

PCA application for identification of NDVI and precipitation patterns in the Pernambuco state

Milena P. Dantas*, Célia C. Braga*, Wanessa L. de B. Costa*, Argemiro L. Araújo*

* Universidade Federal de Campina Grande – UFCG. Rua Aprígio Veloso, 882, Bairro Universitário. CEP: 58429-900. Campina Grande – PB, Brasil.

Received 31 July 2017; accepted 6 August 2017

Abstract

The present study aimed to interpret and analyse the spatial pattern of NDVI and rainfall on Pernambuco state. We used monthly average data of rainfall and NDVI, which we obtained from Terra/MODIS satellite, with spatial resolution of 1km, from 2003-2013 period. We applied the Principal Component Analysis (PCA) to determine the spatial pattern of the variability of variables. Our results showed that, generally, there is a relation between the rainfall distribution and the state relief, not necessarily in amount, but in the spatial distribution. In addition, the vegetation reacts according with the region rainfall.

Keywords: vegetation, images, principal component

1. Introduction

The state of Pernambuco is located in the center-east of the Northeast of Brazil, with climatic conditions depending on the amount and distribution of rainfall. Throughout the state, rainfall decreases east-west and, to a lesser extent, south-north. Thus, there are three variations of the climate in Pernambuco: tropical humid climate predominant in the Coastal and Zona da Mata, tropical subhumid climate that predominates in the Agreste and tropical semiarid climate. The average annual temperatures recorded for the territory vary from 26 ° C to 31 ° C (Dantas et al., 2016).

The variability of precipitation and temperature in the state is related to vegetation and relief and are also influenced by meteorological systems that, interacting with each other, give the state peculiar characteristics. Its vegetal diversification is between the mangrove and the tropical forest, located closer to the coast, and the caatinga located in the interior of the territory, with a climate characterized by semi-arid. There are between these two regions, respectively called Zona da Mata and Sertão, a transition area, known as Agreste. The vegetation is directly related to the climatic variability of a given region (Lexer et al., 2002).

Multitemporal data obtained from remote sensing of different meteorological and environmental satellites have been widely used for different purposes all over the world. Monthly IVDN series were used by Gutman and Ignatov (1998) to produce the vegetation fraction and to compose it in numerical models of weather and climate prediction. Ha et al. (2001) analyzed the variability of IVDN, LAI and estimated

surface temperature from AVHRR / NOAA data in Korea. The authors observed that the interannual variability of LAI strongly depends on vegetation type and that LAI changes are not related to variations of the NIDI.

Factorial Analysis in Principal Components (PCA) is widely used in studies involving a large number of variables, by this technique capable of reducing them without deforming the original data. For this reason it is often applied time series of meteorological data. Sousa et al. (2014) used monthly data from the IVNA of the NOAA and MODIS satellites in the 2010 decade to find a relationship with rainfall in the state of Paraíba-Brazil. The results showed that the correlations are higher in the drier months than in the rainy season. Dantas et al. (2016), used ACP to analyze the time vegetation response to rainfall in the state of Pernambuco-Brazil.

Taking into consideration the above, the objective since work is to improve previous studies for the state of Pernambuco, making use of the statistical technique of the ACP, allowing to identify the spatial patterns of IVDN and precipitation, associating them with the atmospheric systems that produce Rains in the region.

2. Material and methods

2.1 Study area

Pernambuco is among 7°15'45" and 9°28'18" south latitude and between 34°48'35" and 41°19'54" west longitude, limiting with the states of Paraíba, Ceará, Piauí, Bahia, Alagoas and east with the Atlantic Ocean. Data were used 81 stations / meteorological points distributed in an area of

98,149,119 km² (IBGE, 2014) in the period from 2003 to 2013.

