analyzed, the correlations between the

municipalities were more heterogeneous in relation to the 2012/2013 crop year. It is identified, therefore, that from the 2nd ten-day period of January, the crop enters the phase of senescence and harvest, which corroborate Araújo et al. (2014).

For crop years 2011/2012 and 2012/2013, there was an inverse spatial association of yield and Cw (Figure 8), from the 1st ten days of August to the 2nd ten-day period of October. That is, in this period, the soil was with water deficit, which affected the seeding. In 2010/2011, there was less variation of the indices between these two variables, with values close to 0, from the 2nd ten-day period of September to October 2nd. For these two variables, both low (close to zero) and negative values of the bivariate spatial correlation index were identified from November 1st to December 2nd, as well as from January 1st to February 1st. However, for this crop year, this result did not affect soybean yield even though they occurred during a period of crop development. However, this result influenced the crop year 2011/2012, which affected the development of critical phases of the plants and soybean yield, which corroborates Caetano et al. (2018), who concluded that the climatic conditions influence soybean productivity.

The results obtained between yield and Gr, as well as between yield and Te, bivariate spatial correlation indices were lower in crop year 2010/2011 compared to other crop years (Figure 8).

In relation to crop year 2011/2012, affected by the negative water balance, there were high rates of spatial correlation between yield and the agrometeorological variables Gr and Te in the State (Figure 8). This indicates the dependence of these agrometeorological variables on the final crop yield.

Crop years 2011/2012 and 2012/2013 presented similar values of spatial correlation index between yield and temperature (Te), as well as higher values of this index when compared to crop year 2010/2011. Since 2010/2011, the values of this index close to zero indicate smaller variations between municipalities and variable Te.

3.3 Cluster Analysis

To perform the cluster analysis, the dimensionality was firstly reduced by principal components (PCs). Five main components were generated for the 2010/2011 and 2011/2012 crop years, considering 70% of the total variability of the studied variables; and three main components for crop year 2012/2013, considering 76% of the total variability of the studied variables. Considering the relevance of each variable in the formation of CPs and the values of the scores of each CP in each group, Table 1 presents a profile of each group in each crop year.

Table 1. Classification of the values of the main results obtained by interpreting the scores of the main components in the groups according to the relevance of the variables in the main components (CPs) (variables with the highest estimated coefficients in the CPs)

Group	Crop Year 2010/2011	2011/2012	2012/2013
1	**High:** *Te (in crop year)* *EVI (Dec. - Jan.),* *Gr (in crop year)* **Low:** *EVI (until Oct.)*	**High:** *Te (in crop year), Gr (in crop year), EVI (Nov. – Dec.)* **Low:** *EVI (Sep., Feb.), Cw (mostly Dec.)*	**High:** *Yield, Te (in crop year),* EVI (*Nov. – Dec.*) *Gr (mostly Dec.)* **Low:** *EVI (until Nov.)*
2	**High:** *Te (in crop year)* *EVI (Mar.),* *Gr (mostly Dec.)* **Low:** *EVI (until Jan.)*	**High:** *Te (in crop year), Gr (mostly Dec.)* **Low:** *EVI (Nov. – Dec.), Cw (mostly Dec.)*	**High:** *Te (in crop year),* Gr (mostly Dec.) **Low:** *EVI (up to Dec.)*
3	**High:** *EVI (Dec. – Jan.)*	**High:** *Te (from Jan.),* **Low:** *Gr (Jan.),*	**High:** **Low:** *EVI (Nov. – Dec.),*
4	**High:** *EVI (Sep. – Nov.)* **Low:** *Te (crop year)*	**High:** *EVI (Sep. – Nov.) Cw (mostly Dec.)* **Low:** *Te (crop year)*	**High:** *EVI (Sep. – Dec.) Gr (mostly Dec.)* **Low:** *Te (crop year)*
5	**High:** *EVI (until early Nov.)* **Low:** *Te (crop year)* *EVI (Dec. - Jan.)*	**High:** *EVI (until early Nov.) EVI (Feb.)* **Low:** *Te (crop year)*	**High:** *EVI (until early Nov.)* **Low:** *Te (crop year) EVI (Dec.)*

According to this profile and the groups formed (Table 1 and Figure 9), in the three crop years there was a well-defined formation of Western Paraná (group 1, Figure 8). Cluster analysis showed that this region has higher temperatures in common for all crop years, which is favorable to crop development. Thus, especially in crop year 2012/2013, the West region stood out for the concentration of municipalities with high yield values, and this was also observed by the Moran scattering map (Figure 3).

The West region also had high solar radiation values in the 2010/2011 and 2012/2013 crop years, as well as low vegetation index values in the first ten-day periods of each crop year (until November) to high EVI values in the December to January for the 2010/2011 crop year, and from November to December in the two subsequent crop years. However, in 2011/2012, in addition to high temperature values, low Cw values and high Gr values were identified in the West in most of the ten-day periods (from crop implantation to fruiting phase). Thus, this set of climatic results in 2011/2012 was responsible for a drop in yield in these Western municipalities (Figure 3), which are considered the largest producers in the state, that corroborates Kahlon et al. (2018).

Similarly to the Western region, the group of municipalities in the Northwest region of the state (and part of the Northern region in the 2011/2012 and 2012/2013 crop years) (group 2, Figure 9, and Table 1) is characterized by high values of temperature and solar radiation. However, this region differs from the Western region in the period of occurrence of low and high EVI values, which can be explained by the differences between seeding dates and vegetation cover in these two regions. These results corroborate the results presented by the Moran scattering maps of these variables (Figures 3 and 5).

The cluster analysis also shows that the North Central, North Pioneer and West Central regions formed group 3 in the three crop years (Figure 9). There were high EVI values in this region from December to January (in crop year 2010/2011); high temperature values (from January) and low solar radiation values (from January) in crop year 2011/2012; and low values of EVI in the period of November and December of the 2012/2013 crop year.

Two other groups that were formed (groups 4 and 5, Figure 9) correspond to the Southern and metropolitan region of the state of Paraná. These regions have lower temperatures, which were not ideal for seeding, and during the development period the high radiation values (Gr) in crop year 2012/2013 impaired the yield obtained in several municipalities of these regions (Figure 3). In addition, in most of the municipalities in these regions, there were high EVI values until November (Table 1), according to Rosa et al. (2017).
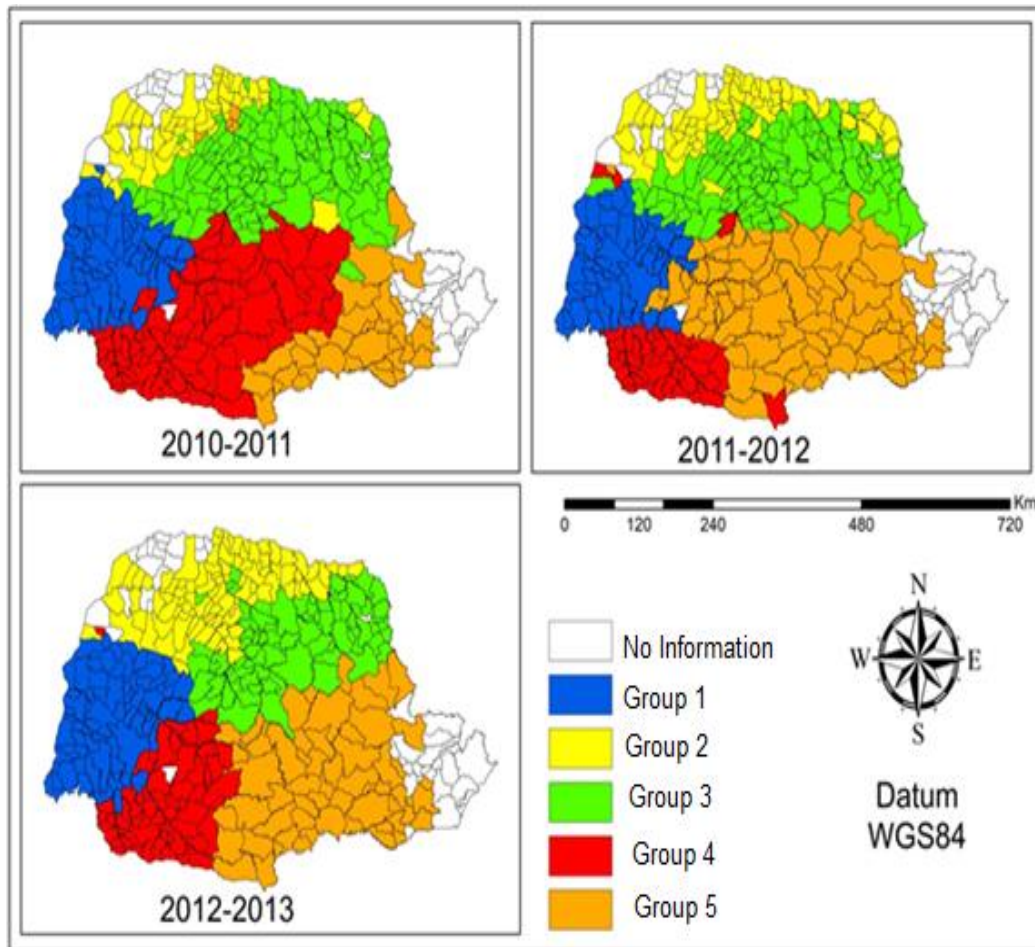
Figure 9. Map of Paraná with groups formed by cluster analysis considering yield, agrometeorological variables and vegetation index in each crop year