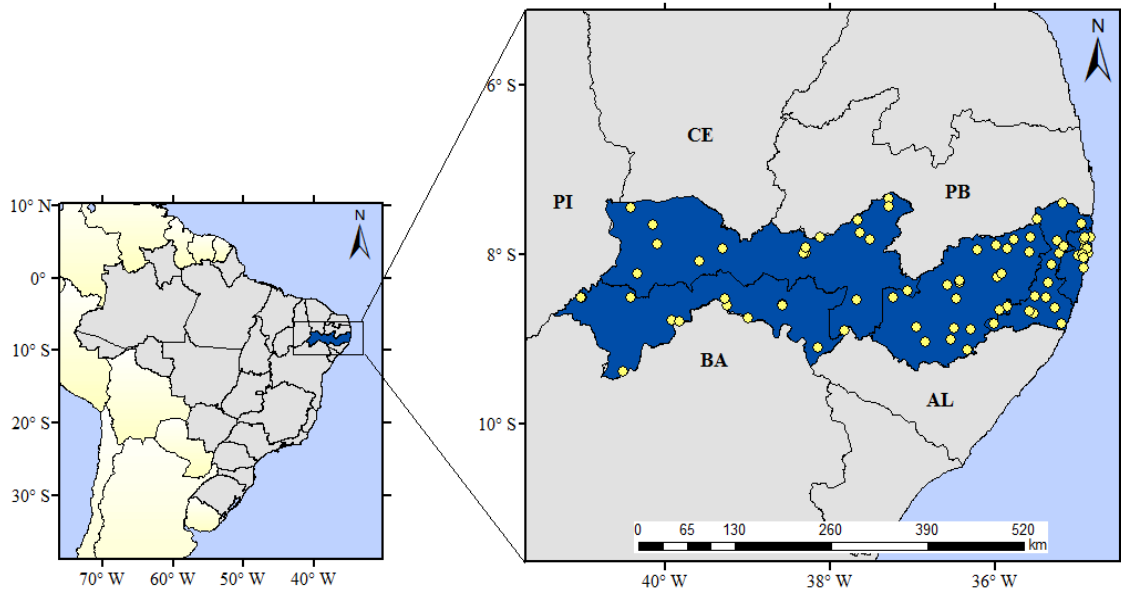


Figure 1 - Location of the state of Pernambuco on the map of Brazil.

2.2 Rainfall data

Precipitation data were obtained from the Pernambuco State Water and Climate Agency (APAC) of 81 meteorological stations distributed in the state of Pernambuco from 2003 to 2013. The relief of Pernambuco is moderate, most of the state is below 600 m. In the Coastline the relief is almost all to the average level of the sea. As it distances from the coast, one observes a coastal plain with altitude between 0 and 10m. Between Zona da Mata and Agreste, one can find the Borborema Plateau, with an average elevation of 600m, passing from 1000m to the peaks. It is

observed that the altitude grows from São Francisco to Sertão. The Serra do Araripe is also highlighted on the border with Ceará, with an altitude of approximately 800 m (Figure 2).

The volume of precipitation of a locality is related to the characteristics of the relief. Figure 3 shows the average annual precipitation totals for the period from 2003 to 2013 in the state of Pernambuco. The highest indices are seen on the coast, with values higher than 1800 mm. Meanwhile, the lowest values are observed in the São Francisco and Sertão regions, with values lower than 600 mm.

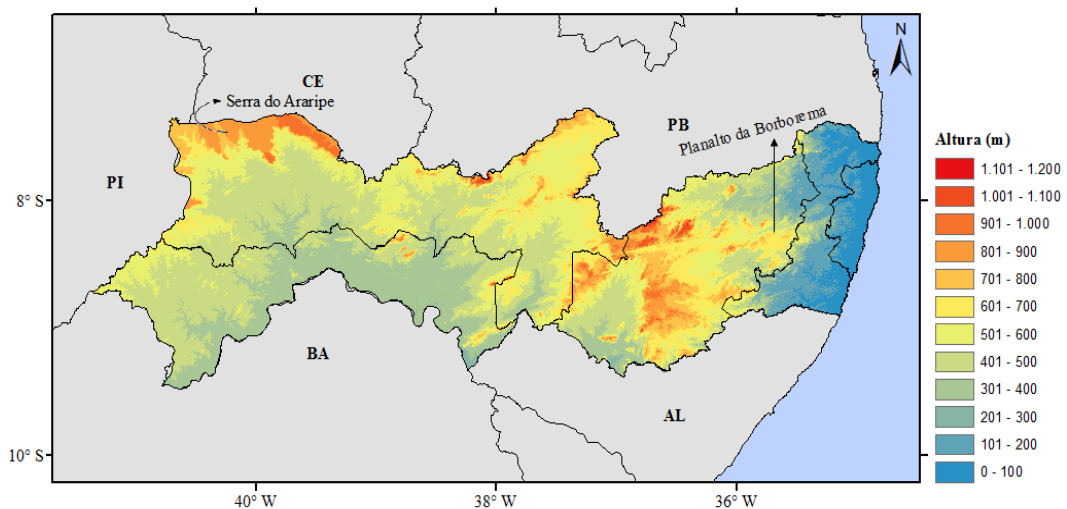


Figure 2 - Relief (meters) from the state of Pernambuco (Miranda, 2005).