Compared to the crop years analyzed, the importance of identifying climate anomalies during the crop cycle is highlighted. Differences are seen for crop years affecting yields mainly in the Western region of the state. The lack of rainfalls associated with high temperatures and radiation caused losses in yield in this region. This fact can also be attributed to the climate differences in the growth seasons, being delimited by clusters obtained in the different crop years studied.

Therefore, the clusters analysis made it clear that soybean productivity depends on climatic conditions, mainly on precipitation associated with different temperatures at different growing seasons, which will determine the profitability of that crop.

4. Conclusions

Greater similarities were identified in the crop years 2011/2012, with lower productivity from

the analysis of significant positive spatial autocorrelation ($I_{MG} = 1$), with p-value < 0.05, that

is, due to climate variations, the spatial autocorrelation indices were similar in the municipalities of Paraná. Another finding was in relation to the Northwest region, where municipalities with low yield were identified in the three crop years studied, thus highlighting

the low productive potential of the region. In the bivariate analysis, the essential period, for the entire states, was from 3rd ten-day period of October to the 2nd of January in the cycle of the culture, showing a significant positive spatial correlation (p-value $<$ 0.05). In the cluster analysis, the importance and variability of climate for the State were observed, causing losses mainly in crop year 2011/2012.

The results also showed that in the three crop years there was a well-defined formation from the Western region of Paraná-Brazil. The cluster analysis made it clear that this region presented higher temperatures for all crop years, which is ideal for the development of the soybean culture. Finally, it is evident that in the 2012/2013 crop years, the West region stood out for the concentration of municipalities with high productivity.

Acknowledgement

References

Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, *27*(2), 93-115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., Bongiovanni, R., & Lowenberg-Deboer, J. (2004). A spatial econometric approach to the economics of site-specific nitrogen management in corn production, *American Journal of Agricultural Economics*, *86*(3), 675-687, Oxford University Press. https://doi.org/10.1111/j.0002-9092.2004.00610.x

Araújo, C. E., Uribe-Opazo, M. A., & Johann, J. A. (2014). Modelo de regressão espacial para a estimativa da produtividade da soja associada a variáveis agrometeorológicas na região oeste do estado do Paraná. *Engenharia Agrícola, 34*(2), 286-299. https://doi.org/10.1590/S0100-69162014000200010

Araújo, E. C., Uribe-Opazo, M. A., & Johann, J. A. (2013). Análise de agrupamento da variabilidade espacial da produtividade da soja e variáveis agrometeorológicas na região Oeste do Paraná. *Engenharia Agrícola*, *33*(4), 782-795. https://doi.org/10.1590/S0100-69162013000400018

Bailley, T. C., & Gatrell, A. C. (1995). Interactive spatial data analysis. Essex: Longman Scientific.