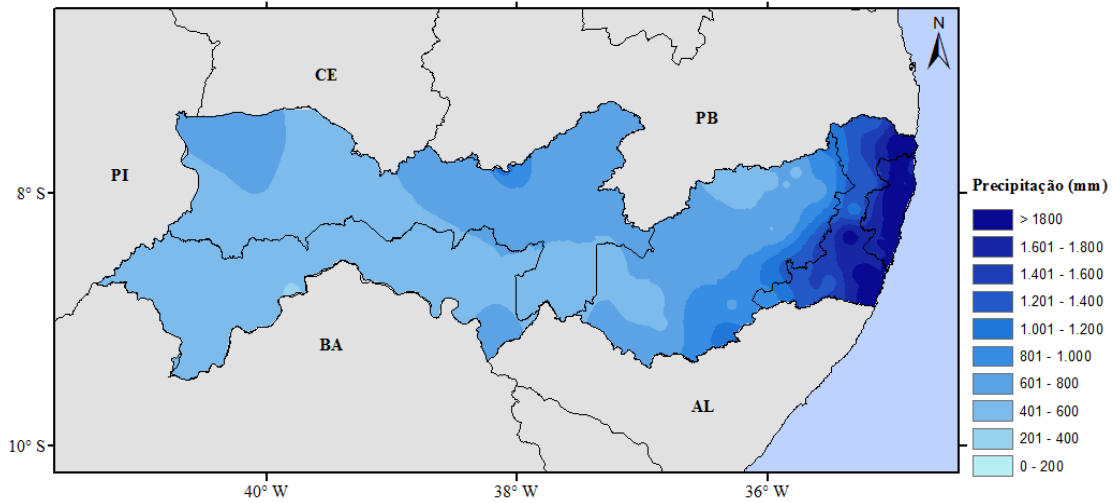


Figure 4 - Spatial distribution of total average annual precipitation (mm) in the state of Pernambuco, period 2003-2013.

2.3 NDVI data from Terra/MODIS

NDVI is a model resulting from the combination of reflectance levels in satellite images, the near infrared (0,725 - 1,10 μm) and the visible (0,58 - 0,68 μm). Soon, the NDVI is determined by the following equation:

$$\text{NDVI} = \frac{(\text{NI} - \text{VIS})}{(\text{NI} + \text{VIS})} \quad (1)$$

The MODIS product for the IVDN, more specifically the product MOD13A3, with spatial resolution of 1 km for the period from 2003 to 2013, was purchased on the REVERB/NASA website. There was a conversion of the images in .HDF to .IMG, the tiles (h13v09 and h14v09) were grouped. In addition, the correction factor (0.0001) was multiplied by the product MOD13A3, with the aid of a software developed to extract information from digital images. Thus, with the processed images, the IVDN value was extracted for each pixel

2.4 Factor Analysis in Principal Components

For an analysis of principal components (ACP), we need an original data matrix X, of p variables for n individuals, in order to obtain an array of variance and covariance S through:

$$S = \frac{1}{n} \cdot X \cdot X^t \quad (2)$$

Since X is the matrix with centered values, X^t is the transposed matrix and n is the number of individuals. The correlation matrix R will be equal to the matrix of variance and covariance, thus:

$$R = S = \frac{1}{n} \cdot X \cdot X^t \quad (3)$$

The matrix R is a symmetric and positive correlation matrix of dimension (p x p), diagonalizable by an orthogonal matrix A, of base change called eigenvectors, thus:

$$D = A^{-1} \cdot R \cdot A = A^t \cdot R \cdot A \quad (4)$$

Where D is the diagonal matrix and A⁻¹ is the inverse of matrix A. Since A is the base change matrix for a new reference system composed of the eigenvectors of R, the principal components (CP) U₁, U₂, ... , U_p are obtained by linear combinations between the transpose of the eigenvectors (A^t) and the standardized observation matrix (X), as described below:

$$U = A^t \cdot X \quad (5)$$

To estimate X_i n-th local values, is used:

$$X_i = a_{j1}U_1 + a_{j2}U_2 + \dots + a_{jk}U_k + \dots + a_{jp}U_p \quad (6)$$

Where a_{ij} is the set of eigenvectors of X in descending order of the most significant eigenvalues of a in A.

The percentage of explained variance of the eigenvalues in descending order is given by:

$$\% \text{Var}(X) = \frac{\sum_{i=1}^n \lambda_i}{\sum_{j=1}^n \lambda_j} \quad (7)$$

The correlation between the i-th original variable and the i-th major component is:

$$Coor(X_i Z_j) = a_{ij} \cdot \sqrt{\lambda_i} \quad (8)$$

Being, a_{ij} the j th element of the i th eigenvector and λ_i the i th eigenvalue. For this analysis, adequate statistical software was used

3. Results and discussion

3.1 Spatial analysis of precipitation

The following are presented and analyzed the spatial and temporal patterns of precipitation and IVDN for the state of Pernambuco in the period 2003 to 2013.

For the spatial analysis of the precipitation, the first 5 common rotational factors were considered, which explained 89.5% of the total variance of the series. Only the first three components of larger weights will be analyzed.

The first spatial pattern (Figure 4a), which explained 27.82% of the variance, has correlations higher than 0.6 in a core isolated in the Sertão, and in much of São Francisco and Agreste. Isolated nuclei higher than 0.8 in the south central region (São Francisco) stand out. The lowest correlations are observed in the north of Zona da Mata. The time

series (Figure 5) associated with this factor identifies the rainy years of 2005 and 2010. And the dry seasons are 2003 and 2012 with little rainfall in the eastern sector.

The spatial configuration of the second component explaining 25.65% of the variance is illustrated in Figure 4b. Positive correlations higher than 0.4 are observed in the east of the state and in some nuclei of the same order in the interior. The temporal series of this factor presents maximum scores in 2007, 2011 and 2013 (Figure 5), indicating that in those years the most relevant rainfall indexes were observed from Zona da Mata to Litoral and minimums in years 2006, 2008 and 2009, showing a Reduction of rainfall in the Zona da Mata.

The third spatial pattern (Figure 4c), explaining 16.83% of the total data variance, presents a positive correlation (> 0.6) in small distinct nuclei throughout the state and above 0.4 in western Pernambuco. Figure 5 shows the time series for this factor, highlighting a maximum in 2004, showing that in this period more rainfall occurred in the west compared to other years, while the lows were observed in 2010 and 2011. Results found similar in the literature by Nicácio et al. (2009) and Silva et al. (2014).

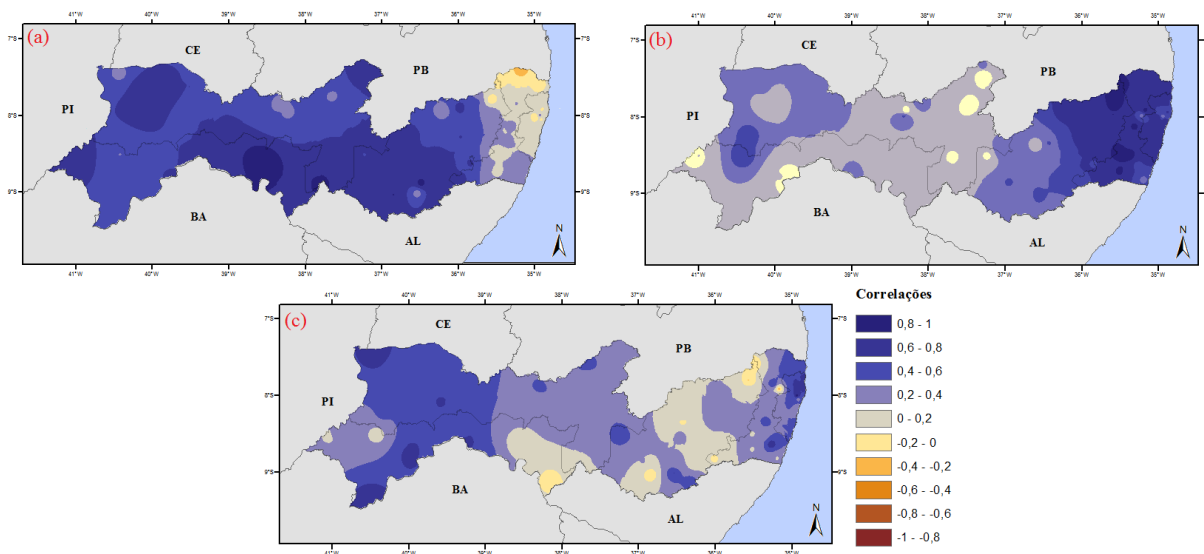


Figure 4 - Spatial patterns (correlations) for the first three common precipitation factors in Pernambuco explaining 70.30% of the total variance