Bergamaschi, H. (1992). Desenvolvimento de déficit hídrico em culturas. In: Bergamaschi, H. (Coord.). *Agrometeorologia aplicada à irrigação*. Porto Alegre: Ed. UFRGS, p. 25-32.

Berlato, M. A., & Fontana, D. C. (1999). Variabilidade interanual da precipitação pluvial e rendimento da soja no Estado do Rio Grande do Sul, *Revista Brasileira de Agrometeorologia*, Santa Maria, *7*(1), 119-125.

Bilbas, H. T. A., Mahmood, S. H., & Omer, C. A. (2017). A Comparison results of factor analysis and cluster analysis to the migration of young people from the Kurdistan Region to Europe. *Journal of Pure and Applied Sciences, 29*(4), 44-55. http://dx.doi.org/10.21271/ZJPAS.29.4.5

Buss, R. N., Silva, R. A., Siqueira, G. M., Leiva, J. O. R., Osmann, C. C., & Oliveira. França, V. L. (2019). Spatial and multivariate analysis of soybean productivity and soil physical – chemical attributes. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *23*(6), 446-453. https://doi.org/10.1590/1807-1929/agriambi.v23n6p446-453

Caetano, J. M., Tessarolo, G., Oliveira, G., Souza, K. S., Deniz-Filho, J. A. F., & Nabout, J. C. (2018). Geographical patterns in climate and agricultural technology drive soybean productivity in Brazil. *Journal PLOS ONE, 13*(1), 1-16. https://doi.org/10.1371/journal.pone.0191273

Cima, E. G., Uribe-Opazo, M. A., Johann, J. A., Rocha-Junior, W. F., & Dalposso, G. H. (2018). Analysis of spatial autocorrelation of grain production and agricultural storage in Paraná. *Revista Engenharia Agrícola,* Jaboticabal, *38*(3), 395-402. https://doi.org/10.1590/1809-4430-eng.agric.v38n3p395-402/2018

Dalposso, G. H., Uribe-Opazo, M. A., Mercante, E., & Lamparelli, R. A. C. (2013). Spatial autocorrelation of NDVI and EVI indices derived from landsat/tm images for soybean crops in the Western of the state of Paraná in 2004/2005 crop season. *Engenaharia Agrícola*, *33*(3), 525-537. https://doi.org/10.1590/S0100-69162013000300009

Dunn, J. C. A. (1973). Fuzzy Relative of the ISODATA Process and its Use in Detection Compact Well-Separate Clusters. *In. Journal of Cybernetics, 3*(3), 32-57. https://doi.org/10.1080/01969727308546046

ECMWF – European Centre For Medium- Range Weather Forecasts. (2012). Available at: http://old.ecmwf.int/about/corporate_brochure/leaflets/Corporate-brochure-2012-en.pdf. Consulted in January, 2019.

Embrapa Informática Agropecuária. (2010). Desenvolvimento do banco de produtos MODIS na base estadual brasileira. Comunicado Técnico, n. 100. Campinas, SP, 2010. Available at: <http://www.cnptia.embrapa.br/files/ct100_ 10.pdf>. Access in: July 25, 2019.

Esri (2018). ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Felema, J., Medeiros, E. R., Ferreira, C. R., Camara, M. R. G., & Nascimento, S. P. (2016). Um estudo sobre a produtividade do feijão, do milho e da soja na agricultura paranaense nos anos de 2000 e 2010: uma análise especial. *Revista Ensaios FEE*, Porto Alegre, *36*(4), 817-842.

Furtado, F. D. (1996). ANÁLISE MULTIVARIADA. Lavras-MG Available at: https://sites.google.com/site/posufrj/ApostilaMultivariada-UniversidadeFed.pdf. Access in: October 30, 2019.

Grzegozewski, D. M., Johann, J. A., Uribe-Opazo, M. A., Mercante, E., & Coutinho, C. A. (2016). Mapping soya bean and corn crops in the State of Paraná, Brazil, using EVI images from the MODIS sensor, *International Journal of Remote Sensing*, *37*(6), 1257-1275. https://doi.org/10.1080/01431161.2016.1148285

Grzegozewski, D. M., Uribe-Opazo, M. A., Johann, J. A., & Guedes, L. P. C. (2017). Correlação espacial da produtividade da soja, índice de vegetação realçado (EVI) e variáveis agrometeorológicas. *Engenharia Agrícola 37*(3), 541-555. https://doi.org/10.1590/1809-4430-eng.agric.v37n3p541-555/2017

IAPAR – Instituto Agronômico do Paraná. Zoneamento agrícola. Londrina-PR. (2011). Available at: http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=1043. Consulted in May, 2015.