The first spatial pattern (Figure 4a), which explained 27.82% of the variance, has correlations higher than 0.6 in a core isolated in the Sertão, and in much of São Francisco and Agreste. Isolated nuclei higher than 0.8 in the south central region (São Francisco) stand out. The lowest correlations are observed in the north of Zona da Mata. The time series (Figure 5) associated with this factor identifies the rainy years of 2005 and 2010. And the dry seasons

are 2003 and 2012 with little rainfall in the eastern sector.

The spatial configuration of the second component explaining 25.65% of the variance is illustrated in Figure 4b. Positive correlations higher than 0.4 are observed in the east of the state and in some nuclei of the same order in the interior. The temporal series of this factor presents maximum scores in 2007, 2011 and 2013 (Figure 5), indicating

that in those years the most relevant rainfall indexes were observed from Zona da Mata to Litoral and minimums in years 2006, 2008 and 2009, showing a Reduction of rainfall in the Zona da Mata.

The third spatial pattern (Figure 4c), explaining 16.83% of the total data variance, presents a positive correlation (> 0.6) in small distinct nuclei throughout the state and above 0.4 in western

Pernambuco. Figure 5 shows the time series for this factor, highlighting a maximum in 2004, showing that in this period more rainfall occurred in the west compared to other years, while the lows were observed in 2010 and 2011. Results found similar in the literature by Nicácio et al. (2009) and Silva et al. (2014).

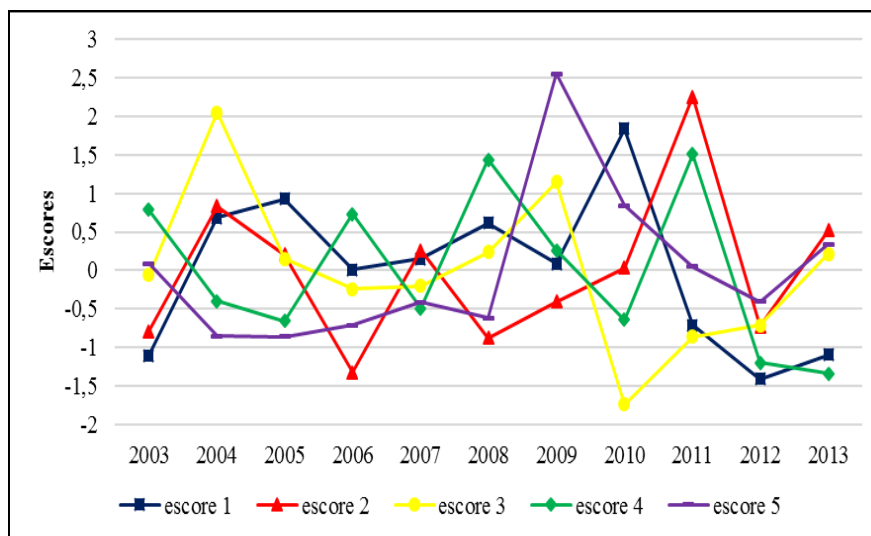


Figure 5 - Principal components of the main five common spatial precipitation factors for the state of Pernambuco.

3.2 Spatial analysis of NDVI

The first three common spatial factors explaining 89.64% of the total variance of the annual data are shown in Figure 6 (a, b, c). For the first factor that accounts for 52% of annual IVDN data, correlations are higher than 0.6 in all of Sertão and even in the Zona da Mata and Ageste. The lowest correlations are found in the southeast and northeast of the state. The annual variability of the IVDN of this factor presents maximum in the years 2007 to 2010 and in 2012, and minimums in 2003 and 2013 (Figure 7). This factor is showing mainly dry years (APAC, 2014), harming vegetation development. According to Weng et al. (2008), vegetation behavior and growth is strongly influenced by rainfall.

The spatial variability of the correlations of the second factor, which explains 20.38% of the total IVDN variance, has positive correlations

throughout the region, but the lowest values are observed in the extreme northwest of Sertão, in the north of Zona da Mata and in the East Coast And south. The time series of this factor shows high weights in 2003 and 2011 and minimums in 2007 and 2012.

Finally, the spatial variability of the correlations of the third factor (Figure 6c), which explains 16.75% of the variance, has positive correlations throughout the state, decreasing from the coast (> 0.8) to the interior where the correlations are minimal positive. In the southwest and central northwest (near zero). The annual temporal variability of this factor has maximum contribution in 2013. In this year, the highest IVDN was verified from the central part of Agreste to the Litoral. These results corroborate the work of Silva et al. (2014) and Dantas et al. (2016).

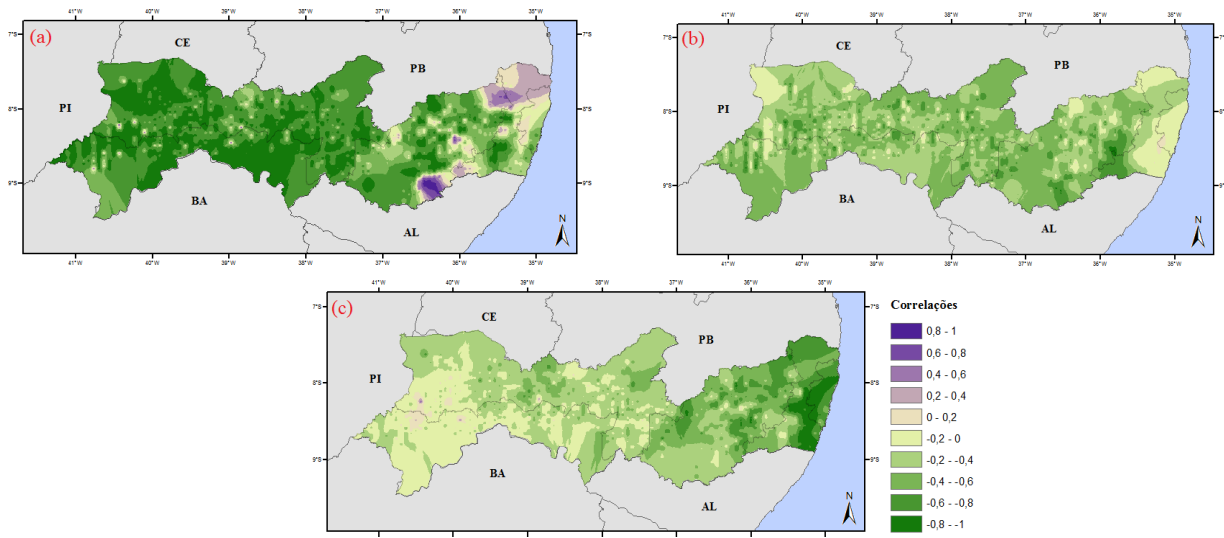


Figure 6 - Spatial patterns (correlations) for the first three common factors of IVDN in Pernambuco explaining 89.64% of the total variance.

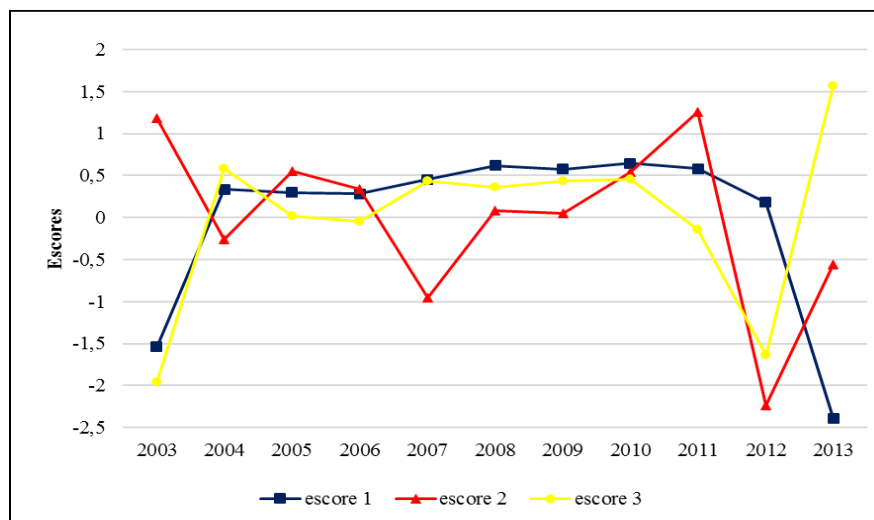


Figure 7 - Principal components of the first three common spatial factors (scores) of IVDN for the state of Pernambuco.