IBGE - Instituto Brasileiro de Geografia e Estatística. (2015). Banco de Dados Agregados - Sistema IBGE de Recuperação Automática – SIDRA. 2015. Available at: http://www.sidra.ibge.gov.br. Accessed on April 15, 2015.

Kahlon, S. C., Li, B., Board, J., Dia, M., Sharma, P., & Jat, P. (2018). Clusters and Principale Component Analysis of Soybean Grown at Various Row Spacings, Planting Dates and Plant

Populations. *Open Agriculture, 3*(1), 110-121. https://doi.org/10.1515/opag-2018-0011

Lee, S. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems 3*(4), 369-385. https://doi.org/10.1007/s101090100064

Lee, S. (2004). A generalized significance testing method for global measures of spatial association: an extension of the Mantel test. *Environment and Planning A*, *36*(9), 1687-1703. https://doi.org/10.1068/a34143

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena, *Biometrika, 37*, 17-23. https://doi.org/10.1093/biomet/37.1-2.17

Nunes, F. G. (2013). Análise Exploratória Espacial de Indicadores de Desenvolvimento Socioambiental das Regiões de Planejamento do Norte e Nordeste Goiano. *Revista Ateliê Geográfico*, *7*(1), 237-259. https://doi.org/10.5216/ag.v7i1.19809

Prudente, V. H. R., Souza, C. H. W. D., Mercante, E., Johann, J. A., & Uribe-Opazo, M. A. (2014). Spatial Statistics Applied to Soybean Production Data from Paraná State for 2003-04 to 2009-10 Crop-Years. *Engenharia Agrícola*. *34*(4), 755-769. https://doi.org/10.1590/S0100-69162014000400015

R Development Core Team, *R* (2018). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-90005107-0. Available at: http://www.R-project.org. Access in: May, 2018.

Rosa, B. T., Borges, L. A. C., Pereira, S. P., Antonialli, L. M., Chalfoun, S. M., & Baliza, D. P. (2017). Estudo sobre Boas Práticas Agrícolas em uma associação de Cafeicultores familiares por meio da análise de clusters. *Coffee Science, 12*(1), 49-59. https://doi.org/10.25186/cs.v12i1

Rousseeuw, P. J. S. (1986). A Graphical Aid to the Interpretation and Validation of Cluster Analysis. In: *Journal of Computational and Applied Mathematics*, *20*(1), 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Seab-Deral. (2015). Secretaria da Agricultura e do Abastecimento do Paraná / Departamento de Economia Rural – Banco de Dados da Produção Agropecuária no Paraná. Situação mensal de plantio, colheita e comercialização de produtos agrícolas no Paraná. Available at: http://www.agricultura.pr.gov.br/ Consulted in January *12*, 2015.

Souza, C. H. W., Mercante, E., Johann, J. A., Lamparelli, R. A. C., & Uribe-Opazo, M. A. (2015). Mapping and discrimination of soya bean and corn crops using spectro-temporal profiles of vegetation indices. *International Journal of Remote Sensing*, London, *36*(7), 1809-1824. https://doi.org/10.1080/01431161.2015.1026956

Vidigal, V. G., Vidigal, C. B. R., & Parré, G. R. (2018). Distribuição espacial da produtividade da soja no Rio Grande do Sul: um estudo exploratório. *Acta Scientiarum. Human and Social Sciences, 40*(2), 1-9. https://doi.org/10.4025/actascihumansoc.v40i2.33652

Zeffrin, R., Araújo., E. C., & Bazzi, C. L. (2018). Análise espacial de área aplicada a produtividade de soja na região oeste do Paraná utilizando o software R. *Revista Brasileira de Geomática*, Curitiba, *6*(1), 23-43. https://doi.org/10.3895/rbgeo.v6n1.5912