4. Conclusion

Although the GAP applied to the rainfall time series has retained five factors, it can be observed from the results already discussed that only the first three common factors represent aspects of rainfall climatology in Pernambuco. The first common factor showed, throughout almost the entire state, high correlations of IVDN and precipitation. A nucleus located in Agreste Pernambuco can be seen in both spatial distributions of this factor. In addition, the north of the Zona da Mata presents negative correlations of IVDN and rainfall.

The second precipitation factor has high correlations in the northwest of the state and the Agreste to the Coastal. For the IVDN, the first factor exhibits positive correlations throughout the state,

although some of them are close to zero mainly in the portions of the East and South Coast, north of the Zona da Mata and northwest of the state.

The third common factor related to rainfall, which showed maximum scores in 2004, allowed us to observe that the vegetation index varies according to the occurrence of rainfall, that is, rainy / dry regions cause an increase / decrease in IVDN.

There is a relationship between precipitation distribution and state relief, not necessarily in quantity but in spatial distribution. The ACP was effective for the proposed treatment

Acknowledgements

The authors thank CNPq and CAPES for the grant of stock.

References

- APAC. Agência Pernambucana de Águas e Climas, 2014. Disponível: <http://www.apac.pe.gov.br>. Acesso: 11 nov. 2014.
- Dantas, M.P., Braga, C.C., Sousa, E.P., Barbieri, L.F.P., Vilar, R.A.A., 2016. Temporal analysis of rainfall and vegetation index using satellite images in Pernambuco State. *Journal of Hyperspectral Remote Sensing* 6, 99-107.
- Gutman, G., Ignatov, A., 1998. The derivation of the green vegetation fraction from NOAA/ AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing* 19, 1533-1543.
- Ha, K., Oh, H., Kim, K., 2001. Inter – Annual and Intra – Annual Variabilities of NDVI, LAI and Ts Estimated by AVHRR in Korea. *Korea Journal of Estimated by AVHRR in Korea. Korea Journal of Remote Sensing* 2, 111-119.
- IBGE. Instituto Brasileiro de Geografia e Estatística, 2014. Disponível: <http://www.ibge.gov.br/estadosat/perfil.php?sigla=pe>. Acesso: 15 jan. 2014.
- Lexer, M.J., Hönniger, K., Scheifinger, H., Matulla, C., Groll, N., Kromp-Kolb, H., Schaudauer, K., Starlinger, F., Englisch, M., 2002. The sensitivity of Austrian forests to scenarios of climate change: a large-scale risk assessment based on a modified gap model and forest inventory data. *Forest Ecology and Management* 162, 53–72.
- Miranda, E.E.de (Coord.), 2005. Brasil em Relevo. Embrapa Monitoramento por Satélite, Campinas. Disponível: <http://www.relevobr.cnpm.embrapa.br>. Acesso: 21 fev. 2015.
- Nicacio, R.M., Araujo, L.L., Gonzaga, E.P., Libos, M.I.P.deC., Oliveira, L.M.T.de, 2009. Relação NDVI e precipitação na bacia do rio Coxim-MS. *Geografia* (Rio Claro. Impresso) 34, 667-680.
- Silva, M.T., Silva, V.P.R., Silva, M.M.M.A., Silva, H.C.D., Oliveira, N.F., 2014. Space time variability of surface temperature in the semi-arid Pernambuco based image TM/Landsat. *Journal of Hyperspectral Remote Sensing* 4, 111-120.
- Silva, M.D., Galvncio, J.D., Pimentel, R.M.M., Silveira, J.A.R., 2014. Cobertura Vegetal no estado de Pernambucano sob regimes de precipitação pluvial distintos. *Revista Brasileira de Geografia Física* 7, 584-596.
- Sousa, L.F., Braga, C.C., Braga, R.C., Dantas, P.M., 2014. interrelationship between rainfall and vegetation index by remote sensing. *Journal of Hyperspectral Remote Sensing* 4, 87-99.
- Weng, Q., Hu, X., Lu, D., 2008. Extracting impervious surfaces from medium spatial resolution multispectral and hyperspectral imagery: a comparison. *International Journal of Remote Sensing* 29, 3209-3232.